Wireless Sensor Data Processing for On-site Emergency Response

by

Yanning Yang

A Doctoral Thesis

Submitted in partial fulfilment of the requirements

for the award of

Doctor of Philosophy of Loughborough University

(June 2011)

Loughborough University

CERTIFICATE OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgments or in footnotes, and that neither the thesis nor the original work contained therein has been submitted to this or any other institution for a degree.

……………………………………………. (Signed)

……………………………………………. (Date)

ACKNOWLEDGEMENTS

There are numerous people who have provided support and guidance throughout the process of completing this thesis. Particular thanks are due to:

My supervisors Prof. Shuang-Hua Yang and Dr. Andrew May, who have both provided insightful suggestions and ongoing encouragement over many years. I am grateful that they were always available whenever I needed help. The technical support they offered helped me focus and stay on the right direction to achieve my goals. They helped me understand I was not doing the PhD alone.

My director of research Dr. Eran Edirisinghe, who provided useful guidance and support. The examiner of my end-of-year review meetings, Dr. Val Mitchell, who provided constructive feedback.

My father, Jianping Yang, and my mother, Suju Wang, who constantly supported me and encouraged me to continue. My uncle, Zhanhou Wang, who shared with me his valuable working experience in the area and clarified my confusion on a number of specific implementation issues. They have been a great strength to me. This thesis would not have been possible without their love and encouragement.

My friend Dace Rozenberga, who was with me during the most difficult times of my PhD and tried her best to help. My friend Yunqiu Li, who offered continuous friendship. My friend Raj Prasanna, who gave me lift to the interviews for data collection. All my friends in Holywell Park, especially Alexandra Alecu, Jatinder Gill, Khusvinder Gill, Hesham Abusaimeh, Yao Fang, HuanjiaYang, Xin Lu, Zaid Bin Ahmad, Ran Xu, and Carlos Insaurralde.

A number of members of staff from the Computer Science Department, who have provided assistance over the last three years. Particularly, Mr Gurbinder Singh Samra, Mrs Judith Poulton, Mrs Christine Bagley, and Dr. Roger Knott for doing their best to help. A member of staff in Holywell Park, especially Sharon Cook, Collette Nicole, and Jackie Cockburn.

Thank you all.

ABSTRACT

This thesis is concerned with the problem of processing data from Wireless Sensor Networks (WSNs) to meet the requirements of emergency responders (e.g. Fire and Rescue Services). A WSN typically consists of spatially distributed sensor nodes to cooperatively monitor the physical or environmental conditions. Sensor data about the physical or environmental conditions can then be used as part of the input to predict, detect, and monitor emergencies. Although WSNs have demonstrated their great potential in facilitating Emergency Response, sensor data cannot be interpreted directly due to its large volume, noise, and redundancy. In addition, emergency responders are not interested in raw data, they are interested in the meaning it conveys. This thesis presents research on processing and combining data from multiple types of sensors, and combining sensor data with other relevant data, for the purpose of obtaining data of greater quality and information of greater relevance to emergency responders.

The current theory and practice in Emergency Response and the existing technology aids were reviewed to identify the requirements from both application and technology perspectives (Chapter 2). The detailed process of information extraction from sensor data and sensor data fusion techniques were reviewed to identify what constitutes suitable sensor data fusion techniques and challenges presented in sensor data processing (Chapter 3). A study of Incident Commanders' requirements utilised a goal-driven task analysis method to identify gaps in current means of obtaining relevant information during response to fire emergencies and a list of opportunities for WSN technology to fill those gaps (Chapter 4). A high-level Emergency Information Management System Architecture was proposed, including the main components that are needed, the interaction between components, and system function specification at different incident stages (Chapter 5). A set of state-awareness rules was proposed, and integrated with Kalman Filter to improve the performance of filtering. The proposed data pre-processing approach achieved both improved outlier removal and quick detection of real events (Chapter 6). A data storage mechanism was proposed to support timely response to queries regardless of the increase in volume of data (Chapter 7). What can be considered as "meaning" (e.g. events) for emergency responders were identified and a generic emergency event detection model was proposed to identify patterns presenting in sensor data and associate patterns with events (Chapter 8). In conclusion, the added benefits that the technical work can provide to the current Emergency Response is discussed and specific contributions and future work are highlighted (Chapter 9).

Keywords: sensor data processing, Data Fusion, Emergency Response, emergency information system, Wireless Sensor Networks

PUBLICATIONS

Journal Paper

Yang, Y., May, A. and Yang, S-H. (2010), Sensor data processing for emergency response, *International Journal of Emergency Management*, Vol. 7, Nos. 3/4, pp. 233-248.

Yang, Y., May, A. and Yang, S. H., A Generic State Model with Neighbourhood Support from Wireless Sensor Networks for Emergency Event Detection, *International Journal of Emergency Management* – **Submitted**

Conference Papers

Yang, Y., May, A. and Yang, S. H., Sensor Data Meaning Extraction for Emergency Response, In Proceedings of *18th TIEMS Annual Conference 2011*, pp. 304-313.

Yang, Y., Prasanna, R., May, A. and Yang, L. (2010), Opportunities for WSN for Facilitating Fire Emergency Response, In Proceedings of *IEEE International Conference on Information and Automation for Sustainability (ICIAfS) 2010*, pp. 81-86.

Yang, Y., May, A. and Yang, S. H. (2009), *Making Sensor Data Work for Emergency Response*, In Proceedings of *16th TIEMS Annual Conference 2009*, Istanbul, Turkey, June 9-11, 2009, pp. 231-240.

Yang, Y., May, A. and Yang, S. H. (2007), *Emergency Information Management System Architecture*, In Proceedings of *14th TIEMS Annual Conference 2007*, Split, Croatia, June 5-8, 2007.

Abbreviations

ACTA	:	Applied Cognitive Task Analysis
ART	:	Adaptive Resonance Theory
CADET	:	Critical Action and Decision Evaluation Technique
CIM	:	Computation Independent Model
DEM	:	Department of Environmental Management
DRA	:	Dynamic Risk Assessment
DSM	:	Data Storage Mechanism
EIMS	:	Emergency Information Management System
EM	:	Emergency Management
EM	:	Expectation Maximization
ER	:	Emergency Response
ERM	:	Emergency Response Management
ERS	:	Emergency Response Systems
FEMA	:	Federal Emergency Management Agency
GDCTA	:	Goal Directed Cognitive Task Analysis
GDIA	:	Goal Directed Information Analysis
GDTA	:	Goal-Driven Task Analysis
GIS	:	Geographical Information System
GPS	:	Global Positioning System
HTA	:	Hierarchical Task Analysis
IC	:	Incident Commander

ICS	:	Incident Command System
IESD	:	Information Extraction from Sensor Data
IMAS	:	Influence Modelling and Assessment Systems
ISIS	:	Incident-Site Information Space
JDL	:	Joint Directors of Laboratories
KDSD	:	Knowledge Discovery from Sensor Data
LMS	:	Least-Mean-Square
MDA	:	Model Driven Architecture
MDAS	:	Model-Driven Architecture Stack
NAND	:	Negated AND
NMSE	:	Normalized Mean Square Error
NS	:	Neighbourhood Support
NSSL	:	National Severe Storms Laboratory
OAET	:	Operator Action Event Trees
PIM	:	Platform Independent Model
PSM	:	Platform Specific Model
RI	:	Rhode Island
RMSE	:	Root Mean Square Error
RQ	:	research question
TAM	:	Task Analysis Method
VR	:	Virtual Reality
WSN	:	Wireless Sensor Network

TABLE OF CONTENTS

List of Figures

List of Tables

List of Symbols

Symbols	Meanings
a	A value that is less than the pre-defined consecutive time T
A	The state transition model that applies to the previous state
A^T	The transpose of A
B	The control input model that applies to the control input vector
C_L	The contour map extracted from the live sensor reading
C_E	The pre-defined contour map event patterns
cf	The confidence associated to the neighbourhood level alarm
CF	The level of severity of the abnormal state at the network level
$d(t)$	Neighbourhood level alarm
$dis(s_j, s_i)$	The distance from sensor node s_j to s_i
D	The set of measured data from sensor network
$D(t)$	The network level alarm
f	The current sampling frequency
$f_A(t)$	The storage cost of the partitioned database at each minute t
$f_B(t)$	The storage cost of the database without partitioning at each minute t
f_L	The lowest acceptable sampling frequency
f_U	The highest acceptable sampling frequency
F	Fire spread model
FN	The number of actual (true) events that are not detected
FP	The number of detected events which are not actual (true) events

g: the pre-defined compression granularity

G_k	Kalman Gain at the k^{th} time instance
H	The observation model which maps the state space into the observed space
H^T	The transpose of H
I	The identity matrix
l_s	Level of support
m	The number of sensor types
n	The total number of warning levels
N_{s_i}	The neighbourhood of sensor node s_i
N_{S_2}	The number of neighbours who are in suspicious state
N_n	The total number of neighbours
NS	Neighbourhood Support
N	Network size, which is the length of the square $N \times N$ grid
NE	The estimation of the sensor reading under normal conditions
o	Outlier rate
O	Outliers
O_l	The lower bound of the pre-defined normal range for an incoming observation
O_u	The upper bound of the pre-defined normal range for an incoming observation
$p_1, p_2, \cdots, p_r$	r number of partitions by the range of sensor readings

$p_1, p_2, \cdots, p_l$	l number of partitions by the range of location
$p_1, p_2, \cdots, p_t$	t number of partitions by the range of time
P_c	The compressed partition
P_w	The working partition
$\hat{P}_k$	The a posteriori error covariance matrix at the k^{th} time instance
P_k	The error covariance at the k^{th} time instance
P_L	The pattern extracted from live sensor data
P_N	The pre-defined normal pattern of behaviour
Q	The process noise covariance
Q_a	The compressed data cell
Q_b	A data cell in the working partition
r	The radius from sensor node s_i
R	The measurement noise covariance
R_l	The lower bound of the pre-defined normal range for the rising rate
R_u	The upper bound of the pre-defined normal range for the rising rate
RR_i	The rising rate of sensor i
RR_{smoke}	The rising rate of smoke sensor
RR_{CO}	The rising rate of CO sensor
RR_{CO_2}	The rising rate of CO_2 sensor
s	The state of a sensor
s_i	Sensor node
$S1$	Normal state

S2	Suspicious/checking state
S3	Abnormal state
S4	False alarm state
t	The time when an event of interest occurs
t	Simulation time duration
T	A pre-defined consecutive time
th	The pre-defined threshold for the level of support
TH_i	The pre-defined threshold for the rising rate of sensor *i*
TH_{smoke}	The pre-defined threshold for the rising rate of smoke sensor
TH_{CO}	The pre-defined threshold for the rising rate of CO sensor
TH_{CO_2}	The pre-defined threshold for the rising rate of CO_2 sensor
TP	The number of correctly detected events
u_{k-1}	The control input vector
v_k	The measurement noise
v: (t,l,r)	The data volume in the range of time, area of location and range of sensor readings
w	A random noise
w_{k-1}	The process noise
W	The level of warning (the Level of Severity)
x	The state of a Kalman Filter system
x_k	The current state of a Kalman Filter system at the k^{th} time instance
x_{k-1}	The previous state of a Kalman Filter system at the $k\text{-}1^{th}$ time instance

$\hat{x}_k$	The a posteriori state estimation at the k^{th} time instance
z_k	The measurement at the k^{th} time instance
μ	The mean of normal readings according to the domain knowledge
μ_{temp}	Average room temperature under normal conditions
μ_{smoke}	Average smoke level under normal conditions
μ_{CO}	Average CO reading under normal conditions
μ_{Flame}	Average Flame reading under normal conditions
$=^{\sim}$	Pattern match
$\neq^{\sim}$	Pattern does not match

Chapter 1. Introduction

1.1 Background to the Research

Emergency Response (ER) has always been important for minimizing the damage and the loss of life and property as natural and man-made disasters continue to occur. Nowadays, the importance of ER has received increasing recognition throughout the world. The demand for efficient ER has been highlighted by "the responses to the 9/11, terrorist attack on the World Trade Centre, Hurricane Katrina in the southern U.S., and the July 7 London bombings" (Yang, L. et al., 2009). In recent years, especially after 9/11, ER has become an area that has been receiving rapidly growing interest from researchers worldwide. Researchers from a diversity of backgrounds, including public administration, control engineering, and computing and communication technologies, have all put effort into developing means of improving the efficiency and effectiveness of ER.

The prevalent problems that may affect the efficiency and effectiveness of ER can be classified into three categories: sociological, organizational and technological, as summarized by Manoj and Baker (2007). Sociological issues consist of many aspects, including 1) the lack of understanding on human behaviour models and the impact of such models on emergency response technology system design, 2) the interoperability issue among different responding agencies, and 3) the issue of trust both in terms of establishing trust among different responding groups to share and communicate critical information and in terms of the lack of trust from emergency responders in terms of adoption of new technology, etc. Organizational issues occur as an organisation of people different from what the responding groups commonly practise emerges and must be adopted during emergency response and recovery.

A number of technological issues related to ER have been recognised and addressed by researchers over the recent decade, e.g. lack of communication systems within one responding agency and among different responding agencies, lack of decision support systems, etc. Out of all the specific technological issues addressed by researchers, the most well recognised one is the lack of means of rapid gathering and communicating of critical information by the emergency responders during their responding to emergencies.

The importance of the availability of information and the means of communicating information has been recognised in the literature for over a decade. “In the absence of data and information, emergency response is simply well-intended guesswork that will most likely result in significant loss of human life” (Erickson, 1999).

In the modern world, successful emergency response not only demands effective management and command, coordination and cooperation from different responding crews, but also requires substantial support from technologies to provide the required information to the right person at the right time in the right format (Carver and Turoff, 2007; Manoj and Baker, 2007; Prasanna et al., 2007; Turoff, 2002; Yang, 2007). The people within an emergency responding agency can be classified into two categories: control centre staff and first responders.

Although different technology aids to ER have been proposed, most of the existing research focuses mainly on providing technology support to the control centre staff. Examples of technology support for control centre staff include: cascading web map services and Geographical Information System (GIS) web services (Vasardani and Flewelling, 2005), and remote sensing (Hutton and Melihen, 2006). The main advantage of such technology aids is that they enable better understanding of the whole picture of view within a region regarding incidents and available resources. As a result, better command and coordination of the available resources within the region can be achieved.

First responders are the teams offering immediate help to victims in case of an emergency. They are “the primary link in the chain of information exchanges that leads to making critical, perhaps lifesaving, decisions” (Sawyer et al., 2004).

However, there is a lack of technology support for first responders, who are usually deployed to the premises as soon as an incident occurs and act as the front-line soldiers in responding to incidents. Examples of first responders include fire-fighters, paramedics and police. Yang and Fredrick (2006) revealed that “during fire incidents, when the first responders arrive on site, they have very limited information about the building, occupants and/or the location of the hazard”. Such lack of information could result in a delay in making decisions about whether or not to enter the building, whether it is safe to enter, whether the priority is to search and rescue occupants or other tasks and how to most efficiently deal with the hazard. The result of wrong decisions due to the lack of information could be fire-fighter deaths and casualties. For example, the death of up to four fire-fighters (BBC news, 2007) while tackling a warehouse blaze in Warwickshire highlighted the dangers of having limited information on the building. According to the latest published fire statistics in the UK (Fire Statistics, 2009), there were 6 fire-fighter deaths and 268 fire-fighter casualties in 2007, 2 fire-fighter deaths and 350 fire-fighter casualties in 2006. These numbers clearly demonstrate the requirement for better first responder support.

Researchers have identified that both “outside building information” and “inside building information” are required to provide a full picture of view in the event of an emergency (Hansen, 2007). Some of the latest Emergency Response Systems (ERS), which are technical systems that provide support to emergency responders, have integrated technologies such as Global Positioning System (GPS) to help emergency responders arrive on site as quickly as possible. However, little information from inside of the building is provided by current ERS. On arrival at an incident premise, Incident Commanders (ICs) have to assess the situation by observing from the outside, asking people who have been in the building or people who know, or checking the facilities provided by commercial buildings such as operation panels to identify which zone a fire incident is in. However, very limited information is typically available and these ways of obtaining information about the nature of an incident inside buildings are often time-consuming and inaccurate.

Wireless Sensor Network (WSN) is an emerging technology that has demonstrated its potential in providing the “inside building” information (Yang and Fredrick, 2006). They can provide increased situation awareness that is important for emergency responders (Yang, L. et al., 2009). WSNs typically consist of battery powered, low cost, resource limited sensor nodes that can be deployed across areas of varying sizes to autonomously form wireless networks (Tanenbaum, 2003). Sensor nodes can even be deployed in harsh environment where human access is difficult, and they can transfer real-time data about the occurrence and spread of an incident (such as a fire) out of the network. The raw data collected from WSNs must be processed and converted into information, which is data that has meaning to an individual within a context of use. It is not until the raw data is converted into meaningful information about the environment being monitored that it can be interpreted by first responders and used by them to better plan their response to incidents.

This thesis focuses on first responders. It investigates the process and techniques of using data from WSNs to meet the information needs of first responders to facilitate the overall response to incidents.

1.2 Research Challenges

Using data from WSNs to satisfy the information needs of first responders presents a number of challenges which must be addressed if they are to be at practical use.

Although WSNs have promising and already successful applications in areas such as habitat monitoring, supply chain management, etc., they are known for having energy constraints and limited resources. Moreover, data from WSNs has different characteristics compared to those of traditional data. The constraints of WSNs and the special features of sensor data raise challenges in sensor data processing approaches, e.g. data storage, data cleaning and meaning extraction, because algorithms and approaches different from those designed for traditional data processing may be required.

The on-site ER environment presents a potential challenge as well. The nature of an incident is dynamic and highly demanding. Real-time or near real-time data retrieval, processing and management is required. However, the resources for computation and communication may be limited at an incident site. Therefore, meeting the demanding performance requirements under resource constraints could be a challenge.

In addition, this thesis aims to establish a link between technology and human factors perspectives, through an investigation of making sensor data 'work' for emergency response. Identifying technology capabilities of WSN that can ultimately impact on better situation awareness and other usability aspects of emergency responders presents potential challenges and this is tackled by incorporating a user-centred design perspective within this thesis (Noyes and Baber, 1999).

1.3 Motivation

This research was driven by the motivation to make a contribution to knowledge in both human factors and sensor data processing. Although this research wanted to focus on technological development, it wanted to base this on a clear understanding of the relevant human factors issues, and particularly the user needs of the ICs.

Most of the existing research in the area of Emergency Management (EM) is either from a pure technology perspective or from a pure human factor perspective. Taking a more multidisciplinary approach and establishing a clear link between the two will provide benefit to researchers from both a technology perspective and a human factor perspective. In particular, identifying the real gaps existing in information gathering in current ER and opportunities for technology to fill in the gaps links users' need with technology capabilities.

The motivation for this research was also driven by the potential benefits that sensor data processing technology can bring to ICs in facilitating the response to emergencies and reducing the risk of first responders being injured. As a result, a more effective ER may be accomplished, which can also bring benefits to the people who are under risk during an emergency, e.g. fire-fighters and building occupants.

The study of how a specific technology (sensor data processing) can bring benefit to users (first responders) prior to, during and after the design and implementation of the technology itself can provide considerable benefits to both researchers and users.

1.4 Research Questions and Objectives

This research aims to establish a link between what WSNs can provide and what first responders require. It focuses on the design and implementation of sensor data processing approaches that are suitable for facilitating first responders during their on-site ER.

The list of research questions (RQ) considered, and the specific research objectives associated with each research question to obtain the research aim, are described as follows:

RQ1: Out of all the potential user groups existing in the ER domain, which is the user group that can benefit the most from the information provided by sensor data?

Objective associated with RQ1: Investigate the existing literature available on ER from a human factors' perspective to understand the generic theory and practice in ER as well as to choose a targeted user group in the first responders.

RQ2: What are the goals, tasks, information requirements of the targeted user group?

RQ3: What are the opportunities for WSN technology to address the gaps related to retrieving the required information?

Objective associated with RQ2 and RQ3: Understand the goals, tasks and information requirements of the chosen user group through interviews and observations, with the purpose of identifying gaps in the current ER practice and analysing potential opportunities for the use of technology.

RQ4: How can the technology capabilities of WSN be implemented in an on-site ER system?

Objective associated with RQ4:

- Research the existing literature available on information extraction from sensor data, and the associated process and techniques of information extraction from sensor data.
- Design a suitable system architecture for a WSN-based Emergency Information Management System (EIMS); specify the system functions and the associated system components.

RQ5: What constitutes a suitable sensor data storage mechanism for on-site ER?

Objectives associated with RQ5:

- Accomplish a comprehensive analysis of the features of sensor data

- Evaluate and propose a suitable data storage mechanism for the WSN-based EIMS, with an emphasis on managing and maintaining the query efficiency regardless of the increase of data volume.
- Evaluate the efficiency of the proposed data storage mechanism and its cost-effectiveness.

RQ6: What constitutes a suitable sensor data cleaning approach for on-site ER?

Objectives associated with RQ6:

- Analyse and propose a suitable data cleaning approach for the WSN-based EIMS, with the emphasis on separating outliers from real environmental changes and dealing with them separately.
- Evaluate the effectiveness of the proposed data cleaning approach in comparison to the existing data cleaning approaches.

RQ7: How to extract the defined "meaning" from sensor data?

Objectives associated with RQ7:

- Analyse and specify what "meaning" can be in the context of WSN-based on-site EIMS, and propose a meaning extraction approach.
- Evaluate the performance of the proposed meaning extraction approach and discuss its application.

1.5 Research Road Map

In order to establish the link between human factors and technology, an on-going cycle of identifying problems with a human factor or user-centred analysis, proposing technical solutions and verifying the solutions was undertaken.

The initial cycle was mainly based on a literature review. This phase focused on establishing the general link between human factors and technology (the selection of a user group and a specific technology). The initial human factors analysis identified the targeted user group, ICs, based on the literature review undertaken in order to understand the current theory and practice in ER and to identify and analyse potential user groups. The literature review of existing types of technologies in ER suggested that WSN has great potential in providing the “inside” building information that may benefit the ICs. This initial proposal of using WSN to facilitate the ICs was positively received in interviews undertaken later.

The cycle was continued by undertaking visits to the Derbyshire fire and rescue service, observations and interviews with senior officers who can act as ICs in case of an incident. This phase further investigated the targeted user group and the specific technology. The aim during this phase was ultimately to answer the question “How can WSN be utilised to facilitate ICs?” The information requirements of ICs were identified through the analysis of their goals and tasks. Gaps in retrieving relevant information were identified and a list of opportunities for the use of technology was proposed. Each proposed opportunity targets an identified gap relating to the needs of the IC, and feedback on the proposed technical solutions was gathered from formal interviews and informal email contacts.

Further technical research was undertaken based on some of the verified technology opportunities. During this phase, the existing research on information extraction from sensor data was investigated, in order to understand the process and evaluate the available data fusion techniques. A system architecture was proposed and a detailed technique for each component of the system (sensor data storage, sensor data cleaning and meaning extraction) was proposed and evaluated in simulations and experiments.

A road map of the research, which demonstrates the link between technology and human factors during these cycles, is shown in Figure 1-1.

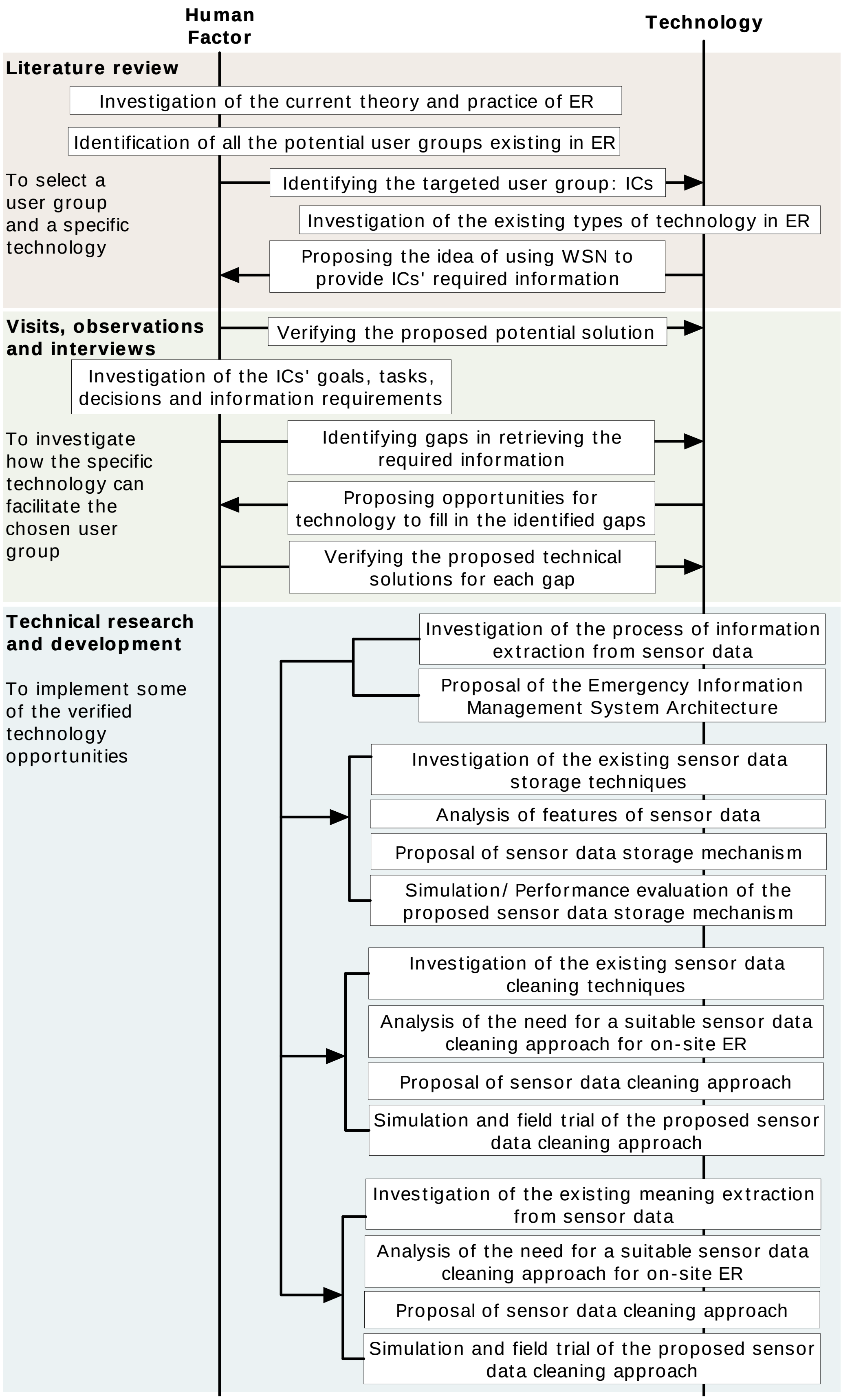


Figure 1-1: Research road map

1.6 Contributions of the Research

The contributions of the research detailed in this thesis are composed of five parts. They are summarized here as follows:

1. A comprehensive analysis of ICs' goals, tasks/decisions, information needs, and their feedback on the proposed technology opportunities. This established a link between user requirements and technology capabilities, with the emphasis on the added benefits that the technology can provide to the users.

2. The proposal of an emergency information management system architecture that addresses the dynamic nature of ER and can facilitate ER before, during and after incidents. The proposed architecture can also be flexibly applied to develop emergency detection systems, risk assessment systems, decision support systems, historical statistics or an integrated multi-purposes emergency response support system according to different contexts of use.

3. A proposed data storage mechanism that accommodates the challenges presented by the features of sensor data as well as maintaining efficiency for the entire ER process. Simulations demonstrated that the proposed data storage mechanism can provide the benefit of efficient querying with neither additional updating cost nor the introduction of unacceptable storage costs.

4. The proposal of a state-aware data cleaning approach for ER. This can fill in the gap relating to the fact that existing data cleaning approaches only produce a compromise between the observations and the system estimation regardless of the state. In comparison, the proposed data cleaning approach can not only reduce noise, but also remove outliers and quickly detect the real environmental changes.

5. A generic model for emergency event detection. This utilises the temporal-spatial correlations typically existed in WSNs in the form of Neighbourhood Support to improve the detection accuracy. The model can apply in both threshold-based event detection and tempo-spatial pattern-based event detection. Simulations demonstrated that applying this model can improve the

efficiency and accuracy of the existing event detection algorithms in both of the two categories of event detection.

All of the contributions aim to create suitable sensor data processing approaches to help satisfy the information requirements for ICs, and consequently facilitate ER.

1.7 Organization of the Thesis

The thesis was organized as follows:

Chapter 1 (this chapter) has introduced the background to the research and put the focus of the thesis into context. It has also identified the potential challenges to and the motivation behind the undertaking of the research. The scope of the thesis has been outlined, and the aims and objectives of the research have been identified. In addition, the road map of the thesis demonstrating the link between technology and human factor has been described, and the contributions of the thesis have been summarized.

Chapter 2 was a literature review. The two main purposes of the literature review were to understand the theory and practice of current ER, and to identify how the existing types of technologies can assist the information needs of users in ER. As a result of the review of the current theory and practice in ER, different user groups were identified and ICs were selected as the targeted user group. Then the existing types of technologies were reviewed with the emphasis of how they can assist the required information needs, resulting in the selection of WSN to be further investigated.

Chapter 3 investigated the existing literature on information extraction from sensor data, as it was identified that it still remains a challenge to make sense of the large amount of data collected from WSN. The process and techniques of information extraction from sensor data were reviewed. Arguments on what constitute a suitable sensor data processing approach were summarized, and general recommendations of suitable sensor data processing approach were derived.

Chapter 4 presented the study of the ICs' requirements, with the aim of identifying opportunities for WSN for facilitating fire emergency response, and determining the focus of further technical work. A goal-driven task analysis method was adopted and further extended to extract goals, tasks, information requirements and opportunities for technology to provide added benefits to ICs. A list of eight technology opportunities was identified and some of them were taken forward for further technical work.

Chapter 5 proposed an architecture for an on-site EIMS, for incorporating WSN capabilities to provide Emergency Information. Its scope and characteristics were defined, the design requirements were summarized. The main components of the system and the interaction between components were discussed. In addition, system functions before, during, and after incidents were specified.

Chapter 6 proposed a data storage mechanism designed for storing and managing data for on-site ER purposes. A comprehensive analysis of the features of sensor data was described, and the requirements of ER applications were evaluated. The detailed design of the data storage mechanism was discussed. Simulation results demonstrated that the proposed data storage mechanism can maintain the query efficiency as the real-time sensor data continues feeding into the database, and this was achieved with neither additional updating cost nor introduction of unacceptable storage costs.

Chapter 7 proposed a data cleaning approach. The feature of the typical existence of ‘noise’ in sensor data was addressed in the chapter. A set of state-awareness rules was proposed in order to address the issue that the existing data cleaning approaches do not take states into consideration and thus cannot separate outliers from real environment changes. The data cleaning experiments demonstrated that by integrating the state-awareness rules with a Kalman Filter, the resulting state-aware Kalman Filter can reduce noise and remove outliers, as well as quickly detect real changes in the environment.

Chapter 8 proposed an approach for meaning extraction from sensor data. One of the examples of “meaning” in the context of on-site ER can be defined as the occurrence and development of a fire incident. Therefore the meaning extraction problem can be converted into an event detection problem. The sensor state model was further developed and applied to two categories of event detection: threshold-based event detection and tempo-spatial pattern based event detection. Simulation results demonstrated that the proposed event detection approach can improve the detection efficiency and accuracy in both categories of event detection. False alarm rate can be reduced to 30%.

Chapter 9 concluded the thesis. This stated what are felt to be the most important contributions of the thesis, and identified future research questions that have emerged from the research contained in this thesis.

The structure of the thesis is shown in Figure 1-2. The figure demonstrates the relationships between the existing knowledge, the research questions posed and the studies undertaken in each chapter to address the research questions.

1.8 Acknowledgements

The work reported in this thesis was entirely undertaken by the author except Chapter 4.

The work contained in Chapter 4 is a study of the ICs' requirements. The data analysis contained in the chapter was designed, managed and completed by the author. The analysis and findings in the chapter were partly based on the author's data collection from the visits to Derbyshire and Leicestershire Fire and Rescue Services and 2 interviews with ICs. The analysis and discussion was also based on additional, interviews and data collection undertaken by Raj Prasanna in the Business School at Loughborough University. The pictures included in the chapter were produced by Raj Prasanna and are used with permission. The chapter was written in its entirety by the author.

All work within the rest of the chapters was undertaken and written solely by the author, and specifically for this thesis.

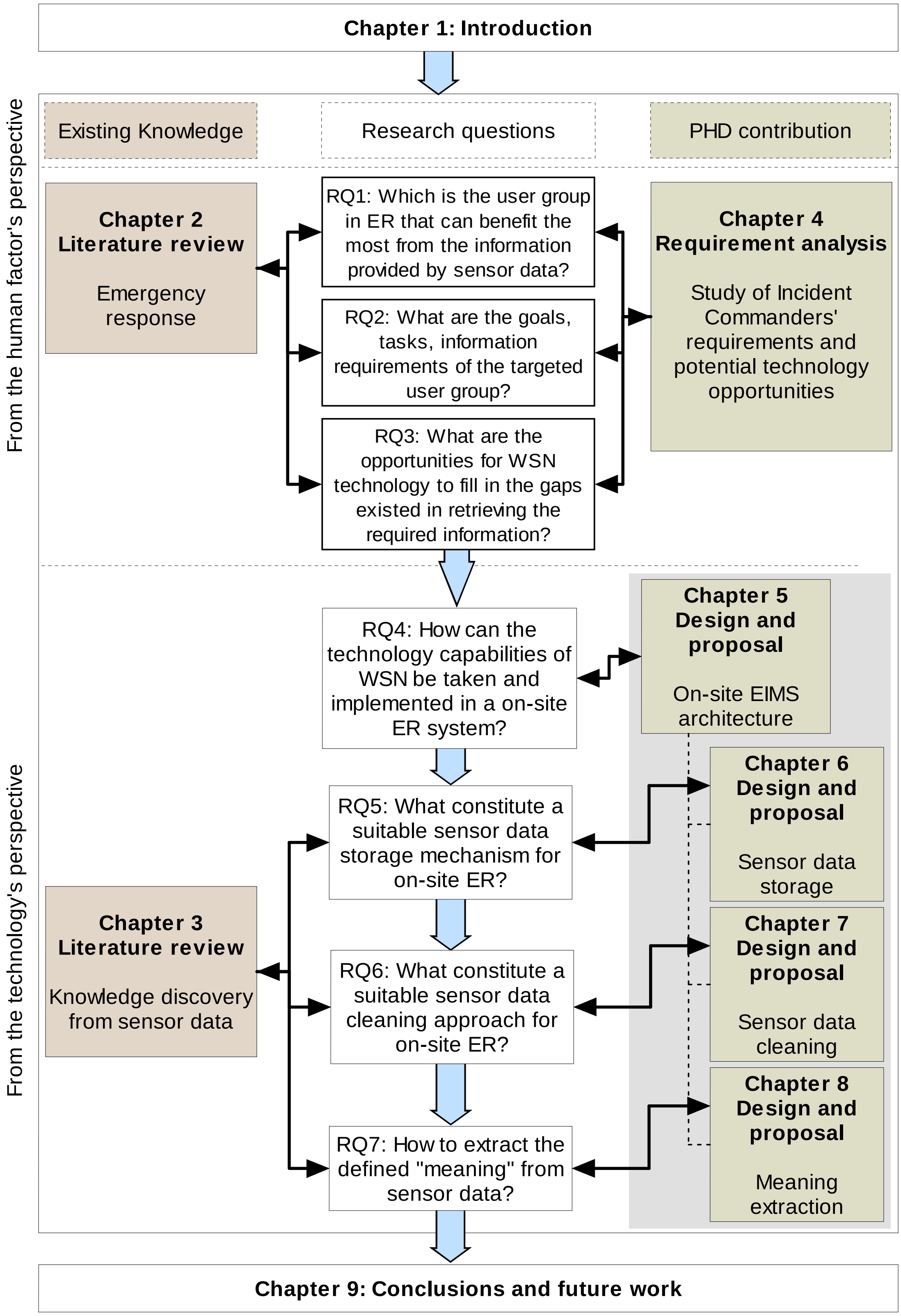


Figure 1-2: Structure of thesis

Chapter 2. Literature Review: Emergency Response

	Research questions addressed in this chapter:
1	What are the theoretical perspectives and real life practices in current Emergency Response?
2	Which is the user group that can benefit the most from sensor data, out of all the potential user groups existing in the domain of Emergency Response? What are their goals, tasks and information requirements?
3	What technology aids are available in existing Emergency Response Systems?
4	How is the existing technology meeting users' information requirements?
5	What general recommendations can be made for Emergency Response Systems design?

This chapter provides a comprehensive review of Emergency Response (ER), with a special emphasis on how the existing types of technologies can assist the information needs of users in ER. The review of the current theory and practice in ER identifies different potential user groups existing in ER, which leads to the subsequent selection of the targeted user group, verification of their information requirements, and evaluation of the strengths and weaknesses of the existing types of technologies in providing the required information.

2.1 Emergency Response: Theory and Practice

2.1.1 Emergency management and emergency response

Throughout history, there have been natural and man-made hazardous events occurring and people putting effort in to reduce the risk they propose to human life and safety. This effort laid the foundation for the discipline of Emergency Management (EM), which has become well known and respected in the recent 20 years. EM is defined as the profession and academic discipline of the management of disasters (prepare for, respond to, recover from disasters and mitigate their consequences to a certain degree) (Haddow et al., 2007). In other words, EM typically consists of four phases: mitigation, preparedness, response and recovery, as shown in Figure 2-1.

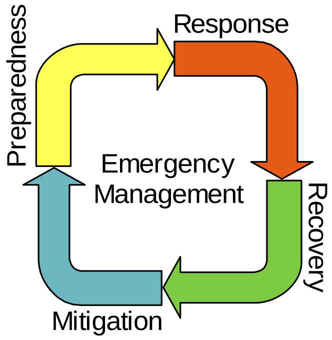


Figure 2-1: The process of emergency management

An appropriate starting place for EM is before the impact of incidents occurs. Mitigation refers to long-term minimising of the destructive effects of disasters, by the means of both structural (e.g. technologies) and non-structural measures (e.g. legislation, planning) (Christoplos et al., 2001). It is an important component of disaster risk management.

The other important component of disaster risk management is emergency preparedness. Its achievement takes place through a process of planning, training and exercising, and documenting response measures and protocols, which are generated by the planning process and rehearsed via training and exercises, in the written plan (Perry and Lindell, 2003).

Post-event response is an execution of the plans and operational responses to the incidents. There is often a theoretical debate over when the response function ends and the recovery function begins. Haddow et al. (2007) classified the response action as the “immediate actions to save lives, protect property, and meet basic human needs”, whereas the recovery function is concerned with issues and decisions that must be made after immediate needs are addressed.

In case of emergencies, service demands escalate tremendously, and survival mainly depends on the effectiveness of ER. Therefore, ER is a phase of particular importance to EM. Nowadays ER (especially response to catastrophic disasters) is characterised by shared authority, dispersed responsibility, scattered resources. Therefore, collaborative processes are essential (Waugh and Streib, 2006). These features of the ER present the requirement of collaboration and coordination between different agencies and between people playing varied roles.

2.1.2 Different roles and responsibilities in emergency response

A review of literature identified four categories of people involved in ER: government/management based roles, first responders, volunteers and technical expertise. Each of them can be considered as a potential user group with unique characteristics.

Government/management based roles

"Inasmuch as disasters are geographically localized, county and municipal authorities are most often required to assume primary responsibility for emergency management" (Waugh and Hy, 1990). Citizens rely first on their local government for timely, coordinated, and comprehensive responses during an emergency. They expect that known potential hazards are included in the emergency management plans, and mitigation actions be taken whenever possible. McLoughlin (1985) concluded that every level of government in the USA is involved in emergency responses at different levels and different phases. Local governments' main role is preparedness and response; state governments' role is to lead, support, and coordinate; federal governments have extensive resource, they provide policy, guidance, technical and financial assistance. Handmer and Parker (1991) assessed the change of focus of British disaster planning and management from 1) planning for wartime emergencies to 2) civil emergencies prevention, planning and management. Emergency Management has become a central activity of public administration and an important function of worldwide government especially after 9/11.

First responders

First Responders are members of organisations and agencies such as emergency communications centres, emergency medical services, fire, rescue and hazardous material response teams, law enforcement agencies, the Red Cross, and other disaster relief organisations (Sawyer et al., 2004). First responders routinely face dangerous environments and situations, ones ranging from fires to natural disasters to terrorist attacks (Betts et al., 2006). They are the teams offering immediate help to victims in case of an emergency, and maintaining an active link to policy makers. They are usually the "prime evaluators of threat and risk to homeland security", and "the primary link in the chain of information exchanges that leads to making critical, perhaps lifesaving, decisions" (Sawyer et al., 2004).

First Responders are characterised by the division of specialised roles and a hierarchical structure. There are various roles in the first responders group, each of which represents different responsibilities. To fulfil these responsibilities, their goals will be different. For example, in the same fire fighting system, "the goal of the gas analyser is to detect the level of toxic gas in the environment, the goal of the fire fighter is to determine the level of risk present in the environment and the goal of the Incident Commander is to decide on the appropriate response for his/her crew" (Stanton et al., 2006). Different first responding agencies with different expertise (e.g. fire and rescue service, police, ambulance) are required to collaborate together when emergencies become serious enough. As a result, the concept of unified command has been used to meet the demand of improved overall management when multiple and diverse agencies are involved (Irwin, 1989). The common practice in ER is to implement an Incident Command System (ICS), defined as "the standardized on-scene incident management concept designed specifically to allow responders to adopt an integrated organisational structure equal to the complexity and demands of any single incident or multiple incidents without being hindered by jurisdictional boundaries" (U.S. National Response Team, 2000).

Volunteers

Stallings and Quarantelli (1985) named the groups of private citizens carrying out important emergency response tasks on a volunteer basis as "emergent citizen groups". They perform damage assessment tasks (e.g. search and rescue immediately after the impact to assess damage, and report to public officials with first information about the actual extent and location of damage), operational tasks (such as distributing food, clothing, shelter or clearing the street), and coordination tasks. They are typically characterised by informal organisation, flat hierarchy, fluid boundary, emergent inner relations, and emergent goals and tasks. Roberts (2004) summarised the past experiments in direct citizen participation in all aspects (e.g. forms, challenges, consequences) and concluded that there is a trend of greater direct citizen involvement as the societies become more decentralized and networked. However, there are still challenges in practise related to size, expertise, risk, time consumption of direct citizen participation.

Technical expertise

A diversity of simulation and modelling systems have been proposed to assist emergency response practise. However, apart from the issue of interoperability among different systems, "typical emergency response organizations usually do not have the technical expertise or the time for building simulation models" (Jain and McLean, 2003). Modern emergency response requires the involvement of technical experts, including developing technical systems to facilitate ER and assisting in the ER training process (Ford and Schmidt, 2000). In addition, during emergencies demanding high domain expertise, e.g. the emergencies in nuclear power industry (Crichton and Flin, 2004), biological terrorism attacks (Rotz et al., 2002), collaboration with domain technical experts is essential.

Summary

In summary, during the response phase, governments' role is mainly a coordinator and supporter, although they have more responsibilities in policy making and command and control during the rest of EM process. First responders are trained to be "front-line soldiers" during the response phase. In case of emergencies, they have the primary responsibilities on-site and are presented with the most risks. Volunteers and technical expertise are also involved in ER activities. They can play a supporting role to first responders, for example by undertaking assigned tasks.

After reviewing different roles involved in emergency response, it can be stated that first responders have the most important role in on-site responses to emergencies. Therefore, the next section reviews further into the first responders group.

2.1.3 Incident commander in the first responders group

As described in section 2.1.2, specialized roles division and a hierarchical structure are the typical characteristics of the first responders group. The common practice of first responders across countries is to implement an ICS for on-site ER, under the concept of unified command.

"Incident Commander is the highest-ranking position within the ICS and the one functional position always filled" in the first responders' hierarchy (Bigley and Roberts, 2001). The initial person occupying this position is usually a senior officer in the first arriving crew. When the scale of the incident expands and becomes not manageable for the initial Incident Commander (IC), further resources will be deployed. The position of IC may be occupied by a later arriving more senior officer, and sectors can be introduced in the incident scene. In this case, the initial IC usually becomes a sector commander. No matter what the scale and complexity of the incident is, as shown in Figure 2.2, the IC remains at all times responsible for: the command and control, deployment of resources, tactical planning and co-ordination of the sector operations (Incident Command - 3rd edition, 2008).

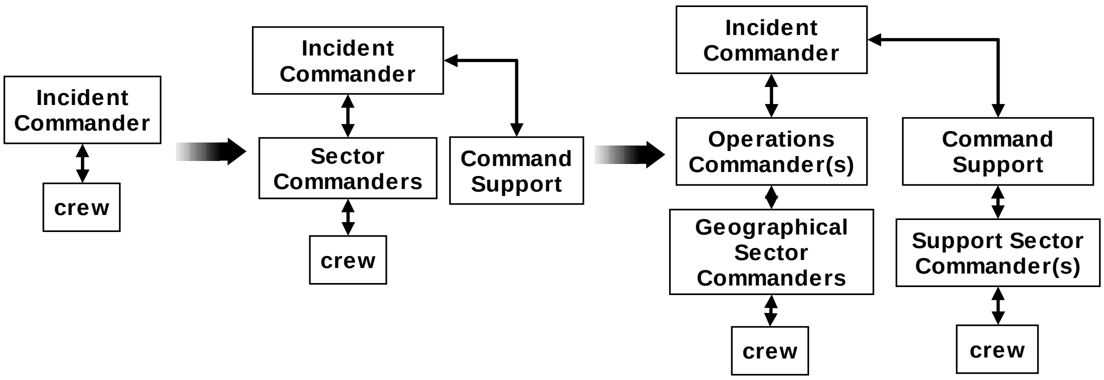


Figure 2-2: ICS structure as the scale and complexity of the incident increases

"The IC is an information intensive position, which involves coordinating the overall response strategy to an emergency and managing available people and resources in real time" (Jiang et al., 2004a). Therefore, decisions made by ICs would have great impact on the overall efficiency of on-site ER. Technology support providing the required information to ICs would likely have benefits for the overall performance of on-site ER.

Goals and priorities

Although the terms used may be slightly different, the principles in ER are similar across countries. Turoff (2002) claimed that "past and future objectives remain the same in crises, providing relevant communities collaborative knowledge systems to exchange information".

For different types of incident, the focus of goals could be different. The primary goals of incident command in the Fire Service are "to ensure that all is done that can be done to protect people – the civilians at risk and the fire-fighters – and that everything is done to preserve and protect property by confining the fire and extinguishing it as quickly as possible" (Carter and Rausch, 2008), whereas in public health incidents, the primary goal would be to prevent epidemic and ensure the health and safety of the general public (Qureshi et al., 2006).

Although different ER agencies have their individual responsibilities, they share common objectives: "save life; prevent escalation of the disaster; relieve suffering; safeguard the environment; protect property; facilitate investigation/inquiry; and restore normality as soon as possible" (Hill and Long, 2001). A similar set of ICs' goals and priority were published by the U.S. Federal Emergency Management Agency (FEMA): "reducing the immediate hazard, saving lives and property, establishing situational control, and restoring normal operations. Lifesaving and responder safety will always be the highest priorities and the first objectives in the Incident Action Plan" (FEMA, 2008). Therefore, it can be summarized that ICs goals during emergencies (in a descending priority) are to save life, stabilize the incident, and protect property and environment.

In the emergency response plan of the Department of Environmental Management (DEM), State of Rhode Island (RI), the key goals of the IC are defined as: "Establish incident response objectives and strategies; Acquire and apply the most accurate, up-to-date assessments of the situation; Supervise an effective, safe, and efficient ICS organization; Deploy and monitor resources; Keep stakeholders and staff well-informed; Demobilize Incident Command" (DEM RI U.S., 2008). Compared to the sets of goals described above, this set of goals focuses on the role that ICs play in an emergency situation, instead of the ideas they are required to keep in mind when they play their roles. This is most like a sequence of common tasks that ICs are required to carry out from the beginning to the end of an emergency.

Tasks/Decisions

Perry (2003) summarized the duties of the IC throughout an incident as follows:

- conduct initial situation evaluation and continual reassessments;
- initiate, maintain and control communications;
- identify incident management strategy, develop an action plan and assign resources;
- call for supplemental resources, including Emergency Operation Centre activation;
- develop an organizational command structure;
- continually review, evaluate and revise incident action plan;
- provide for continuing, transferring and terminating command.

The initial situation evaluation and continual reassessments mentioned above are defined as Dynamic Risk Assessment (DRA) in the fire service guidance in the UK (Incident Command - 3rd edition, 2008). It is called a Dynamic Risk Assessment because the process of risk assessment is carried out in a changing environment, where what is being assessed is developing as the process itself is being undertaken.

It is a standard operational procedure of the local emergency responders (West Yorkshire Fire and Rescue Service, 2007) that the IC will carry out a DRA on arrival at an incident. The principal elements of the DRA are termed "the operational risk assessment process" and include the following areas:

- To look for and identify hazards (hazard spotting).
- To decide who might be harmed and how.
- To evaluate the risks (possibility and severity) arising from the hazards.
- To decide what precautions are necessary.

In response to the hazard and risk evaluated, the appropriate tactical mode in any sector or incident which has not been sectorised can be declared to be defensive or offensive. Where the risk to crews is excessive, defensive mode will be declared. Where safe systems or work are deployed and sufficient control measures are implemented, the tactical mode is likely to be offensive. Defensive mode usually means the crews will stay outside of the premises, whereas offensive mode may involve deployment of the crew into the premises. Where an incident is sectorised, the tactical mode can vary between sectors, in this case, the incident is in transitional mode. The IC is required to include the tactical mode in the regular (about every 20 minutes) communication both to the fire crew and to the control centre.

After a tactical plan has been initiated on the basis of a DRA, it is important that this is reviewed and confirmed as quickly as practicable, and further reviewed and confirmed at regular intervals. As the incident develops, there may be dynamically generated risks, e.g. there may not be a staircase anymore, or part of building structure may have collapsed, temperature of chemical storage may rise to a critical point. Such dynamically generated risks must be included in the DRA. Any changes on the tactical mode or resource deployment, or any further request for resources as a result of the dynamically generated risks must be updated in the IC's communication to the fire crew and to the control centre.

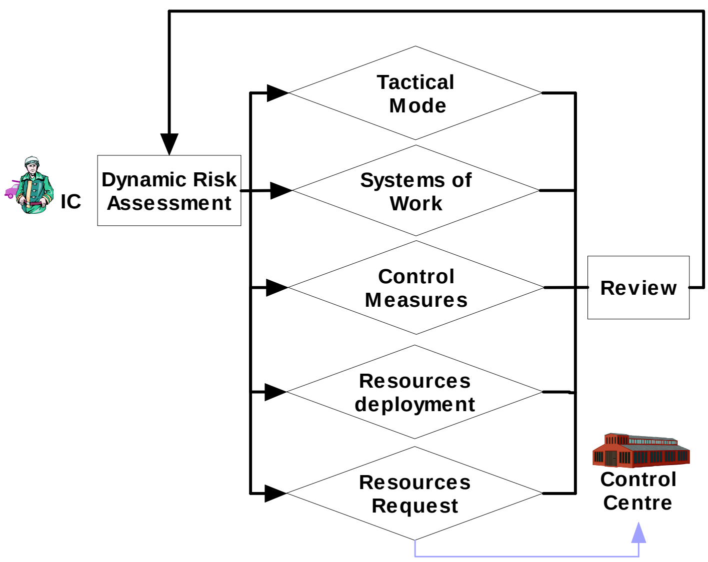


Figure 2-3: ICs' tasks and decisions

As summarized in Figure 2-3, ICs make critical decisions about tactical mode, control measures to minimise the risks, resource deployment, etc. based on the result of DRA.

In case of an incident, DRA is carried out by the first arriving IC as soon as the first fire crew arrives on site. It is continued by whoever is in the IC's position during the incident at regular intervals, and lasts until the "stop" message is issued by the IC declaring the end of an emergency response. Therefore, DRA is the most important function and task of ICs.

Information Requirements to achieve the tasks

It has been recognised that good decision-making relies on the information available and the ability of decision makers to cope with the demands imposed upon them by the management of an emergency response situation (Danielsson, 1998).

The faster the emergency responders are able to gather, analyse, share, and act on key information, the more effective their response will be, the better the needs will be met, and the greater the benefit to all affected people (Van de Walle and Turoff, 2007).

It has been widely recognised in the emergency communities (Carver and Turoff, 2007; Manoj and Baker, 2007; Prasanna et al., 2007; Turoff, 2002; Yang, 2007) that on-site dynamic information retrieving, sharing and presenting in the right format at the right time and to the right person will assist in improving initial key decision making.

To execute DRA on-site, "upon arrival at an incident the first task of the IC must be to gather all available information relating to the incident" (Incident Command - 3rd edition, 2008). Unfortunately, in the case of fires in and around large scale structures, when the first responders arrive at the site of an incident they have very limited access to on-site real time dynamic information (Yang and Frederick, 2006). Examples of such information include environmental conditions within the buildings, status of the casualties, resource requirements and the locations of various hazards.

There is also a lack of a comprehensive understanding about how the required information is gathered by ICs in real practice, and whether there are gaps in getting the required information at the right time in the right format. Therefore, a further analysis of ICs' tasks with the aim of identifying these gaps and the underlying requirements for technology is needed, which is described in chapter 4.

Task analysis methods

The available task analysis methods are reviewed in this section as a preliminary step to the further study of ICs' requirements. The aim of this review is to understand the existing task analysis methods and find whether there is one that can identify the underlying requirements for technology.

A wide variety of different task analysis methods exist. Embrey (2000) classified task analysis techniques into two categories: action oriented techniques and cognitive task analysis approaches. The former mainly focus on observable aspects of operator behaviour, whereas the later focus on the mental processes underlying observable behaviour. Examples of action-oriented techniques are Hierarchical Task Analysis (HTA), Operator Action Event Trees (OAET), and Decision/Action Flow Diagrams. Examples of cognitive task analysis techniques are Critical Action and Decision Evaluation Technique (CADET), and Influence Modelling and Assessment Systems (IMAS).

A guide to task analysis (Kirwan and Ainsworth, 1992) divided task analysis methods into five broad categories: task data collection, task description, task simulation, task behaviour assessment, and task requirement evaluation, as shown in Table 2-1.

Table 2-1: A summary of task analysis methods

Categories of Task Analysis Methods	Example methods	
Task data collection	• Activity sampling • Critical incident technique • Observation	• Questionnaire • Structured interview • Verbal protocol.
Task description	• Charting and networking methods • Decomposition methods • Hierarchical task analysis	• Link analysis • Operational sequence diagram • Timeline analysis.
Task simulation	• Computer modelling and simulation • Simulator and mock-up	• Table top analysis • Walk-through and talk-through
Task behaviour assessment	• Barrier and works safety analysis • Event tree • Failure modes and	• Fault tree • Hazard and other operability study • Influence

	effects analysis • Diagram,	• Management oversight and risk tree
Task requirement evaluation	• Ergonomics check-lists	• Interface surveys

Hierarchical Task Analysis (HTA)

HTA was introduced by Annett and Duncan (1967) to evaluate an organization's training needs. The underlying technique, hierarchical decomposition (Annett et al., 1971), analyses and represents the behavioural aspects of complex tasks such as planning, diagnosis and decision making (Annett and Stanton, 2000). HTA breaks tasks into subtasks and operations or actions. These task components are then graphically represented using a structure chart. HTA entails identifying tasks, categorizing them, identifying the subtasks, and checking the overall accuracy of the model.

Although the emphasis was put more on hierarchical decomposition, HTA, in fact, identifies both observable tasks/behaviours and underlying goals.

HTA is a useful analytical framework for complex tasks. However, it only represents a hierarchy of goals and tasks; therefore there is no link that leads from tasks to requirements for underlying technologies.

Goal-Driven Task Analysis (GDTA)

Goals have been recognised to be essential to understand the decision-making process. Albers (1998) argued that in the unstructured environment (meaning jobs that require dynamically adjusting of tasks as new information presents itself) the user's goal is not just completing a specific task, but decision-making or problem-solving. As such, the user is goal-driven and not task-driven. Therefore, data collection and the analysis method for complex problem-solving processes should be goal-driven.

The GDTA methodology presented by Albers (1998) also put emphasis on the information needed to achieve the goals. The methodology was designed to improve situation awareness for complex problem-solving by providing information to help solve a problem, "a goal which many systems fail to meet". The result is a goal/information diagram linking the user's goals and information needs.

Albert (1998) concluded that "Goal-driven analysis is not a means of identifying individual steps within a process or debating the relative merits of various approaches. Rather, it identifies the informational needs required to distinguish paths in a problem-solving situation."

This method links from goals to tasks to information, which is one step closer to identifying the requirements for underlying technologies.

Goal Directed Information Analysis (GDIA)

Prasanna et al. (2009) proposed GDIA as a cognitive task analysis protocol to capture the information requirements of emergency first responders. It combined the capabilities of two existing methods, Goal Directed Cognitive Task Analysis (GDCTA) (Endsley et al., 2003) and Applied Cognitive Task Analysis (ACTA) (Militello and Hutton, 1998), so that in combination it can form a better tool to address the requirements-gathering in emergency domains such as fire and rescue. According to Prasanna et al. (2009), the comparison can be summarized in Table 2-2.

Table 2-2: Comparison of GDCTA, ACTA and GDIA

GDCTA	ACTA	GDIA
Top-down method	Bottom-up method	Combined method
Aim to design information system interfaces	Aim to design training programs and instructional material	Aim to design information system interfaces
Data collection based on highly unstructured interviews using "why" and "how" as the main probes	Provide guidelines on the appropriate data collection methods (e.g. simulations, interviews, observations) to implement each step of the method	Provide clear guidelines on how to implement each application steps of the protocol
No indication of any criteria of interviewee selection	Interviewees of different level of experience (experienced and novice) are selected to avoid bias.	Interviewees of different level of experience (experienced and novice) are selected to avoid bias.

The GDIA method gathers data in the order of context, scenario, tasks, goals and information. It is a more comprehensive information gathering protocol. However, the aim is to gather requirements for information system interface design rather than the underlying technology.

Therefore, a method that is adapted based on GDTA and GDIA, with an extension link towards the underlying technology is utilised in the further study of ICs' requirement analysis in Chapter 4.

2.2 Technology Aids in Emergency Response

Researchers have developed various new mechanisms for responding to emergencies using computer and communications technologies. Attention was also paid to a diversity of aspects in ER, including system architectures, geographical mapping and Geographical Information System (GIS), routing services and logistics management, communication services between ER personnel, hazard materials detection, etc. This section provides a comprehensive review of the existing research in those areas, which leads to subsequent evaluation of the strengths and weaknesses about the existing types of technologies in providing the information to facilitate ER.

2.2.1 Architecture proposals

Hinton et al. (2005) proposed a flexible and scalable future wireless emergency response system architecture, consisting of a fixed part as well as a portable part, to meet the requirements as the capacity and connectivity needs rise. They pointed out that building a global wireless infrastructure for the worst-case scenario could result in a waste of limited spectrums and resources, whereas a scalable and fast-deployable architecture is crucial. In their system architecture, the fixed part consists of base-stations, whereas the portable part consists of portable base-stations and mobile gateways. This enables control of the integration of fixed and portable networks. Deployment procedures of the mobile gateway were also specified. Their proposed system architecture is hardware and infrastructure focused.

Zlatanova (2005) summarized three possible types of system architecture for ER in urban areas - centralized, federated and dynamic collaboration - based on how data is managed. He stated that according to the dynamic nature of ER, dynamic collaboration is the most appropriate type. The proposed architecture provides dynamic collaboration through a key data middleware layer, which sits between a user layer (wire or wireless accessed users) and a distributed data server layer. The data middleware layer acts as an intelligent data fusion system, which manages data stored in its original representation rather than under a common schema and provides data and answers to queries according to different user dynamics automatically. This proposal captured some important needs of ER, e.g. the dynamic nature of ER, the importance of data and communication of data, and the critical component (middleware) of the dynamic collaboration system architecture. However, it states that the responsibility to create such critical component lies with geospatial researchers and developers, no details on what may constitute the middleware and how it may be created were suggested. Moreover, the scenario used to describe how the system architecture may work indicates an initialization period of “not more than three or four hours” to gather information, before decisions can be made. This seems unrealistically long in a real ER scenario.

2.2.2 Geographical information and mapping technologies

Many of the national Emergency Response Systems (ERS) are based on or partly based on geographical information and mapping technologies, especially when the country itself has a large total area, for example, China Earthquake Forecast Centre utilizes a national Forecast and Emergency Response System based on a GIS. GIS is also used in local area ER services to provide the first responders with a visual view of the city. Centralized GIS has been widely used in emergency response centres, where data from various sources is collected and maintained in a central database. However, it has been argued that it is difficult for centralized GIS to keep all the data updated frequently, and it is subject to the risk of single point of failure. Research has suggested using the GIS Web Services and the Cascading Web Map Servers to optimize centralized GIS and existing mapping technologies for Emergency Response Management (ERM) (Vasardani and Flewelling, 2005)

GIS has the advantage of providing a geographical representation of nation-wide incident distribution or risk distribution. It is widely used in national control centres to facilitate decision making. For example, the control centre staff, who manage and control all the resources available in the area, can have an intuitive whole picture of all the resources available and the distribution of incidents in the area, with the current status of the deployed resources. Abnormal situation can then be identified and command issued in a timely manner. Data can be stored in a geo-database and organized by locations. Spatial relationship and emergency situation trends of a specific address can be identified, therefore providing responders with an intuitive picture of a location of interest. Computer-generated maps can be shared between different responding crews, hence potentially increasing the interoperability between them.

However, applying GIS in first responders' daily activities on-site is still a challenge. First, important tactical information such as the nature of the fire, hydrant locations, digital aerial orthography, floor plans, and construction materials cannot be provided by the GIS. It "only provides access to 'outside the building' information, such as street maps and public works", therefore, first responders only have "half of the picture in the event of an emergency" (Hansen, 2007). In addition, in contrast to control centre staff who have easier access to the diverse information presented by GIS and less time constraints to assimilate them, on-site first responders such as a fire ground commander make decisions under severe time pressure while performing life-critical tasks (Klein, 1999). Therefore accessing and assimilating such diverse of information is not appropriate for first responders under on-site conditions.

2.2.3 Modelling and simulation

A variety of modelling and simulation techniques have been employed to facilitate training ER personnel. Jain and McLean (2006) proposed a framework to facilitate application of modelling and simulation to incident management, which addresses incident management on three axes – incident, domain and lifecycle phase. Pimentel (2002) described a program called the Weapons of Mass Destruction Decision Analysis Center developed by Sandia National Labs as a way to simulate a war-room environment in the event of a terrorist attack. Its aim is to train public officials' response to a bio-terror attack. Hanson (2000) reviewed a system called BioSimMER, which is "a prototype virtual reality (VR) system designed to train first responders to nuclear, biological, and chemical acts of terrorism".

Simulation tools are also used for the identification and early warning of disaster events. The Warn-on-Forecast project being undertaken by the National Severe Storms Laboratory (NSSL) in the U.S. proposes to use weather prediction modelling to explore ways to better detect tornadoes and extend the warning lead times (Lakshmanan et al., 2007).

Modelling and simulations have the advantage of providing a sense of virtual reality in emergency situation development projections. It has the potential of identifying the possibility of an occurrence prior to the event, thus mitigating the event impact and increasing preparedness. However, the available simulation tools are mainly for standalone use, and they require a considerably large amount of computing resources. Therefore, applying modelling and simulation tools in an on-site ER environment would be challenging due to the limited resources available, the high time pressure, and the lack of expertise in such environment.

2.2.4 Communication based messaging services

Communication-based messaging services focused on (1) enhancing interoperability, which is the communication between different emergency response agents involved in an incident such as police, fire brigades and ambulance, and (2) enhancing controllability, which is the communication between first responders and their control centre, and between national control centre and local control centres.

An Electronic Message Management System for Emergency Response has been developed to better track the commands issued and their procession (Andersen et al., 1998). It was claimed in the paper that keeping a consistent communications log is useful for coordination between ER personnel, resources management and the preparedness of EM areas. The system was designed for emergency personnel to communicate with each other. It can be considered as an email system with special features that were designed for multiple ER agencies including EM organizations. The system enables the decision makers to send tasks as an email request to the relevant people, and monitor the status of any task.

However, such pure text-based communication amongst on-site first responders may not be appropriate due to possible low visibility, high risks and time critical hands-on tasks at emergency scenes.

2.2.5 Wireless sensor networks for emergency response

Wireless Sensor Networks (WSNs) consist of large arrays of battery-powered nodes, each of which can carry different types of sensors to monitor the environment and transmit data wirelessly. Sensor nodes are typically small-size, low-cost and low-power consumption. WSNs have been used in emergency medical care (Welsh et al. 2004), in-home healthcare (Stankovic et al., 2005), civil infrastructural health monitoring (Kottapalli et al., 2003), emergency evacuation (Barnes et al., 2007). These applications have proved the capabilities of WSN in improving the efficiency of ER.

The existing research that applies WSN in Emergency Response can be classified into two main categories: emergency monitoring and emergency navigation.

Emergency monitoring

Researchers have envisioned sensor network nodes playing a variety of emergency response roles. For instance, simultaneous physical environment monitoring, health monitoring, and location tracking.

The FireBug project (Doolin and Sitar, 2005) designed a wildfire monitoring system. The system collects temperature, relative humidity and barometric pressure with an on-board GPS unit attached to a wireless, networked mote. Wildfire behaviour was analysed based on patterns identified from the data collected during burns. For instance, the result showed temperature increasing, and barometric pressure and humidity decreasing as the flame front advanced. Such fire behaviour was claimed to be potentially useful for advancing fire science, for helping fire-fighters, and for designing future generations of sensors and sensor platforms.

CodeBlue is an on-going project intended to provide a wireless infrastructure for emergency medical care (Welsh et al., 2004; Gao et al., 2008). Wearable sensors are used to monitor vital signs (e.g. pulse oximetry, blood pressure, temperature, electrocardiogram) of the patients. The data collected by sensors is then relayed to handheld computers carried by emergency medical technicians, physicians, and nurses, to improve the ability of medical first responders to triage and treat patients.

Siren is a WSN-based communication system for fire-fighters directly engaged in structural fires. It provides fire-fighters with contextual data, such as location and temperature, and alerts them about imminent dangers using tacit means, i.e. messages exchanged between handheld devices carried by fire-fighters (Jiang et al., 2004b).

Civil infrastructural health monitoring (Kottapalli et al., 2003) uses wireless sensors to measure acceleration, linear displacement, strain, angular displacements during an extreme event like an earthquake in a near real-time manner. For long-term periodic monitoring, how the structural properties respond to environmental variables like temperature and humidity are also measured in order to get an accurate picture of the health of building structures.

Emergency navigation

Emergency navigation is used to describe applications or services that provide navigation information to people in the building in case of emergency.

Li et al. (2003) designed their navigation algorithms using the artificial potential fields concept. The exit generates an attractive potential, pulling sensors to the exit, while each obstacle generates a repulsive potential, pushing sensors away from it. Each sensor calculates its potential value and tries to find a navigation path with the least total potential value.

However, Tseng et al. (2006) argued that Li et al.'s algorithm could cause message overhead and could lead users outside the emergency region to exit through emergency region. Tseng et al. (2006) proposed a protocol based on temporarily order routing algorithm, and applied the concept of emergency region in their navigation algorithm. The result of this was reduced message packets and safer navigation.

Barnes et al. (2007) took into account predictions of hazards spread (3 Dimensional), such as fires, and evacuees' movements, as opposed to a static emergency region, to ensure the evacuees stay safely ahead of hazards.

The LifeNet project (Klann, 2009) proposed the concept of using sensor nodes to create an electronic "lifeline", instead of the physical lifeline that is typically used in search and rescue practice to guide fire-fighters in complex structural buildings. The electronic lifeline was designed as a wearable computing system and micro display to compute and display navigation guidance for fire-fighters in the buildings.

Challenges and opportunities in applying WSN in ER

One of the well recognised in-network challenges is the limited resources that sensors typically have. Lorincz et al. (2004) analysed that the limited communication and computation capabilities available at sensor nodes leads to the following challenges: discovery and naming, robust routing, prioritization of critical data, security, tracking device locations. Limited network lifetime because of the energy constraints of sensor nodes is also well recognised. A variety of energy-saving methods were proposed. E.g. Merrett (2005) proposed a resource awareness WSN system called IDEALS, "a system to manage a wireless sensor network using a combination of information management, energy harvesting and energy monitoring, which we label resource awareness", to enhance network lifetime and maximize the information throughput.

Although in-network challenges have been well-researched, less attention has been paid to challenges that sensor data presents to data processing. Research has stated that sensor data often contains errors (due to sensor function) and noise (due to other environmental interference) (Elnahrawy and Nath, 2003). Therefore, the quality of sensor data needs to be improved. However, there is a lack of comprehensive analysis about the nature of sensor data and what challenges they present to sensor data processing.

2.2.6 Robots search and rescue

Kumar et al. (2004) combined a network of Mote Sensors with mobile robots team and radio tags to provide an integrated view for situation awareness, guide fire-fighters to targets, and warn them of potential dangers. Their experiments included location tracking of mobile robots (Kantor and Singh, 2002), temperature gradient graph (Kantor et al., 2003), and direction guidance generation based on relative position (Kantor et al., 2003), which has demonstrated the potential that robots can interact with sensors, to help acquire information from inside of the building. This identified potential areas of risk, thus facilitated fire-fighters in Emergency Response scenarios. However, their research was based on an initial scenario that sensors and radio tags are deployed into the burning building by robots before they can be organized into networks, in other words, the robots have to be deployed into the building before the network can be setup, however, the required setup time which could cause delay in responding to fire emergencies was not mentioned.

2.3 Lessons Learned

This chapter reviewed Emergency Response from the perspectives of both the human factor and the technology. The lessons learned are as follows:

Incident commanders' decisions are vital for the result of emergency response.

Out of all the phases (mitigation, preparedness, response and recovery) typically involved in EM, ER is of particular importance. It manifests how effective the post-event mitigation and planning was, and its effectiveness determines survival or death in case of incidents. Among all the groups of people involved in the ER phase, first responders are the "front-line soldiers", therefore they have the most important responsibilities and the highest risk presented to them. Incident Commanders (IC) in the first responders' group have ultimate responsibility on the incident ground (Bigley and Roberts, 2001), they are in charge of all the responding activities on the incident ground, hence, their decisions have a great impact on the overall performance of ER. Moreover, IC is an information intensive position (Jiang et al., 2004a), and ICs' position on-site is usually out of the premises. In addition, their hands can usually be free, therefore they can have more information displayed to them to facilitate decision-making processes. As a result, IC is selected as the targeted user for further research.

The importance of information to ICs' decision making had been well recognised. The lack of information on-site had been drawn to attention. However, there is a lack of comprehensive understanding about how the required information is gathered by ICs in real practice, whether there are gaps in getting the required information at the right time in the right format. Therefore, a further analysis of ICs' tasks with the aim of identifying such gaps and the underlying requirements for technology is needed, which will be described in Chapter 4.

Wireless sensor networks have great potential in facilitating ER

A variety of technologies have been proposed to facilitate ER. In terms of how they can fulfil the information needs of the selected users, GIS can provide "outside" building information, whereas WSN has the potential for providing "inside" building information. Hence, a GIS is more suitable for the control centre staff to monitor and coordinate resources over a large area, whereas WSN is more suitable for an IC on-site to gather tactical information that can facilitate decision making. Visualization tools and virtual reality provide intuitive information and support during a training period, but it may take too long to run such virtual reality simulations during on-siteER. Communication-based messaging services can facilitate the communication amongst different responding agencies and between the control centre and the IC on-site, however, it may not be suitable to have messaging communication between the IC and front-line responders deployed at the incident scene. This is because front-line responders usually need their hands free to carry out rescuing tasks, or they may work in conditions where they cannot see any text messages. Robot search and rescue is a good concept. Robots have the potential to be part of ICs' search and rescue team on-site. They can be deployed into places that are too dangerous for humans to enter, and replace the damaged sensors in the building. However, this requires an initial setup time which might be not acceptable in an on-site ER scenario.

Compared to other technologies that have been applied in ERS, using WSNs to collect data has great potential in facilitating the ER process. This is because:

- It is accurate, quick and can operate in harsh environments.

Data collected by humans through human sensing is unreliable and time-consuming, whereas data collected from WSN is reliable and quick. Sensors can operate in tough environments where humans are not able to be, and they can still operate when the main power is cut due to emergency. Therefore, the environment inside of the incidents can still be observed. This data is valuable to the decision makers to make a better response plan.

- It is of low cost and low energy consumption.

Sensors are designed to be battery powered, and have a long lifetime. As the technology develops, sensors are becoming smaller and less energy consuming as well. The second generation of Berkeley Motes was only approximately 2mm by 2.5mm in size (Pister et al., 2001). It can be predicted that such a high-integrated sensor node will become of low cost in the near future.

General recommendation for system design

- Satisfy information needs

Good decision-making relies on the information available and the ability of decision makers to cope with the demands imposed upon them by the management of an emergency response situation (Danielsson, 1998). At an emergency scene, decision making is at a very high level, ICs typically handle many decision points under high time pressure. This requires information gathering to be quick, accurate and relevant. Hence, it can be recommended that an ERS should provide the right information at the right time to the right person in the right format.

- Address the dynamic nature of ER

Emergency scenes are ever-changing environments. Working under such conditions, ICs are required to carry out a Dynamic Risk Assessment. Researchers have also suggested that flexible and scalable system architectures would be suitable for ER. Thus, it can be recommended that an ERS should address the dynamic nature of ER to accommodate different requirements presented by changing risks in the environment and changing conditions of the technical system.

Chapter 3. Literature review: Information Extraction from Sensor Data

	Research questions addressed in this chapter:
1	What are the existing concepts, theory and research about information extraction from sensor data?
2	What are the challenges in research and development?
3	What has been considered as "information" in the existing research?
4	What are the general performance requirements for sensor data processing?

It was revealed in Chapter 2 that providing information needs of Incident Commanders (ICs) is important for the overall performance of on-site Emergency Response (ER), and Wireless Sensor Networks (WSNs) have great potential in providing critical information from inside of buildings for ER. WSNs can be deployed to monitor natural or man-made environments and detect emergencies with minimum attention and maintenance. During the period of monitoring, a large amount of data can be collected. However, "it remains a major challenge to make sense of the collected data, i.e., to extract the relevant knowledge from the raw data." (Römer, 2008) This chapter will review the process and techniques of Information Extraction from Sensor Data (IESD).

3.1 Information Extraction from Sensor Data

Information extraction focuses on the automatic process of obtaining information, which can be described as "knowledge", from raw data. Tan (2006) classified the process of information extraction from sensor data into three stages: data pre-processing, data mining, and data post-processing.

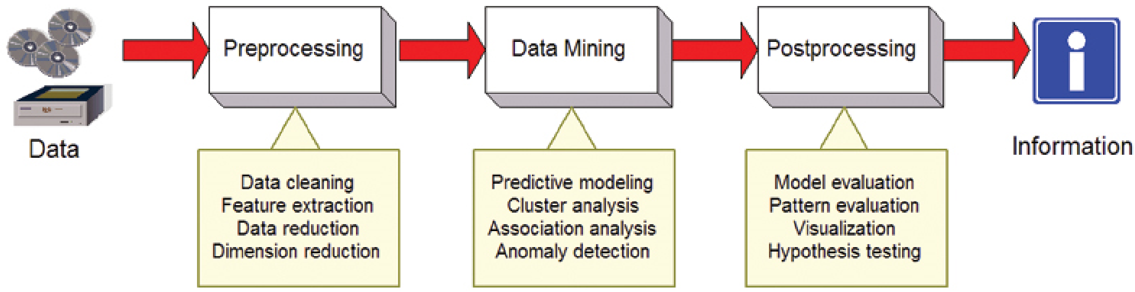


Figure 3-1: The overall process of information extraction from data (Tan, 2006)

3.2 Sensor Data Pre-Processing

The data quality problem in sensor networks is an issue that has been receiving increasing interest recently. Sensor data often contains noise (Elnahrawy and Nath, 2003), outliers (Basu and Meckesheimer, 2007), and missing values (Davidson and Ravi, 2005). The causes of such data quality problems include 1) sensors' internal errors, 2) the harsh environment in which sensors are deployed, and 3) damage or loss during wireless transmission, as shown in Figure 3-2.

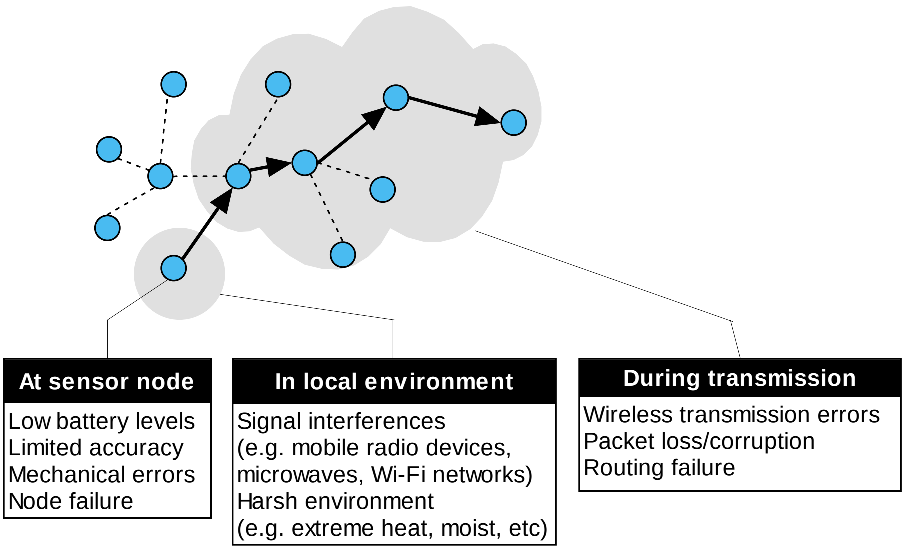


Figure 3-2: Possible causes of data quality problems in sensor networks

Data pre-processing includes data cleaning, outlier detection, missing values recovery, data reduction, dimension reduction, and data prediction, etc. It is a crucial step in the process of IESD. In applications that sensor data is used to picture the real situation and make crucial decisions, the noisy or miss-leading data may result in impractical or even harmful decisions. In addition, many traditional data mining methods do not have good tolerance with noisy or missing values. Therefore, data pre-processing is necessary for improving the quality of data before applying data mining technologies.

In this section, three main branches of data pre-processing are reviewed in more details: sensor data cleaning, missing values recovery from sensor data, and sensor data reduction.

3.2.1 Sensor data cleaning

A variety of filtering methods has been proposed to implement data cleaning in sensor network, including Bayesian Theory (Elnahrawy and Nath, 2003), Neural-fuzzy Systems (Petrosino and Staiano, 2007), Wavelet Transform (Zhuang and Chen, 2006), Kalman Filter (Tan et al., 2005) and Weighted Moving Average (Zhuang et al., 2007).

Jeffery et al. (2006) proposed a pipeline framework for sensor data cleaning, which contains five sequential processing stages: point (filter out obvious outliers from individual sensor readings), smooth (aggregate sensor readings within a temporal granule), merge (aggregate sensor readings within a spatial granule), arbitrate (remove conflictions or duplicates between different spatial granules) and virtualize (combine readings from different types of devices or data sources). This framework covered the possible steps involved in sensor data cleaning in general. It can be fully or partially implemented according to specific scenarios. However, the detailed cleaning tasks were proposed to be achieved by simple smoothing within temporal and spatial windows.

Elnahrawy and Nath (2003) proposed a more sophisticated data cleaning method based on Bayesian theory. The method takes the noisy sensor data (the observation *o*), error model for a sensor (assumed to be normally distributed with zero mean and a known standard deviation $\sim N(0,\delta^2)$ and prior knowledge (true reading distribution *p(t)* and conditional probability *p(o|t)*) as input, and outputs the uncertainty model of sensor readings (the posterior probability of *t* given *o*, denoted as *p(t|o)*) using the Bayes' Theorem $p(t|o)=\frac{p(o|t)p(t)}{p(o)}$. With the assumption that the reading of a specific sensor follows a Gaussian distribution, and using some properties of Gaussian distribution, it can be concluded that the posterior probability *p(t|o)* also follows

Gaussian distribution. The mean and standard deviation of *p(t|o)* can be calculated from those of *p(t)* and those of the error model. The result of the data cleaning in general was a compromise between the prior knowledge and the observed noisy sensor data.

Petrosino and Staiano (2007) presented a regression based sensor network data cleaning method. This method proposed to model each sensor's behaviour using a type of regression (Neural-fuzzy models were adopted for regression in their experiments), and estimate sensor reading according to the regression. If the difference between the estimated sensor readings ($\hat{y}$) and the newly-arrived sensor readings (y) are bigger than the estimated error, approximated by the Root Mean Square Error (RMSE), on the training sample, the readings are considered unreliable. Unreliable readings will be replaced by the model estimates plus or minus the RMSE, whereas reliable readings will be kept. In this way the cleaned readings will fall between [$\hat{y}$-RMSE, $\hat{y}$+RMSE].

Wavelet transform is another method proposed to clean sensor data. Zhuang and Chen (2006) argued that wavelet transform can reduce noise in time series data generated by sensors. In addition, they stated that by transmitting wavelet coefficients rather than raw sensing series, data traffic can be considerably reduced.

Tan et al. (2005) implemented Kalman Filter and Linear Regression in the sensor cleaning toolkit designed. Their experimental results demonstrated that Kalman Filter has better filtering performance as well as estimation closer to the real trend than Linear Regression. Although it showed promising results, Linear Regression worked poorly for data with high variability according to their experiments. Compared to the Bayesian theory based method and Neural-fuzzy method, Kalman Filter is also more lightly weighted, since it does not require the training process that the other two methods do in order to learn the model parameters before they can be used.

Zhuang et al. (2007) proposed a smart weighted moving average based sensor data cleaning approach, which consists of three steps.

- Step 1: "locate important values by range prediction"
- Step 2: "gain confidence for important values through sensor testing and neighbour testing at individual sensors"

- Step 3: "perform weighted moving average at the sink"

This approach used Kalman Filter and Linear Regression for range prediction. Values outside the predicted range would be considered as the "important" values, and their confidences would be calculated in Step 2. Finally, the weighted moving average at the sink combines the temporal average and the spatial average together. The simulation demonstrated a better performance than the simple moving average algorithm. However, in their simulation, the performance of Kalman Filter and Linear Regression was almost the same.

The characteristics of the existing sensor data cleaning methods can be briefly summarized in Table 3-1. Kalman Filter has high flexibility with low implementation complexity.

Table 3-1: The characteristics of the existing sensor data cleaning methods

	A priori conditions	Complexity	Flexibility
Bayesian theory	Training process to learn the model parameters	high	high
Neural-Fuzzy Systems	Training process to learn the model parameters	high	high
Wavelet Transform	Data samples large enough to perform down sampling	high	high
Kalman Filter	Parameter tuning	low	high
Weighted Moving Average	Weight assignment model	low	low

3.2.2 Missing values recovery from sensor data

Davidson and Ravi (2005) tested the packet loss with a Berkeley mote network in an indoor environment. They found that approximately 3% of packets were lost over 3 hours of testing time when the motes were only 10 feet away from the base station with no walls or structures in between, and the packet loss increased to 23% when the same experiment was conducted with motes placed in a separate room from the base station. Their experiments proved that missing values is a problem seriously affecting the quality of sensor data.

To handle the problem of network packet loss, the traditional way is to wait for a predefined period of time before the receiver sends a retransmission request to the sender, or the sender automatically retransmits if no acknowledgement has been received from the receiver. However, there are two major drawbacks of applying this approach in sensor networks: 1) "increased power consumptions by the sensors (they should listen for requests and resend data if needed)", and 2) "increased latency of the produced result by the query (time spent for transmitting a request and waiting for a response)" (Halatchev and Gruenwald, 2005).

Therefore, the existing research on handling missing sensor data focused on estimating or recovering missing values using available values from sensors related to the sensors of missing values.

A variety of estimation methods were proposed, such as Expectation Maximization (Davidson and Ravi, 2005), Association Rules (Halatchev and Gruenwald, 2005), Belief Propagation (Chu et al., 2005).

Davidson and Ravi (2005) stated that in order to estimate missing values, Expectation Maximization (EM) is a common approach to converge to local maxima of the complete data likelihood (i.e., the likelihood of both the observed and missing data). The E-step calculates the expectation (or possibilities) of the missing nodes values ($P(Y \mid X, \theta)$), where X represents the observed data and Y represents the missing values. Given the expectations for the missing values, the M-step calculates the value for θ that maximizes the expected complete data likelihood. However, they argued that the typical implementation of EM parametric is not suitable for distributed deployment in sensor networks, mainly for two reasons:

- sensor nodes do not have sufficient memory to implement either the E-step or the M-step with parametric models.
- the M-step requires collecting and transmitting the expected values for missing data to the base station for aggregation, and it may take iterations to converge, which causes overwhelming energy consumption at sensor nodes.

Therefore, they proposed a non-parametric EM method that minimizes power consumption and computation at sensor nodes. The method maps sensor nodes and its neighbours into an undirected graph. Then the EM method will interactively fill in the expected missing values and add/remove nodes in the graph to maximize the probability. The resulting graph contains the estimated missing sensor values. It was claimed that their method allows EM method to be deployed in sensor nodes.

Halatchev and Gruenwald (2005) addressed the energy consumption problem of applying traditional association rule mining in sensor networks, and proposed the following modifications to solve the problem:

- “instead of generating all association rules between sensors, generate all the association rules between pairs of sensors only”
- “use additional data structures to reduce the number of memory access operations”
- “evaluate frequent item sets and association rules between pairs of sensors always with respect to a particular state of sensors”
- “use the sliding window concept in the data structures that store the data arriving from the sensors and in the additional data structures”

Simulation results showed that the proposed method requires more memory space and takes longer to produce an estimation than the considered alternative approaches, but it achieves better accuracy of the estimated value than the alternative approaches do.

Chu et al. (2005) addressed the strong dependencies between sensor readings, and proposed a way to recover missing values by modelling data dependencies with Markov networks. Belief propagation is used to efficiently compute the marginal or maximum posterior probabilities, so as to infer missing values or to correct errors.

In summary, the challenges presented to applying traditional methods in sensor data include limited memory, limited energy resources at individual sensor node, and strong dependencies between sensor nodes.

3.2.3 Sensor data reduction

“Having a large amount of redundant data may slow down or confuse the knowledge discovery process” (Han and Kamber, 2006). In-network aggregation of redundant data can reduce the total data flow over the sensor network, thus it can extract the most representative data using the minimum resources (Akcan and Brönnimann, 2007), and effectively reduce power consumption (Santini and Römer, 2006). Therefore, a branch of sensor data pre-processing research focused on sensor data reduction in WSNs.

Akcan and Brönnimann (2007) proposed a weighted in-network sampling algorithm to obtain a deterministic (much smaller but representative) sample instead of raw redundant data. Compared with random sampling, the advantage of weighted sampling algorithm is "it can guarantee that each node's data has the same chance to belong to the final sample, independent from its provenance in the network". The limitation of this algorithm is that it was designed for arbitrary network topologies. The final sample may lose its representativeness in the case of loss of connection in the aggregation tree structure, because an entire sub-tree may no longer contribute to the final sample due to a link or node failure.

Instead of selectively sampling the network nodes, Santini and Römer (2006) proposed a prediction-based data reduction strategy. It is to have prediction methods deployed both at sensor-level and base station-level, so that sensors only need to send data that deviates from the prediction. Compared to the existing techniques under the same strategy, their proposed Least-Mean-Square (LMS) adaptive algorithm claimed to be light-weight, because it enables sensor nodes to predict expected values without a-priori knowledge about statistical properties of the observed phenomena. Their experiments demonstrated the algorithm's effectiveness on reducing communication cost as well as ensuring the accuracy of reconstructed original data.

3.3 Sensor Data Mining

Data mining aims to extract patterns from data. Traditional data mining technologies include Decision Trees, Rule-based Classifiers, Artificial Neural Networks, Nearest Neighbour, Naive Bayes, Support Vector Machines, Logistic regression, etc. Most of them were initially developed to be applied in central data warehouse. Based on a different dataset, the recent research on mining sensor data mainly focused on distributed in-network data mining.

3.3.1 From traditional data mining to sensor data mining

Researchers have applied some traditional data mining technologies in mining sensor data. For instance, Kulakov and Davcev (2005) demonstrated how a popular artificial neural networks algorithm called Adaptive Resonance Theory (ART) model can be adapted in the field of WSNs. Other examples include distributed Bayes algorithm

(Krishnamachari and Iyengar, 2004), and Hill Climbing algorithm (Krivitski et al., 2007).

Researchers have demonstrated how traditional data mining technologies can be integrated within a sensor network structure. Bontempi and Borgne (2005) suggested that the architecture of a sensor network must take into account not only the energy and transmission issues but also criteria related to the accuracy and quality of the data mining task. This general recommendation means that the organization of the same sensor network may change according to the type of data mining task and the required quality precision, or the other way around, different sensor data mining tasks may require different sensor network structures.

Most people suggested a form of hierarchical network topology for their proposed sensor data mining. Bontempi and Borgne (2005) proposed a two-level architecture for sensor data mining, as shown in Figure 3-3.

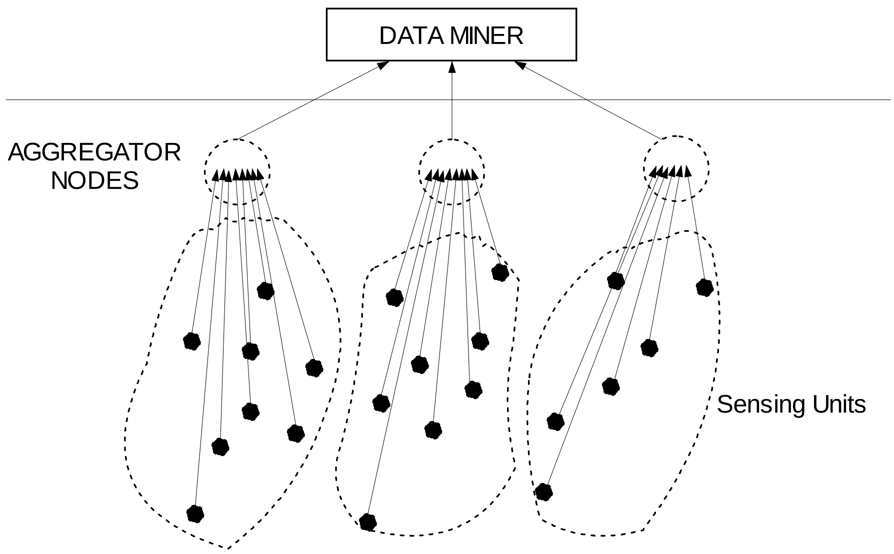


Figure 3-3: An adaptive modular architecture of sensor data mining (Bontempi and Borgne, 2005)

The lower level consists of aggregator nodes (dotted circles) that perform modular aggregation of the neighbouring sensing units (black dots). The aggregated signals are then sent up to the upper level data miner, where the required sensing tasks (e.g. classification, regression or prediction) are performed. This architecture introduced a layer of aggregator nodes in the WSN topology, each of which acts as a cluster head of a number of sensor nodes.

Three network architectures were proposed by Kulakov and Davcev (2005) to incorporate ART1 and Fuzzy ART artificial neural network algorithms into sensor networks to perform auto classification tasks. ART1 was designed as an algorithm for unsupervised learning of binary input patterns, whereas Fuzzy ART is an analog version of the ART1 algorithm. The first architecture used one cluster head to collect all sensor data from its cluster of units. And the Fuzzy ART model is only implemented in the cluster head. The second architecture employed redundant cluster heads collecting data at different levels of details. Fuzzy ART model is implemented at all sensing units, each of them classifies and represents over the same data at a different level of detail. The third architecture consisted of a top layer cluster head (ART1 implemented) to collect and classify the data after they are once classified at the lower level sensing units (Fuzzy ART implemented). It can be expanded into hierarchical cascades of neural-network classifiers implemented in units of a sensor network, as shown in Figure 3-4.

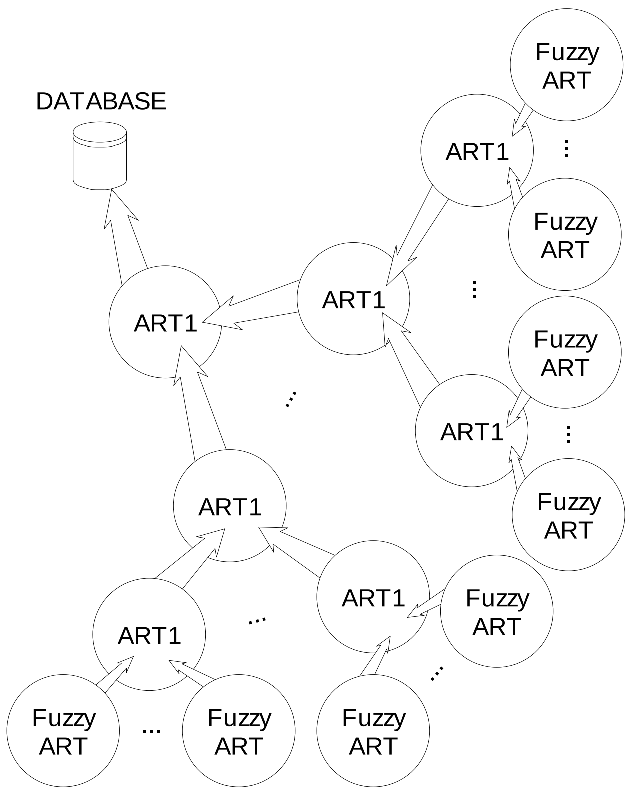


Figure 3-4: Hierarchical cascades of ART neural-network classifiers implemented in units of a sensor network (Kulakov and Davcev, 2005)

Compared to Bontempi and Borgne's architecture, Kulakov and Davcev's architecture demonstrates the same idea of using a clustered network topology. The difference is about what data mining algorithm is implemented in each cluster head, and how many levels of cluster hierarchies there are in the network. Bontempi and Borgne's architecture consists of two hierarchies: the lower level performs aggregation, the upper level performs prediction. Kulakov and Davcev's architecture can consist of cascades of hierarchies, each level performs classification.

McConnell and Skillicorn (2005) suggested distributed prediction based on voting. In their framework, the sensor networks are configured in a tree structure, where the root is a more capable computation device and the leaves are sensors with limited resources and computation capability. Local predictors are deployed in each sensor node, where local information is combined to generate the target class prediction. Then the root will choose a global prediction by voting from all the received local predictions.

Krivitski et al. (2008) demonstrated another research effort in applying existing data mining algorithms in a local distributed manner. They modified the traditional hill-climbing algorithm to find a locally optimized location of *k* facilities in sensor networks.

3.3.2 What constitutes a suitable sensor data mining algorithm

In terms of what constitutes a suitable sensor data mining algorithm, the following characteristics can be extracted from the existing sensor data mining algorithms.

- Distributed

Kulakov and Davcev (2005) argued that sensor networks require "simple parallel distributed computation, distributed storage, data robustness and auto-classification of sensor readings". McConnell and Skillicorn (2005) stated that as sensors become active devices with their own processing capability, distributed algorithms in sensor networks become a new possibility. Bontempi and Borgne (2005) argued that the rationale behind distributed in-network data mining is that in-network aggregation can have two benefits: saving communication cost as well as reducing data dimensionality. Krivitski et al. (2008) also stated that in-network local algorithms have superb message pruning capabilities, thus they are "better than centralized algorithm both in terms of message efficiency and of convergence time".

- Hierarchical topology based

On top of the argument that sensor data mining should be distributed in-network, the topology of the sensor network must be considered with data mining techniques. The organization of sensors into hierarchies of clusters has been widely supported. Kulakov and Davcev (2005) stated that limited communication band width, limited computing resources, limited power supply and the need for fault-tolerance are typical constraints for data mining in sensor networks. By organizing sensors into clusters, dimension reduction can be achieved, thus the communication costs of clustering-based algorithms are significantly smaller. This idea of implementing hierarchical topology in sensor networks for data mining tasks has been supported by a number of researchers (e.g. Bontempi and Borgne, 2005; McConnell and Skillicorn, 2005; Krivitski et al., 2007).

- Data-driven/event-driven/service-driven rather than synchronization

Kulakov and Davcev (2005) noted that "all previous work on distributed clustering assume tight cooperation and synchronization between the processors containing the data and a central processor that collects the sufficient statistics needed in each step of the hill-climbing heuristic". However, they argued that such synchronization controlled by a central point may not be suitable in sensor networks. Instead, they believed that data-driven is one of the most important features which qualify an algorithm for sensor networks.

Bontempi and Borgne (2005) used the Lazy Learning algorithm in the upper level prediction of their two-level modular adaptive architecture. They claimed that such an algorithm assumes no a priori knowledge on the process underlying the data, only driven by "available information represented by a finite set of input/output observations", which makes it appealing in the sensor network context.

Krivitski et al. (2008) also stated that the local majority voting algorithm they used has good performance in terms of message load and convergence time because it is event-driven and requires no form of synchronization.

Silberstein et al. (2007) proposed data-driven processing in sensor network. Rezgui and Eltoweissy (2007) introduced service-driven query routing.

In summary, the concept of data-driven/event-driven/service-driven rather than synchronization has been widely supported by research in sensor data mining.

3.3.3 Sensor data mining applications

Sensor data mining algorithms have been proposed to accomplish typical data mining tasks such as prediction (Bontempi and Borgne, 2005), classification (Kulakov and Davcev, 2005), and optimized location deployment (Krivitski et al., 2007). The patterns extracted from data for these mining tasks are for example prediction model, classification, and location for facilities.

There are research efforts that use data mining algorithms for a specific application area, e.g. fault detection (Krishnamachari and Iyengar, 2004), anomalies detection (Palpanas et al., 2003; Subramaniam et al., 2006). However, they are based on assumed scenarios rather than real world application requirements.

Therefore, it can be argued that there is a lack of research on how the mined patterns can benefit real world applications. Or in other words, there is a gap between such mined “information” and the information needs of real world applications.

3.4 Sensor Data Post-processing

Data post-processing includes pattern evaluation, model evaluation, data visualization/presentation, etc. This step can link the result of sensor data mining to specific applications, or in generic research, this step is often integrated with sensor data mining.

3.4.1 Pattern evaluation

Several works described pattern evaluation for a specific application.

Kamphuis et al. (2008) presented the process of using sensor data patterns to classify abnormal and normal milk. Three categories of pattern descriptor were proposed: level, variability and shape, each of which includes a number of pattern descriptors that can be used to analyse sensor data patterns. The results have been used to classify abnormal milk from the normal.

Heierman and Cook (2003) proposed a data mining technique that discovers regularly occurring device usage patterns from sensing human interactions with home appliances (e.g. light on, light off, video on, video off, etc.). They suggested the possibility of using such patterns to automate human interaction with home appliances.

The evaluation of the mined device usage patterns (candidate episodes) were analysed according to their significance, determined by the potential amount of compression on the length of sensor data sequence that a pattern provides. The most significant candidates would be suggested as the targeted objects for home automation.

There is also generic research on pattern evaluation. An algorithm named FP-mine (FP: Frequent Pattern) was developed by Cheng and Ren (2007) to discover frequent moving patterns based on the idea of pattern growth (FP-growth algorithm) and pattern storage structure (P-tree). The pattern evaluation includes determining the frequencies of the mined patterns.

3.4.2 Data visualization

Data visualization can be based on computer graphics, statistical methods, and user interaction techniques.

Koo et al. (2006) created a software in order to analyse the multi-sensor data for gas transmission pipeline inspection, in which heterogeneous sensor data was displayed on a virtual 3D pipeline generated to help users get a realistic view when performing pipeline inspection and facilitate rapid and precise decision.

Pattath et al. (2006) presented an interactive visual analytic system using a PDA to visualize network and sensor data from Purdue’s Ross-Ade Stadium during football games. Attendees can be monitored by mobile devices enabling the detection of crowd movement and event activity and insightful information can be provided to network monitoring personnel, safety personnel and analysts.

Sparchholz et al. (2005) considered laser scanners and cameras as sensors deployed in a castle. An approach creating a three dimensional virtual world by utilizing high resolution multi-sensor data was presented by them.

3.5 Sensor Data Fusion

Contrary to information extraction that focuses on “the organized process of identifying valid, novel, useful and understandable patterns from large and complex data sets” (Maimon and Rokach, 2010), Data Fusion focuses on the combination of multiple sources, the result of which can be low-level improved raw data or high-level information.

Hall and Llinas (1997) reviewed the earliest and probably most well-known model of data fusion, JDL data fusion process model, proposed by the Data Fusion Sub-Panel of the Joint Directors of Laboratories (JDL), established by US Department of Defence (DoD). The JDL model classified data fusion process into 4 levels: object refinement, situation refinement, threat refinement, and process refinement. The layout of the JDL framework is demonstrated in Figure 3-5.

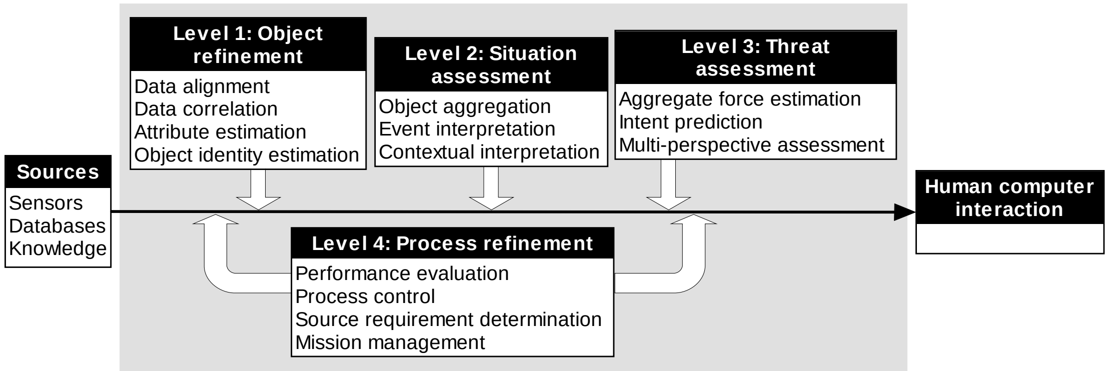


Figure 3-5: The JDL data fusion framework

Hall and Llinas (1997) also summarized the example techniques used in data fusion, which include probabilistic association, Kalman Filter, Neural Networks, Fuzzy logic, cluster algorithms, pattern recognition. They are similar to the example techniques reviewed earlier in section 3.1 for the process of information extraction from sensor data. Hence, there is an overlap between information extraction and Data Fusion.

In the context of sensor data processing, information extraction from sensor data is typically based on combinations of data from multiple types of sensors, considered as multiple data sources. Therefore, the term Sensor Data Fusion is used interchangeably with Information Extraction from Sensor Data in this thesis.

3.6 Summary

This chapter reviewed the process of IESD and the techniques involved in the process.

In summary, the process of IESD consists of three stages: data pre-processing, data mining and data post-processing. The aim of data pre-processing is to deal with data quality problems such as noise, outliers and missing values. Data mining aims to find patterns from data, which can be interpreted as “knowledge”. Data post-processing include pattern evaluation, visualization, etc. Post-processing of sensor data can be bound together with data mining to find patterns by analysing the result of data mining algorithms, or act as a separate function module. Information extraction from sensor data streams usually requires integration of multiple types of sensors, or integration of sensor data with images or other data sources. The concept sensor data fusion is used to describe such integration of data from multiple data sources. The two concepts are used interchangeably in the context of this thesis.

A summary of representative example techniques used in different stages of information extraction from sensor data is shown in Table 3-2.

Table 3-2: Example techniques in information extraction from sensor data

IESD process	Process function	Example techniques
Stage1: Pre-processing	Sensor data cleaning	• Bayesian Theory (Elnahrawy and Nath, 2003) • Neural-fuzzy Systems (Petrosino and Staiano, 2007) • Wavelet Transform (Zhuang and Chen, 2006) • Kalman Filter (Tan et al., 2005) • Weighted Moving Average (Zhuang et al., 2007)
	Missing values recovery	• Expection Maximization (Davidson and Ravi, 2005) • Association Rules (Halatchev and Gruenwald, 2005) • Belief Propagation (Chu et al., 2005)
	Sensor data reduction	• Deterministic sampling (Akcan and Brönnimann, 2007) • Prediction-based reduction (Santini and Römer, 2006)
Stage 2: Mining	Classification	• Artificial Neural Network (Kulakov and Davcev, 2005) • Bayes algorithm (Krishnamachari and Iyengar, 2004)
	Prediction	• Voting algorithm (McConnell and Skillicorn, 2005) • Lazy Learning algorithm (Bontempi and Borgne, 2005)

	Solution optimization	• Hill Climbing algorithm (Krivitski et al., 2007)
Stage 3: Post-processing	Pattern evaluation	• Frequent pattern mining (Cheng and Ren, 2007) • Pattern descriptors (Kamphuis et al., 2008) • Episode Discovery algorithm (Heierman and Cook, 2003)
	Data visualization	• Computer Graphics (Koo et al., 2006) • User interaction (Pattath et al., 2006)

The lessons learned from the literature review are as follows:

Lesson 1: It remains a challenge to make sense of the collected sensor data.

It has been addressed in the literature that sensor data often contains noise, outliers, and missing values, which presented challenges for sensor data fusion. As a result, sensor data pre-processing techniques are necessary to solve the data quality problems associated with data collected from WSNs.

A variety of data mining techniques have also been proposed to find patterns from sensor data. Typical data mining tasks are such as classification, prediction, location optimization. The results of these data mining tasks are extracted prediction models, classification, or location for facilities. However, these patterns, although considered as "knowledge" from a technology perspective, are not the knowledge that users require to facilitate their tasks.

There is some research that describes pattern evaluation for specific application areas. However, they are based on assumed scenarios rather than requirements generated from real applications. Therefore, it can be argued that there is a need to build the link between what has been considered as "information" in the existing research and the required information from a real application.

Therefore, although WSNs can be deployed as a data collection method for many different applications, as Römer (2008) stated, "it remains a major challenge to make sense of the collected data, i.e., to extract the relevant knowledge from the raw data."

Lesson 2: What constitutes suitable sensor data processing techniques?

Interesting arguments have been found in the literature about what constitute suitable sensor data processing techniques.

- Online or offline?

Online data processing means to process as data arrives. It is based on the understanding that sensor networks produce data streams rather than static data (Elnahrawy and Nath, 2003; Jeffery et al., 2006). Sensor data is streaming data, therefore, offline data processing, usually used by conventional data warehouses in the form of a separate database function, is considered not suitable for sensor data processing.

- Deployment at sensor-level or base station-level?

Some researchers believed that data processing should be performed at sensor level. Sensor-level processing preserves the behaviour of each sensor. Elnahrawy and Nath (2003) argued that “the reading of each individual sensor is important”, therefore, their sensor data processing functionality works on every single sensor. Sensor-level processing saves energy consumption on transmitting outliers. For instance, Zhuang and Chen (2006) argued that outlier cleaning must be done in-network to accommodate the limited battery power and costly data transmission in sensor networks.

On the other hand, base station-level sensor data processing can include in-network aggregation of readings from a set of sensors to reduce the effect of noise (Yao and Gehrke, 2002). Yao and Gehrke (2002) argued that base-station level processing with in-network aggregation can save energy consumption of transmitting data from every individual sensor, thus increase network lifetime.

It can be argued that sensor data processing functionality that is flexible and can be performed at both sensor level and base station level would be desirable.

- Distributed or centralized?

Some researchers argued that sensor networks require distributed data mining. This argument is based on two understandings. 1) As sensors become active devices with their own processing capability, distributed algorithms in sensor networks become possible (McConnell and Skillicorn, 2005). 2) Distributed sensor data mining can save communication cost as well as reduce data dimensionality (Bontempi and Borgne, 2005; Krivitski et al., 2007). Therefore, instead of transmitting raw data to the base, the data is pre-processed or integrated using local models, and only the resulting coefficients will be transmitted.

Others argued that "performing cleaning at the sensor, query processing at the database level has no advantages" (Elnahrawy and Nath, 2003), because data pre-processing models developed for individual sensors usually require data communication between the sensor neighbours to generate the result, which could result in more data communication than only transmitting individual sensor data to the base station. In addition, it increases storage and computation cost at sensor level.

Consequently, the advantages and disadvantages of distributed and centralized data mining of sensor data is summarized in Table 3-3. Whether to use distributed or centralized data mining can be chosen according to application needs.

Table 3-3: Comparison of distributed and centralized sensor data mining

Type of data mining	Advantages	Disadvantages
Distributed	• Reduce network congestion • Maintain certain level of data redundancy to improve reliability • Reduce energy cost on wireless transmitting sensor data	• Increase individual sensor node cost • Increase sensor node complexity due to the need of complicated data management and processing under resource and power constraints • Potential processing overhead on sensor node
Centralized	• Less resource constraints on data mining algorithm • Better efficiency, accuracy and quality of data processing • Timely user query processing	• Have to effectively maintain large amount of raw data • Potential performance bottle neck

- Data-driven/event-driven or synchronization?

Previous works on distributed network computing were argued to be based on synchronization (Kulakov and Davcev, 2005). However, synchronization controlled by a central point may not be suitable in sensor networks. Instead, data-driven was argued to be one of the most important features which qualify an algorithm for sensor networks. And the concept of data-driven/event-driven/service-driven rather than synchronization has been widely supported by recent research in sensor data mining (Bontempi and Borgne, 2005; Krivitski et al., 2007; Silberstein et al., 2007).

Lesson 3: General performance requirements for sensor data fusion approaches

A number of performance evaluation metrics were used by recent research, based on considerations of effectiveness and cost.

Effectiveness has been measured by absolute error such as Normalized Mean Square Error (NMSE) (Bontempi and Borgne, 2005; Chu et al., 2005; Zhuang et al., 2007). It has also been measured by relative metrics, for example, precision and recall (Chu et al., 2005; Elnahrawy and Nath, 2003; Zhuang and Chen, 2006), percentage improvement (Cubica and Moore, 2003; Zhuang and Chen, 2006).

Cost has been measured communication-wise and time-wise. Examples of communication cost measurement include transmission bytes (Zhuang et al., 2007) and message load (Krivitski et al., 2007). Examples of time cost has been measured by convergence time (Krivitski et al., 2007).

The general performance requirements for sensor data fusion approaches are high in effectiveness, low in cost.

Chapter 4. Study of Incident Commanders' Requirements

	Research questions addressed in this chapter:
1	What are the goals and priorities of goals of the Incident Commanders?
2	What are the tasks required to achieve those goals and decisions they need to make during the tasks?
3	What information is required and what information is not required by the Incident Commanders in order to best facilitate their decision making?
4	What are the current means of obtaining the required information?
5	Are there any gaps in retrieving the required information?
6	What technological solutions can be proposed to fill in the identified gaps?

The literature review in Chapter 2 revealed the importance of information to Incident Commander's decision making and the lack of information on-site. However, how the required information is gathered by Incident Commanders (ICs) in real practice, whether there are gaps in getting the required information at the right time in the right format is not clear. The literature review also revealed that Wireless Sensor Network (WSN) has great potential in facilitating Emergency Response (ER). However, not much research described how the information extracted from sensor data can contribute to ER. Therefore, it was necessary to conduct a study of ICs' requirements.

This chapter described the study of ICs' requirements, with the aim of identifying opportunities for WSN to facilitate fire Emergency Response, and determining the focus of further technical work. The analysis and findings in this chapter were partly based on the author's data collection from the visits to Derbyshire and Leicestershire Fire and Rescue Services and 2 interviews with ICs, and more based on interviews and data collection done by Raj Prasanna at Business School in Loughborough University. The pictures included in the chapter were produced by and used here with the permission from Raj Prasanna. This chapter was based on the author's published work (Yang et al., 2010b)

4.1 Introduction

As specified in Chapter 2, ICs play an extremely important role in responding to emergencies (Bigley and Roberts, 2001; Jiang et al., 2004a), thus the targeted user was identified to be ICs in the first responders group.

Although different types of incidents have their unique characteristics, the responses to different types of incidents share common goals, phases, commanding hierarchies, etc. Therefore, it can be argued that lessons learned from studying the response to one type of incident can be applied to others.

Fire is one of the disasters that can occur most frequently and cause serious damage and loss of lives, including the lives of first responders. According to the statistics published by the UK Home Office, every year there were more than 40,000 accidental house fires in England, resulting in an average of 285 deaths and 9,000 burn injuries. (Yang and Frederick, 2006) Therefore, studying fire ICs' needs and opportunities to provide technology support to facilitate their responses to emergencies would have great benefits.

The literature covered common goals of all emergency response agencies, and goals of ICs from the perspective of specific tasks. The procedures of incident command were described in detail in Incident Command System, so the role and responsibilities of IC could be comprehended. However, how the required information is gathered and utilised by ICs to achieve their operational goals was not clear in the literature.

After the literature reviews, the plan for future technical work was initiated. Since emergency navigation was considered to be a helpful research topic, the initial focus of my research was to be on producing incident seriousness distribution in the building for situation awareness and fire fighting goals, and producing safe exit routes calculation for fire-fighters deployed in the building to better protect them during the completion of their rescuing tasks. Whether this technology support would be helpful to ICs or not needed to be verified.

Therefore, visits to Derbyshire Fire and Rescue Service were organized to further study the requirements of ICs and how technology can provide support. In this chapter, the requirement analysis study was described in detail and findings were analysed.

4.2 Aims and Objectives

The purposes of the visits were:

- To provide a comprehensive understanding of fire ICs' goals, tasks, and information needs
- To identify the link between the information requirements and the opportunities for WSN for facilitating fire ER

The objectives are:

- To understand the current practise of fire Emergency Response in the UK
- To confirm the goals and priorities of ICs described in the literature
- To understand the information requirements of fire ICs
- To understand the current technology aids available to fire ICs
- To identify gaps/opportunities for technology to support ICs

4.3 Methods

A number of existing task analysis methods were analysed in literature review to find the most appropriate one for establishing the link between users and technology capabilities. As a result, the goal-driven task analysis methodology (Albers, 1998) was adopted to determine ICs' goals, tasks, and information needs during first response. The goal-driven methodology was further expanded to study the underlying technology opportunities to support the information needs. The expanded goal-driven methodology aimed to relate goals, tasks/decisions, information needs and technology opportunities.

Visits to the Derbyshire fire and rescue service were organized, and the ICs in the fire crews were chosen as the subjects of the case study of ICs' requirements. The expanded goal-driven methodology was applied throughout the study.

A number of investigation methods, both formal and informal, were employed to understand the current practice of fire ER and determine fire ICs' requirements. These included visiting the control centre, interviewing senior officers, observing fire exercise, analysing obtained training videos, and emailing contacts to follow up queries.

The first phase of visits focused on understanding the current theory and practice of fire ER. Starting from visiting the control centre, where all the emergency calls are handled, and where the response to fire emergency begins, field visits included the demonstration of national FireControl project at regional fire and rescue service, and fire exercises at commercial buildings such as Westfield shopping centre in Derby. Although direct visits to training centres were not possible due to schedule and transport limitations, videos of the training were obtained, and the research student who visited the training centre was interviewed to better understand what the training involved.

The second phase of visits narrowed down the focus to ICs' requirements and potential opportunities for the use of technology. Two senior officers were interviewed to understand their roles as IC in practice. Each interview consisted of a brief introduction to the purposes of the interview, followed by open-ended questioning, and discussion of key points. The questions can be classified into 4 sections: 1) the goals and priorities of IC, 2) the role as an IC and current practice, 3) information needed to achieve ICs' goals, 4) feedback about how technology can help ICs. Each interview took 1.5 hours, followed by informal contact to obtain feedback on additional proposals to use technological assistance and to resolve any queries that had arisen.

The interviewees were selected for:

- their expertise in this area of fire ER

- their many years of experience working in different roles (including IC) in the fire and rescue services, and in different regions

This maximized the validity of the data collected.

4.4 Analysis and Results

Using the expanded goal-driven task analysis method, the framework of the analysis is shown in Figure 4-1. The analysis started from identifying the fire ICs' goals, to understanding the tasks required to achieve those goals and decisions they need to make during the tasks, which in turn results in information needs. Then the current means of obtaining the information were evaluated. As a result, gaps were identified and technology opportunities to fill the gaps were proposed.

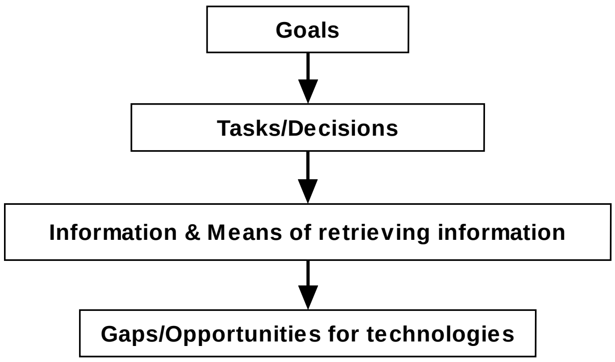


Figure 4-1: Framework of user requirement analysis

4.4.1 Goals and priorities

Data collection with fire ICs revealed that their most important goals are to ensure safety of crews, saving life (including casualties and the general public), and extinguishing the fire. In general, safety of crews is the highest priority of the three, then the safety of the public, then extinguishing the fire. The current basic principle that ICs work to is: they will risk life (of crew) if there is danger to life (of the public); otherwise risks taken will only depend on the dangers present, and they will not commit their crews into hazard situations for no benefit of life.

Working to this guideline, the goals and priorities may vary according to the particular circumstances of each incident instance. When an incident occurs, an incident commander will perform a "Dynamic Risk Assessment (DRA)", which identifies:

- the risks present
- the severity and possibility of the risks
- people at risk
- what control methods are necessary
- whether the risks are worth the benefits.

ICs review and confirm the initial risk assessment as soon as practicable and further review and confirm it at regular intervals throughout the incident.

Although it was mentioned in the literature that one of the ICs' goals was saving lives and protecting fire-fighters (Carter and Rausch, 2008), it was clearly apparent that safety of crews was put in such high priority in their practice and training: higher than the safety of the public, and higher than extinguishing the fire. One of the ICs interviewed said it was part of his responsibility to always try to make the job as safe as possible.

4.4.2 Decisions required to achieve the goals

DRA is the most important and continuous task that ICs need to do on-site. Based on the result of DRAs, ICs make decisions on

- *The Tactical mode* (defensive or offensive). Where safe systems of work are deployed and adequate control measures implemented, the mode of operation is likely to be 'offensive'. However, where the risk to crews is excessive, defensive mode will be declared.
- *Tasks to be adopted to minimize the risk* (e.g. whether or not to send crews into the building to undertake search and rescue, fire-fighting, or ventilation). If the risk is worth the potential benefits, ICs may command crews into the building, otherwise, they will not risk crews for no benefit of life.
- *Whether additional resources are required* (e.g. equipment/experts for a specific type of chemical hazard, additional crews to support fire-fighting,

other agencies). When the resources available are not sufficient to minimize the risks and perform the tasks, ICs will ask for additional resources.

This analysis was in line with the tasks and responsibilities of ICs as specified in the literature (Incident Command - 3rd edition, 2008).

4.4.3 Summary of information needed

To perform the task of DRA, ICs need information about the incident, hazards, water resources and personnel resources.

4.4.3.1 Nature of incident

Magnitude of the incident

The ICs usually perform an assessment of the incident on their arrival at the incident ground, to determine the potential magnitude of the incident. Properly assessing the potential magnitude of an incident provides a basis for implementation of their procedures. Examples of the information required were described by ICs: "what type of fire it is, whether there are chemicals involved, whether there is any chemical storage around that we need to be aware of, how the fire has started and developed, whether any actions have been taken to respond to the incident, etc. It is just like a forever on-going list….."

Whether it is life-involved

Whether it is life-involved is vital for ICs' decision making. As stated by one of the ICs interviewed, "we will not commit our crews into hazard situations for no benefit of life". The follow up action plans and commands to be employed can be completely different (offensive or defensive mode), depending on whether there are people trapped in the building. If it is life-involved, the IC needs to know the number of people involved, their locations, and the number and location of casualties if possible.

4.4.3.2 Hazards

During the assessment, the IC evaluates the hazards present and the *potential* hazards. They need to identify: what type of hazard there is (structural, chemical, etc.); who are in danger; the severity (injuries that the hazards cause) and likelihood of the hazard, and location of the hazard/potential hazard. The time of this evaluation must also be recorded. After assessing the hazards present and the potential impact of the incident, the IC will determine if additional resources and support are required. An IC described their DRA as "constantly asking: What do I need? What are the risks? Are the risks worth the benefits?"

4.4.3.3 Water resources

To determine whether the available water supply is adequate, the IC needs to know information about water resources nearby, including facilities designed for fire-fighting purposes (e.g. water pumps, inlets/outlets, hydrants), as well as other public or private water resources (e.g. lakes, rivers, swimming pools).

The information needed about fire-fighting facilities includes location and current condition (e.g. whether it is functioning). The information needed about other public or private water resources includes location and volume of water available.

4.4.3.4 Staff deployment

After perceiving the potential impact of the incident, ICs needs to know how many fire-fighters are available, their skills and equipment, and whether additional resources are required.

ICs also would like to have information about people on the incident site. They confirmed that it would be helpful if they can track location of staff, vehicles, equipment, and any people from other responding agencies if present.

4.4.4 Summary of information not needed

During interviews, ICs pointed out that floor plans are not necessary for *fire-fighters*. The fire-fighters themselves do not use floor plans because they are difficult to obtain, and difficult to carry during search and rescue. Paper floor maps for some premises are available but even so, are not always needed. Emergency responders have usually conducted fire exercises at the building before an incident occurs, so would be familiar with the design and layout of the building. Fire-fighters find their own way entering and returning from the incident.

ICs confirmed that digital floor maps *integrated with essential information* (e.g. temperature, smoke) would be useful for them (as opposed to the fire-fighters). In this case, ICs require only essential elements of the building structure (e.g. entrances, walls, staircases, rooms, doors, windows or ventilation control points, etc.). In addition, it is desirable if location of hydrants and location of any storage of chemical or hazardous materials can be integrated on the floor map. More detailed information relating to the content of rooms was said to be unhelpful and confusing. Displaying the individual locations and readings of individual sensors were also said to be unhelpful.

4.4.5 Current means of information acquisition

When ICs arrive on site, their current methods for acquiring information are: direct observation, asking people with the appropriate knowledge, and checking the facilities provided by the building. An IC is usually not the first one to arrive on site. When he or she arrives on site, he takes his time to assess the situation before assuming control. He can conduct a 360˚ walk around the scene and observe the fire or smoke from the outside. He can obtain a brief from the senior fire-fighter (who has been acting as a first-arriving IC and has been in charge of initial supervision and deployment of the crews by the time he arrives), or the security officers of the building, or local people who have witnessed the incident, as and when needed. Large commercial buildings, industrial sites or residential home are usually split into zones, and facilities are provided such as the panels that identify in which zone the indicator has been activated and where a zone is. An IC can check these facilities and concentrate their resources on the zone where the indicator has been activated.

If there is nobody present when an IC arrives, they will predict the possible risks present according to the time and the place, then make a decision according to their prediction, with worst-case scenarios. For example, if an incident was at a large shopping centre at 2 o'clock in the morning, most likely there would be no life-involved; if it was a shopping centre at 2 o'clock in the afternoon on a Saturday, it probably would be life-involved.

Apart from only a few available tools (such as simple facilities to identify fire zones in the building, CCTV images from the security offices, paper-based floor maps), the gathering of the information required for initial assessment is not supported by any technology.

The current method of tracking staff deployment is using plastic name tags (Figure 4-2). This enables ICs to keep a track of who is in and who is out, but once staff are deployed into the premises, there is currently no means of obtaining their location, health condition, etc. The fire-fighters deployed in the premises are required to exit the building from where they entered, unless in extreme cases where exiting from where they entered is impossible.

Figure 4-2: Using name tags at entry control

In addition, the communicating of the required information is mainly based on verbal communications with the assistance of quick sketching (as shown in Figure 4-3). Currently, little technology is used – suggesting that there may be opportunities for supplementing existing practices with the capabilities provided by new technologies.

Figure 4-3: Mean of communicating information

4.4.6 Opportunities for technologies

Based on interviews with ICs, observation of fire exercises, and subsequent validation, a list of opportunities for technology was revealed. Each of the opportunities targets a gap discovered from the field study. The possible technological solution to each problem was proposed based on the judgment of the authors. The 'usability aspect' is an analysis of what specific end user added value is provided by the technology. Feedback was gathered by discussing each of these ideas with ICs, both during the visits and through subsequent follow up – this enabling limited validation.

Opportunity 1 – Seeing inside of the building

Gap Discovered: The lack of real-time information on the status of an incident.

Problem Summary: When the IC arrives on site, he has almost no knowledge of what the situation is inside the building. The current ways of retrieving this information are by observing from outside of the building, asking someone who has been in or checking operation panels. However, these are time-consuming and very limited information can be obtained. Therefore, it is difficult for the IC to monitor the real-time situation inside the building during the response to the incident.

Proposed technological solution: The use of WSN deployed in the building to provide a colour-coded distribution of the real-time fire development in the building.

Usability aspect: This information inside the building can provide a more complete view of the situation on-site (impacting ICs' situation awareness at a *perception* level). It can also provide better understanding of risks, influencing decisions on whether it is safe to enter the building (impacting ICs' situation awareness on *comprehension* level).

Feedback: ICs agreed that such information from inside the building would be very useful. They also highlighted that this information is more useful for ICs than fire-fighters. The ICs' position is usually outside the building, and they can have more resources available; in contrast, fire-fighters often work in an environment with low visibility, and they usually want to keep their hands free to carry out their operational tasks.

Opportunity 2 – Transmission of information to fire crew during approach to the incidents

Gap discovered: The lack of information during approach to incidents.

Problem summary: ICs and fire-fighters usually have very little information about the actual incident status during the mobilisation phase (routing to the incident site). A brief description about the fire incidents may be obtained by the control centre staff answering the emergency call and passed on to the fire crews, but they mainly depend on information gathered on arrival. However, the current way of assessing the situation on site is time-consuming and the information that can be retrieved is limited.

Proposed technological solution: Provide information (e.g. the occurrence and characteristics of the incident, real-time fire development inside the building) based on data from WSN to fire crews on their way to the incidents.

Usability aspect: Based on this information, ICs can plan ahead and respond to incidents more quickly (impacting on situation awareness and the goal of preparedness). Precautions can be taken based on risk analysis of the information (impacting on the goal of ensuring fire-fighters' safety).

Feedback: It was confirmed to be beneficial. As stated by a fire station manager, "before arrival we are looking to prepare both physically and mentally ourselves for the job to be done, therefore it is crucial we receive some useful information which supports us to make a picture of what is going on at this particular incident".

Opportunity 3 – Routing within the premises

Problem summary: Search and rescue team members can be almost "blind" in the building. Smoke, flame and moisture can affect visibility where search and rescue takes place. Often it can be dark due to time and lighting system shutdown. The protective equipment worn by search and rescue teams, for example a mask and breathing apparatus, also restrict vision. Thus, it is difficult to navigate by sight.

They often work in an environment with low visibility, noise, and unknown risks. Therefore, equipment that is worn and searching techniques trained are designed with the priority of protecting fire-fighters, whereas these can reduce the efficiency of search and rescue.

Figure 4-4: Fire-fighters performing Left Hand Search

Fire-fighters typically search by going along a wall as their reference, touch the top, scan the front to check for obstacles and step down to check stability of building structure. They walk with short, deliberate steps (as shown in Figure 4-4). This procedure is designed to ensure their safety; however, it is difficult and time consuming, and fire-fighters can miss areas or lose their reference points.

For buildings where the structure typically consists of large storage rooms with large area in the middle, such as warehouses, a "guideline" is used (as shown in Figure 4-5), which is a rope deployed in the building with specially tied knots at critical positions (such as turnings). When fire-fighters enter a building, they attach themselves to the "guideline" and follow its direction. When they leave the building, the knots are used as reference points to confirm their positions and to indicate their way out. Again, this is time consuming, and the guideline could be tangled or burnt, resulting in fire-fighters losing their reference.

Proposed technological solution: Routes can be calculated based on the integration of the real-time fire development information, potential hazardous areas, and the floor map, to direct fire-fighters into or out of the premises and avoid hazardous areas.

Usability aspect: This information can navigate fire-fighters inside the building under working conditions with low visibility and audibility (impacting on search and rescue efficiency). This information can help fire-fighters avoid the hazardous areas inside the building (impacting on the goal of ensuring the safety of the crew, which is the highest priority of all the goals).

Figure 4-5: Fire-fighters using guideline (Source: Fire Service Technical Bulletin 1/1997: Breathing Apparatus Command and Control Procedures)

Feedback: ICs and trainers did not think additional support for safe escaping route was necessary; they believed that their current training and usage of "guidelines" already enabled fire-fighters to find their way into and out of premises, and they were trained to avoid hazardous areas. The quickest possible route to the incidents would be helpful, but not necessary. However, *fire-fighters*, in contrast to the views of ICs, found it difficult to use "guidelines" and they were less confident about their effectiveness. They thought it would be helpful if they could be directed in the building using additional means.

Opportunity 4 – Projecting the future

Gap discovered: Difficult to know the history of an incident and to project the future development of the fire.

Problem summary: ICs do not have a way to monitor the incident during the time between the incident happening and the first crew arriving. In addition, they only have limited information about the incident between the first crew arriving and the IC arriving. ICs integrate the information that they obtain on-site with their understanding of it, in order to project what will happen in the future. For the same situation perception on arrival, different histories of incident development result in different projection of the future (as shown in Figure 4-6).

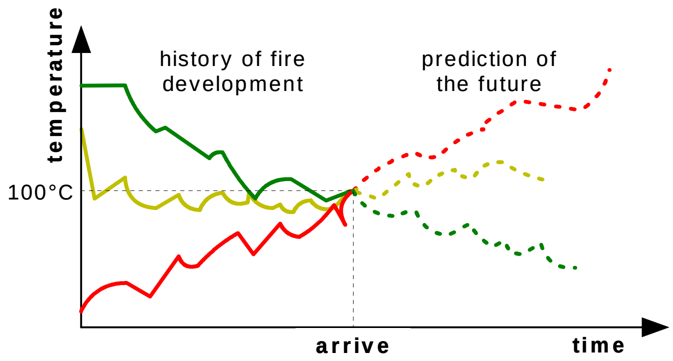


Figure 4-6: Projection of the future based on history of fire development

Therefore, without the history of incident development, knowing only the situation on arrival is not sufficient to understand the fire development trends, and predict the future. The difficulty of projecting the future results in difficulty in assessing risks and deciding on appropriate plans of action.

Proposed technological solution: Based on sensor data, a period of fire spread history can be kept, the trend can be calculated and the incident status in the near future can be forecasted.

Usability aspect: Better risk assessment can be done based on accurate, complete historical data (impacting on ICs' situation awareness). Better identification of potential hazards in the future (impacting on ICs' situation awareness at projection level).

Feedback: ICs confirmed that assisted prediction can save them valuable time collecting information, decision making and projections will be made easier and more accurate.

Opportunity 5 – Track fire-fighters' location

Gap discovered: Difficulty in tracking fire-fighters' location.

Problem summary: ICs only know fire-fighters in the building by name tags at the entry control. There is currently no way for them to know fire-fighters' exact location and status during their tasks in the premises.

Proposed technological solution: Use Wireless Sensor Networks to track mobile fire-fighters in the premises, and integrate the information with real-time fire distribution on the floor map.

Usability aspect: Tracking fire-fighters location can provide the IC with better control over his crew (impacting on the goal of ensuring the safety of the crew).

Feedback: ICs said it would be very useful. ICs would like to have information about the people on site, including fire-fighters, staff from other agencies, and casualties if possible. They do not like 'unknown' people walking around on site. All participants confirmed it would better ensure a fire-fighter's safety.

Opportunity 6 – Provide real-time hazards information

Gap discovered: Lack of efficiency in retrieving information about hazards.

Problem summary: The current way of obtaining hazards information is time-consuming. Moreover, there is currently no way to record dynamic hazards generated during the fire incident, e.g. building structure may become instable as the fire develops, there may not be a staircase in-situ anymore, or the temperature of the storage of hazardous material may change, resulting in an increased risk.

Proposed technological solution: Any known potential hazards information (such as chemical storage) can be pre-stored in a database and made available as incidents occur, real-time hazards information as the incident develops can be updated into the system: building structure hazards can be updated by human input, temperature on hazardous storage can be monitored by sensors. All hazard information can be recorded.

Usability aspect: The hazards information can provide the IC a more accurate incident ground DRA (impacting on ICs' situation awareness). Correct decisions can be made based on the accurate risk assessment (impacting on ICs' decision making).

Feedback: ICs confirmed it would be very useful to have dynamic hazard information. ICs currently expect to get such information through briefing from Sector Commanders, however, Sector Commanders have struggled to provide the support to the level expected by the IC. Therefore, having such information provided by technical systems would also improve the efficiency of their team dynamics.

Opportunity 7 – Integration of relevant information within a unified system

Gap discovered: Difficulty in efficient integration of different categories of information.

Problem summary: ICs have to obtain different categories of information from different sources, which is time consuming.

In addition, the information about incidents is not organized in an incident-site-centred way, thus it is difficult to access some information. For instance, information about nearby public and private water resources, and nearby storage of hazardous materials does exist. However, it is not part of the fire-fighting 'system', therefore ICs cannot easily obtain this information.

Proposed technological solution: Organize an Incident-Site Information Space, and integrate information about incident, hazards, water resources, human resources, and other relevant information within a unified system.

Usability aspect: The integration of information can provide a more complete picture of view on the incident more efficiently (impacting on ICs' situation awareness).

Feedback: ICs confirmed that a unified system, which can be accessed by an officer who is either a novice or experienced, would be one of the key characteristics for future system, since it improves the efficiency and accuracy of decision making.

Opportunity 8 – Provide vertical temperature distribution across a door

Gap discovered: Difficulty in predicting potential risks at the doorway

Problem summary: When fire-fighters come to a door during their left/right hand search, they have to follow the doorway procedure (as shown in Figure 4-7). One of them sprays water on the door; the other touches the door, to predict the vertical distribution of temperature on the door. This perception will lead to the conclusion of whether it is likely to cause a flashover or backdraft when the door is opened. Necessary precautions will be taken and control methods applied to minimize the risk (e.g. open the door to a gap, spray water through the gap and immediately close the door, and repeat until it's safer to enter).

Figure 4-7: Fire-fighters carrying out the doorway procedure

Proposed technological solution: Use sensor data to provide the vertical distribution of actual temperature across the door.

Feedback: ICs thought this information would be helpful to the fire-fighters. However, they were concerned about the way to present this information to fire-fighters. As mentioned before, fire-fighters worked in environment with visibility and audibility limitations.

4.4 Conclusions

The study of fire ICs' requirements provided a comprehensive understanding of their goals, tasks/decisions, information needs, and their feedback on the proposed technology opportunities. This chapter established link between user requirements and technology capabilities, and provided an expanded goal-driven method to establish these links. This relates goals, tasks/decisions, information needs and opportunities for technology.

Some interesting differences between literature and practice were also discovered. Although it was mentioned in the literature that one of ICs' goals is to save lives and protect fire-fighters, it was found in their practice and training that safety of crews was given highest priority, higher than safety of the public, and higher than extinguishing the fire. Although it was recognised in the literature that ICs' decisions are important for the outcomes of ER (Bigley and Roberts, 2001), and information is important for

ICs' decisions (Van de Walle and Turoff, 2007), what information is required to achieve ICs' goals was not clear. The study identified four categories of information needed (nature of incident, hazards, water resources, staff deployment), and detailed information under each category was provided to form a comprehensive picture of information required to achieve ICs' goals.

The study confirmed ICs' lack of real-time information on incident identified in literature, and also discovered other gaps such as search and rescue difficulties. The findings confirmed that sensor data could contribute significantly to fire ER, and suggested how it could address the opportunities discovered.

The findings of the study lead to the change of the initial focus of the further technical work. ICs did not approve that using sensor data to find safe escaping routes for fire-fighters was necessary, instead the study revealed that it was not part of the information that ICs require to make decisions. Therefore, more technology opportunities that can provide the required information and help ICs better achieve their goals were explored. Out of all the technology opportunities discovered, sensor data can contribute to most of them. However, based on sensor data, ICs would benefit the most from the opportunities of seeing inside the building and getting this information on their way to the premises. Therefore, these two opportunities were taken as the targeted opportunities to my further technical work.

Some human issues were also revealed during the visits and interviews. ICs do not trust technology. In fact, they try very hard to avoid relying on technology in order to accomplish their tasks. They are also very concerned about the cost of technology. This presents a challenge to support systems based on any technology. In addition, having only an easy-to-use interface is not enough. A successful system must demonstrate the added value the technology can provide for user outcomes, and must have high reliability, require low maintenance, and be cost efficient, in order to be able to convince ICs to use them.

Chapter 5. On-site Emergency Information Management System Architecture

	Research questions addressed in this chapter:
1	What are the features of an on-site Emergency Information Management System (EIMS) based on WSN?
2	What main components are necessary for the on-site EIMS architecture, and how do the system components interact with each other?
3	How can the dynamic nature of Emergency Response be addressed?
4	What are the benefits and trade-offs associated with the proposed EIMS design?

The study of Incident Commanders (ICs)' requirements in Chapter 4 identified a comprehensive set of information needed and information not needed, as well as a list of opportunities for WSN technology. Out of all the technology opportunities discovered, ICs would benefit the most from the opportunities of seeing inside the building and transmitting this information to the fire crew on their way to the premises. Therefore, these two opportunities were further investigated in the technical work contained in this thesis.

The information required to facilitate Emergency Response (ER) is referred to as Emergency Information. Providing Emergency Information can enhance the situation awareness of ICs (Yang, L. et al., 2009), and thus facilitate their decision making. Having extracted the opportunities for WSN technology, the next step was to design a system architecture suitable for incorporating WSN capabilities to provide Emergency Information. A system that collects and processes data, extracts information and manages the extracted information is referred to as Emergency Information Management System (EIMS). This chapter describes the proposed on-site EIMS architecture. Its scope and characteristics are defined and the enhanced goal-driven architecture design is demonstrated. The resulting on-site EIMS architecture is further analysed. The main components of the system and the interaction between components are discussed. In addition, system functions before, during and after incidents are specified. This chapter was based on the author's published work (Yang et al., 2007).

5.1 The Scope of On-site EIMS

A typical Emergency Response System (ERS) contains three layers from the bottom to the top: a data collection layer, an emergency information management layer, and a command and control layer, as shown in Figure 5-1.

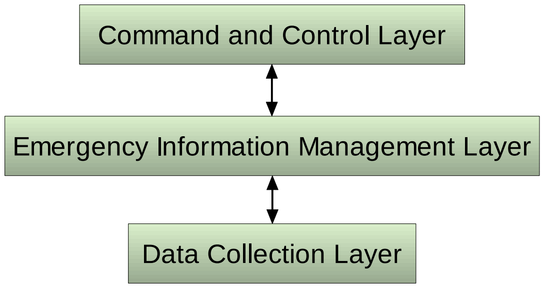


Figure 5-1: Typical ERS framework

The importance of information to decision making has been highlighted in the literature review. And the requirement of addressing the dynamic nature of ER was recommended. Kyng (2006) stated one of the challenges in designing interactive systems for ER is that "systems change with every situation and even with specific situations unfold". Therefore, without quick and reliable real-time data retrieval, the decisions made by emergency responders are likely to be neither timely nor optimum. Based on all the collected data, the Emergency Information Management Layer, i.e. EIMS, could analyse the situation, extract the information summary, generate risk assessments, and project what is likely to happen in the near future. This information can be shared within the emergency response management team, and provide decision support to improve the efficiency and accuracy of their decision making. Finally, an efficient command and control system is required to execute the response plans. My research mainly focuses on the requirements and design of EIMS in the middle layer, although system interactions with the information source layer and the command and control layer are also considered.

According to the targeted users, EIMS can be classified in two types: on-site EIMS and off-site EIMS.

The targeted users of on-site EIMS are the IC and his command support team, whose position will be at the scene of emergency sites in the case of an emergency. On-site EIMS manages information in the case of an emergency, and organizes the information centred by the location of emergency throughout the duration of the emergency. Therefore, on-site EIMS are typically characterized by event-driven operation mode, real-time performance and adaptability to the dynamic nature of emergencies.

The targeted users of off-site EIMS are the staff in command and control centres. Contrary to the IC and his crew, the position of control centre staff is not at the scene of emergencies. They will stay in the control centre, monitor the entire region, stay in communication with the ICs on-site, and coordinate and provide additional support if it's required. Off-site EIMS manages information within a larger region, where multiple incidents are possible, and associates the information in the region with its distributed location. Off-site EIMS are typically characterized by constant monitoring, and analysing and synthesizing capabilities.

Not enough attention has been paid to system design to support frontline first responders. Some system architectures designed for emergency response management have been suggested by researchers (e.g., Turoff et al., 2006; Zlatanova, 2005). However, they emphasized the requirements for coordination among responding agencies and communication of commands, and focused on systems for emergency centres. In other words, they are off-site EIMS. Although there has been an attempt to study system design for ICs (Jiang et al., 2004a), the prototypes proposed focused on interface design rather than underlying technology capabilities.

Hence, this chapter focus on the system architecture for on-site EIMS.

5.2 Goal-Driven Architecture Design

It has been recognised in the literature that "tailorability, adaptability, composability are critical for large information systems" because the demands from the stakeholders are ever-changing, and needs from the end users vary (Meertens et al., 2010).

Motivated by such demands, Meertens et al. (2010) proposed a goal and model-driven architecture design method and demonstrated how the goals of the stakeholders can be captured in goal models and then incorporated into the design of a care service platform. This method presented in detail the Model-Driven Architecture Stack (MDAS), consisted of Computation Independent Model (CIM), Platform Independent Model (PIM), Platform Specific Model (PSM), and application code. The activities required and languages available at each level within the MDAS were also described.

The method has the advantages of 1) providing a link between users' goal space and design space by incorporating user goals in the service architecture design and 2) architecture modelling at different levels of abstraction. It demonstrated how a single goal of the targeted users (e.g. elderly patients who need a reminder service) can be incorporated in a single service design under the scenario of patient care service. However, it did not consider the scenario where users are responsible for tasks with high complexity, such as Incident Commanders making critical decisions under time constraints. In this type of scenario, the goals of the users are better modelled as correlated hierarchies rather than individuals. In addition, the specific requirements in the problem and solution domain should be integrated. Therefore, the following two improvements are proposed to enhance the goal and model-driven architecture design method for the scenarios with complex user tasks:

- Capture goals in goal-action diagrams

Goal-action diagrams represent the result of goal modelling in a hierarchical way. It starts with a major goal, which can be broken into sub-goals, each of which requires information. Each needed piece of information would in turns require a system action, which can be interpreted as a service.

- The integration of requirements in the problem and solution domain

It is argued that the domain-specific requirements must be considered in the implementation of the enhanced goal-driven architecture design method in EIMS architecture design. Fouad et al. (2010) demonstrated the importance of including requirements engineering paradigm within Model Driven Architecture (MDA) design, and suggested a way of extending the MDA framework by adding a link from CIM to PIM that integrates the problem and solution domains. However, there is no extensive study on how such a link can be built and how the domain-specific requirements can be included in the architecture design for a particular domain. This chapter will demonstrate how the requirements in ER can be integrated in the architecture design.

The enhanced goal-driven architecture design method was implemented in this chapter for the system architecture design for on-site EIMS.

5.3 The Enhanced Goal-Driven Architecture Design

The literature review on ER revealed two general recommendations for any emergency response system design (Chapter 2), satisfying information needs and addressing the dynamic nature of ER. The literature review on Information Extraction from Sensor Data (IESD) revealed that sensor data fusion is important to make sense of large amounts of collected sensor data, in order to benefit from WSN technology (Chapter 3). As a result, it can be argued that a good EIMS requires the following features:

- adaptive to different incident stages
- fusion of multiple data sources.

This section will discuss the two aspects of domain-specific requirements and the goal-actions capturing in different incident stages in more details.

5.3.1 Adaptive to different incident stages

The existing architectures suggested by researchers (Turoff et al., 2006; Zlatanova, 2005) showed little consideration of the different actions the system should take at different incident stages.

The life cycle of an incident consists of three stages: before, during and after an incident. Therefore, a successful EIMS must take this into account and enable different actions to be taken at different incident stages. “Successful emergency management requires comprehensive emergency planning and preparedness before an effective response to the inevitable disaster can be implemented” (Tufekci and Wallace, 1998). During the incident, effective response means: (1) immediate warnings generated at the early stage, followed by (2) quick actions taken to control the development of the incident. Emergency Management (EM) does not end when the incident ends, post-event analysis and recovery is important as well. As Cutter (2003) said, an emergency response cycle includes rescue and relief actions immediately following an event, and long-term stages of recovery and preparedness for the next unexpected event. Since emergency responses may well differ during these three stages, EIMS which aims to provide support to emergency responders may perform different actions throughout different phases of an incident.

Irrespective of pre-event planning, on-event response or post-event analysis, data and information are important throughout all the 3 incident stages. “In the absence of data and information, emergency response is simply well-intended guesswork that will most likely result in significant loss of human life” (Erickson, 1999). Therefore, good information management can facilitate effective ER. The most crucial requirement of an EIMS is to provide the right information in the right format at the right time. This would typically mean providing this capability to the ICs since they would be in overall command of incidents. Typical goals of the ICs (in descending priority) would be to: “save life; prevent escalation of the disaster; relieve suffering; safeguard the environment; protect property; facilitate investigation/inquiry; and restore normality as soon as possible” (Hill and Long, 2001).

The goal-action diagrams were used to capture the goals and analyse what information and system actions are needed before, during and after an incident.

Before incident

The main goal before an incident is to prepare for incidents and as far as possible to prevent them occurring. The major goal can be broken into sub-goals, for example, to predict potential incidents, which need information about any abnormalities that can be monitored – this therefore requires EIMS to generate reports on faulty parts and abnormal phenomena detected during diagnosis. The full hierarchical breakdown of the goal-action diagram before an incident is shown in Figure 5-2.

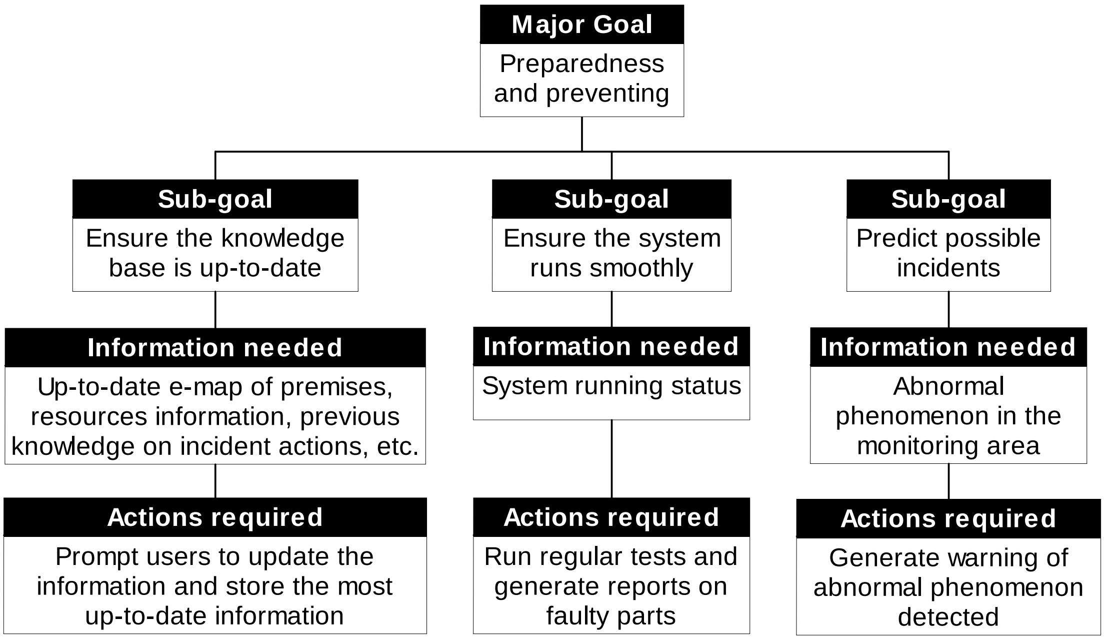


Figure 5-2: Goal-action diagram before an incident

During incident

The major goal during an incident is a quick and effective response, including situation assessment and efficient use of the available resources. Situation assessment can be further broken down into incident identification and forward projection, which require different decisions to be made. For example the incident commander will have to identify the nature of the incident, which demands information about incident occurrence and spread, and in turn requires data such as temperature, smoke, etc. at specific locations and also requirements about data accuracy and collecting frequency. As a result, actions that the system should take consist of alarm generation, real-time monitoring of the incident, and making historical trend diagrams available on request. The full breakdown of the goal-action diagram is shown in Figure 5-3.

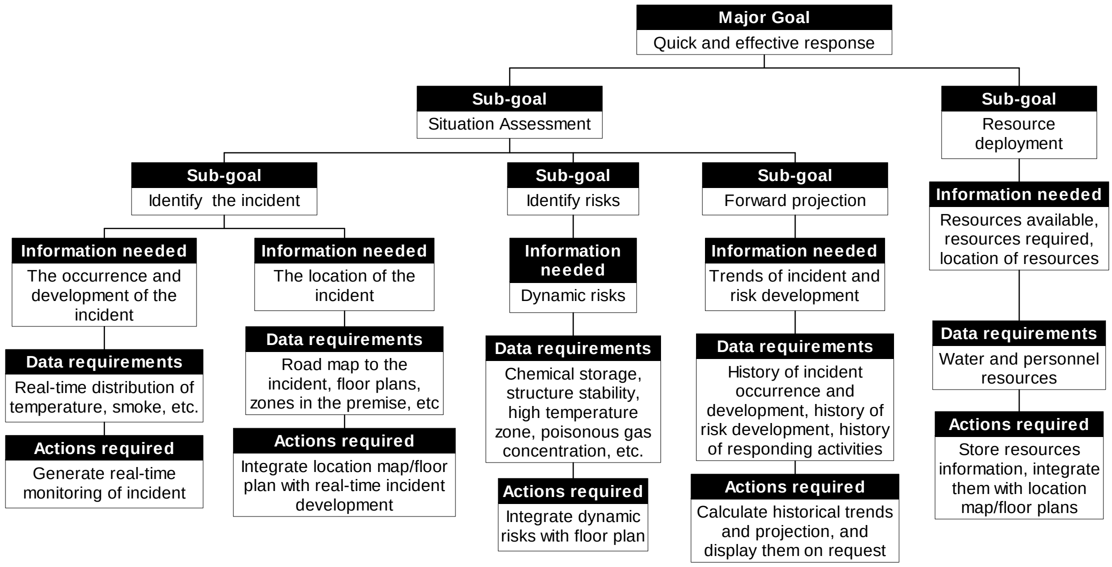


Figure 5-3: Requirements breakdown during an incident

After incident

The major goal after an incident is to collate and deliver statistical information on the incident. Post-incident recovery may require long-term work and the involvement of multiple agencies. However, only the goal of the immediate analysis after an incident is discussed here as shown in Figure 5-4 due to the focus on the goals and requirements relevant to an EIMS.

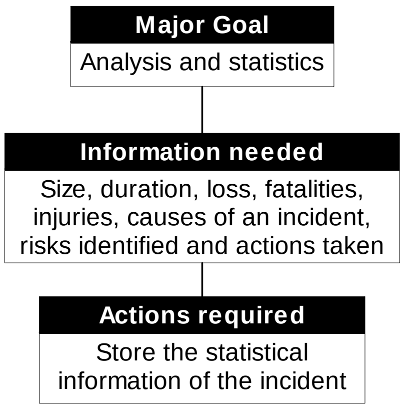


Figure 5-4: Requirements breakdown after an incident

In summary, the requirement analysis demonstrated that:

- In order to prepare for emergency incidents, the system should maintain up-to-date information and run regular diagnostics to ensure the system works normally.

- To ensure efficient response to incidents, the system should attempt to predict possible incidents and generate warnings on detection of abnormal phenomenon.
- To assist response to incidents that are occurring, the system should monitor real-time hazard conditions, dynamic risk information and track responder location during incidents.
- To help post-event analysis, the system should record the incident's nature, size, duration, loss, fatalities, injuries, causes, risks present and actions taken after incidents for future reference and response optimisation.

5.3.2 Fusion of multiple data sources

The importance of data fusion has been well addressed in the literature. It is generally recognised that combining multiple data resources to gather information can achieve better efficiency and potentially better accuracy than using single resource (Llinas and Hall, 1998; Foresti, 2002). However, fusion of multiple data sources in the context of on-site EIMS has not been studied before according to the author's best knowledge. In this section, fusion of multiple data sources for on-site EIMS is studied and three types of data sources are identified.

Data source 1: Wireless Sensor Networks

Capabilities: WSNs have demonstrated their capability of collecting data about the environment being monitored in a wide variety of applications. Wireless sensor nodes are designed to be light-weight, cheap, and energy-efficient, so that they are easy to be deployed. In case of an emergency such as a fire incident, CCTV cameras may not be functioning since the main power is often switched off to prevent risks. However, the battery-powered sensor nodes will not be affected and they can form a vital part of data sources in case of an emergency. Sensor networks deployed in large commercial buildings, where higher risk to people would be present in case of fire emergencies, have the potential of providing valuable inside-building information about incident occurrence and its real-time development. Multiple types of sensors can be used to improve system performance in identification, detection, and tracking of a phenomenon, and improve the situation awareness of the targeted user group ICs.

Challenges: Sensor data often contains noise, outliers and missing values. In statistics, an outlier is an observation that is numerically distant from the rest of the data (Barnett and Lewis, 1994). In the scenario of ER applications, an outlier is considered as an extreme sensor reading that is not caused by a real environmental change. These data quality problems may affect the accuracy of identification, detection, and tracking of emergencies.

System actions required: In order to incorporate the capabilities of WSN and overcome the challenges, data cleaning must be performed to reduce noise and outliers. Missing values can indicate the seriousness of a fire emergency; or an anomaly in the network, therefore it may not be appropriate to recover the missing values which existing research proposed to do.

Data source 2: Static pre-stored data.

Capabilities: Static pre-stored data includes electronic floor plans, hydrant locations and properties, construction materials, chemical storage, resources nearby, etc. This static information provides an important context for the data from WSN.

Challenges: How to organize different types of data and keep all the pre-stored data updated and valid.

System actions required: It is proposed to generate such information on separate layers, and integrate them with the layer of information from WSN data. Such information can be stored in a central database, and downloaded in case of a fire emergency. During and after the incident, any invalid or out-of-date information can be updated. The updated information can be uploaded when the crew deployed to the incident site return.

Data source 3: Human input of dynamic risks.

Capabilities: The importance of information on dynamic risks was strongly emphasized from interviews with ICs. Examples of dynamic risks include damaged stairs or other parts of the building structure during an incident, temperature rise on chemical storage, etc. Temperature can be monitored by sensors; however, risks such as building structure instability or damage cannot be so easily monitored by technological means. Human input, for example, from a fire-fighter who has just been in the premises, has the capability of covering and updating this type of data.

Challenges: Keeping the updating task simple and efficiently sharing this information.

System actions required: Enabling quick update and sharing of dynamic risk information, and generating such information on a separate layer so that it can be integrated with other layers.

In summary, the EIMS should consist of two levels of fusion, fusion within the WSN, and fusion among different types of data sources.

5.4 EIMS Architecture

Each system action, which can be interpreted as a service, is modelled in an enterprise modelling language ArchiMate (The Open Group, 2009). Then the models of individual services (Appendix II) are integrated into one system architecture. The integration of separate services consists of redundancy removing, elements grouping and optimization. The integration is necessary to understand the fundamental technical components needed for a system, which lead to the research in the following chapters on how to better provide technical capabilities required in each component.

The resulting abstract level architecture of on-site EIMS is proposed as shown in Figure 5-5.

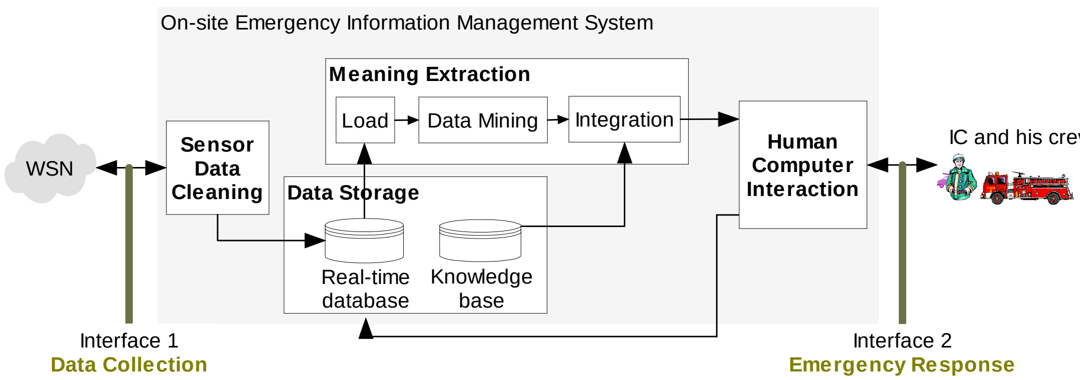


Figure 5-5: Diagram of emergency information system architecture

The system consists of four fundamental technical components:

- Sensor data cleaning

The sensor data cleaning component is the bridge between WSN, located in the lower layer of information source layer, and the middle layer of the emergency information management system. The sensor data cleaning component is necessary to reduce noise and outliers typically existing in sensor data. The processed sensor data with improved data quality can then be stored in the real-time database, ready to be used by other system components. The detailed implementation of the sensor data cleaning component will be described in Chapter 6.

- Data storage

The data storage component consists of two types of database. The real-time database stores data on dynamic occurrence and real-time development of an incident generating from sensor networks, whereas the knowledge base contains pre-stored static information about the incident premise, resources nearby, and relevant knowledge. The details of the proposed data storage mechanism are provided in Chapter 7.

- Meaning Extraction

The meaning extraction component is responsible for loading and processing the real-time sensor data, integrating the result with other types of data sources. More specifically, it can be integrated with pre-stored data such as a road map or floor plan to generate interactive incident monitoring, or integrated with the dynamic risk information to provide better support for ICs' dynamic risk analysis. The research on meaning extraction is described in more details in Chapter 8.

- Human Computer Interaction

The human computer interaction component is the interface between the targeted users (ICs and his crew) and the system. This displays the required information in the right format at the suggested level of detail, as well as providing the users the freedom to choose the desired level of detail and the information combination. Dynamic risks information during an incident is updated through the HCI component. This component is important for the ultimate benefits the system can provide to the targeted users. However, interface design is not the main focus of this thesis.

Interaction with lower Data Collection Layer

In the proposed architecture, the on-site EIMS interacts with the lower data collection layer in two ways. One communication link is between the system and WSN, the other communication link is between the system and ICs and his crew.

Interaction with upper Command and Control Layer

The interaction with upper Command and Control Layer is mainly through the existing communication network of the fire and rescue teams and according to their operation procedure. The information sharing can be proposed in two directions, one is from the IC to his crew on-site, and the other is from the IC to the control centre.

5.5 Function Specification of EIMS

The EIMS can operate in two modes, event-driven mode and demand-driven mode. The event-driven mode is a bottom-up process that is initiated by sensors reporting a detected anomaly. In this mode, the sensors are configured to be in sleep mode by default, and a “wake up” mechanism is activated when one or more sensors detect a possible abnormal situation about a phenomenon. On the other hand, the demand-driven mode is a top-down process initiated by the users of the system, either regularly or in case a system check is necessary or a particular interest has been generated.

The functions of the on-site EIMS are specified in terms of different stages in the life cycle of an incident.

5.5.1 Before incident - System diagnosis and early detection

- System diagnoses in demand-driven mode

System diagnosis before an incident serves the purpose of ensuring the system functions properly.

The system diagnosis regularly checks the EIMS together with WSN in the data collection layer. It receives actual data from wireless sensor nodes as input, and checks the system output against the expected output to identify any abnormal situations, e.g. non-responding sensor nodes, faulty sensors that generate invalid values, communication link error, system function error, etc. Further actions - ignoring, checking or response - can be decided by users based on the system report. Note that the IC is a role being filled when an incident occurs, therefore, at this stage, the users of the system could be a local security officer or control centre staff. Using the EIMS system during emergency responders' training can also be considered as a system diagnosis; in this case the IC would be the user. The flowchart of the demand-driven system diagnosis is shown in Figure 5-6.

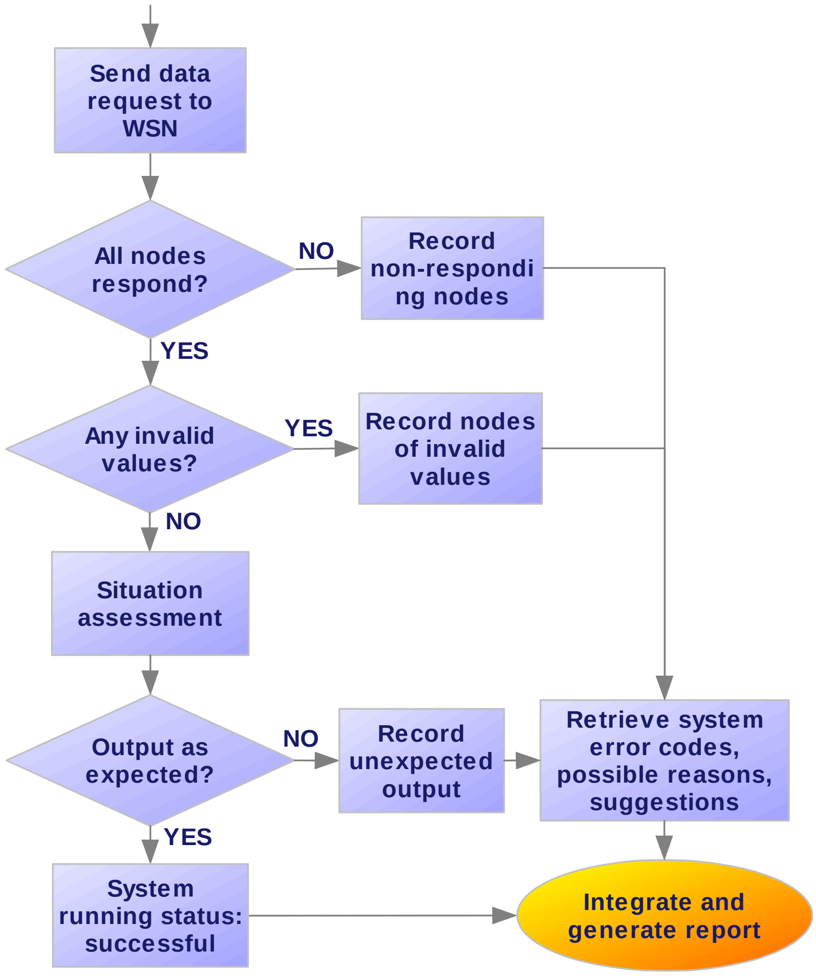


Figure 5-6: System diagnoses flowchart

- Early detection in event-driven mode

In order to enable event-driven detection, the sensor nodes are required to be configured in beacon-enabled mode. In large networks, sensor nodes are organized as a cluster of clusters. The cluster heads send periodical beacons to the nodes in the cluster to confirm their presence and put them on duty for a fraction of time, while sensor nodes may sleep between beacons. Sensor nodes in the same cluster can be rotated to be on duty, and the sensed value will be checked against rules. Data transmission between a sensor node and the cluster head only commences when the sensed value becomes suspicious. A suspicious data received from sensor node will trigger the wake up mechanism. The warning generated can be filtered either by automatic procedures implemented in the system or by the user. The system utilizes the automatic filtering schema to check whether it is likely to be a false alarm or real alarm. It also allows the user to access the information with the desired level of detail to double check the judgement if they want. If an incident is confirmed, the system will switch to monitoring mode. The process of event-driven early detection of possible incidents is depicted in Figure 5-7.

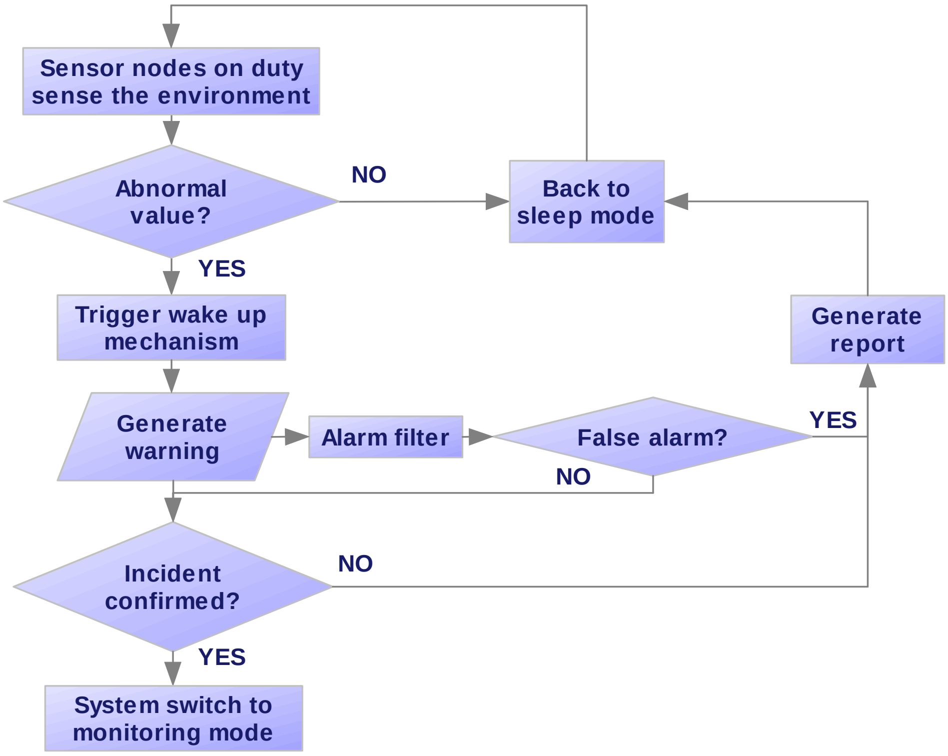


Figure 5-7: Early detection of possible incidents flowchart

5.5.2 During incident - Monitoring the incident

Once an incident is confirmed, the system is switched to monitoring mode. All the sensors will wake up and start transmitting data. At the same time, the system mounted on the vehicle, which carries the deployed responding team, is switched on. It downloads the pre-stored information about the specific incident, and receives the real-time data from the WSN during their mobilisation phase (approach to the incident). The data from WSN is integrated with pre-stored information such as floor plans and location of facilities, to enable monitoring of the incident development. To assist risk assessment, sensor data can be integrated with pre-defined risk levels, and output the location and levels of risks.

On arrival at the incident premise, the IC will carry out dynamic risk analysis and make decisions on the tactical mode and action plan accordingly. The result of dynamic risk analysis during the incident can be updated into the system, and integrated with other layers of information.

The detailed behaviour is demonstrated in Figure 5-8.

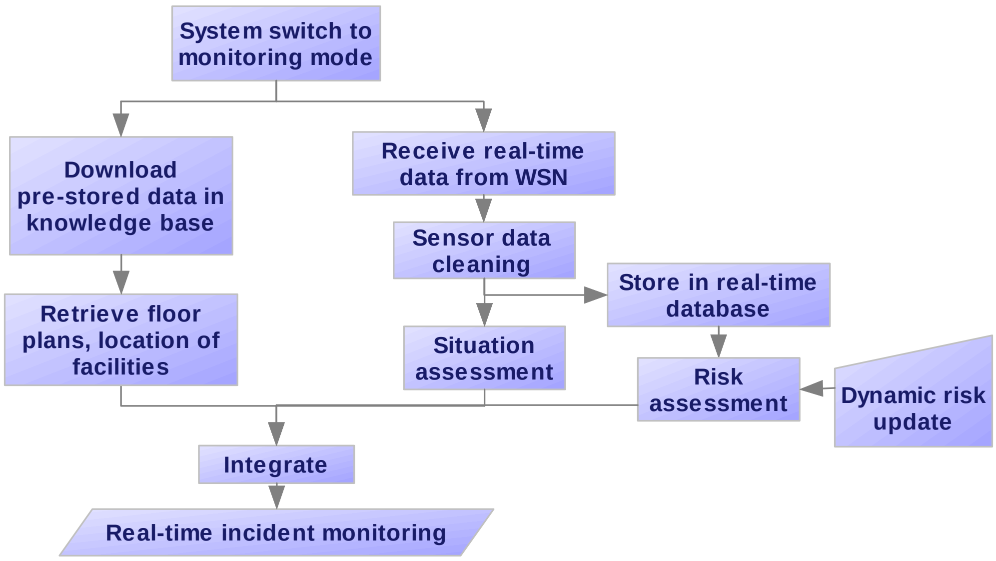


Figure 5-8: Real-time data monitoring flowchart

5.5.3 After incident – Update and upload knowledge base

The task of EIMS in this stage is a simple but important database interaction with incident commanders. The incident commander fills in a post-event analysis form after an incident has been dealt with. The content should include the following information: Extent, Duration, Loss, Fatalities, Injuries, Causes, Emergency level, Actions taken, and any suggestions on amending the existing emergency plan. Some information such as duration of the incident could be filled in automatically by the system, whilst the rest has to be entered manually. This information will be added to the knowledge base for future statistics and integration purposes. The updated knowledge base will be uploaded to the central storage when the responding team has returned from an incident.

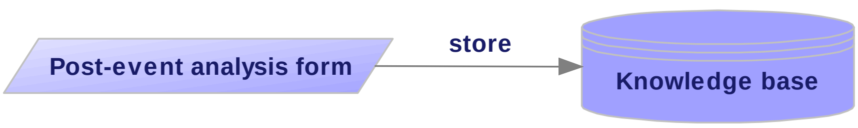


Figure 5-9: Update and upload knowledge base flowchart

5.6 Discussion and Conclusions

This chapter proposed an EIMS architecture which facilitates responding to emergencies before, during and after an incident. Requirements analysis and overall system structure design have been discussed, potential benefits and design trade-offs of the system have been highlighted. The findings demonstrated that: (1) in order to prepare for emergency incidents, the system should run regular diagnoses to ensure the system works normally; (2) to ensure efficient response to incidents, the system should attempt to predict possible incidents and generate warnings on detection of abnormal phenomenon; (3) to assist responses to incidents that are occurring, the system should monitor real-time hazard conditions, dynamic risk information and track responder location during incidents; (4) to help post-event analysis, the system should record the incident's nature, extent, duration, damage, fatalities, injuries, causes, risks present and actions taken after incidents for future reference.

The benefits and trade-offs associated with the proposed EIMS design are discussed below.

Benefits

- Quicker and better Emergency Response

The nature of emergency response is that there is often very limited information during the early stages of the incident, especially the information about the inside of buildings. Traditional ways of getting information are by direct observation, verbal communication with personnel on the incident ground, or checking operational panels. However, these methods are not efficient and the ways information can be gathered are very limited. The proposed EIMS enables event-driven early detection of possible incidents, and provides real-time information on incident development during the mobilisation phase. As a result, quicker ER can be achieved and the first responders can be better prepared. In addition, the real-time incident development can be monitored throughout the response to the incident, and risk assessment is facilitated by integrating human updates on dynamic risks into the system. Thus, the proposed EIMS facilitates information gathering, and addresses the dynamic nature of incidents.

- Flexible application

Although the specified system behaviours are mainly based on requirements gathered from fire ICs, the proposed system architecture can be flexibly applied to a variety of applications. Based on this architecture, depending on different definitions of “meaning” according to the context, the architecture can be used to develop emergency detection systems, risk assessment systems, decision support systems, historical statistics or an integrated multi-purposes emergency response support system. Depending on different contexts of use, it could be located in the local security office, an emergency control centre, or as a vehicle mounted system as described in this chapter, or a web-based service which allows access from mobile devices.

Trade-offs

There are some trade-offs which should be taken into consideration when designing such an EIMS.

- Diagnoses frequency/system cost trade-off

The research has demonstrated that the system should run regular diagnostics before the incident, but running diagnostics too often could be a waste of time and system resources. In the case of running synthesized diagnoses with a wireless sensor network as bottom layer, too frequent diagnostics might cause the sensors to run out of battery power, thus causing them to be inoperative.

- Quicker response/reducing false alarms trade-off

The EIMS aims to facilitate a quicker response to emergencies; however, higher sensitivity may cause a higher rate of false alarms. The quickest way is to generate the alarm on receiving any data out of normal expectation without filtering, but these alarms might not be the symbol of a real incident. To filter and reduce the false alarms requires extra time and system resources consumption, for example the activation of more sensor nodes in the wake up mechanism before an alarm condition is notified. This is a trade-off to be taken into account during the early stage of an incident.

- Historical trend/running cost trade-off

Another consideration is that during an incident, a historical trend diagram showing the situation from the beginning of an incident could help operators to understand the incident situation and project what would happen in the near future. However, if the incident situation changes rapidly over time, to maintain a trend diagram could result in an overwhelming request for the amount of data storage space and data analysis time.

- Automatic suggestion/manual decision trade-off

Computer-assisted risk analysis and decision support based on reliable data sources and efficient data mining could help the IC judge the situation and quickly issue the control commands. As described by Danielsson (1998), the key to incident command is the quick implementation of a fast strategic response. However, the suggestions made by the system may not be appropriate to the specific situation. Therefore, such a system should not over-automate its response: the original information that is used to generate risk analysis/decision support results should be available to enable the incident commanders to make their own decisions if they so wish.

On the basis of the abstract level system architecture proposal, the detailed sensor data fusion methodology required for each main component of the system will be studied in the following chapters.

Chapter 6. Data storage Mechanism

Research questions addressed in this chapter:	
1	What constitute a suitable data storage mechanism for WSN-based on-site EIMS?
2	What impacts do the features of sensor data have on data storage mechanism?
3	What requirements do ER applications have on data storage mechanism?
4	What is the detailed design of the proposed data storage mechanism?
5	How does the proposed data storage mechanism perform in experiments?

As proposed in the on-site Emergency Information Management System (EIMS) architecture (see Section 5.4), data storage is one of the necessary system components. Data Storage Mechanism (DSM) is the way that data is stored and managed in the data storage component. It has a major impact on the efficiency of further data processing for the purpose of meaning extraction. Consequently, what constitute a DSM for WSN-based on-site EIMS should be analysed first.

It can be argued that what constitute a DSM for WSN-based on-site EIMS largely depend on the features of sensor data and the requirements of Emergency Response (ER) applications. Using WSN as the data collection layer, the EIMS will pre-process, save, analyse and present the information to the targeted user: the Incident Commander (IC) and his command support team. Considering the EIMS as a whole, it has two interfaces. The data collection interface communicates with the WSN to retrieve sensor data streams, the Emergency Response interface communicates with the IC and his command support team to dynamically update risks and monitor the incident development during ER (as shown in Figure 5-5). Therefore, the features of sensor data and the features of ER decide the requirements for the design of data storage mechanism and data processing.

The objective of this chapter is to 1) provide a comprehensive analysis of the features of sensor data, 2) evaluate the requirements of ER application, and 3) propose a DSM that accommodates the challenges presented by the features of sensor data as well as maintains efficiency for the entire ER process. The work presented in this chapter was based on the author's published work (Yang, Y. et al., 2009)

6.1 The Need for a Different Data Storage Mechanism

The majority of the existing research on sensor data storage focused on purely technical issues such as storage placement (Ganesan et al., 2003) or energy consumption (Mathur et al., 2006). Distributed storage placement in the network has been argued to provide benefits such as shared storage load and collaboration among sensor nodes, as well as reduced impacts of single node failure. Some researchers declared that distributed storage at random locations is not suitable for sensor data as sensor data is location-specific, it should be stored at its origin instead. However, Ledlie et al. (2005) argued that sensor data generated at one location could have uses at a variety of sites elsewhere in the network, therefore sensor data should be stored near where it is primary used, together with its provenance (the history of how and when it came to be). Based on the understanding of the limited energy resources typically existing in WSNs, Mathur et al. (2006) evaluated a variety of flash-based storage options for sensor platforms, and concluded that surface-mount parallel Negated AND (NAND) flash can achieve 100-fold decrease in per-byte energy consumption in comparison with the MicaZ on-board serial flash.

Although the existing research addressed some important issues introduced by WSN, such as the ad hoc nature of WSN and energy constraints in WSN, not enough attention has been paid to the impact that the nature of sensor data has on DSM. A data storage mechanism that addresses the features of sensor data can benefit sensor data storage regardless of where data is stored. In addition, on-site ER presented different information retrieval and storage requirements compared to other WSN applications. WSN applications such as habitat monitoring or patient health monitoring typically require the full storage of all the data generated. Such applications can be characterized by constant monitoring with a lower data sampling rate, and offline statistics and analysis. On contrary to such constant monitoring applications, on-site

ER applications mainly focused on the period that covers an emergency. The sensor data generated consists of real-time sensor data streams during a comparably intensive period of time. Consequently, a different DSM is required for on-site ER applications.

6.2 Features of Sensor Data

In comparison with traditional data, the data from WSN has special features, which bring a list of challenges in the management and processing of sensor data. The features of sensor data are analysed as follows:

The streaming nature of data

Sensor data is best modelled as continuously arriving data streams rather than persistent relations (Arasu et al., 2004). Data streams differ from traditional data in the following ways:

- Sensor data is automatically generated and arrives in a multiple, continuous, time-varying manner. Therefore, the volume of sensor data increases along time, and the total volume of data is potentially unlimited. In contrast to this, traditional data typically consists of entries input by human, permanently or persistently stored in databases. The volume of traditional data is relatively stable.

- "Data stream is time ordered data, either explicit with time-stamp or implicit based on arrival order." (Kim et al., 2005) However, traditional data usually is not time-ordered unless explicitly specified.

As a result, the streaming feature of sensor data presents challenges in sensor data processing such as storing data with unboundedly increasing volume, continuous loading and continuous queries.

Existence of high tempo-spatial correlations

Sensors are usually deployed at a certain density so that they can cover the entire monitoring field. As a result, "most sensor-nets likely exhibit temporal and spatial correlations among node readings" (Silberstein et al., 2007). More specifically, the high temporal and spatial correlations existing in sensor data exhibits in this way: "the readings observed at one time instant are highly indicative of the readings observed at the next time instant, as are readings at nearby devices" (Jeffery et al., 2006).

The high tempo-spatial correlation provides potential benefits. They can be incorporated to estimate missing or corrupted data (Chok and Gruenwald, 2009), detect outliers and improve the quality of sensor data, data suppression (Silberstein et al., 2007), reduce data transmission in network, thus reduce energy consumption (Yoon and Shahabi, 2007). However, challenges exist in identifying correlations and modelling correlations, and keeping correlation information updated, etc.

Generation of redundant data

Significant data redundancy in a database can result from the strong spatial and temporal correlations typically present in sensor data. However redundancy can be used to predict missing values and to detect outliers, and a certain level of redundancy can improve the accuracy of database query results. Resolving redundancy in sensor data is not simply a case of removing redundant data, but rather maintaining it at a level that provides confidence in the data without producing unnecessary storage demands.

Sensor data contains 'noise'

Sensors are designed to have low power consumption and to have low cost. However, this design focus can result in the accuracy of sensors being limited. As well as design issues, sensors are normally deployed in harsh environments with the possibility of background interference, and consequently sensors can experience internal faults or damage during emergencies such as fires. Research has shown that sensor data often contains errors (due to sensor function) and noise (due to other environmental interference) (Elnahrawy and Nath, 2003). These characteristics indicate that sensor data should be cleaned before being stored in any database.

Sensor data is meaningless unless associated with time and location

Sensor readings are meaningless if they are not associated with time. E.g. knowing that there is a temperature of 25°C means nothing unless it is associated with time, such as it is 25°C now or it was 25°C an hour ago. Similarly, sensor data has meanings only if it is associated with location information. Therefore, the data storage mechanism should provide support to associate data with the time and location.

6.3 Features of Emergency Response Applications

As mentioned in Section 2.1, modern emergency response typically requires collaboration and coordination between different ER agencies. Apart from the requirement from the management side's perspective, ER also requires the efficient and accurate supply and communication of information during the dynamic development of incidents.

The need for timeliness despite the dynamic nature of ER

Due to the nature of emergency response, although dynamic risks will develop, and situation will be ever-changing, timely and efficient response to incidents is always required, because any delay may result in the loss of life and property. Van de Walle and Turoff (2007) state that: 'In the practitioner community, emergency managers have learned and stated that accurate and timely information is as crucial as is rapid and coherent coordination among the responding organizations'. It is essential therefore that WSN-based ER applications be resilient, real-time systems, and that they can support swift queries.

Sufficient accuracy to support decision making

The accuracy of the sensor data may influence the correctness of any decisions made by incident commanders. Therefore sufficient (but not excessive) accuracy of sensor data is essential for ER applications. For example, if alarms generated by noises or outliers are not filtered, response to the resulting false alarms will result in a significant waste of time and personnel resource. However, if a real emergency is filtered by mistake, the result could be the loss of life or property. The decisions made based on sensor data can be a matter of life or death, hence it is important to maximise the quality of sensor data.

6.4 The Proposed Sensor Data Storage Mechanism

As a result of the analysis of the features of sensor data and ER applications, the designed sensor DSM for ER applications aims to

- Store and manage sensor readings, time-related information and location-related information (impacted by the importance of associating time and location with sensor data);
- Support timely response to queries in spite of the increase in the volume of data (impacted by the streaming feature of sensor data);
- Be dynamic and flexible to accommodate different needs at different incident stages (impacted by the ER application requirements).

6.4.1 Database schema for building fire safety

The data that needs to be stored in WSN based EIMS can be classified into three categories: sensor information, time-related information and location-related information. Sensor information includes sensor identifiers and real-time streams of sensor readings. Considering the scenario that WSNs are deployed to detect structural fire emergencies in premises, the time-related information would be timestamps associated with each sensor reading. The location-related information in this case would include the position where each sensor is deployed in the premises, and the location of the premises.

The design of the overall database schema is as shown in Figure 6-1. The SensorReading table stores and manages the streaming data received from the WSN and the associated timestamps. The SensorDeployment table stores and manages the relative sensor location in a building. Building information is stored in two tables, Building and FloorMap.

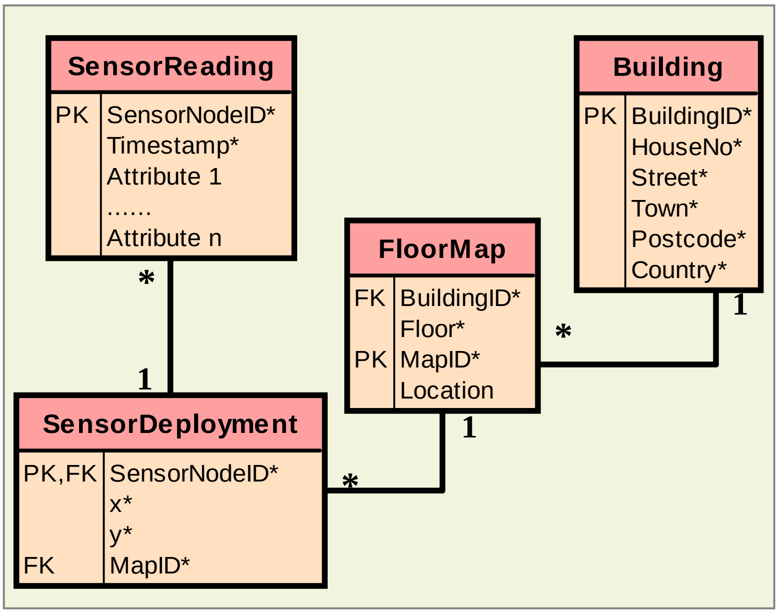


Figure 6-1: Overall database schema

Each tuple in the SensorReading table specifies the source of the sensor data (SensorNodeID), the time instance that the sensor data is generated (Timestamp), and readings from each type of sensor carried on a sensor node (denoted as Attributes 1 to n without losing generality). Each sensor node in a WSN can carry a number of sensors, each of which is considered to generate an attribute of the sensor node and assigned an attribute name, such as temperature, smoke, flame, etc. Atrribute1 to n will be replaced by attribute names in real applications. A sensor node ID and timestamp uniquely determine its readings, denoted as [SensorNodeID, Timestamp] → [A1, A2…An]. SensorNodeID is the foreign key that links to the primary key on the SensorDeployment table.

Each tuple in the SensorDeployment table specifies a sensor node (SensorNodeID), where it is deployed (MapID), and its relative coordinates on the map (x, y). A sensor node ID uniquely determines where it's deployed, denoted as [SensorNodeID] → [MapID, x, y].

Building information is organized as a table containing entries for each building address, and a table of floor maps. Each tuple in the Building table specifies a building (BuildingID), and its address (HouseNumber, Street, Town, Postcode, Country). A building ID uniquely determines its address, written as [BuildingID] → [HouseNumber, Street, Town, Postcode, Country]. Each tuple in the FloorMap table specifies what a map describes (BuildingID, Floor) and where the map is stored (MapID). A building and floor pair uniquely determines its map, denoted as [BuildingID, Floor] → [MapID].

In the case of an emergency, the building data can be retrieved by searching according to address and postcode. The required floor map can be retrieved and displayed, on which sensor data can be integrated. Temporal correlations of sensor data are maintained in the SensorReading table. Spatial correlations are maintained by associating real-time sensor readings with its location on the floor map through the bridge of the SensorDeployment table.

6.4.2 Time-driven sensor data management

The streaming feature of sensor data means that, the volume of sensor data increases along time, and the total volume of data is potentially unlimited. Consequently, the query efficiency may drop as the incident develops. However, to provide support for ER personnel to make quicker decision, information need to be presented to them in a real-time manner, which requires the data storage to support efficient response to data queries throughout the incident. Therefore, a time-driven sensor data management is proposed to maintain the query efficiency.

First of all, the real-time sensor data can be modelled in a three dimensional space, with the dimensions representing time, area of location and range of sensor readings, as shown in Figure 6-2. Each cell of this three dimensional space represents the data volume in the range of time, area of location and range of sensor readings, denoted as *v: (t,l,r).*

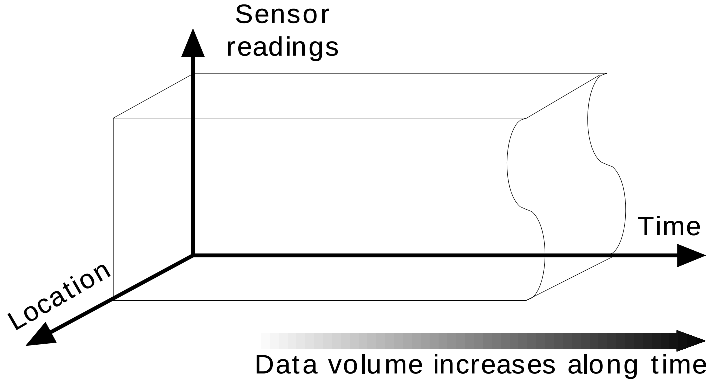


Figure 6-2: The real-time sensor data volume space

A typical data management method to improve the query efficiency is by partitioning. Partitioning is the division of a database or its element into smaller, more manageable parts, according to the partition key, so that only the partitions that contain the answer to the query will be scanned rather than the whole dataset. Partitioning provides benefits such as manageability, performance and scalability. However, different partitioning schemes can result in different performance. Therefore, a suitable partitioning scheme has to be decided according to different application requirements.

Within the ER scenarios of concern, typical queries during the incident are tempo-spatial range based queries such as:

'What is the distribution of sensor readings in the area S between time A and time B?'

Therefore, possible choices of partitioning scheme are:

- partitioning by range of sensor readings, where the complete dataset *D* is partitioned in *r* number of partitions $p_1, p_2, \cdots, p_r$
- location range partitioning, where the complete dataset *D* is partitioned in *l* number of partitions $p_1, p_2, \cdots, p_l$
- time range partitioning, where the complete dataset *D* is partitioned in *t* number of partitions $p_1, p_2, \cdots, p_t$

Partitioning by the range of sensor readings is not suitable. The distribution of sensor readings covers the whole range of sensor reading, thus, all partitions (e.g. $p_1, p_2, \cdots, p_r$ as shown in Figure 6-3 (a)) are required to be scanned to answer the query even after partitioning by the range of sensor readings. No reduction on the amount of data that has to be scanned can be achieved from this type of partitioning, therefore, the query efficiency may still drop as the incident develops.

Location range partitioning can achieve some reduction on the amount of data that has to be scanned to answer the query. In case the interested area *S* in the query is a region smaller than the whole network coverage (e.g. when users zoom in), only partitions that consist of *S* (e.g. $p_2, \cdots, p_{l-1}$ as shown in Figure 6-3 (b)) are required to be scanned to answer the query. However, in case the interested area *S* is the whole network coverage, no reduction would be achieved. In addition, the size of the location range partitions will grow as the time duration grows. Therefore, the amount of data that needs to be scanned will still increase as the incident develops, although not as much as that without partitioning.

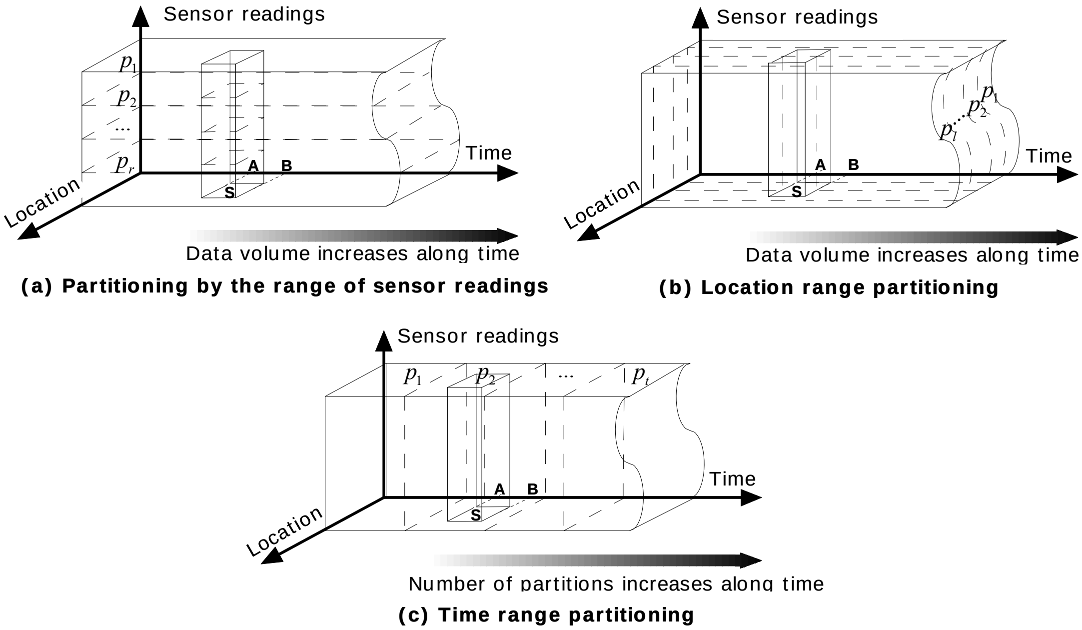


Figure 6-3: Different partitioning scheme on the real-time sensor data space

In comparison, partitioning the real-time sensor data space by time range achieves the most reduction on the query work load. The number of partitions will increase as the size of the dataset grows. Nevertheless the number of rows that need to be scanned for any query will be kept relatively stable, since only partitions that contain answers to queries (p_2) will be scanned instead of the whole dataset, as demonstrated in Figure 6-3 (c). Therefore, the database can still maintain query efficiency as the volume of data increases as the incident develops.

6.4.3 Adaptive support for different incident stages

It has been recognised in the context-aware computing literature that adaptive data management can address the challenge of computing for a changing environment (Haghighi et al., 2007). Adaptive behaviour aims to automatically and dynamically adjust data retrieving parameters according to changes in context or the availability of resources. In order to address the dynamic nature of ER, the author proposed to provide adaptive support for different incident stages. The proposed adaptive support included adaptive data sampling and adaptive data aging and data compression.

To accommodate different needs at different incident stages, the adaptive sampling to support different incident stages is designed in association with the overall database schema. Before an incident, data collection can be configured at a low frequency. Sensors can be in sleep mode most of the time and only wake up once in a while to monitor the environment. When a suspicious state is detected, data collection frequency can be increased. During incidents, sensor readings will be saved in real-time. Finally, the data collection frequency can be set back to normal after the incident.

The benefits of the adaptive behaviour are twofold:

- It reduces the energy consumption of the sensor network.
- It reduces the amount of data storage space that is required (as demonstrated in Figure 6-4).

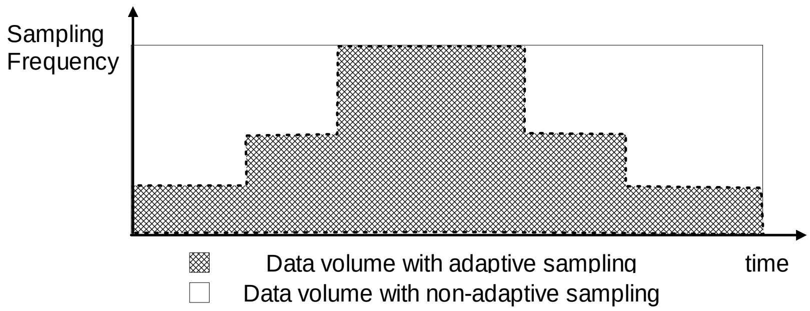


Figure 6-4: Adaptive data sampling

More specifically, the adaptive data sampling behaviour can be controlled by the Level of Severity (LoS). LoS can be defined in different ways. For simplicity, LoS is introduced under the scenario of WSN with single type of sensor in this section. LoS can be defined as the magnitude of sensor reading, mapped to a warning level. (Definitions of LoS under other scenarios can be found in Chapter 8.) The detailed adaptive data sampling algorithm (pseudo code) is shown in Table 6-1.

Table 6-1: Adaptive data sampling algorithm (Pseudo code)

f : the current sampling frequency f_L: the lowest acceptable sampling frequency f_U: the highest acceptable sampling frequency n: the total number of warning levels W: represents the Level of Severity
Initialization: $f = f_L$; **Adaptive data sampling:** for each acquired sensor reading r; 1. Map r to a warning level W; 2. if $f < f_U$ 3. $f = f_L \times 2^W$ //increase the current sampling frequency 4. else 5. $f = f_U \div 2^{n-W}$ //decrease the current sampling frequency end

In the case that the available data storage space is running low during the incident, an adaptive data aging and data compression mechanism is triggered to free the space. Data aging typically means the removal of data from the media that has aged. However, sensor data reflecting incidents has both real-time value and historical value. After incident, the sensor data collected about an incident may still be useful for the purpose of post-event analysis of what happened during the incident, and for training purposes. Only the summary of sensor data up to the level of detail that can enable post-event playback is required, rather than the complete dataset. Consequently, it can be argued that data compression rather than removal would be more suitable for the aged data in ER.

It is proposed to age and compress the sensor data partition by partition. A First-In-First-Out rule is utilised, which means the oldest partition is always aged first.

Aging Rule 1: A partition becomes "aged" when its time range is outside the user-specified retaining range.

Aging Rule 2: The oldest partition becomes "aged" when the available storage space is lower than a predefined limit.

The aged partitions can be compressed due to the significant data redundancy typically existing in sensor data. Compression reduces the amount of storage space required by losing the redundant data. However, a summary of the original data is still kept to enable post-event playback. For example, a fire occurred in a room on the ground floor of a multi-story building; the rooms on the second floor were not affected by the incident. Therefore, sensor readings from the second floor rooms would be similar to each other and similar along time (redundancy). Hence, keeping only the aggregated sensor reading, including its the time and region, would be sufficient to enable post-event playback, while data storage required would be significantly reduced. The aged partitions would be compressed according to the pre-defined granularity. The granularity defines the size of the data cube in the partitions that need to be compressed. The sensor readings within a data cube would be aggregated. The detailed adaptive data compression algorithm is presented in Table 6-2.

Table 6-2. Adaptive data compression algorithm (Pseudo code)

P_w: The working partition P_c: The compressed partition
Q_b: A data cell in the working partition Q_a: The compressed data cell
g: the pre-defined compression granularity

```
for each Pw that satisfies the aging rules
1.    Pc = compress (Pw, g);
2.    store(Pc, summary); //store the compressed partition in the summary table
3.    delete (Pw);
end
function compress (Pw, g)
1.    for each data cell Qb in Pw
2.        Qa = aggregate (Qb);
3.        Add Qa to Pc;
4.    end
5.    return Pc;
end
```

After the aged partitions are compressed up to the desired granularity, they are stored in a summary table. The structure of the summary table is as shown in Table 6-3.

Table 6-3: Summary

Columns	Description

ID	Identification
Aggregation	Q_a
Region	Spatial granularity
Valid from	This is the time that the aggregation is valid from.
Valid until	This is the time that the aggregation is valid until.

6.5 Performance Evaluation of the Proposed Data Storage Mechanism

A small scale experiment was carried out mainly in the aim of evaluating the query efficiency of the proposed data storage mechanism whilst real-time data continues feeding into the database. The hypotheses are as follows:

1) As a result of the proposed time range partitioning, the query efficiency was expected to remain the same regardless of how large the database was becoming.
2) Partitioning may introduce extra storage cost and extra maintenance complexity.

Therefore, whether the proposed DSM can maintain good query efficiency and whether any additional storage and updating costs would be introduced to achieve better query efficiency were studied in the experiment.

6.5.1 Experiment setup

A data generator, two databases and a query engine were implemented for the experiment, as shown in Figure 6-5. The data generator simulated the data streams generated by temperature sensors in a 5×5 WSN. It generated random data at certain time intervals, assembled the random data with the timestamp and its sensor ID. The real-time assembled data was then stored in both the database partitioned by range of timestamp and a comparison database without partitioning. The commands (in Java statements) used to create the databases are shown in Table 6-4. Both of the two databases were implemented using MySQL server 5.1.45 in a PC with Intel Pentium 4 CPU (3.20GHz) and 2GB of RAM. The query engine was implemented in Java with MySQL connector/J 5.1.1.3. The same tempo-spatial range queries were executed on both databases, and their performance was compared.

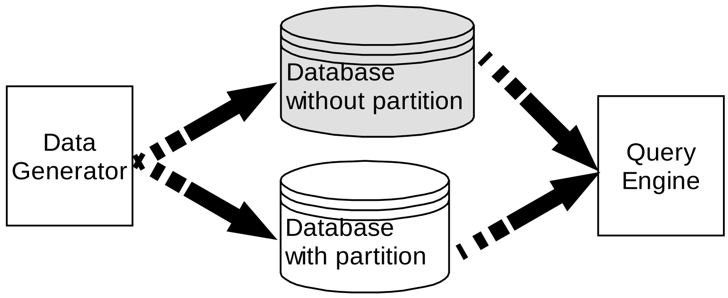


Figure 6-5: Database storage mechanism experiment design

Table 6-4: Create Table Commands

Create table with partitions:

```
String sqlquery = "CREATE TABLE IF NOT EXISTS partitionedtable(\n"
+"sensorid SMALLINT(3) UNSIGNED NOT NULL,\n"
+"datetime TIMESTAMP NOT NULL DEFAULT CURRENT_TIMESTAMP ON UPDATE CURRENT_TIMESTAMP, \n"
+"PRIMARY KEY (sensorid,datetime),\n"+ "temperature SMALLINT(3)) \n"
+"PARTITION BY RANGE (UNIX_TIMESTAMP(datetime)) \n(";
Date tr = new Date();
for (int i = 0; i < 29; i++)//for one hour simulation time
{
    tr = new Date(tr.getTime() + 120000);//2 minutes per partition
    DateFormat dateFormat = new SimpleDateFormat("yyyy-MM-dd HH:mm:ss");
    String itr = dateFormat.format(tr);
    sqlquery = sqlquery + "PARTITION p" + i + " VALUES LESS THAN (UNIX_TIMESTAMP('" + itr + "')),\n";
}
sqlquery = sqlquery + "PARTITION p29 VALUES LESS THAN MAXVALUE);";
```

Create table without partitions:

```
String sqlquery="CREATE TABLE IF NOT EXISTS unpartitionedtable("
+ "sensorid SMALLINT(3) UNSIGNED NOT NULL,"
+ "PRIMARY KEY (sensorid,datetime),"
+ "datetime TIMESTAMP NOT NULL DEFAULT CURRENT_TIMESTAMP ON UPDATE CURRENT_TIMESTAMP, temperature SMALLINT(3))";
```

The data generator and the query engine were implemented as two independent threads running in parallel. The data generator generated data at the rate of one data per sensor per second and repeated it for 3600 times, the result of which was a simulation of the data stream generated by the 5×5 WSN in one hour. The query engine performed a set of queries to the databases at the end of every alternative minute for 30 times across the one hour simulation period, and returned the resulting performance metrics at each point.

6.5.2 Performance metrics

The performance of the databases was evaluated on three aspects: query efficiency, updating efficiency and storage cost.

- **Query efficiency** was measured by the time duration it took to execute the tempo-spatial range query in each database, recorded in the built-in SQL diagnostic facility MySQL profiler.

- **Updating efficiency** was measured by the time duration it took to insert the real-time generated data into each database, recorded in the MySQL profiler.

- **Storage cost** of a database table (in KB) was calculated by the length of data file plus the length of index file, and then divided by 1024.

An example set of queries to retrieve the desired performance metrics is shown in Table 6-5.

Table 6-5: An example set of queries

Query the efficiency of time range queries:

SET profiling = 1;

SELECT * FROM test.partitionedtable WHERE datetime < *B* AND datetime > *A*;

SELECT * FROM test.unpatitionedtable WHERE datetime < *B* AND datetime > *A*;

SHOW PROFILES;

Query the updating efficiency:

SET profiling = 1;

INSERT INTO partitionedtable (sensorid, datetime, temperature) VALUES (*sid*, '*datetime*', *temperature*);

INSERT INTO unpartitionedtable (sensorid, datetime, temperature) VALUES (*sid*, '*datetime*', *temperature*);

SHOW PROFILES;

Query the storage cost:

SELECT table_name, table_rows, ROUND((data_length + index_length)/1024,2) 'KB' FROM information_schema.tables WHERE table_schema='test';

6.5.3 Performance evaluation

The experimental result of query efficiency is as shown in Table 6-6. It demonstrates that as the size of the databases becomes larger with time, the database with partitioning maintains the query efficiency, whereas the query efficiency of the database without partitioning drops very quickly. To execute the same time range query, more rows have to be scanned in the database without partitioning, whereas the number of rows needs to be scanned in the partitioned database remains stable. To find the answer to the same query, only the partition(s) that contains the answer to the query is scanned in the database with partitioning, whereas the whole dataset has to be scanned in the comparison database without partitioning. The time needed to find the answer to the query in the partitioned database decreased from approximately 10% to around 1% of the time needed for the same query in the database without partitioning. This demonstrates that the proposed DSM can achieve a query efficiency that is up to 100 times faster than the database without partitioning as the duration of the incident increases.

Table 6-6: Database query efficiency experimental result

Query Efficiency						
Size of database (rows)	Database without partitioning			Database with partitioning		
	Rows scanned	Partitions scanned	Time needed	Rows scanned	Partitions scanned	Time needed
5078	5078	all	0.31s	1342	P3	0.03s
61024	61024	all	3.86s	1500	P20	0.04s
70139	70139	all	5.97s	3000	P24, P25	0.08s

The experimental result of updating efficiency is shown in Figure 6-6. It can be observed that both databases can maintain the updating efficiency as the time duration increases, although the database without partitioning demonstrated a slightly bigger variation on its updating query duration. The average duration of executing updating queries on the partitioned database was calculated to be approximately 0.0367 second, and its standard deviation was approximately 0.0010 second. Very similar to that, the average duration of executing updating queries on the database without partitioning was approximated to be 0.0507 second, with standard deviation approximates 0.0009. The difference of the average query duration between the two databases was only 0.02 second. As a result of the analysis, it can be concluded that the partitioned database does not require extra processing time to perform updating queries. In fact, it can be slightly quicker than the database without partitioning.

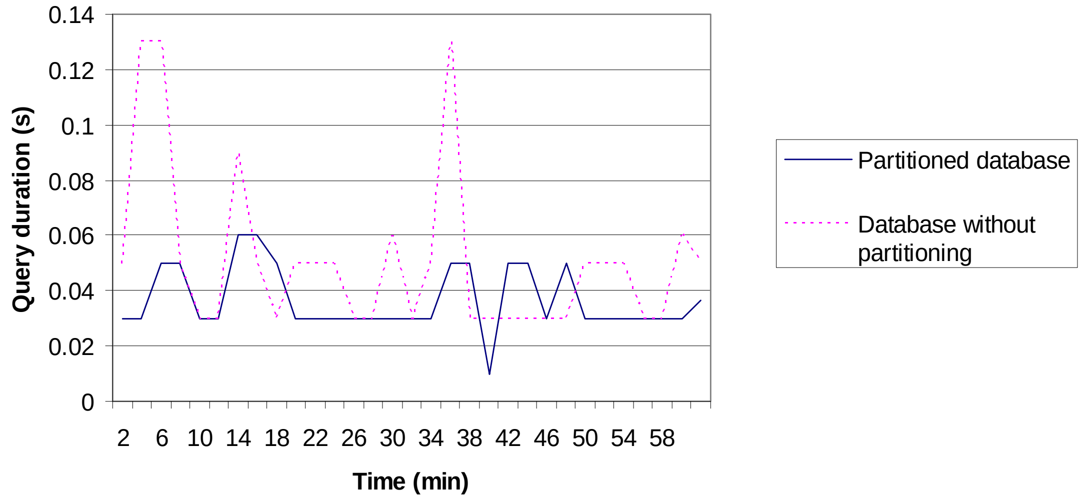


Figure 6-6: The comparison of updating efficiency

The experimental result of storage cost is shown in Figure 6-7. When the databases were empty, the partitioned database did require more storage space (512KB) than what the database without partitioning required (48KB). However, as the time duration increased, the database without partitioning quickly ran out of the initially allocated storage space and had to acquire extra space. As a result of the extra storage space acquisition, the storage space required by the database without partitioning became bigger than what the partitioned database required for the same amount of data. For example, at the end of 22 minutes in the experiment time scale, 1552KB was required by the database without partitioning, which was 432KB more than the 1120KB storage space required by the partitioned database. Another example is at the end of 42 minutes, where the database without partitioning required 752KB more than the partitioned database. In addition, if we denote the storage cost of the partitioned database at each minute t as $f_A(t)$, and the storage cost of the database without partitioning as $f_B(t)$, it can be calculated that the definite integral of $f_B(t)$ over the one hour simulation interval is 743KB more than that of $f_A(t)$, denoted as $\int_0^{60} f_B(t) - \int_0^{60} f_A(t) = 743$. This demonstrates that although the partitioned database requires more initial storage space, the overall storage space across time is less than what is required by the database without partitioning. Consequently, it can be concluded that the partitioning does not introduce unacceptable extra storage cost.

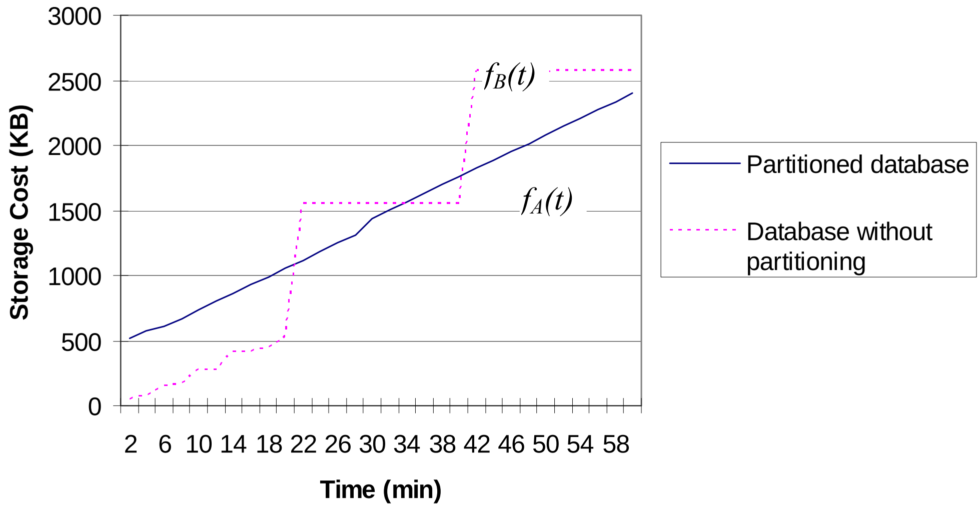


Figure 6-7: The comparison of storage cost

6.6 Conclusions

In this chapter, a DSM for WSN-based on-site EIMS has been proposed based on the understanding that a suitable DSM should address both the features of sensor data and ER application requirements. The features of sensor data include the streaming nature, the existence of high tempo-spatial correlations, redundancy, noises, and the importance of time and location related information. The general requirements of ER applications are accuracy and efficiency despite of its dynamic nature. The findings demonstrate that a suitable DSM for WSN-based on-site EIMS should 1) store and manage sensor readings, time-related information and location-related information; 2) support timely response to queries in spite of the increase in the volume of data; and 3) be dynamic and flexible to accommodate different needs at different incident stages.

The design of the proposed storage mechanism includes a database schema that maintains the tempo-spatial correlations typically existing in sensor data, a time-driven sensor data management, and adaptive support for different incident stages. It is designed to be efficient, flexible and suitable for the data flow, data collection and ER application requirements.

The performance evaluation demonstrated that the proposed data storage mechanism can support timely response to queries (up to 100 times faster than a database without partitioning) in spite of the increase in the volume of data. Since efficient querying can lead to an efficient response to emergencies, this data storage mechanism could be advantageous in Emergency Response Systems. Apart from the benefit of efficient querying, there is neither additional updating cost nor unacceptable storage cost introduced. This suggests that the proposed data storage mechanism would be particularly suitable for on-site ERS, where the resources are limited but high efficiency is required.

Chapter 7. Sensor Data Cleaning

	Research questions addressed in this chapter:
1	What constitute a suitable data cleaning approach for WSN-based on-site Emergency Information Management System (EIMS)?
2	What is the detailed design and characteristics of the proposed sensor data cleaning approach?
3	How does the proposed sensor data cleaning approach perform in experiments?

One of the features of sensor data is that it usually contains 'noise', as analysed in Section 6.2. Many traditional data mining methods do not have good tolerance with data that contains noise and uncertainty. In addition, under the scenario of using sensor data to assist the decision making of the Incident Commander (IC) and his command support team, the noisy sensor data can be misleading. Using the raw sensor data without dealing with the data quality issues can result in impractical or even harmful decisions, which in turns has negative impacts on the overall performance of Emergency Response (ER). Hence, data cleaning is necessary for improving the quality of data before applying data mining technologies.

Sensor data cleaning mainly deals with two data quality problems: noise and outliers. In statistics, an outlier is an observation that is numerically distant from the rest of the data (Barnett and Lewis, 1994). In the scenario of ER applications, an outlier is considered as an extreme sensor reading that is not caused by a real environmental change. Such outliers differ from regular system noise in the way that they tend to appear as strong peaks at a random frequency, whereas regular noise occurs at all frequencies with small intensities. Outliers might have an overwhelming effect on the total measurement and an overwhelming effect on smoothing methods. Although there has been some research effort on cleaning and querying noisy sensor data (e.g. Elnahrawy and Nath, 2003), it did not separate outliers from real environmental changes. The result of that was a compromise between the prior knowledge and the observed noisy sensor data, or in other words, simple smoothing.

In the scenario of ER applications, it is important to separate outliers from real environmental changes, and treat them differently. Therefore, a sensor data cleaning approach suitable for on-site EIMS should aim to not only reduce noises and remove outliers in the sensor data, but also quickly detect any real environmental changes.

This chapter describes the proposed sensor data cleaning approach, namely, state-aware Kalman Filter. The work presented in this chapter was based on the author's published paper (Yang, Y. et al., 2010a).

7.1 The Need for a Suitable Sensor Data Cleaning Approach for on-site ER

Apart from the discussed requirement of separating outliers from real environmental changes and treating them differently, it is also important to seek an efficient and light-weight data cleaning approach. Efficiency means low time cost, light-weight means low resource consumption. These two characteristics are important due to the requirement of timeliness for ER applications and the limited resources on-site. This section reviews the available data cleaning approaches in terms of their suitability for on-site ER.

Researchers have used approaches such as Bayesian Theory (Elnahrawy and Nath, 2003), Neural Network (Petrosino and Staiano, 2007), Wavelets (Zhuang and Chen, 2006), Kalman Filter and Weighted Moving Average (Zhuang, Y. et al., 2007) for sensor data cleaning. However, using probability-based theory to predict the most likely range that sensor sampling falls in has its limitation in application to ER, because emergencies are unpredictable events, and using probabilities learned from data collected from one environment to predict sensor samplings in another environment may not be appropriate. Approaches based on Neural Networks are usually based on theories and simulations, but the possibility of implementing such a complex system in practical situations has not been seen. Wavelet transformation based methods require an initial dataset large enough to be able to separate noise from real data, and is therefore more suitable for analysis of historical data than real time data. Compared to the above approaches, Weighted Moving Average is easy to implement, but its flexibility is limited and the neighbour tests phase might be an issue because it may result in neighbours checking each other, causing overflowed checking requests. Kalman Filter demonstrated both good performance in filtering data and light-weight implementation, consequently, it was chosen for further research.

7.2 The Proposed State-Aware Kalman Filter

Kalman Filter was chosen as the benchmark because it is the most efficient, flexible and light-weight compared to other types of existing data cleaning approaches. Nevertheless, whether it is a satisfactory fit for on-site ER is not clear. Therefore, an initial performance evaluation of Kalman Filter was carried out first. Based on the result of the initial performance evaluation, a set of state-awareness rules is proposed. It is proposed to integrate the state-awareness rules with Kalman filter, and the integrated result is named state-aware Kalman Filter.

7.2.1 Kalman Filter

This section briefly summarizes the iterative cycle of Kalman Filter. Kalman Filter addresses the problem of estimating the state x of a system represented by the model $x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1}$, which calculates the state x_k from its previous state x_{k-1} and control input u_{k-1}, with a measurement model $z_k = Hx_k + v_k$. The random variables w_{k-1} and v_k represent the process and measurement noise respectively. They are assumed to be independent of each other, white, and with normal probability distributions, denoted as: $P(w) \sim N(0,Q)$, $P(v) \sim N(0,R)$. In a high-level overview, the iterative Kalman Filter cycle is as shown in Figure 7-1:

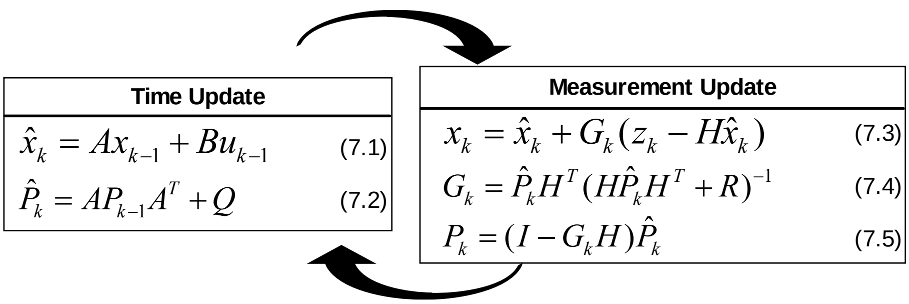


Figure 7-1: The Kalman Filter cycle

Step 1: Time Update. At each time instance, the estimated state $\hat{x}_k$ is calculated from the previous state of the system using equation (7.1), and the error covariance of the estimation $\hat{P}_k$ is calculated by equation (7.2).

Step 2: Measurement Update. The current state of the system x_k is then calculated by updating x_k using the current measurement z_k and Kalman Gain G_k, as shown in equation (7.3), where G_k is calculated by equation (7.4). The updated error covariance P_k is calculated using equation (7.5).

The time update step predicts the next coming measurement, the measurement update step corrects the prediction using the actual measurement. The updated x_k and P_k at the current time instance will then be taken as feedback to calculate the estimation $\hat{x}_k$ and $\hat{P}_k$ for the next time instance.

More about Kalman Filters can be found in the reference (Welch and Bishop, 1995).

7.2.2 The initial simulation of Kalman Filter

Process model

The initial simulation of Kalman Filter considered the scenario of a thermal sensor network that measures the temperature of an indoor environment. The coefficients of the Kalman Filter process model were selected as follows:

The room temperature can be considered stable, which means the temperature at one time instance *k* should be the same as the temperature at the previous time instance *k-1*, therefore coefficient A = 1. There is no control input, therefore $u = 0$. As a result, the system model can be rewritten as $\hat{x}_k = x_{k-1} + w_{k-1}$. Sensors directly measure the actual temperature in the environment, therefore coefficient H = 1. Hence, the measurement model can be rewritten as $z_k = x_k + v_k$. The process noise w_{k-1} and measurement noise v_k were assumed to be white Gaussian noise, and with normal probability distributions $P(w) \sim N(0,Q)$, $P(v) \sim N(0,R)$.

Substituting coefficients A = 1, H = 1 into the filter equations, the resulting equations are:

Time update:

$$\hat{x} = x_{k-1} \tag{7.6}$$

$$\hat{P}_k = P_{k-1} + Q \tag{7.7}$$

Measurement update:

$$x_k = \hat{x}_k + G_k(z_k - \hat{x}_k) \tag{7.8}$$

$$G_k = \hat{P}_k(\hat{P}_k + R)^{-1} \tag{7.9}$$

$$P_k = (I - G_k)\hat{P}_k \tag{7.10}$$

Data simulation

The Kalman Filter was implemented and applied on data simulated in Matlab. The simulated data represents sensor data streams from a 5×5 thermal sensor network for a period of 300 seconds. Gaussian white noise (0 mean, standard deviation 1.5°C) was applied to the pre-set normal room temperature (20°C) throughout the simulation time. The random outliers were simulated as spikes occurring at random locations at random time. The scenario of fire occurrence and development was simulated by a sudden rise of sensor value from normal room temperature to abnormal conditions at a random time, a random location, spreading to the nearby area, and lasting for a random duration. An example plot of the simulated data is shown in Figure 7-2.

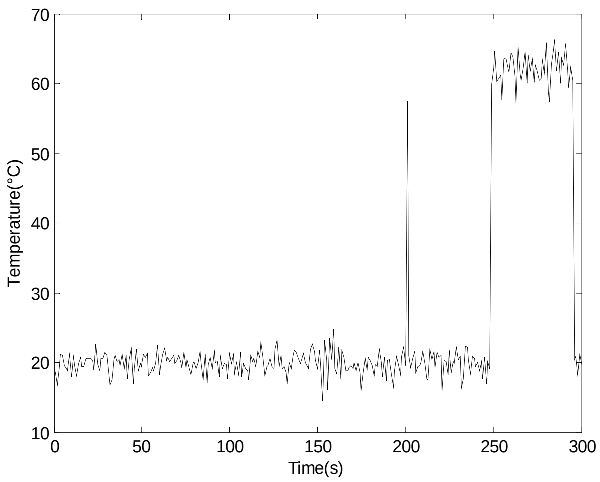


Figure 7-2: An example plot of the simulated data

Performance evaluation

The initial temperature x_0 was set to 20°C, the initial error covariance P_0 was set to 10. Q and R were tuned to achieve the best filter performance, i.e. the best balance between smoothing noises and detecting real changes.

$$Q = \begin{bmatrix} 0.5773 & 0.5413 & 0.5585 & 0.5471 & 0.6222 \\ 0.5313 & 0.5407 & 0.6249 & 0.5311 & 0.5038 \\ 0.6510 & 0.6665 & 0.6891 & 0.5358 & 0.5370 \\ 0.5110 & 0.5682 & 0.5979 & 0.5894 & 0.6122 \\ 0.6185 & 0.5701 & 0.5669 & 0.5250 & 0.6000 \end{bmatrix}$$

$$R = \begin{bmatrix} 1.9857 & 1.9057 & 2.1727 & 2.1957 & 2.2284 \\ 2.2085 & 2.4619 & 2.4967 & 2.0198 & 2.1613 \\ 2.0618 & 2.0046 & 2.1913 & 2.4201 & 2.1583 \\ 2.2221 & 2.2478 & 2.2643 & 2.6023 & 2.5113 \\ 1.9209 & 2.2334 & 2.4676 & 2.3931 & 2.1422 \end{bmatrix}$$

The result of the initial simulation of the Kalman Filter is as shown in Figure 7-3.

The initial simulation demonstrated that the Kalman Filter is efficient and has a good smoothing effect. However, the outlier (at 38 seconds in the simulation time scale) was only reduced but not effectively removed. In addition, it took time (20 seconds) to detect the real change in the environment (at 53 seconds in the simulation time scale).

Thus, it can be concluded that although the efficiency of Kalman Filter has been proved, using only Kalman Filter is not a satisfactory solution, because it does not separate outliers from real environmental changes to be dealt with differently. Therefore, a set of state-awareness rules was proposed and integrated with Kalman Filter to overcome these issues.

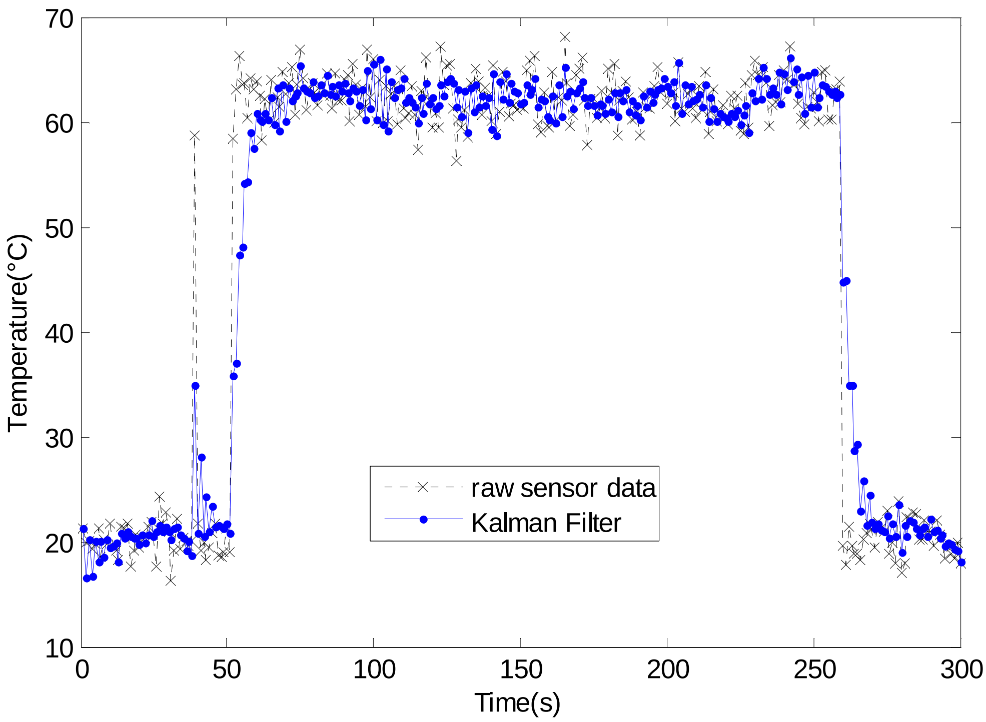


Figure 7-3: The result of the initial simulation of Kalman Filter

7.2.3 State-awareness rules

The fact that most data cleaning algorithms, including Kalman Filter, usually produce a simple compromise between the observed raw data and system estimation (prior knowledge), is because they do not take the state of the system into consideration. In other words, they are not state-aware. In this section, a set of state-awareness rules is proposed to address this issue.

The system can be in normal, outlier or abnormal state. The aim of the proposed data cleaning approach is to reduce noise, remove outliers and quickly detect the abnormal state. Assumptions were made that sensor nodes are organized as clusters in the WSN.

Rule (i): If the incoming observation z of a sensor node is within a pre-defined range $<O_l, O_u>$, and its rising rate r is within a pre-defined range $<R_l, R_u>$, then the sensor node is considered to be in a normal state.

Rule (ii): If either the current observation z or the rising rate r of a sensor node is beyond its pre-defined range, then the sensor node reports to be suspicious.

Rule (iii): If both the sensor node and one of the sensor nodes in the same cluster report to be suspicious, then the sensor node is considered to be in an abnormal state.

Rule (iv): If the sensor node reports to be suspicious, but none of the other sensor nodes in the same cluster report to be suspicious, then the sensor node is considered to be in an outlier state.

7.2.4 State-awareness rules integrated with the Kalman Filter

It is proposed to integrate the set of state-awareness rules with the Kalman Filter. The integrated Kalman Filter adjusts the projected estimation differently according to the state of the sensor, thus can remove outliers as well as quickly detect the environmental changes. As a result, it has been termed a state-aware Kalman Filter. The on-going cycle of the state-aware Kalman Filter as shown in Figure 7-4 can be described as follows:

Step 1: Time Update/Suspicious Detection At each time instance, the estimated state $\hat{x}_k$ and the error covariance of the estimation $\hat{P}_k$ are estimated using equation (7.1) and (7.2) respectively, at the same time detecting for any occurrence of a suspicious state using rule (ii).

Step 2: Measurement Update On detection of a suspicious state, the sensor nodes in the same cluster are checked to confirm it as an abnormal state or classify it as an outlier. If it is an abnormal state, measurement z_k is used as x_k, and P_k is increased α times to quickly detect the change; If it is an outlier state, the last normal data x_{k-1} is used as x_k. If a normal state is found, the normal measurement update step of Kalman Filter is used.

The time update/suspicious detection step estimates the next coming measurement and detects any occurrence of a suspicious state, the measurement update step corrects the estimation differently according to the given state. The updated x_k and P_k at the current time instance will then be taken as feedback to calculate the estimation $\hat{x}_k$ and $\hat{P}_k$ for the next time instance.

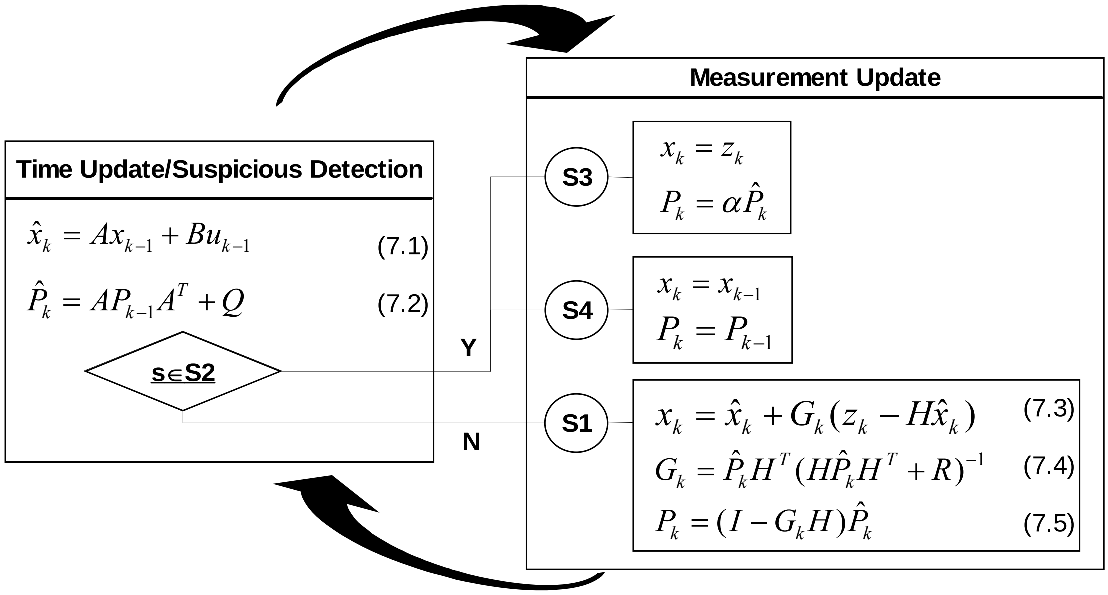


Figure 7-4: The iterative cycle of the state-aware Kalman Filter

7.3 Performance Evaluation

The performance evaluation experiments of the proposed sensor data cleaning approach, known as state-aware Kalman Filter, in comparison with Kalman Filter, were implemented on both simulated sensor network data and sensor data collected from field trails. The experiments aimed to evaluate the performance of the proposed data cleaning approach in comparison with Kalman Filter on the following three aspects:

- Reducing noise
- Removing outliers
- Detecting real changes in the environment.

The hypothesis was that the proposed data cleaning algorithm can reduce noise, remove outliers as well as quickly detect any real change in the environment that indicates a potential fire.

7.3.1 Simulation implementation and results

The simulation of the data cleaning approaches was implemented under the scenario that multiple types of sensors (temperature, smoke, CO and flame) are deployed in an indoor environment to monitor fire emergencies. The block diagram of the simulation system is shown in Figure 7-5.

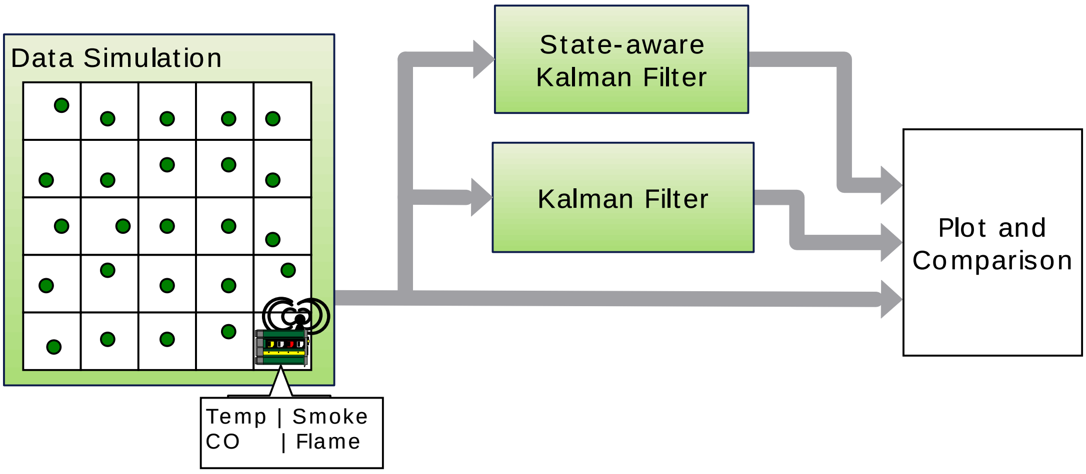


Figure 7-5: The block diagram of the simulation system

Data Simulation

In comparison with the initial simulation of Kalman Filter where only temperature sensor data was simulated, the performance evaluation of the proposed data cleaning approach simulated a WSN that consists of multiple types of sensors.

Data streams from a 5×5 sensor network for a period of 300 seconds were simulated using MatLab. Each node in the network consisted of four types of sensors (temperature, smoke, CO, flame). According to the domain knowledge, the observations of different types of sensors under normal room conditions were assumed to be:

- temperature: 20℃
- smoke: 0.5% obs/m (obscuration per meter)
- CO: 2.5ppm
- flame: 350nm.

Gaussian white noise with zero mean was simulated for each type of sensors, with standard deviations set as:

- temperature: 0.5℃
- smoke: 0.04% obs/m

- CO: 0.2 ppm
- flame: 1.5nm.

1% random outliers (spikes) were added to each type of sensor data at random locations at random time. The scenario of fire occurrence was represented by a sudden change of sensor value from room conditions to abnormal conditions at a random time, a random location, and spreading to the nearby area.

An example plot of the data simulation with multiple sensors is shown in Figure 7-6.

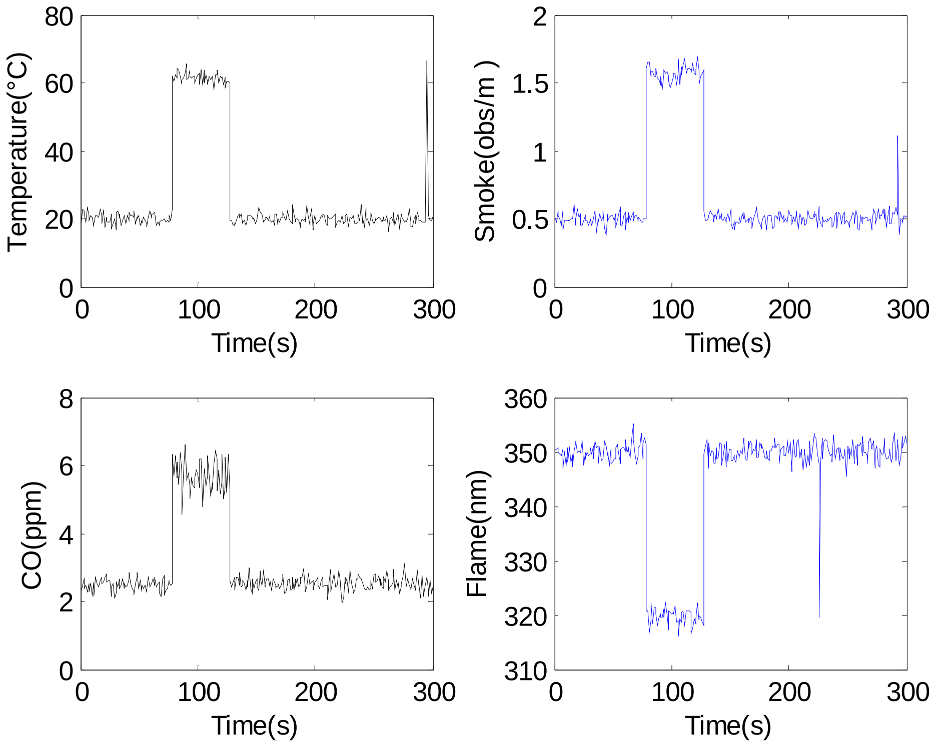


Figure 7-6: An example plot of data simulation with multiple sensors

Parameters selection

The proposed data cleaning approach and Kalman Filter were individually applied on each type of sensor data. The same initial filter parameters were chosen for both of them:

- The initial measurement x_0 was chosen as 20℃ for temperature sensor, 0.5% obs/m for smoke sensor, 2.5ppm for CO sensor and 350nm for flame sensor.
- The initial error covariance P_0 was set to be 10 for all types of sensors.

- Filter parameters Q and R were tuned to achieve the best performance.

Performance evaluation

The performance evaluation of data cleaning approaches on temperature sensors are shown in Figure 7-7. Both data cleaning approaches demonstrated good noise reduction performance. The proposed state-aware Kalman Filter effectively removed the outliers (e.g. at the 40^{th} second), whereas Kalman Filter only reduced them. In addition, there was a delay in detecting the fire occurring (10 seconds from the time the event occurred to the time it was detected) by Kalman Filter, whereas the proposed state-aware Kalman Filter reduced this delay to 1 second.

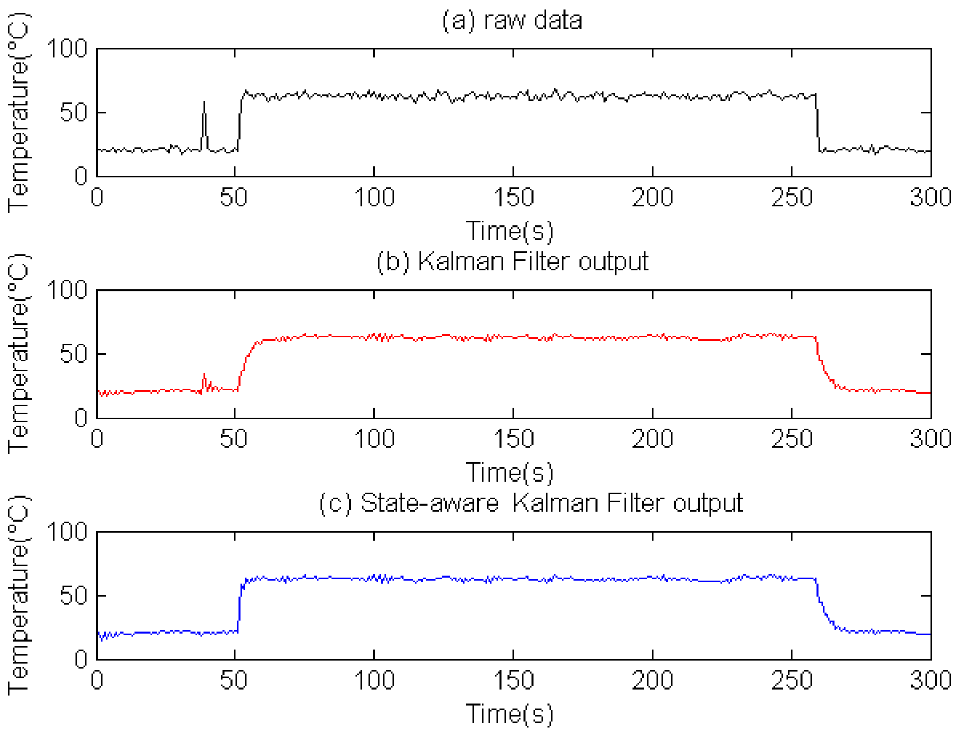


Figure 7-7: Performance evaluation of data cleaning approaches on temperature sensors

The performance evaluation of data cleaning approaches on smoke (Figure 7-8), CO (Figure 7-9) and flame (Figure 7-10) sensors demonstrated similar results to temperature sensors. Noises in all the types of sensor data were reduced by the Kalman Filter, but there was a delay in detecting the fire occurring (12s delay of detecting the fire occurring in flame, 2s delay of detecting it in smoke and CO). These delays were reduced to 1s by the state-aware Kalman Filter. In other words, the state-aware Kalman Filter can reduce the delay of detecting environmental changes on sensors by

50% to 90%. Using Kalman Filter only reduced outliers (at 204 seconds in smoke, at 147 seconds in CO, at 57 seconds in flame), whereas outliers existing in the raw data were effectively removed by the proposed state-aware Kalman Filter.

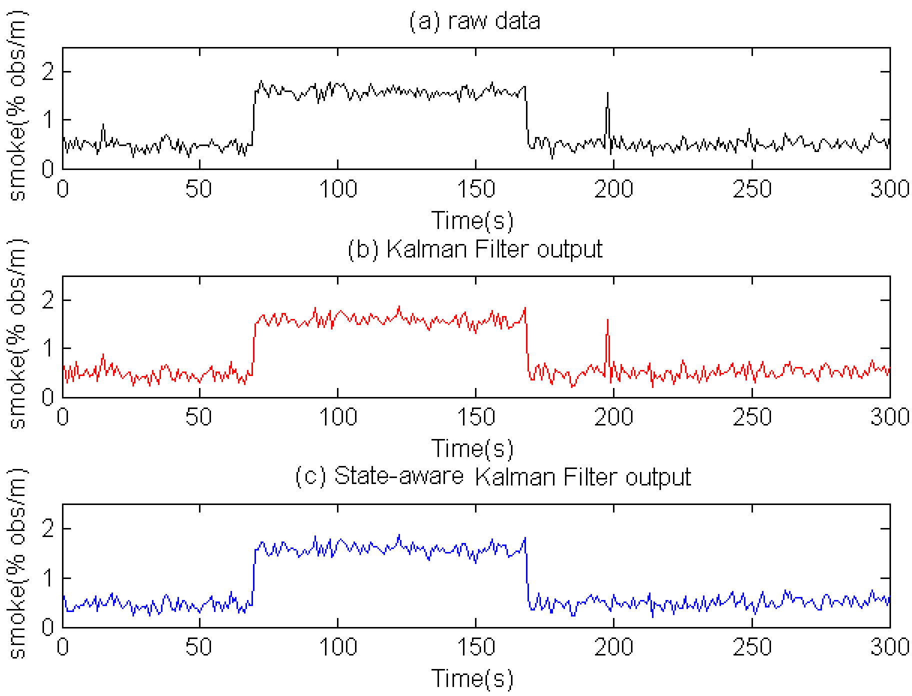


Figure 7-8: Performance evaluation of data cleaning approaches on smoke sensors

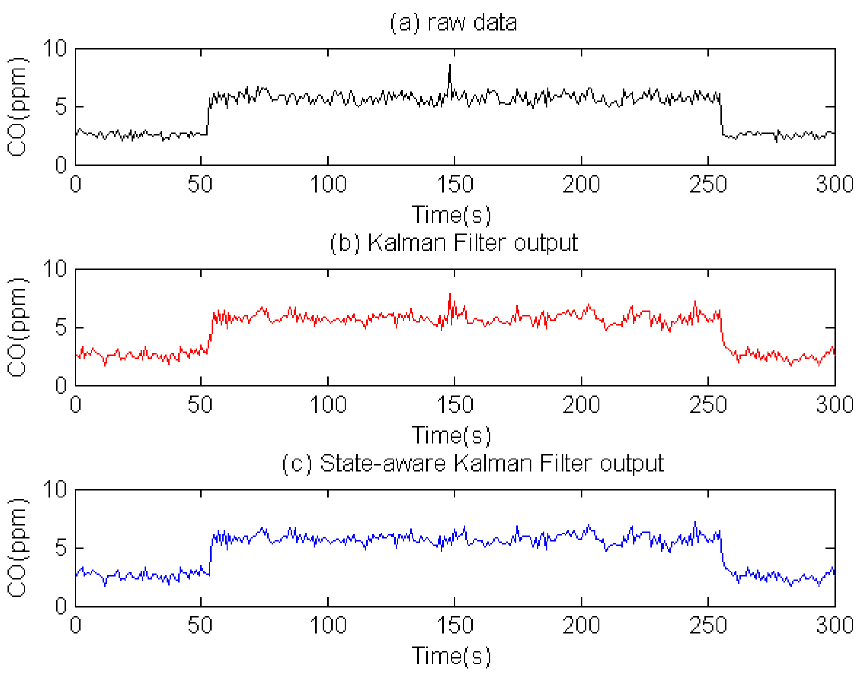


Figure 7-9: Performance evaluation of data cleaning approaches on smoke sensors

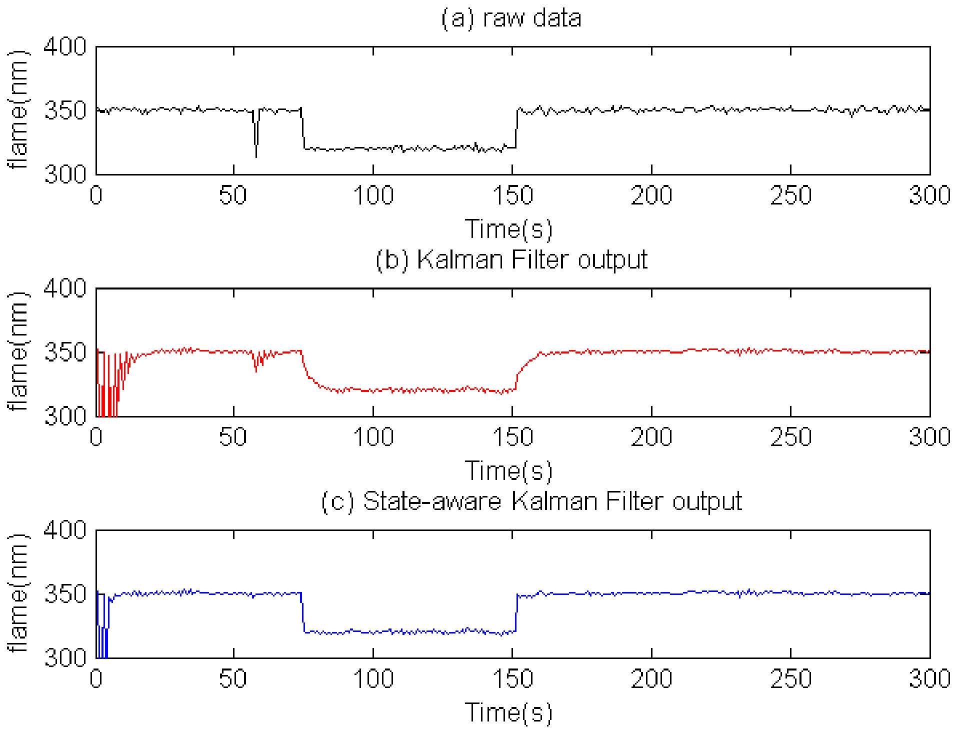


Figure 7-10: Performance evaluation of data cleaning approaches on flame sensors

Consequently, it can be concluded that the state-aware Kalman Filter demonstrated a better performance than using Kalman Filter alone. It can effectively remove outliers as well as efficiently detect real changes in the environment.

7.3.2 Performance evaluation based on data collected from the field trial

To verify the performance evaluation result based on simulation, the two data cleaning approaches were also applied on the sensor data collected from a field trial.

The field trial was carried out in the training centre of a local fire and rescue service station on 17/03/2010, between 11:21 and 11:50 am. Twenty sensor nodes were randomly deployed in each room and on the stairs. Each node consists of four types of sensors: temperature, smoke, CO and flame. A controlled gas fire was switched on during the field trial, and the sensor data collected before, during and after the fire was recorded in MySQL database.

A MySQL database connector was implemented in Matlab, to retrieve the sensor data and feed it into the data cleaning approaches. The block diagram of the test system is shown in Figure 7-11.

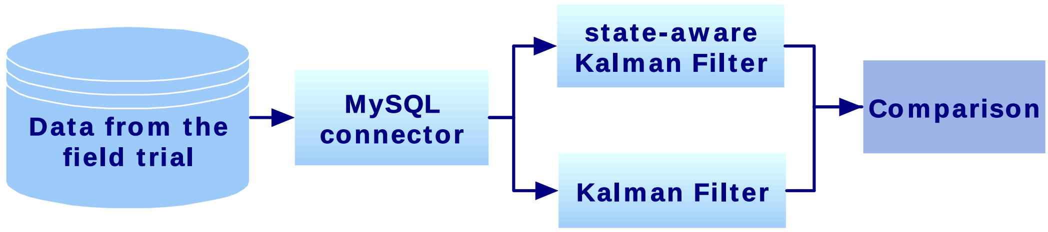


Figure 7-11: The block diagram of testing system based on field trial data

An example plot of the collected raw sensor data of a node is shown in Figure 7-12. Data from node 3 is used as an example since it was the closest node to the gas fire and its data is the most representative of the whole picture of the fire. Each time slice represents 4 seconds of time.

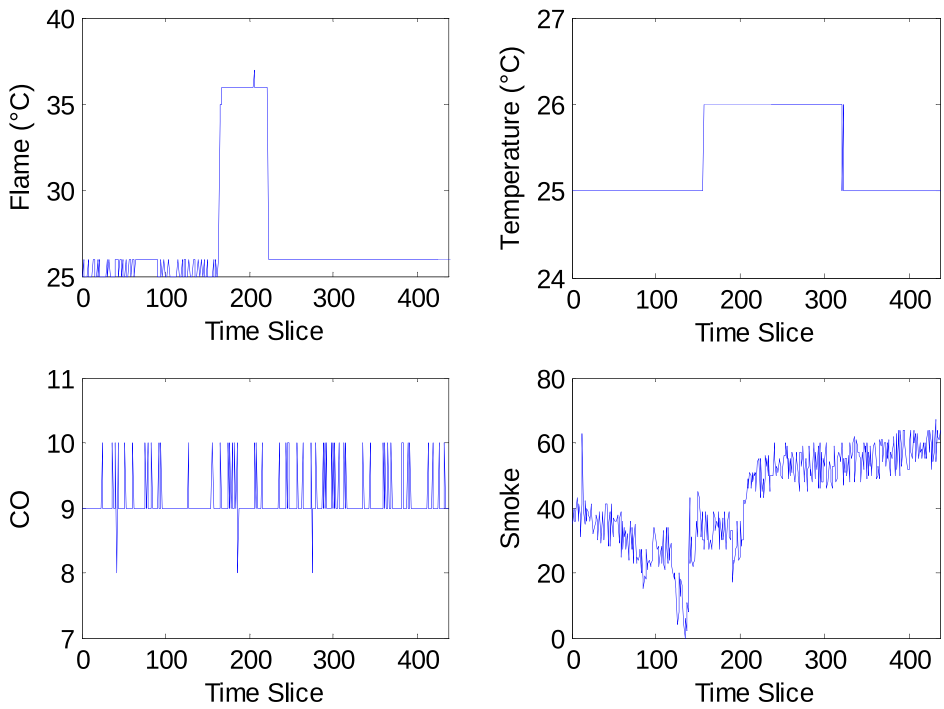


Figure 7-12: An example plot of raw sensor data (NodeID: 3)

It can be observed from the collected data stream that outliers did exist in a real situation. They occurred to all types of sensors. Smoke and CO sensors are more subject to the effects of noise and outliers, whereas data from flame and temperature sensors is relatively “cleaner”.

Some characteristics of a fire can also be observed from the raw sensor data collected from the field trial. Different types of sensors have different levels of sensitivity. The smoke sensor was the first to detect the fire, immediately followed by the flame sensor, then followed by the temperature sensor. The CO sensor didn't detect the fire, possibly because the fire was a controlled flaming gas fire rather than a smouldering fire. There was enough oxygen, hence not much CO was generated from the burning gas.

The result of temperature sensor data cleaning is shown in Figure 7-13. Using Kalman Filter only reduced the outlier at the 318^{th} time slice, whereas the state-aware Kalman Filter removed the outlier. There was an 8 time slice (32 seconds) delay in detecting the fire occurrence using only Kalman Filter, whereas the time delay was reduced to 4 seconds using the state-aware Kalman Filter.

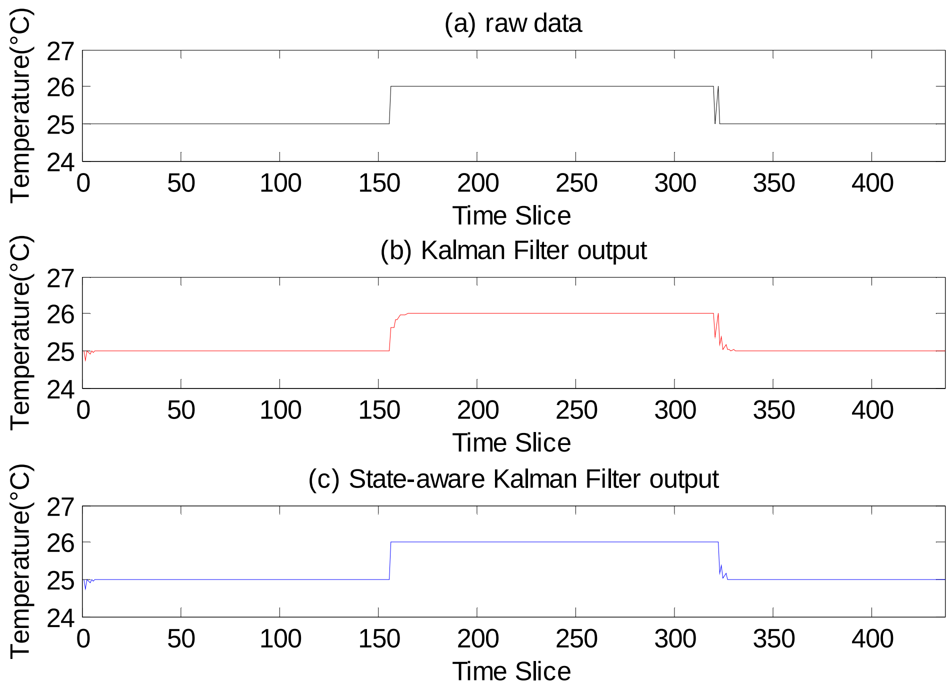


Figure 7-13: Temperature sensor data cleaning

The result of flame sensor data cleaning is shown in Figure 7-14. Both data cleaning approaches smoothed the noise existing in the raw data. The delay of detecting the fire was 13 time slice (52 seconds) using only Kalman Filter, whereas the delay was reduced to 4 seconds using the state-aware Kalman Filter.

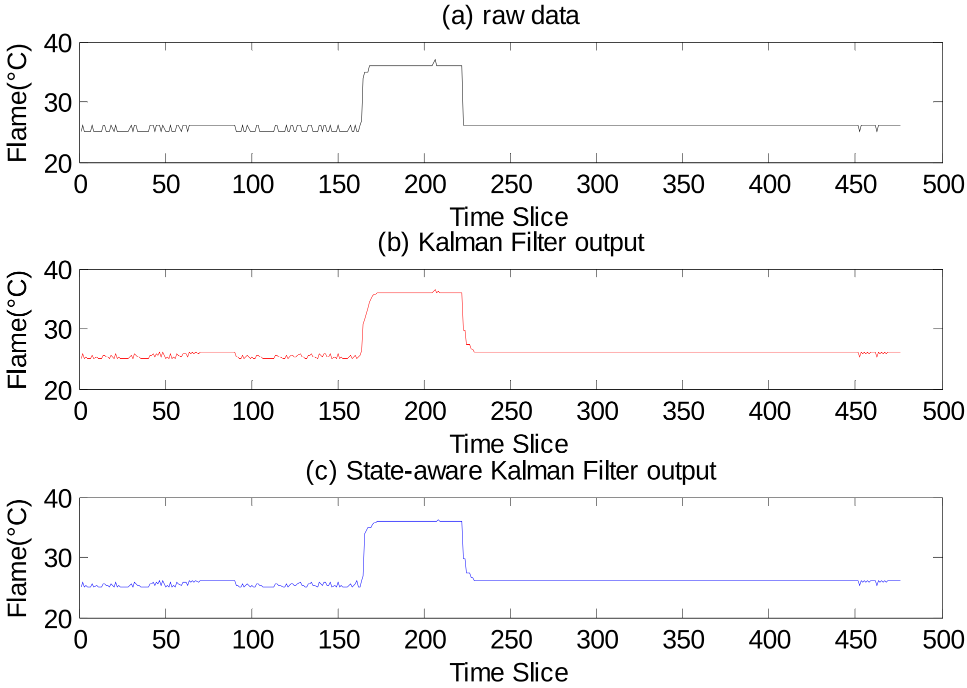


Figure 7-14: Flame sensor data cleaning

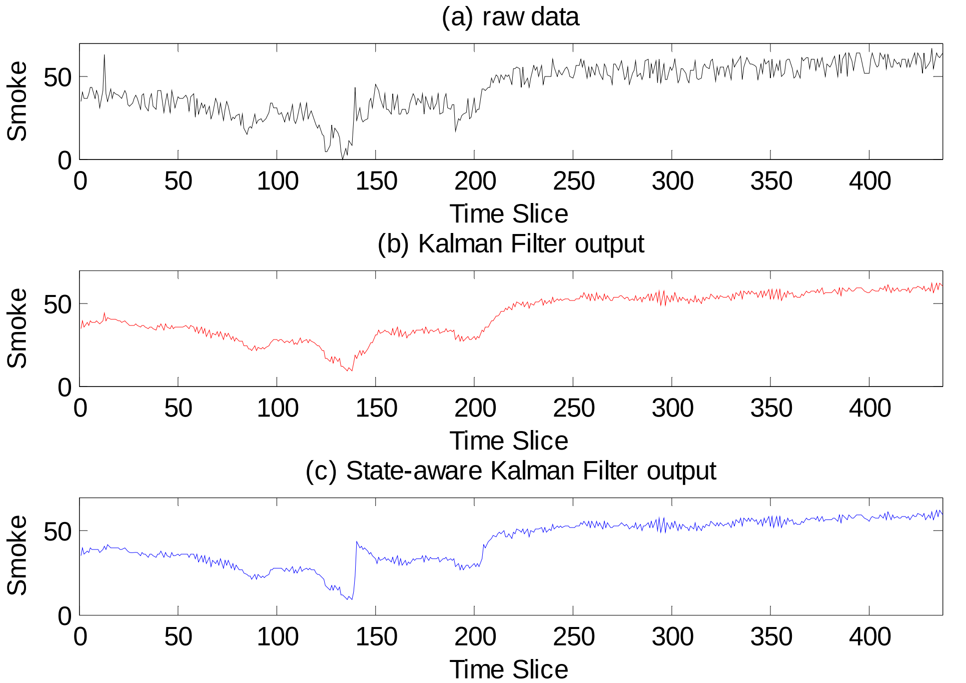


Figure 7-15: Smoke sensor data cleaning

The result of smoke sensor data cleaning is shown in Figure 7-15. It can be observed that smoke sensor data is more subject to noise, despite its efficiency in detecting the fire at its early stage. Due to this feature, tuning the filters to reduce the noise will introduce longer delay in detecting the real change in the environment that represents a possible fire. In this situation, the advantage of the proposed state-aware Kalman Filter in terms of maintaining the balance between noise reduction and quick detection of real environmental changes is more obvious.

The result of CO data cleaning is shown in Figure 7-16. It can be observed that the same noise reduction performance was achieved by the two data cleaning approaches. The controlled gas fire was not detected in the data from the CO sensor. Nevertheless, using the two data cleaning approaches still demonstrated benefits of smoothing out the noise.

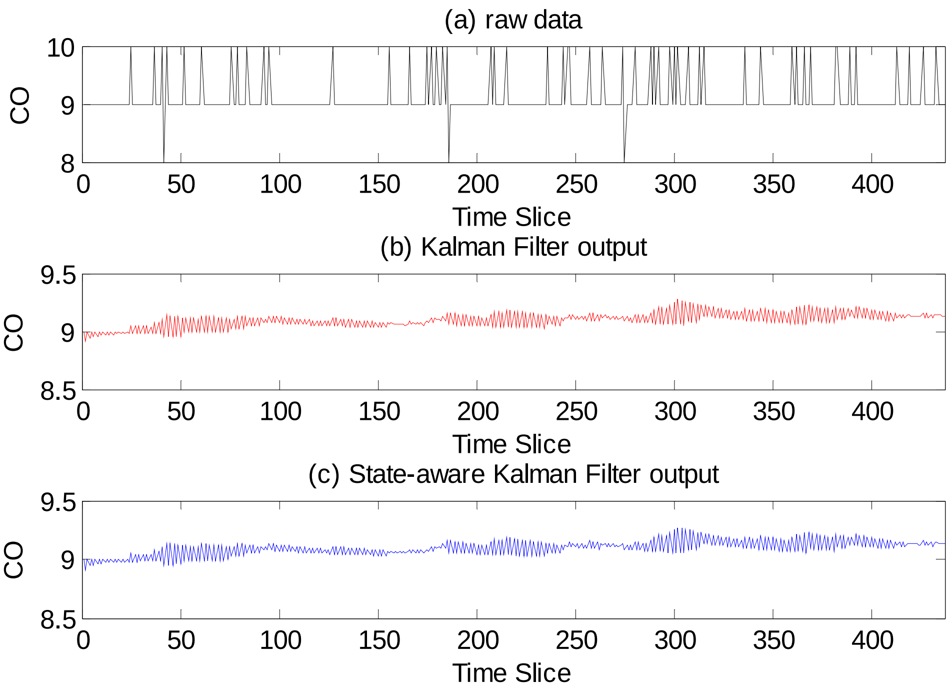


Figure 7-16: CO sensor data cleaning

7.4 Conclusions

This chapter proposed a sensor data cleaning approach for ER applications that not only reduces noise, but also removes outliers and quickly detects the real environmental changes.

The design of the proposed data cleaning approach addressed the noisy feature of sensor data analysed in section 6.2. It also addressed the different states that sensor data can be in, and utilised spatial correlations between sensor nodes to check and confirm whether a suspicious sensor data is an outlier or a real environmental change.

The data cleaning experiments demonstrated that by introducing the state-awareness rules and integrating it with Kalman Filter, the resulting state-aware Kalman Filter can separate outliers and real environmental changes and deal with them differently. It can be argued that being state-aware is a new advantage in addition to those inherited from Kalman Filter: e.g. effectively smoothing of noise and being light-weight. As a result of such advantages, it can be concluded that the proposed state-aware Kalman Filter would be a satisfactory fit for on-site ER applications.

In addition, lessons have been learned from this study that can provide benefits to future research. Interesting characteristics of fire have been discovered from the data collected from the field trial, e.g. the order of the detection of fire in different sensors. These observed characteristics of a real fire can be incorporated into future data simulation.

Chapter 8. Meaning Extraction from Sensor Data

	Research questions addressed in this chapter:
1	What can be defined as “meaning” in the context of on-site ER?
2	What is the detailed design and characteristics of the meaning extraction approach?
3	How does the proposed meaning extraction approach perform in experiments?

Meaning extraction is an important component in the proposed Emergency Information Management System (EIMS) architecture, as well as a necessary step to make sensor data work for Emergency Response (ER). On one hand, as the analysis of the features of sensor data demonstrated in Section 6.2, sensor data is meaningless unless associated with time and location. On the other hand, individual sensor readings, even if they have been cleaned, are not what Incident Commanders (ICs) expected to see (Chapter 4). Requirements from both the features of sensor data and ER applications have demonstrated the necessity of extracting meanings and semantics of sensor data in order to provide support for ICs.

It can be argued that the process of making sensor data work for ER does not complete until all three steps are accomplished: (1) it needs to be properly pre-processed, (2) it must be stored and managed efficiently, and (3) meaning must be extracted from the data prior to its presentation to the emergency responders. A data storage mechanism design has been proposed for the storage and management of real-time sensor data streams and their associated time and location information (Chapter 6). A sensor data cleaning method has been proposed to reduce noise and data outliers and quickly detect real environmental changes (Chapter 7). These works have laid a promising basis in the process of making sensor data work for emergency response. This chapter aims to investigate the next step in the process, which is meaning extraction from sensor data.

Meaning extraction is an emerging technology that identifies elements of information contained in datasets that imply meaning in the context of application and can be interpreted by the users to facilitate their tasks. This chapter described a comprehensive analysis on what can be defined as meaning in the context of ER, and the study undertaken for extracting an example of the possible meanings: the occurrence and characteristics of an incident. Focusing on the extraction of the occurrence and characteristics of an incident from sensor data converts the meaning extraction problem into an event detection problem. A generic model for event detection from sensor data is proposed and simulation results are discussed.

8.1 The Need for a Meaning Extraction Method for on-site ER

Meaning extraction has been applied in a wide variety of domains, e.g. natural language processing, web semantics analysis, as well as text mining. In the context of natural language processing, "meaning" has been defined as the part of sentence, e.g. Subject part, verb part, object part, and adverb part (Bajwa, 2010). In the context of web semantics analysis, "meaning" has been defined as the similarity distances of literal objects (Cilibrasi and Vitanyi, 2006). In the context of text mining, "meaning" has been defined as acronyms and their meanings (Kempe, 2006). According to the various contexts of applications, different information from data has been considered as "meaning". This demonstrated that meaning extraction is highly domain-specific. As a result, the meaning extraction methods utilised are highly related to the specific meaning extraction problems.

In the context of meaning extraction on sensor data streams, the typical way of defining "meaning" is to represent it in the form of meta-data, features or frequently occurred patterns. Kariya and Kiyoki (2005) used output from taste sensors to compute meta-data of taste impression, which implies the relations between different food and preferences. Hunter and Colley (2007) proposed an online unsupervised learning method to analyse human behaviour in real-time by extracting features that can represent places visited and routes taken between places from sensor streams. Dong and Calvo (2009) addressed the problem of the interestingness of the automatically-extracted patterns (association rules) and proposed to integrate user-specified interests to filter the large amount of statistically significant association rules.

Despite the wide applications of meaning extraction, there is very limited research on meaning extraction in the context of ER, even less research on meaning extraction from sensor data for ER. Wickler and Potter (2009) proposed to derive features from given sensor data which will be or be very close to information to fire-fighters, e.g. the height of smoke layer in a room, as a necessary step to provide decision support. However, there is a lack of a comprehensive analysis of what can be defined as "meaning" that applies to ER in general and the detailed technology proposal to extract the defined meaning. Hu et al. (2007) focused on extracting flood area for emergency response of flood disaster. However, the proposed method is based on radar data, which has different features from sensor data, therefore it is not suitable. There is no other research in this category according to the author's best knowledge

The existing meaning extraction methods typically utilised statistical methods, including Markov Logic (Bajwa, 2010), Bayesian networks, decision trees, logistic regression, neural network etc. However, statistically significant patterns may not be of significance in the context of ER because emergencies such as fires are usually events of low probability. There is also a lack of research on meaning extraction for on-site ER.

8.2 Ways of Conveying Meaning in the Context of ER

Chapter 4 identified a list of opportunities for technology that can provide the required information and help ICs and their command support team better achieve their goals. Although sensor data can contribute to most of the discovered opportunities for technology, ICs and their command support team would benefit the most from the ability of seeing inside the building and getting this information during the mobilisation phase (on their way to the premises). Therefore, these two opportunities were further analysed to extract what information can be considered as "meaning" in terms of extracting it from sensor data.

It is proposed the following pieces of information can be considered as "meaning" in the context of ER:

- The occurrence and characteristics of an incident

The issue of the availability of information immediately after an incident has been revealed both in the literature and in interviews with ICs. "There is a significant lack of information about the scale of a disaster in the immediate aftermath" (Manoj and Baker, 2007). Therefore, extracting the occurrence and characteristics of an incident based on sensor data and providing them to the emergency responding crew during the mobilisation phase can fill in the gap of the lack of information about an incident in the immediate aftermath. It can also help them see inside the building and facilitate early planning and preparation.

More specifically, it is proposed that the occurrence of an incident can be represented by an alarm associated with confidence, whereas the characteristics of an incident can be represented by the time, the affected area, the type and the severity of the incident.

- Real-time development of an incident

The lack of available information about an incident in the immediate aftermath is followed by large amounts of imprecise information (Manoj and Baker, 2007). The interviews with ICs revealed that the imprecise information is due to lack of ability to retrieve the required information via reliable technology means. Hence, providing the real-time development of an incident can enable ICs to see inside the building, and facilitate their dynamic risk analysis.

More specifically, it is proposed that the real-time development of incident can be represented in three levels, perception level (multiple resolution of the distribution of sensor readings in real-time), comprehension level (the distribution of level of seriousness in real-time), projection level (the direction, speed of incident development; the projected distribution of level of seriousness; the projected potential risks).

Meaning can be conveyed by a number of ways. However, this chapter mainly concentrates on the first one: the size and characteristics of an incident. By defining the occurrence, size and characteristics of an incident as meaning, the meaning extraction problem is converted to event detection problem.

8.3 Event Detection

The problem of event detection can be introduced more formally as follows:

Given a set of measured data arriving over time, denoted as $D = \{z_t \mid t = 1, 2, 3, \ldots n\}$, event detection is to find the time t when an event of interest occurs (where the data is different from normal pattern of behaviour).

Therefore, the common goals of event detection are:

- To identify whether an event of interest has occurred
- To characterize the event (e.g., the time, the affected area, the type and the severity of the event)
- Accurate and early detection

Kerman et al. (2009) stated that the most common challenges in event detection are: situational dependence, criticality of application, numerous and diverse data sources, and network topology. They suggested that event detection algorithms should overcome those challenges as well as meet the main requirements: timeliness, a high true detection rate, and a low false alarm rate. In the case of emergency event detection from multi-sensory data, the requirements on detection timeliness and accuracy still apply, in addition, special challenges (e.g., temporal spatial information incorporation, information integration from multiple sensor streams, computational complexity) presented by sensor data should be taken into consideration.

Two categories of event detection approaches have been identified in sensor network applications: threshold-based event detection and tempo-spatial pattern based event detection.

8.3.1 Threshold-based event detection

Threshold-based event detection method is based on the underlying intuition that an event occurring will result in changes in the sensor readings, e.g., an object moving will result in an increased acceleration reading, a fire will result in an increased temperature reading. Therefore, normal behaviour can be defined as a threshold (e.g., maximum values, rates of increase and combination thereof from multiple sensors) based on statistics of historical data (or domain knowledge), and alarms can be raised if the predefined threshold is exceeded. Examples in this category include Abadi et al. (2005), Chen et al. (2007), Gehrke and Madden (2004), etc.

The advantage of threshold-based event detection is its simplicity of implementation and low computation complexity. However, tuning the threshold is highly dependable on the specific detection problem and the environment that sensors are monitoring, and some events cannot be fully captured by discrete threshold values. The accuracy of detection is limited.

8.3.2 Tempo-spatial pattern based event detection

In contrast to threshold-based event detection, the underlying intuition of tempo-spatial pattern based event detection is that an event occurring in the monitoring field usually results in some tempo-spatial patterns in the sensor readings of networked nodes. For instance, a gas leakage event can be characterized as a spatial distribution of sensor readings following a gradual decreasing trend from the source to the surrounding area nearby. The event of interest can be defined as temporal (Mukherji et al., 2008), spatial (Xue et al., 2006), or tempo-spatial patterns, then the event detection problem is converted into a pattern-matching problem.

The intuition of an example spatial pattern-matching is shown in Figure 8-1. The pattern extracted from live sensor data in the left column is compared with the pre-defined pattern for an event in the middle column. If it doesn’t match, the detection result is normal; otherwise, an event has been detected.

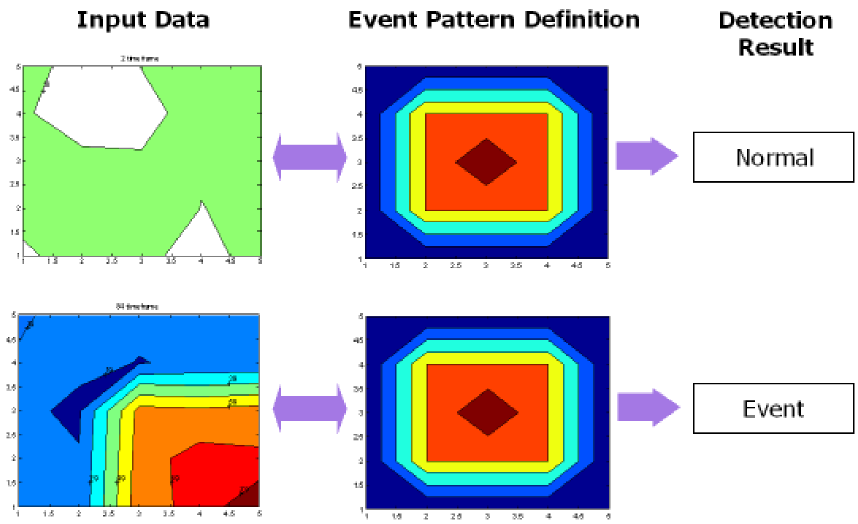


Figure 8-1: The intuition of spatial based event detection

The advantage of tempo-spatial pattern based event detection is that it takes context information into account, and incorporates tempo-spatial correlations typically existing in sensor data to improve the accuracy of event detection.

However, the disadvantages are:

- Increased complexity, because it incorporates data from the whole network to do pattern matching.
- Difficulty in defining suitable patterns to represent the events of interest. If a static pattern is defined, its flexibility is low; if the pattern is defined with user-specified factors, tuning the factors results in similar difficulties as tuning thresholds.

8.4 A Generic State Model for Emergency Event Detection

Various methods have been proposed for event detection using sensor networks. However, they have been proposed in accordance with specific event detection scenarios. In this section, a generic state model for emergency event detection is proposed, which can model the behaviour of any sensor network based emergency event detection systems.

It can be argued that the key requirements of emergency event detection based on sensor network are:

- Event-driven operational mode in contrast to constant monitoring

Due to the energy-constraints in WSNs, constant data transmission for monitoring is not suitable. Event-driven operational mode means that sensor nodes are configured to be in sleeping mode most of the time, and wake up when a suspicious event occurs. A sensor node consists of a sensing unit, a computation unit and a communication unit. The communication unit is the most energy consuming of the three. Sleeping mode means that the communication unit of a sensor is asleep, whereas the sensing unit is still functioning. Once an event is detected, the communication unit will be woken up. The detection of events starts from the bottom level up.

- Improvement of reliability

Sensor data typically consists of errors (due to sensor malfunction) and noise (due to other environmental interference) (Elnahrawy and Nath, 2003), therefore, a mechanism to improve the detection reliability is necessary.

- Integration of temporal spatial correlations typically existing in sensor network

Sensor nodes are usually deployed at a certain density so that they can cover the entire monitoring field. As a result, “the readings observed at one time instant are highly indicative of the readings observed at the next time instant, as are readings at nearby devices” (Jeffery et al., 2006). The high temporal spatial correlations can be incorporated to improve the reliability of event detection.

- Maintain the trade-off between early detection and detection reliability

In emergency event detection, early detection is highly demanded. Therefore, there is a need for minimizing the detection time delay. Early detection requires high sensitiveness of sensors. However, this may cause detection reliability issues. Therefore, the trade-off between early detection and detection reliability must be maintained.

The proposed generic event detection model addresses the above requirements. It consists of three parts: sensor level state model, neighbourhood support and network level fusion.

8.4.1 Sensor level state model

The proposed Sensor State Model addressed that for each sensor node being monitored, there are four states: normal, suspicious/checking, abnormal, and false alarm, as shown in Figure 8-2.

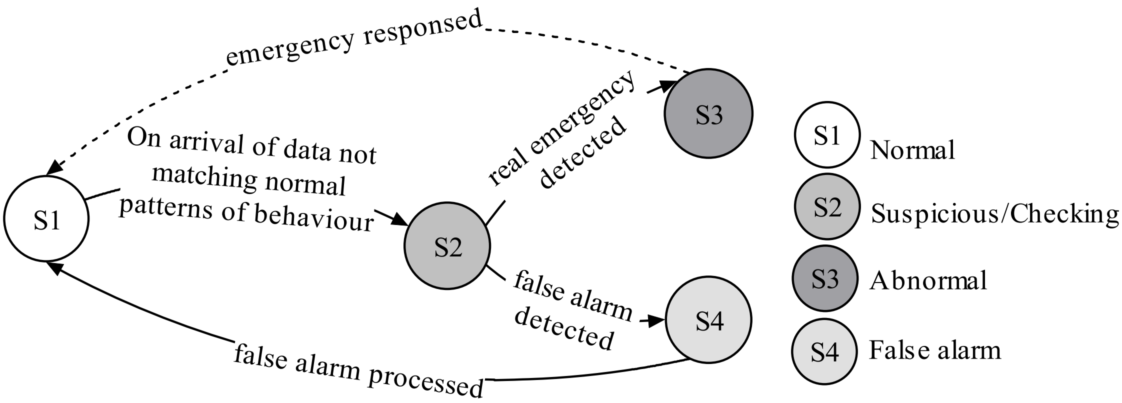


Figure 8-2: Sensor level state model

S1 (Normal): A state s is considered to be normal when the pattern extracted from live sensor data P_L matches the predefined normal pattern of behaviour P_N, denoted as

$$s \in S_1 \Leftrightarrow P_L \tilde{=} P_N \qquad (8\text{-}1)$$

S2 (Suspicious/Checking): A suspicious state appears when P_L does not match the normal pattern of behaviour P_N, denoted as

$$s \in S_2 \Leftrightarrow P_L \tilde{\neq} P_N \qquad (8\text{-}2)$$

When a suspicious state occurs, the system will enter into a checking state, to check the level of support from its temporal or spatial correlations, to improve the detection precision.

S3 (Abnormal): If the level of support l_s is above a threshold th, it indicates that an incident is detected, denoted as

$$s \in S_3 \Leftrightarrow s \in S_2 \wedge l_s > th \qquad (8\text{-}3)$$

S4 (False alarm): If the level of support is not big enough, it is considered to be a false alarm, denoted as

$$s \in S_4 \Leftrightarrow s \in S_2 \wedge l_s < th \tag{8-4}$$

False alarms will be logged, and the sensor reading will go back to its previous state.

8.4.2 Neighbourhood support

The proposed sensor state model introduces a suspicious/checking state to improve the reliability of event detection.

There are different ways of implementing the checking process in an application, e.g. checking different types of sensors at the same location or checking the same type of sensor at different locations, checking limited nodes or checking all the nodes in the network, checking the temporal support or checking the spatial support.

When a sensor node detects a suspicious/checking state, some event detection approaches check the temporal support to confirm whether it is a real event or outlier. An example of temporal support checking would be to check whether the suspicious state lasts for a consecutive time of T . Although this method can filter out short temporary outliers, it improves detection reliability at the cost of detection time, and it does not filter out the case when a sensor node is generating out of range value all the time due to malfunctioning.

The concept of Neighbourhood Support is more formally defined here as an example of spatial support checking.

Neighbourhood: In an area where a density of sensor nodes are deployed, the neighbourhood of sensor node s_i consists of all the sensor nodes that are deployed within the radius r from s_i , denoted as $N_{s_i} = \{s_j \mid dis(s_j, s_i) \le r\}$, where $dis(s_j, s_i)$ is the distance from sensor node s_j to s_i , assuming that sensor nodes know their geographical location information, either during the deployment stage or through RF-based beacons.

Neighbourhood Support is defined as the level of support a sensor node gets from its neighbourhood region on its detected suspicious state, denoted as NS .

Intuitively, sensor nodes in the neighbourhood behave as witnesses to confirm or deny the suspicious state that a sensor node detects. It can be implemented in the format of thresholds or in the format of contour map.

8.4.3 Network level fusion

In the context of sensor networks, voting algorithms have been recommended as the mechanism for fusing decisions of multi-sensors (Parhami, 1994; Klein, 1993), e.g., majority voting, m out of n voting. However, voting over the network could reject an event that affects a small region but with high severity, or an earlier stage of a big event. It can be argued that in the context of emergency event detection, any local-level abnormal state confirmed by Neighbourhood Support should be alarmed. The role of network level fusion is to characterize the event (e.g., the affected area, the type and the severity of the event) rather than to vote to agree or disagree on a detection of an event. The proposed alarm generation mechanism is as follows:

At neighbourhood level, if the state of a sensor node becomes abnormal at time t, an alarm $d(t)$ is set to 1, and t is the time of event detection, denoted as $s(t) \in S_3 \Rightarrow d(t) = 1$, with associated confidence $cf = NS$.

At network level, for a $N \times N$ network, alarm $D(t)$ is set if one or more sensor nodes become abnormal at time t, denoted as $D(t) = d_1(t) \vee d_2(t) \vee \cdots \vee d_{N^2}(t)$, with an associated confidence CF representing the level of severity of the abnormal state.

8.5 Application of State Model with NS to Threshold-based Event Detection

In the scenario of threshold based event detection, P_L and P_N in the generic state model shown in equation (8-1) and (8-2) are presented in the format of thresholds.

8.5.1 Single node system with multiple types of sensors

Taking the fire detection algorithm using smoke and gas sensor proposed by Chen et al. (2007) as a typical example in this category, the proposed algorithm can be presented in the form of sensor state model as follows:

S1 (Normal): A state s is considered to be normal when the rising rate of smoke sensor RR_{smoke} is within a preset threshold TH_{smoke}, denoted as

$$s \in S_1 \Leftrightarrow RR_{smoke} \le TH_{smoke} \qquad (8\text{-}5)$$

S2 (Suspicious/Checking): A suspicious state appears when RR_{smoke} exceeds TH_{smoke}, denoted as

$$s \in S_2 \Leftrightarrow RR_{smoke} > TH_{smoke} \tag{8-6}$$

When a suspicious state occurs, the system will check the rising rate of CO sensor and CO_2 sensor, denoted as RR_{CO} and RR_{CO_2} respectively.

S3 (Abnormal): If the rising rate of any of the two gas sensors is above the predefined threshold, TH_{CO} and TH_{CO_2} respectively, it indicates that an incident is detected, denoted as

$$s \in S_3 \Leftrightarrow s \in S_2 \wedge (RR_{CO} > TH_{CO} \vee RR_{CO_2} > TH_{CO_2}) \tag{8-7}$$

S4 (False Alarm): If none of the rising rates of gas sensors exceed the thresholds, it is considered to be a false alarm, denoted as

$$s \in S_4 \Leftrightarrow s \in S_2 \wedge RR_{CO} \le TH_{CO} \wedge RR_{CO_2} \le TH_{CO_2} \tag{8-8}$$

This method was intended for fire detection in aircraft cargo compartments, which can be considered as a single node system with multiple types of sensors. The rising rates of gas sensors behaved as the level of support to any suspicious rising rate of smoke sensor. In the scenario of fire detection in commercial high-rise buildings where a network of sensor nodes with multiple types of sensors is deployed, our proposed concept of Neighbourhood Support and network level fusion can be applied on top of the node level detection.

8.5.2 A network of sensor nodes

The general threshold based event detection in sensor networks using the proposed state model with NS is presented as follows:

S1 (Normal): At each node with m types of sensors, the state of each sensor z_i is normal when its rising rate RR_i is within its threshold TH_i, denoted as $z_i \in S_1 \Leftrightarrow RR_i \le TH_i, i = 1, 2, \cdots, m$. The overall state of a node s is considered to be normal when all m types of sensors are normal, denoted as

$$s \in S_1 \Leftrightarrow z_1 \in S_1 \wedge z_2 \in S_1 \wedge \cdots \wedge z_m \in S_1 \tag{8-9}$$

S2 (Suspicious/Checking): The state of each sensor is suspicious when its rising rate exceeds its threshold TH_i, $z_i \in S_1 \Leftrightarrow RR_i > TH_i, i = 1, 2, \cdots, m$ A suspicious state of a node appears when the state of any of the m types of sensors is suspicious, denoted as

$$s \in S_2 \Leftrightarrow z_1 \in S_2 \vee z_2 \in S_2 \vee \cdots \vee z_m \in S_2 \tag{8-10}$$

When a suspicious state occurs, the system will check the Neighbourhood Support. Neighbourhood Support can be defined as the ratio of the number of neighbours who are in suspicious state N_{S_2} to the total number of neighbours N_n, denoted as $NS = \frac{N_{S_2}}{N_n}$.

S3 (Abnormal): If the Neighbourhood Support is above a threshold *th*, it indicates that an incident is detected, denoted as

$$s \in S_3 \Leftrightarrow s \in S_2 \wedge NS > th \tag{8-11}$$

S4 (False alarm): If the Neighbourhood Support is not big enough, it is considered to be a false alarm. False alarms will be logged, and the sensor will go back to its previous state, denoted as

$$s \in S_4 \Leftrightarrow s \in S_2 \wedge NS < th \tag{8-12}$$

Any neighbourhood level abnormal state confirmed by Neighbourhood Support will be alarmed, an alarm $d(t)$ is set to 1, as described in section 8.4.3. At network level, for a $N \times N$ network, the level of severity associated with alarm $D(t)$ can be defined as the ratio of the number of local alarms over the total number of sensor nodes in the network, denoted as $CF = \frac{\sum_{i=1}^{N^2} d_i(t)}{N^2}$, with the assumption that the data sources are mutually exclusive.

8.6 Application of state model with NS to tempo-spatial pattern based event detection

In the scenario of temporal pattern based event detection, P_L and P_N in the generic state model shown in equation (8-1) and (8-2) are presented in the format of a sequence of sensor readings. In the scenario of spatial pattern based event detection, P_L and P_N in the model are presented in the format of contour map, which is the distribution of sensor readings over the network.

Using contour map matching based event detection (Xue et al., 2006) as a typical example in this category, the idea can be presented in the form of sensor state model in the following way:

S1 (Normal): A state s is considered to be normal when the contour map extracted from the live sensor reading C_L does not match any of the predefined contour map event patterns C_E, denoted as

$$s \in S_1 \Leftrightarrow C_L \neq^{\sim} C_E \tag{8-13}$$

S2 (Suspicious/Checking): A suspicious state appears when C_L matches the contour map event pattern C_E, denoted as

$$s \in S_2 \Leftrightarrow C_L =^{\sim} C_E \tag{8-14}$$

When a suspicious state occurs, the system will check whether the match continues for a consecutive time of T.

S3 (Abnormal): If the suspicious state lasts for a consecutive time of T, an incident is detected, denoted as

$$s(t) \in S_3 \Leftrightarrow s(t-T+1,\cdots,t) \in S_2 \tag{8-15}$$

where t is the detection time.

S4 (False Alarm): If the suspicious state does not last for a consecutive time of T, it is considered to be a false alarm, denoted as

$$s(t) \in S_4 \Leftrightarrow s(t-T+1,\cdots,t-a) \in S_2 \wedge s(t-a-1,...,t) \in S_1, a \le T-1 \tag{8-16}$$

This method integrated the feature of spatial correlations in sensor network. However, there are two drawbacks:

Although the proposed in-network construction of contour map is energy efficient compared to transmitting all the raw data out of the network, constantly constructing contour maps from the live sensor data over the network results in too much unnecessary usage of network resources.

Checking temporal support improves the detection reliability at the cost of delay in detection.

Therefore, the following improvements are suggested:

- To use event-driven operation mode instead of constant monitoring over the network
- On-event construction of Neighbourhood Support

The improved contour map based event detection is presented as follows:

Sensor nodes are configured to be in sleep mode by default. Sleep mode means that sensor nodes in the same cluster can be configured to wake up in turns for a time slice each to check whether a suspicious state occurs, but they do not transmit data out (sleep) unless there is a demand or there is a suspicious state detected (wake up).

Equations (8-9) and (8-10) can be used to detect S1 (Normal) and S2 (Suspicious/checking) respectively, to avoid the unnecessary usage of network resources caused by constantly constructing contour maps from the live sensor data over the network.

When a suspicious state is detected by a sensor node, it will wake up its neighbourhood. The system will construct the Neighbourhood Support in the format of contour map centred from the suspicious node, and check whether it matches the predefined spatial pattern.

S3 (Abnormal): If the constructed Neighbourhood Support matches the predefined spatial pattern C_E, it indicates an incident is detected, denoted as

$$s \in S_3 \Leftrightarrow s \in S_2 \wedge NS =^{\sim} C_E \tag{8-19}$$

S4 (False Alarm): If the constructed Neighbourhood Support does not match the predefined spatial pattern, it is considered to be a false alarm, denoted as

$$s \in S_4 \Leftrightarrow s \in S_2 \wedge NS \neq^{\sim} C_E \qquad (8\text{-}20)$$

False alarms will be logged, and the sensor will go back to its previous state.

8.7 Performance Analysis

For each identified category of event detection, the chosen example event detection method and the method using our proposed generic state model with NS were implemented in Matlab environment, and their performances were compared.

8.7.1 Performance metrics

Computation efficiency, reliability, robustness and detection time of the event detection methods were used as the metrics of performance evaluation.

- *Computation efficiency* measures the computation costs of the event detection methods.

- *Reliability* measures the effectiveness of the event detection methods. A traditionally used metric for detection reliability is *accuracy*, defined as the fraction of all the detection results that are correctly reported as "event" and "non-event". However, in emergency event detection scenarios, the event frequency is typically very low, which means there are many more non-events than events. Because of this imbalance, predicting "non-event" all the time results in good accuracy, therefore it is not suitable to measure reliability by detection accuracy for "event" and "non-event". Two sub metrics: precision and recall were adopted to measure the reliability of emergency event detection. "Precision is the fraction of reported events that are actual (true) events. Recall is the fraction of all events that are reported correctly." (Kerman et al., 2009) More formally, " $precision = \frac{TP}{TP+FP}$, $recall = \frac{TP}{TP+FN}$ " (Olson and Delen, 2008), where TP is the number of correctly detected events, FP is the number of detected events which are not actual (true) events, FN is the number of actual (true) events that are not detected. In ideal situations, precision and recall would be 1.

- *Robustness* measures the reliability when the uncertainty in the data (outlier rate) increases.
- *Detection time* measures the time that is required from the occurring of an event to the detection of it.

8.7.2 Parameters considered

Network size (N), which is the length of the square $N \times N$ grid. It was varied from 5 to 20, to evaluate the computation efficiency of the selected event detection methods.

Outlier rate (o), which simulates temporary errors caused by signal conflictions in the network or sensor malfunctioning over the quality of the sensor data. The outlier rate was varied from 1% to 20%, to evaluate how robust the event detection methods can be when processing data with low to high uncertainty.

Fire spread model (*F*), which simulates the characteristics of an occurrence of fire. Two fire spread models were used, to evaluate the detection time required under different fire scenarios. Fire spread model A represents fire spreading that reflects in all types of sensors at the same time, whereas fire spread model B represents fire spreading that reflects in different types of sensors at different time.

8.7.3 Simulation setup

The simulation system considers the scenario of detecting fire emergency events from a $N \times N$ network of wireless sensor nodes. Each node has four types of sensors (temperature, smoke, CO, flame). The block diagram of the simulation system is shown in Figure 8-3.

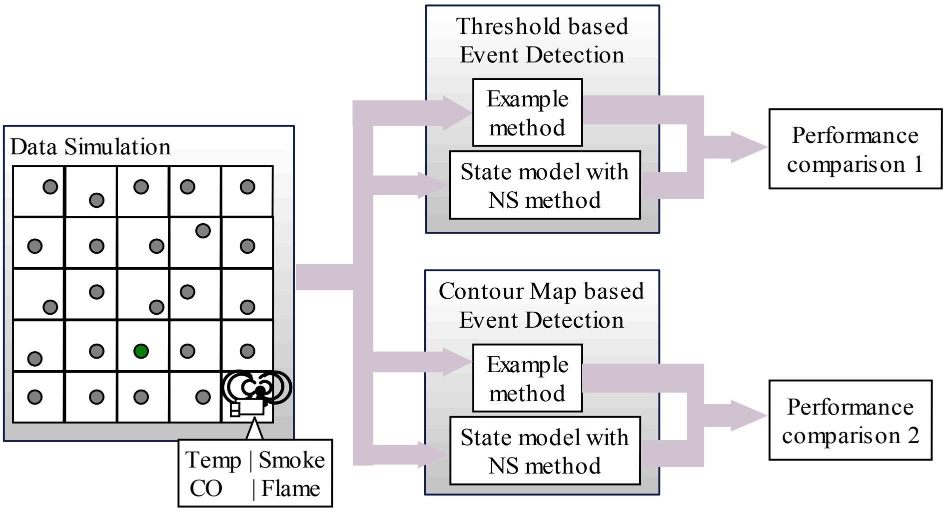


Figure 8-3: The block diagram of the simulation system

Data stream from the network for a time duration of t was simulated. For each type of sensor, its reading z is generated according to the formula $z = NE + O + F$, where:

NE is the estimation of the sensor reading under normal conditions, generated by adding a random noise (w) to the mean (μ) of normal readings according to the domain knowledge, denoted as $NE = \mu + w$. For example, a normal room temperature μ_{temp} can be set to 20°C, a normal smoke level μ_{smoke} can be set to 0.5% obs/m (obscuration per meter), a normal CO level μ_{CO} is set to 2.5ppm, and a normal flame value μ_{flame} is set to 350nm.

O represents random outliers added at random places, the number of which is decided by outlier rate (o) multiplied by the simulation time duration (t).

F represents fire spread model, simulated as a sharp linear increase of sensed value occurring at a random location, random time, and spreading for a random duration. The increase of sensed value occurs in data from all types of sensors at the same time in fire spread model A, whereas in Fire spread model B the values increased in the order of flame, smoke, CO then temperature.

An example plot of simulated data from one sensor node (including temperature, smoke, CO and flame readings) in a 5×5 network for the time duration of 300s is shown in Figure 8-4.

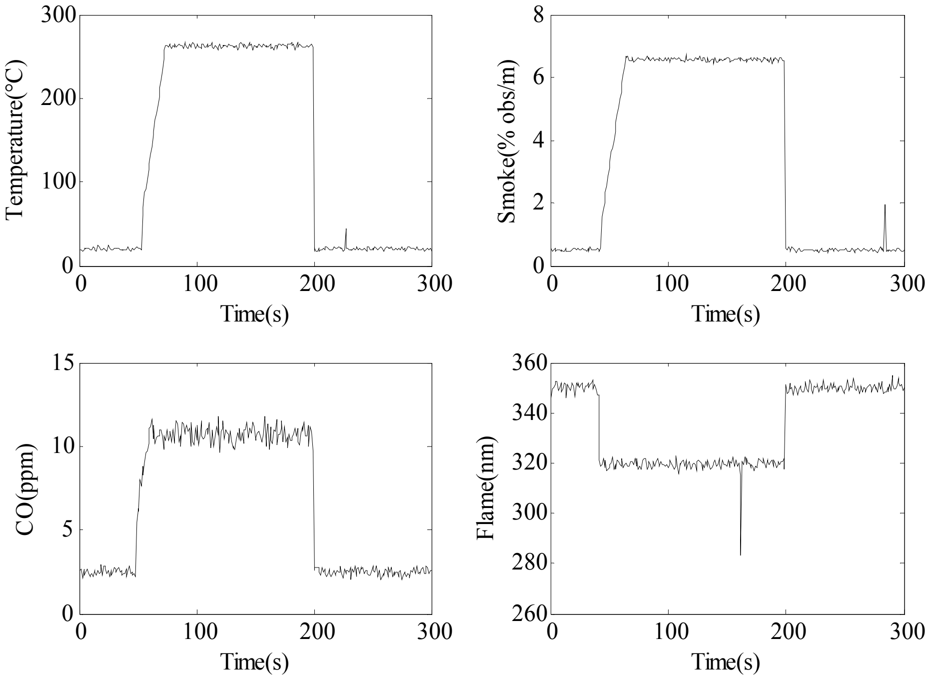


Figure 8-4: An example simulated data for a sensor node in a 5×5 network for the duration of 300s

8.7.4 Performance evaluation

Simulation 1: Threshold-based event detection

The first set of performance evaluation was in the category of Threshold-based Event Detection (TED). The example TED (Chen et al., 2007) was chosen as the bench mark in this category because it utilised data fusion from different types of sensors to improve detection reliability. Its performance was the best in the field. The chosen example and the proposed State Model with NS (SM/NS) TED were compared in the following aspects.

- Computation efficiency

In this simulation, the network size was varied from 5×5 (5m in length) to 20×20 (20m in length), while the outlier rate was fixed to 1% and the simulation time was fixed to 300s, to investigate the computation efficiency of the two approaches and the effect of network size on their computation efficiency.

As shown in Figure 8-5 (a), SM/NS TED required lower computation efficiency than the example TED in all cases. The efficiency of both approaches was affected as the network size increased. However, the computation efficiency of the example TED increased to 14 times as high as the value for the 5×5 network, whereas the computation efficiency of SM/NS TED only increased to 6 times as high. This demonstrated that the network size has less effect on the running cost of SM/NS TED than that of the Example TED. Therefore, SM/NS TED is more efficient and it has the scalability to better accommodate different network sizes, compared to the example TED.

- Reliability

To evaluate the reliability of the two approaches, they were tuned respectively to the best performance (generating the highest precision while still maintaining recall=1). The precision of them was calculated as the simulation time varied from 5 minutes to 30 minutes, whilst the network size was fixed to 5×5 and the outlier rate was fixed to 1%. The same dataset and the same thresholds were used by both of the approaches.

As shown in Figure 8-5 (b), the precision of the example TED was approximately 70-80%, whereas the precision of the SM/NS TED was approximately 80-90%. SM/NS TED improved the precision of the existing example by about 10%.

- Detection Time

As shown in Figure 8-5 (c), under the conditions of fire spread model A, fixed network size (5×5), fixed outlier rate (1%), fixed sampling frequency (once per 5 seconds), statistics over the simulation time of 250 minutes did not show a significant difference on the detection time of the two approaches. The average detection time of the Example TED was 4.75s, the average detection time of SM/NS TED was 4.72s.

However, under the same conditions as above except using fire spread model B, the detection time required by the example TED increased 5 times whilst the detection time required by SM/NS TED almost remained the same. The Example TED required 30s to detect a fire incident, whereas the SM/NS TED only required 5s. In fact, fire spread model B is closer to a real fire scenario, where the detection of change by flame and smoke sensors will be quicker, whereas the detection of change by temperature sensor will be slow as it takes time to heat the air. This demonstrated the advantage of using Neighbourhood Support for suspicious checking over using other types of sensors.

- Robustness

In this simulation, the outlier rate was varied from 1% to 20%, whilst the network size was fixed to 5×5. Statistics over 30 minutes simulation time showed that although the difference on the precision of both approaches was not significant when the outlier rate was low, the precision of the example TED dropped quickly when the outlier rate increased (as shown in Figure 8-5 (d)). This demonstrated that SM/NS TED has better robustness than the example TED.

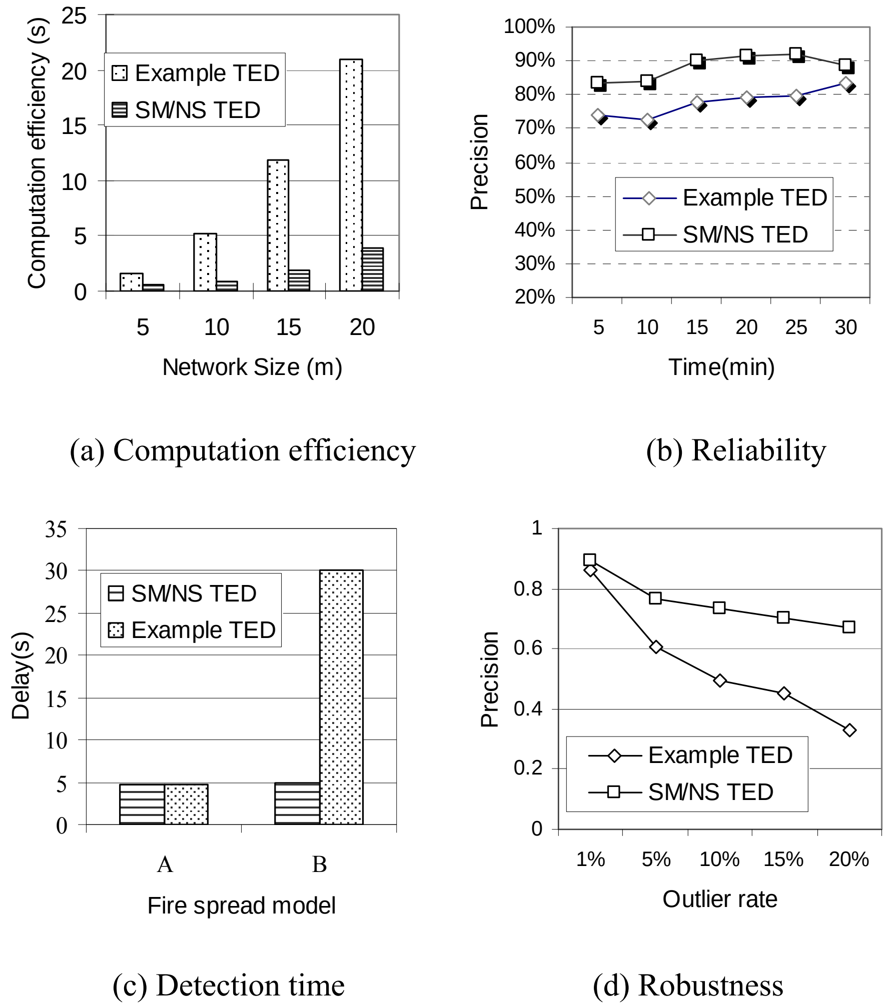


Figure 8-5: Performance comparison 1

Simulation 2: Contour-map-based event detection

The other set of performance simulation used Contour-map-based Event Detection (CED) as an example in the category of tempo-spatial pattern based event detection. The example CED (Xue et al., 2006) was chosen as the benchmark in this category because it is the best in the field and the most cited. The performance of the proposed SM/NS CED was compared to that of the example CED.

- Computation efficiency

In this simulation, the network size was varied from 5×5 (5m in length) to 20×20 (20m in length), while the outlier rate was fixed to 1% and the simulation time was fixed to 300s, to investigate the computation efficiency of the two approaches and the effect of network size on their computation efficiency.

The result demonstrated that SM/NS CED was approximately 10 times faster than the example CED, as shown in Figure 8-6 (a). The computation efficiency of both approaches increased around 3 times comparing to their own values in a 5×5 network. However, comparing with each other, the increase of the example CED was approximately 10 times as big as the increase of SM/NS CED, as the network size varied.

- Reliability

The reliability test was carried out under the conditions of fixed network size (5×5), fixed outlier rate (1%), and both approaches tuned to their best performance (generating the highest precision while still maintaining recall=1).

As simulation time varied from 5 minutes to 30 minutes, statistics of precision of both approaches (Figure 8-6 (b)) showed that their precision was very close to each other, although on average the example CED was slightly better (around 0.1%).

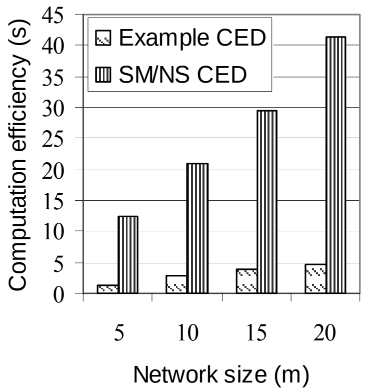
- Robustness

In the robustness test, the network size was fixed to 5×5, whilst the outlier rate varied from 1% to 20%. The simulation was run for 30 minutes.

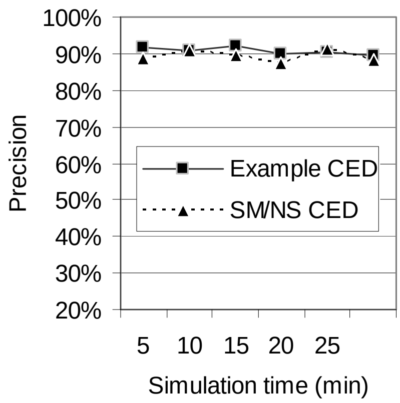
As shown in Figure 8-6 (c), the precision of both approaches were around 90% when the outlier rate was 1%. However, the precision of the example CED dropped as the outlier rate increased and quickly became unacceptable (<50%), whereas the precision of SM/NS CED was maintained above 70%. Therefore, the SM/NS CED demonstrated a better robustness than the example CED.

- Detection time

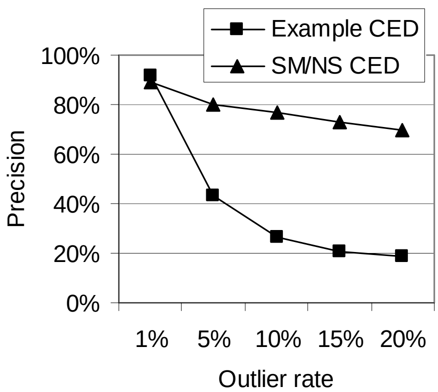
Since the example CED uses temporal support check, it checks whether the suspicious state lasts for a consecutive time of *T* before an alarm is generated, the detection time of the example CED will be the duration of *T* sampling intervals, whilst simulation results demonstrated that the detection time of SM/NS CED was approximately 1 sampling interval. Therefore, the detection time required by the example CED is approximately *T* times of what is required by SM/NS CED.



(a) Computation efficiency



(b) Reliability



(c) Robustness

Figure 8-6: Performance comparison 2

8.8 Conclusions

This chapter analysed that "meaning" in the context of ER can be defined as the occurrence and characteristics of an incident, and real-time development of an incident. Taking the occurrence and characteristics of an incident as an example of meaning, the problem of meaning extraction was converted to a problem of event detection. A generic state model of emergency event detection in Wireless Sensor Networks was proposed, which addresses the key requirements of emergency event detection based on sensor network.

The model operates in an event-driven mode, in contrast to constant monitoring. Therefore, it can save the unnecessary usage of network resources and storage space caused by constant monitoring. The bottom-up event-driven detection process involves 1) sensor node level state model for suspicious behaviour detection, 2) Neighbourhood Support checking to confirm or deny the event, and 3) network level fusion to characterize the event (e.g. the affected area, the type and the severity of the event).

The concept of Neighbourhood Support integrated the high temporal spatial correlations typically existing in sensor data to separate real alarms from outliers, thus improving the reliability of emergency event detection.

The model is also designed to be light-weight, and maintain the trade-off between early detection and detection reliability.

Such characteristics as being event-driven, reliability-improving and light-weight make the model suitable for on-site emergency response system where the requirement of computation efficiency and reliability is high whilst resources are limited.

The model was applied to both threshold based event detection and tempo-spatial pattern based event detection. In each category of event detection, an example approach was chosen, and the performance of the State Model with Neighbourhood Support event detection was compared with the chosen example. Their performance in terms of efficiency, reliability, robustness and detection time was analysed. The results demonstrated that applying the proposed model in event detection improved detection reliability of the example threshold-based event detection, and computation efficiency of the tempo-spatial pattern based event detection. It also improved scalability over network size and robustness over data quality in both event detection categories.

The simulation result demonstrated that applying the proposed model in event detection can reduce false alarm rate to 30% (precision 70%). The reduction of false alarms can improve the reliability of WSN-based emergency event detection systems. The improvement in emergency event detection reliability can save the unnecessary waste of resources responding to false alarms, and reduce the risk of not having sufficient resources for deployment in real incidents – thereby enabling more effective use of emergency response resources. The improvement in emergency event detection reliability should also improve emergency responders' trust in technical systems, meaning that they are more confident and better prepared when responding to incidents.

The research so far is based on simulated data, although characteristics of real fire incidents are incorporated to ensure the simulated data is as realistic as possible. Therefore, to overcome this limitation, future work will study the performance using sensor data collected from field trails.

The research so far only considered one example of meaning in the context of on-site emergency response, which is the occurrence and characteristics of an incident. Due to time constraints, the extraction of the real-time development of incident from sensor data, which will also improve ICs' situation awareness, was not included in the thesis. However, the research undertaken and sensor data meaning extraction methods developed in this thesis provide a promising basis for future work in the field of sensor data processing to provide real benefits for on-site ER.

Chapter 9. Conclusions and Future Work

9.1 Summary

The effectiveness of Emergency Response (ER) depends significantly on the availability of accurate and reliable information about an incident and the efficient gathering and communication of such critical information in real-time. Wireless Sensor Networks (WSNs) have demonstrated potential for addressing the information gap in existing ER - the lack of information to support first responders. Data collected from WSNs deployed inside premises can provide Incident Commanders (ICs) with a 'view' of an incident from inside the building. Hence they can improve the overall effectiveness of ER. However, although large amounts of data can be collected from WSNs and be made available to emergency responders, the meaningfulness of the raw sensor data stream is low, due to the features of sensor data, e.g. its data quality issues, streaming feature, existence of tempo-spatial correlations, etc. The full analysis of the features of sensor data can be found in Section 6.2. The limitation of resources in an on-site ER environment and the high demand of real-time or near real-time system behaviour such as sensor data processing further increases the challenge of making sense of the large amount of data collected from WSNs.

This thesis addressed the challenge of making sensor data work for on-site ER by investigating the use of sensor data processing technology. The sensor data that has meaning to ICs within the context of on-site ER is considered as "emergency information". An architecture of Emergency Information Management System was proposed. Three main necessary steps - sensor data storage, sensor data cleaning, and sensor data meaning extraction - were further investigated in details.

This thesis has primarily had a technical focus. However a key theme throughout the thesis has been to consider the link between the human factors perspective on what benefits sensor data can bring to on-site ER, and the design and implementation of technology support to bring such benefits.

9.2 Research Objectives Revisited

The research questions and objectives stated in the introduction are revisited here as follows:

RQ1: Out of all the potential user groups existing in the ER domain, which is the user group that can benefit the most from the information provided by sensor data?

Objective associated with RQ1: Investigate the existing literature available on ER from a human factors' perspective to understand the generic theory and practice in ER as well as to choose a targeted user group in the first responders.

Objectives Revisited: Through the literature review on ER, the IC and his command support team were identified as the targeted user group in the scenario of on-site ER. This was because of the ICs' special position in first responders:

1) The IC is the highest rank in the first responders' hierarchy and is responsible for all activities on-site (Bigley & Roberts, 2001), thus his decisions have significant impacts on the overall performance of ER.

2) The IC is an information intensive position (Jiang et al., 2004). The position of the IC and his command support team on-site is usually out of the premises and their hands are usually free, therefore they can have more information displayed to them to facilitate decision-making processes.

RQ2: What are the goals, tasks, information requirements of the targeted user group?

RQ3: What are the opportunities for WSN technology to address the gaps related to retrieving the required information?

Objective associated with RQ2 and RQ3: Understand the goals, tasks and information requirements of the chosen user group through interviews and observations, with the purpose of identifying gaps in the current ER practice and analysing potential opportunities for the use of technology.

Objectives Revisited: Through the requirement analysis of ICs, a comprehensive analysis of what information is required by the ICs, and the current means of retrieving such information was undertaken. It was initially assumed after the literature review that using sensor data to find safe exit routes for fire-fighters and building occupants would be useful. However technical support for this activity was proved unnecessary during the interviews and analysis. The requirement analysis of ICs managed to identify the real gaps existing in gathering the required information for ICs using their current means, and the true usefulness of some technical proposals to address these gaps. As a result, a list of gaps and technology opportunities were identified. Sensor data can contribute to all the identified technology opportunities discussed in Section 4.4.6. However, ICs will benefit the most from the opportunity of 'seeing inside' the premises and receiving information during the mobilisation phase (i.e. when travelling to an incident). Consequently, those two opportunities were taken forward to technical development.

RQ4: How can the technology capabilities of WSN be implemented in an on-site ER system?

Objective associated with RQ4:

- Research the existing literature available on information extraction from sensor data, and the associated process and techniques of information extraction from sensor data.
- Design a suitable system architecture for a WSN-based Emergency Information Management System (EIMS); specify the system functions and the associated system components.

Objectives Revisited: Through the literature review on information extraction from sensor data, the process of information extraction and the available sensor data processing techniques in the process were analysed. The challenge of making sense of large amount of collected sensor data was revealed. What constitutes suitable data processing techniques for sensor data was evaluated.

Through the proposed on-site Emergency Information Management System (EIMS) architecture, the essential characteristics of a suitable system architecture for the WSN-based EIMS were identified to be 1) adaptive to different incident stages, and 2) incorporating fusion of multiple data sources. The proposed architecture demonstrated these two essential characteristics and showed how the technology capabilities can be integrated in a system. The essential components of the system were identified as data storage, data cleaning, and meaning extraction. How multiple data sources can be fused in the system was discussed in detail. Each component and the interaction among different system components were specified. In addition, the adaptive system behaviours were specified for different incident stages.

RQ5: What constitutes a suitable sensor data storage mechanism for on-site ER?

Objectives associated with RQ5:

- Accomplish a comprehensive analysis of the features of sensor data
- Evaluate and propose a suitable data storage mechanism for the WSN-based EIMS, with an emphasis on managing and maintaining the query efficiency regardless of the increase of data volume.
- Evaluate the efficiency of the proposed data storage mechanism and its cost-effectiveness.

Objectives Revisited: A comprehensive analysis of the features of sensor data in comparison to traditional data was provided. Through the proposed data storage mechanism, what constitutes a suitable data storage mechanism was evaluated. The proposed data storage mechanism can accommodate the challenges presented by the features of sensor data. The design of the proposed storage mechanism includes a database schema that maintains the tempo-spatial correlations typically existing in sensor data, a time-driven sensor data management, and adaptive support for different incident stages. The performance evaluation demonstrated that the proposed data storage mechanism can support timely response to queries (up to 100 times faster than a database without partitioning) in spite of the increase in the volume of data.

RQ6: What constitutes a suitable sensor data cleaning approach for on-site ER?

Objectives associated with RQ6:

- Analyse and propose a suitable data cleaning approach for the WSN-based EIMS, with the emphasis on separating outliers from real environmental changes and dealing with them separately.
- Evaluate the effectiveness of the proposed data cleaning approach in comparison to the existing data cleaning approaches.

Objectives Revisited: Through the proposed data cleaning approach, outliers can be separated from real environmental changes and can be dealt with differently. The design of the proposed data cleaning approach addressed the features of sensor data – typically with the existence of noise in the data. It proposed the use of state-awareness rules, to check and confirm whether a suspicious sensor data is an outlier or a real environmental change. The performance evaluation demonstrated that the proposed data cleaning approach not only reduces noise, but also effectively removes outliers and quickly detects the real environmental changes (reduces the delay of detecting environmental changes on temperature, smoke, CO, and flame sensors by 50% to 90%).

RQ7: How to extract the defined “meaning” from sensor data?

Objectives associated with RQ7:

- Analyse and specify what “meaning” can be in the context of WSN-based on-site EIMS, and propose a meaning extraction approach.
- Evaluate the performance of the proposed meaning extraction approach and discuss its application.

Objectives Revisited: Through the proposed meaning extraction analysis and the study of event detection, the cleaned sensor data is further processed to extract information that can be interpreted by ICs and their command support team. An example of “meaning” in the context of ER can be defined as the occurrence and characteristics of an incident, and real-time development of an incident. Taking the occurrence and characteristics of an incident as an example of key meaning that data must convey, the problem of meaning extraction was converted to a problem of event detection. A generic state model of emergency event detection in Wireless Sensor Networks was proposed, which addresses the key requirements of identifying the occurrence and characteristics of an incident based on sensor networks. It was applied in both threshold-based and tempo-spatial pattern-based categories of event detection. The performance evaluation demonstrated that the proposed event detection model is 2-5 times faster than the example threshold-based event detection, and approximately 10 times faster than the example tempo-spatial pattern-based event detection. The proposed event detection model also increased reliability by 10% in the threshold-based event detection and increased robustness by maintaining greater than 60% precision when the outlier rate was up to 20% in both event detection categories. Therefore, it can improve the efficiency and accuracy of existing examples of event detection method in both threshold-based event detection and tempo-spatial pattern-based event detection.

9.3 Contributions to Knowledge

The research contained in the thesis expanded the current body of knowledge in relation to sensor data processing for emergency response. The contributions to existing knowledge from the research undertaken are considered to be five-fold.

Contribution 1: A list of identified technology opportunities to address the current gaps relating to ICs retrieving the information they require during emergency response

The existing research in the literature revealed the impacts of ICs' decision making on the overall effectiveness of ER, the importance of information to ICs' decision making, and information required in order for them to complete their key tasks. However, how the ICs retrieve and utilise the required information in real practice, and whether there are any gaps related to being able to retrieve the required information were not clear. The literature review also revealed that WSNs have great potential in facilitating ER However, little research was found that described how the underlying technology capabilities of WSN can contribute to ER.

> Therefore, there is a need for a comprehensive analysis of ICs' information requirements that can link ICs' goals to technology capabilities.

This thesis provided a thorough analysis of ICs' requirements for information and communication of the information under the framework of goals, tasks/decisions, information and means of retrieving information, inability to retrieve the required information and technology opportunities. Both information needed by ICs for their decision making and information that would be a distraction to them were analysed. A list of capability gaps was identified and the issue of trust of technology was revealed based on the data collected from observations of fire ER and interviews undertaken with fire ICs. As a result of the analysis, eight technology-based opportunities to address information retrieval limitations were proposed, and the added value of each opportunity was discussed with ICs. All the ideas were validated by the majority of the participants during the validation phase.

The analysis of ICs' requirements undertaken in this thesis differs from traditional task analysis by extending the focus from human factors to technology capabilities. It contributes to the existing knowledge in the way that it highlighted how the underlying technology capabilities can contribute to ICs' goals and bring added benefits to ICs during emergency response. Task analysis methods from a pure human factor perspective have been well implemented in the domain of emergency response, however, they usually focus only on goals and physical and cognitive tasks. Although some researchers attempted to establish the link from tasks to information requirements (e.g. Prasanna et al. (2007)), the information requirements was gathered by Prasanna et al. (2007) with the aim of interface design rather than analysing the required underlying technology capabilities. This research undertaken established a link between user requirements and the desired underlying technology capabilities, with the emphasis on the benefits that the technology can provide to the users. Most of the technology opportunities apply not only in fire ER, but also in other types of emergencies.

Contribution 2: A proposed on-site Emergency Information System Architecture that fuses multiple data sources and is adaptive to different incident stages

Some system architectures designed for emergency response management have been suggested by researchers (e.g., Turoff et al., 2006; Zlatanova, 2005). However, they are off-site EIMS. The existing off-site EIMS architectures emphasized the requirements for coordination among different responding agencies to handle the interoperability issues, and the communication of issued commands both within a responding agency and among different responding agencies. In addition, they are typically systems for *emergency centres*. Although there has been an attempt to study system design for ICs (Jiang et al., 2004a), the prototypes proposed focused on interface design rather than addressing underlying technology capabilities.

> Therefore, there is a need for an on-site system design to support frontline first responders that addresses the capabilities of WSN and the need to take into account the dynamic nature of ER.

The proposed EIMS was designed in order to support ICs and their command support team among first responders on-site. It incorporates the capabilities of WSNs and addresses the requirements of the dynamic nature of ER revealed in the literature (Kyng, 2006). The proposed on-site EIMS architecture consists of four main components: data storage, data cleaning, meaning extraction and data presentation. Each component was specified, and the interaction between components was analysed. The interaction between the EIMS system and its lower data collection layer and upper command and control layer in the context of an Emergency Response system was discussed. The detailed system functions to facilitate ER before, during and after incidents were specified.

The proposed on-site EIMS architecture differs from the existing EIMS architectures because it incorporates the domain-specific requirements to the link between users' goal space and system design space. The requirements to support frontline first responders were captured by goal-action diagram and each system action was modelled and integrated to an abstract level architecture. The proposed EIMS architecture highlighted capabilities of WSN and the essential technical components of an EIMS as well as addressed he dynamic nature of Emergency Response.

The proposed architecture can also be flexibly developed in order to apply it to other application areas. These include: emergency detection systems, risk assessment systems, decision support systems, historical statistics or an integrated multi-purposes emergency response support system according to different context of use.

Contribution 3: A proposed data storage mechanism that addresses the features of sensor data and is suitable for on-site ER applications

The majority of the existing research on sensor data storage focused on purely technical issues such as storage placement (Ganesan et al., 2003) or energy consumption (Mathur et al., 2006). However, little attention has been paid to the impacts that the features of sensor data have on Data Storage Mechanism (DSM). In addition, the in-network data storage placement that has been suggested by most existing research is not a satisfactory fit for on-site ER applications. In spite of the stated benefits, e.g. shared storage load, reduced impacts of single node failure, the data stored solely in sensor nodes within a WSN could be destroyed and lost during an incident such as fire.

> Therefore, there is a need for an alternative sensor data storage mechanism that takes into account the features of sensor data and is suitable for on-site ER applications.

This thesis proposed a data storage mechanism that accommodates the challenges presented by the features of sensor data as well as maintaining the query efficiency for the entire ER process. The design of the proposed storage mechanism includes a database schema that maintains the tempo-spatial correlations typically existing in sensor data, a time-driven sensor data management, and adaptive support for different incident stages. Simulations demonstrated that the proposed data storage mechanism can achieve the benefits of efficient querying (up to 100 times faster than database without partitioning) with neither additional updating cost nor the introduction of unacceptable storage costs.

The proposed data storage mechanism differs from existing research and contributes to knowledge in that it focuses on (1) the impacts that the features of sensor data have on data storage mechanisms and (2) the demands of consistent efficiency of on-site ER. It stores and manages sensor readings, time-related information and location-related information. It supports timely response to queries in spite of the increase in the volume of data. And it is dynamic and flexible to accommodate different needs at different incident stages. These characteristics made the proposed data storage mechanism suitable for on-site ER applications based on WSN.

Contribution 4: A data cleaning approach that is state-aware, and not only reduces noise but also separates outliers from real environmental changes

Existing research has revealed the data quality problems that a data cleaning approach has to deal with, including noise and outliers. Existing research efforts on cleaning and querying noisy sensor data (e.g. Elnahrawy and Nath, 2003) focus mainly on smoothing noise and outliers. However, they did not separate outliers from real environmental changes. This results in a compromise between the prior knowledge and the observed noisy sensor data, or in other words, simple smoothing regardless of the state. Simple smoothing means that an extreme sensor reading caused by a real environmental change could be considered to be an outlier and be smoothed. This would result in a delay in detecting the real environmental change that indicates the occurrence of an incident.

> Therefore, there is a need for a different data cleaning approach for ER scenarios that can separate outliers from real environmental changes, and treat them differently.

The proposed data cleaning approach within this thesis integrated a set of state-awareness rules with a Kalman Filter, the result is referred to as a state-aware Kalman Filter. The proposed state-aware Kalman Filter addresses three different states of a sensor reading: normal, outlier and abnormal and deal with them differently. The data cleaning experiments demonstrated that by introducing the state-awareness rules and integrating it with a Kalman Filter, the resulting state-aware Kalman Filter can reduce noise, remove outliers as well as quickly detect real environmental changes.

The proposed state-aware Kalman Filter differs from the existing data cleaning approaches and contributes to knowledge by providing state-aware smoothing rather than compromising between the observations and the system estimation regardless of the state. It can be argued that being state-aware is a new advantage in addition to those inherited from a Kalman Filter: e.g. effectively smoothing of noise and being light-weight. As a result, it can be concluded that the proposed state-aware Kalman Filter would be a satisfactory fit for on-site ER applications.

Contribution 5: A generic state model for emergency event detection from sensor data as an example of meaning extraction

Meaning extraction is an emerging field that identifies elements of information contained in datasets that imply meaning in the context of application and can be interpreted by the users to facilitate their tasks. The existing meaning extraction research is highly domain-specific. What was considered as “meaning” in the existing research on meaning extraction varied in different contexts. An example of “meaning” in natural language processing has been defined as the part of sentence, whereas the similarity distances of literal objects was considered as “meaning” in web semantics analysis. Thus, the subsequent meaning extraction method was chosen differently according to the specific domain. However, there is a lack of research on the context of meaning extraction from sensor data streams for on-site ER. The existing meaning extraction methods typically utilised statistical methods, including Markov Logic (Bajwa, 2010), Bayesian networks, decision trees, logistic regression, neural network etc., to extract meaning in the form of meta-data, features or frequently occurred patterns. However, statistically significant patterns may not be of significance in the context of ER because emergencies such as fires are usually events of low probability.

> Therefore, there is a need for an analysis of what constitutes meaning in the context of on-site ER, and the detailed technology proposal to extract the defined meaning.

The research undertaken in the thesis determined that ‘meaning’ in the context of on-site ER can be defined as the occurrence and characteristics of an incident, and the real-time development of an incident. Taking the occurrence and characteristics of an incident as a key concept, a generic model for emergency event detection from sensor data was proposed. The bottom-up event-driven detection process involves a sensor node level state model for suspicious behaviour detection, Neighbourhood Support checking to confirm or deny the event, and network level fusion to characterize the event (e.g. the affected area, the type and the severity of the event). The simulation results demonstrated that applying the proposed event detection model improved reliability of the typical example in the category of threshold-based event detection, and improved the computation efficiency of the typical example in the category of tempo-spatial pattern-based event detection. It also improved scalability in terms of network size and robustness in terms of data quality in both event detection categories.

The proposed generic emergency event detection model contributes to the knowledge in that

- the proposed model operates in an event-driven mode, in contrast to constant monitoring - therefore, it can save on unnecessary usage of network resources and storage space caused by constant monitoring;

- the concept of Neighbourhood Support integrated the high temporal spatial correlations typically existing in sensor data to separate real alarms from outliers, thus improving the reliability of existing emergency event detection;
- it maintains the trade-off between early detection and detection reliability.

The characteristics of being event-driven, improving reliability and being light-weight make the model suitable for on-site emergency response systems, where the requirement of computation efficiency and reliability is high whilst resources are limited.

9.4 Future work

A number of further research issues can be identified as a result of the research undertaken for this thesis and a range of potential avenues for further research have come out of this thesis.

Addressing other features of sensor data

This thesis identified the need of addressing the special features of sensor data (as analysed in Section 6.2) in sensor data processing and attempted to address some of the features of sensor data in the research contained therein. The proposed data storage mechanism addressed the streaming feature of sensor data and provided a base for associating sensor data with time and location information. The concept of Neighbourhood Support suggested in Section 8.4.2 utilised the typical existence of high tempo-spatial correlations in WSNs to improve the reliability of event detection based on sensor data. However, further investigation on addressing other identified features of sensor data is still needed. Examples of future work in this area are:

- Location-specific or time-specific information retrieval, which focuses on associating time and location information with sensor data
- Further investigation on the high redundancy typically existing in sensor data and a development of a suitable approach to manage sensor data redundancy

Further investigation on meaning extraction method in the context of on-site ER

As stated by Römer (2008), "it remains a major challenge to make sense of the collected data, i.e., to extract the relevant knowledge from the raw data." This thesis investigated what constitutes meaning in the context of on-site ER. Due to time constraints, it only investigated the occurrence and characteristics of an incident as an example meaning. Further investigation on extracting the real-time development of an incident is needed.

In addition, a generic meaning extraction method for on-site ER would be desirable. The need for a unified system has been raised by the ICs (as discussed in Section 4.4.5). They believe such unified system would be one of the key characteristics for future systems, since it improves the efficiency and accuracy of decision making. Whether different pieces of information that constitute "meaning" are extracted by different methods or a generic method is not important to the users, as long as the desired information is provided in a unified system. However, a generic method if exists would have benefits such as more efficient use of computation resources, which would be an advantage for on-site ER where resources are limited. Therefore, it would be interesting to further investigate whether there is a generic meaning extraction method that applies to all pieces of information that constitute "meaning".

Incident-Site Information Space (ISIS)

This concept was suggested in section 4.4.5 as a technical proposal for integrating all the required categories of emergency information in a unified system. This thesis investigated the steps needed and sensor data processing technology suitable to make sensor data work for ER. It focused on understanding the nature of the incident based on sensor data. However, the integration of the nature of the incident with other categories of required information, e.g. water resources, hazard, staff deployment, is an important characteristic for future systems. It has been proposed to organise the required categories of information into an ISIS. Further research on how the ISIS can be formally defined, modelled and implemented, as well as investigation on the underlying data fusion technology required for organising such ISIS would be beneficial to the ICs.

Broadening sensor data processing to wider ER communities and beyond

The research contained in the thesis could be beneficial to wider emergency response communities by further broadening it to other types of ER. Fire ER was chosen as the representative example of ER for the purpose of this thesis based on the understanding that the requirements gathered and sensor data processing technology developed for one type of ER could apply on other types of ER. Further investigation of the feasibility of applying the proposed sensor data processing methods in other types of emergencies will be beneficial for a wider ER community.

Another way of broadening sensor data processing to wider emergency response communities is to further investigate the feasibility of applying sensor data processing techniques to other possible user groups existing in the ER domain. Different user groups have different responsibilities, and therefore different goals. As a result, information requirements may differ. It would be interesting to further investigate how sensor data processing technology can accommodate varied needs of different types of users.

The sensor data processing techniques developed in this PhD research were for emergency response, future work should investigate the possibility of applying them beyond ER communities.

References

1. Abadi, D. J. et al. (2005), "REED: Robust, efficient filtering and event detection in sensor networks". In *Proceedings of the 31st international Conference on Very Large Data Bases* (Trondheim, Norway, 30 August - 02 September 2005), pp. 769-780.

2. Akcan, H. and Brönnimann, H. (2007), "A New Deterministic Data Aggregation Method for Wireless Sensor Networks". *Elsevier Journal of Signal Processing 2007*, Vol. 87, No. 12, pp. 2965–2977.

3. Albers, M. J. (1998), "Goal-Driven Task Analysis: Improving Situation Awareness for Complex Problem-Solving", In *Proceedings of the 16th annual international conference on Computer Documentation* (Quebec, Canada, 24-26 September 1998), pp. 234-242.

4. Andersen, H. B. et al. (1998), "An Electronic Message Management System for Emergency Response". *IEEE Transactions of Engineering Management*, Vol. 45, No. 2, May 1998, pp. 132-140.

5. Annett, J. and Duncan, K. (1967), "Task Analysis and Training Design". *Occupational Psychology*, Vol. 41, pp. 211-221.

6. Annett, J. et al. (1971), "*Task Analysis*". London: Her Majesty's Stationery Office.

7. Annett, J. and Stanton, N., eds. (2000), "Task analysis". London: Taylor and Francis.

8. Arasu, A. et al. (2004), "Characterizing memory requirements for queries over continuous data streams". *ACM Transactions on Database Systems* (TODS), Vol. 29, No. 1, March 2004, pp. 162-194.

9. Bajwa, I. S. (2010), "Context Based Meaning Extraction by Means of Markov Logic". *International Journal of Computer Theory and Engineering*, Vol. 2, No. 1, pp. 1793-8201.

10. Barnes, M. et al. (2007), "Emergency Evacuation using Wireless Sensor Networks". In *Proceedings of the 32nd IEEE Conference on Local Computer Networks* (LCN 2007, Clontarf Castle, Dublin, Ireland, 15-18 October 2007), pp.851-857.

11. Barnett, V. and Lewis, T. (1994), "Outliers in Statistical Data", 3rd edition, John Wiley and Sons.

12. Basu, S. and Meckesheimer, M. (2007), "Automatic outlier detection for time series: an application to sensor data". *Knowledge and Information Systems*, Vol. 11, No. 2, pp. 137-154.

13. BBC news (2007), "*The ever-present threat of death*". [online] Available at: http://news.bbc.co.uk/1/hi/uk/7076583.stm <Accessed on 22/02/2011>

14. Betts B. J. et al. (2006), "Improving Situational Awareness for First Responders via Mobile Computing". In *Proceedings of EuroSys2006* (Leuven, Belgium, 18-21 April 2006).

15. Bigley, G. A. and Roberts, K. H. (2001), "The Incident Command System: High-Reliability Organizing for Complex and Volatile Task Environments". *The Academy of Management Journal*, Vol. 44, No. 6, pp. 1281-1299.

16. Bontempi, G. and Borgne, Y. L. (2005), "An adaptive modular approach to the mining of sensor network data". In *Proceedings of 1st International Workshop on Data Mining in Sensor Networks as part of the SIAM International Conference on Data Mining* (Newport Beach, CA, 21-23 April 2005), SIAM Press, pp. 3-9.

17. Carter, H. R. and Rausch, E. (2008), "*Management in the Fire Service*". 4th ed. Sudbury: Jones and Bartlett Publishers.

18. Carver, L. and Turoff, M. (2007), "Human-computer interaction: the human and computer as a team in emergency management information systems". *Communications of the ACM*, Vol. 50, No. 3, pp. 33-38.

19. Chen, S. J. et al. (2007), “Fire detection using smoke and gas sensors”. *Fire Safety Journal*, Vol. 42, No. 8, pp. 507-515.

20. Cheng, Y. and Ren. X. (2007), "Mining Moving Patterns Based on Frequent Patterns Growth in Sensor Networks", In *Proceedings of International Conference on Networking, Architecture, and Storage (NAS)*, pp. 133-38.

21. Chok, H. and Gruenwald, L. (2009), “Spatio-temporal association rule mining framework for real-time sensor network applications”. In *Proceeding of the 18th ACM conference on Information and knowledge management* (CIKM '09). ACM, New York, NY, USA, pp. 1761-1764.

22. Christoplos, I. et al. (2001), “Re-framing Risk: The Changing Context of Disaster Mitigation and Preparedness”. *Disasters*, Vol. 25, No. 3, pp. 185-198.

23. Chu, F. et al. (2005), “Data cleaning using belief propagation”. In *Proceedings of the 2nd international workshop on Information quality in information systems* (Baltimore, Maryland, 17 June 2005), pp. 99-104.

24. Cilibrasi, R. and Vitanyi, P. (2006), “Automatic Extraction of Meaning from the Web”. In *Proceedings of 2006 IEEE International Symposium on Information Theory* (Seatle, WA, 9-14 July 2006), pp. 2309-2313.

25. Crichton M.T. and Flin, R. (2004), “Identifying and training non-technical skills of nuclear emergency response teams”. *Annals of Nuclear Energy*, Vol. 31, No. 12, pp. 1317-1330.

26. Danielsson, M. (1998), “The cognitive structure of decision making tasks in major versus minor emergency response”, In Scott, P.A., Bridges, R.S., and Chasterrs, J. ed,*Global Ergonomics: Proceedings of the Ergonomics Conference*, Cape Town, South Africa, 1998, pp. 627-632.

27. Davidson I. and Ravi, S. S. (2005), “Distributed Pre-Processing of Data on Networks of Berkeley Motes using Non-Parametric EM”. *Data Mining and Sensor Network Workshop - colocated with 5th SIAM Data Mining Conference*.

28. DEM RI U.S. (2008), “Emergency Response Plan”. [online] Available at: http://www.dem.ri.gov/topics/erp.htm [Accessed on 12/07/2010]

29. Dong, D. and Calvo, R. A. (2009), "Using component extraction association rules for sensor data". In *Proceedings of Workshop on Advances and Issues in Biomedical Data Mining*, *Pacific-Asia Knowledge Discovery and Data Mining* (PAKDD09, 27-30 April 2009, Bangkok, Thailand).

30. Doolin, D. M. and Sitar, N. (2005), "Wireless sensors for wildfire monitoring". In *Proceedings of SPIE Symposium on Smart Structures and Materials/ NDE 2005* (San Diego, California, 6-10 March 2005),

31. Embrey, D. (2000), "*Task Analysis Techniques*". [online] Available at: http://www.humanreliability.com/articles/Task%20Analysis%20Techniques.pdf [Accessed on 06/11/2010]

32. Elnahrawy, E. and Nath, B. (2003), "Cleaning and querying noisy sensors". In *Proceedings of 2nd ACM international Conference on Wireless Sensor Networks and Applications* (San Diego, CA, USA, 19 September 2003), pp. 78-87.

33. Endsley, M. R., Bolte, B., and Jones, D. G. (2003), "Designing for Situation Awareness: An Approach to Human-Centered Design", Taylor and Francis Ltd., .London.

34. Erickson, P. A. (1999), "Emergency response planning for corporate and municipal managers". Academic Press. San Diego.

35. FEMA (2008), "National Incident Management System". [online] Available at: http://www.fema.gov/pdf/emergency/nims/NIMS_core.pdf [Accessed 17/08/2010]

36. *Fire Statistics - United Kingdom 2007* (2009). Crown copyright 2009. London. UK.

37. Ford, J. K. and Schmidt, A. M. (2000), "Emergency response training: strategies for enhancing real-world performance". *Journal of hazardous materials*, Vol. 75, Iss. 2-3, pp. 195-215.

38. Foresti, G.L. (2002), "Distributed Sensor Networks and Neural Trees for Multisensor Data Fusion in Computer Vision". In Hyder, A. K. et al. eds. *Multisensor Fusion*. The Netherlands: Kluwer Academic Publishers, pp. 183-196.

39. Fouad, A. et al. (2010), "Embedding requirements within Model-Driven Architecture". *Software Quality Journal*, 06/12/2010, Springer.

40. Ganesan, D. et al. (2003), "An evaluation of multi-resolution storage for sensor networks". In *SenSys '03: Proceedings of the 1st international conference on Embedded networked sensor systems* (Los Angeles, California, USA), pp. 89-102.

41. Gao, T. et al. (2008), "Wireless Medical Sensor Networks in Emergency Response: Implementation and Pilot Results". In *Proceedings of 2008 IEEE International Conference on Technologies for Homeland Security (*Waltham, MA, 12-13 May 2008*)*, pp. 187-192.

42. Gehrke, J. and Madden, S. (2004), "Query processing for sensor networks". *Pervasive Computing*, Vol. 3, No. 1, pp.46-55.

43. Haddow, George D. et al. (2007), "Introduction to Emergency Management". 3rd ed. Amsterdam: Butterworth-Heinemann.

44. Haghighi, P.D. et al. (2007), "An Architecture for Context-Aware Adaptive Data Stream Mining," *International Workshop on Knowledge Discovery from Ubiquitous Data Streams (17 September 2007)*, pp. 117.

45. Halatchev, M. and Gruenwald L. (2005), "Estimating Missing Values in Related Sensor Data Streams". In *Proceedings of the International Conference on Management of Data* (January 2005), pp. 83-94.

46. Hall, D. L. and Llinas, J. (1997), "An introduction to multisensor data fusion". In *Proceedings of the IEEE*, Vol. 85, No. 1, pp. 6–23.

47. Han, J. and Kamber, M. (2006), "Data Mining: Concepts and Technologies". USA: Morgan Kaufmann Publishers.

48. Handmer J. and Parker D. (1991), "British Disaster Planning and Management: An Initial Assessment". *Disasters*, Vol. 15, No. 4, pp. 303-317.

49. Hanson, M. (2000), "Beyond first aid: emergency response teams turn to graphics". *IEEE Computer Graphics and Applications.*Vol.20, No. 2, pp. 12-18.

50. Heierman, E.O., III; Cook, D.J. (2003), "Improving home automation by discovering regularly occurring device usage patterns". In *Proceedings of Third IEEE International Conference on Data Mining (*ICDM 2003, 19-22 November 2003), pp. 537-540.

51. Hill, B. and Long, J. (2001), "Performance Modelling in the Domain of Emergency Management". In M.A. Hanson ed. *Contemporary Ergonomics 2001*. London: Taylor and Francis, pp. 165-170.

52. Hinton, D. et al. (2005), "An Architectural Proposal for Future Wireless Emergency Response Networks with Broadband Services". *Bell Labs Technical Journal*, Vol. 10, No. 2, pp. 121-138.

53. Hu, Z. et al. (2007), "Fast flooding information extraction in emergency response of flood disaster". In *Proceedings of ISPRS Workshop on Updating Geo-spatial Databases with Imagery and The 5th ISPRS Workshop on DMGISs* (Urumchi, Xingjiang, China, 28-29 August 2007), pp.173-177.

54. Hunter, J. and Colley, M. (2007), "Feature Extraction from Sensor Data Streams for Real-Time Human Behaviour Recognition". In: Kok, J. N. et al. (Eds.) *PKDD 2007*, Lecture Notes in Computer Science (LNCS), Vol. 4702, Springer-Verlag Berlin Heidelberg, pp. 115-126.

55. Hutton, J. and Melihen, A. (2006), "Emergency Response – Remote Sensing Evolves in the Wake of Experience". *Photogrammetric engineering and remote sensing,* Vol. 72, No. 9, pp. 977-981.

56. Incident Command - 3rd edition (2008), Fire and Rescue Manual - Volume 2 - Fire Service Operations. Crown Copyright 2008, UK.

57. Irwin, Robert L. (1989), "The Incident Command System". In Erik Auf Der Heide, ed. *Disaster Response: Principles of Preparation and Coordination.* St. Louis: Mosby. Chapter 7.

58. Jain, S. and McLean, C. (2003), "A Framework for Modeling and Simulation for Emergency Response". In *Proceedings of the 2003 Winter Simulation Conference* (7-10 Dec. 2003), pp. 1068-1076.

59. Jeffery, S. R. et al. (2006), "A Pipelined Framework for Online Cleaning of Sensor Data Streams". In *Proceedings of the 22nd International Conference on Data Engineering* (03-07 April 2006), pp. 140-143.

60. Jiang, X. et al. (2004a), "Ubiquitous computing for fire-fighters: field studies and prototypes of large displays for incident command". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria, 24 – 29 Apr., 2004), pp. 679-686.

61. Jiang, X., et al. (2004b), "Siren: Context-aware Computing for Firefighting". *Pervasive Computing, Lecture Notes in Computer Science 2004*, Volume 3001/2004, pp. 87-105. Springer.

62. Kamphuis, C. et al. (2008), "Using sensor data patterns from an automatic milking system to develop predictive variables for classifying clinical mastitis and abnormal milk". *Computers and Electronics in Agriculture*, Vol. 62, Iss. 2, July 2008, pp. 169-181.

63. Kantor, G. and Singh, S. (2002), "Preliminary results in range only localization and mapping". *In IEEE Intl. Conf. on Robotics and Automation*, pp. 1819–1825.

64. Kantor, G. et al. (2003), "Distributed Search and Rescue with Robot and Sensor teams". In *Proceedings of Field and Service Robotics* (FSR03).

65. Kariya, H. and Kiyoki, Y. (2005), "A Metadata Extraction Method for Taste-Impression with Sensor Technologies". In *Proceedings of 2005 Symposium on Applications and the Internet Workshops*, pp. 442-445.

66. Kempe, A. (2006), "Acronym-Meaning Extraction from Corpora Using Multitape Weighted Finite-State Machines". Research Report 2006/019, Grenoble Laboratory, Xerox Research Centre Europe. [online] Available at: www.xrce.xerox.com/content/download/6989/52409/file/report.pdf (Accessed on 15/10/2010)

67. Kerman, M. C. et al. (2009), "Event Detection Challenges, Methods, and Applications in Natural and Artificial Systems". In *Proceedings of 14th International Command and Control Research and Technology Symposium: "C2* and Agility" (15-17 June 2009).

68. Kim, C. H. et al. (2005), "Architectures for streaming data processing in sensor networks". In *Proceedings of the 3rd ACS/IEEE International Conference on Computer Systems and Applications*, pp. 59- 64.

69. Kirwan, B. and Ainsworth, L. K. eds. (1992). "*A Guide to Task Analysis"*. London: Taylor and Francis, Ltd.

70. Klann, M. (2009), "Tactical Navigation Support for Fire-fighters: The LifeNet Ad-Hoc Sensor Network and Wearable System". In Loffler, J. and Klann, M. ed. *Mobile Response*, Lecture Notes in Computer Science, Volume 5424, pp. 41-56. Springer.

71. Klein, G. (1999), "Sources of Power: How People Make Decisions". Cambridge, MA: MIT Press.

72. Klein, L.A. (1993), "Sensor and Data Fusion Concepts and Applications". SPIE, Apr. 1993.

73. Koo, S. O. et al. (2006), "Visualization for a Multi-Sensor Data Analysis". In *Proceedings of the International Conference on Computer Graphics, Imaging and Visualisation* (CGIV'06), pp.57–63.

74. Kottapalli, V.A. (2003), "Two-tiered wireless sensor network architecture for structural health monitoring". In *Proceedings of the SPIE - The International Society for Optical Engineering USA 2003*, Vol. 5057, pp. 8-19.

75. Krishnamachari, B. and Iyengar, S. (2004), "Distributed Bayesian Algorithms for Fault-Tolerant Event Region Detection in Wireless Sensor Networks". *IEEE Trans. Comput.*, Vol. 53, No. 3 (Mar. 2004), pp. 241-250.

76. Krivitski, D. et al. (2007), "A Local Facility Location Algorithm for Large-Scale Distributed Systems". *Journal of Grid Computing*, Vol. 5, No. 4, pp. 361–378.

77. Kulakov, A. and Davcev, D. (2005), "Distributed Data Processing in Wireless Sensor Networks Based on Artificial Neural-Networks Algorithms". In *Proceedings of the 10th IEEE Symposium on Computers and Communications* (ISCC '05). IEEE Computer Society, Washington, DC, USA, pp. 353-358.

78. Kumar, V. et al. (2004), "Robots and Sensor Networks for First Responders". *Pervasive Computing,* published by IEEE CS and IEEE ComSoc, Vol. 4, pp. 24-34.

79. Kyng, M. et al. (2006), "Challenges in designing interactive systems for emergency response". In *Proceedings of the 6th conference on Designing Interactive systems* (DIS '06). ACM, New York, NY, USA, pp. 301-310.

80. Lakshmanan, V. et al. (2007), "The warning decision support system - integrated information". *Weather and Forecasting*, Vol. 22, No. 3, pp. 596-612.

81. Ledlie, J. et al. (2005), "Provenance-Aware Sensor Data Storage". In *Proceedings of the 21st International Conference on Data Engineering* (05-08 April 2005).

82. Llinas, J. and Hall, D. L. (1998), "An Introduction to Multi-Sensor Data Fusion". In *Procs. of the 1998 IEEE International Symposium on Circuits and Systems* (May–June 1998), pp. 537–540.

83. Lorincz, K. (2004), "Sensor networks for emergency response: challenges and opportunities". *IEEE Pervasive Computing* USA Oct.-Dec. 2004, Vol.3, No.4, pp. 16-23.

84. Maimon, O. and Rokach, L. (2010), eds. "Data Mining and Knowledge Discovery Handbook", 2nd edition, Springer.

85. Manoj B.S., and Baker, A.H. (2007), "Communication challenges in emergency response". *Communications of the ACM*, 2007, Vol. 50, No. 3, pp. 51-53.

86. Mathur, G. et al. (2006), "Ultra-low power data storage for sensor networks". In *IPSN '06: Proceedings of the fifth international conference on Information processing in sensor networks* (Nashville, Tennessee, USA), pp. 374-381.

87. McConnell, S. M. and Skillicorn, D.B. (2005), A distributed approach for prediction in sensor networks. *Proc. of 1st International Workshop on Data Mining in Sensor Networks as part of the SIAM International Conference on Data Mining*, SIAM Press, Newport Beach, CA (2005), pp. 28–37.

88. McLoughlin, D. (1985), "A Framework for Integrated Emergency Management". *Public Administration Review*, Vol. 45, Special Issue, pp. 165-172.

89. Meertens, L. O. et al. (2010), "Goal and model driven design of an architecture for a care service platform". In *Proceedings of the 2010 ACM Symposium on Applied Computing* (SAC '10). ACM, New York, NY, USA, pp.158-164.

90. Merrett, G. V. et al. (2005), “Resource aware sensor nodes in wireless sensor networks”. *Journal of Physics: Conference Series*, Vol. 15, No. 1, pp. 137-142.

91. Militello, L.G. and Hutton, R.J.B. (1998), “Applied Cognitive Task Analysis: a Practitioner’s Toolkit for Understanding Cognitive Task Demands”. *Ergonomics*, Vol. 41, No. 11, pp. 1618-1641.

92. Mukherji, A. et al. (2008), “SNIF TOOL: Sniffing for patterns in continuous streams”. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management* (CIKM '08, Napa Valley, California, USA, 26-30 October 2008), pp. 369-378.

93. Noyes, J. M. and Baber, C. (1999), “User-Centred Design of Systems”. Springer.

94. Olson, D. L. and Delen, D. (2008), “Advanced Data Mining Techniques”. s.l.:Springer.

95. Palpanas, T. et al. (2003), “Distributed deviation detection in sensor networks”. *SIGMOD Rec.,* Vol. 32, No. 4, pp. 77–82.

96. Parhami, B. (1994), “Voting Algorithms”. IEEE transactions *on Reliability*, Vol. 43, No. 4, pp. 617-629.

97. Pattath, A. et al. (2006), “Interactive visualization and analysis of network and sensor data on mobile devices”. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology* (VAST ’06).

98. Perry, R. W. (2003), “Incident management systems in disaster management”. *Disaster Prevention and Management*, Vol. 12, No. 5, pp. 405-412.

99. Perry, R. W. and Lindell, M. K. (2003), “Preparedness for Emergency Response: Guide-lines for the Emergency Planning Process”. *Disasters*, Vol. 27, No. 4, pp. 336-350.

100. Petrosino, A. and Staiano, A. (2007), “A Neuro-fuzzy Approach for Sensor Network Data Cleaning”. In Apolloni, B. et al. (Eds.) KES2007/WIRN 2007, Part III, LNAI 4694, Springer-Verlag Berlin Heidelberg, pp. 140-147.

101. Pimentel, B. (2002), "Software simulates terror hit – Sandia develops program as tool for public officials". [online] Available at http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2002/08/19/BU133752.DTL&type=business [Accessed on 19/09/2010]

102. Prasanna, R. et al. (2007), "Integrated Information Management System Model for Emergency Response". In *Proceedings. of the 13th International Conference on Automation and Computing* (Staffordshire University, England, September 2007), pp. 59-64.

103. Prasanna, R. et al. (2009), "GDIA: a Cognitive Task Analysis Protocol to Capture the Information Requirements of Emergency First Responders". In Landgren, J. et al. eds. *Proceedings of the 6th International ISCRAM conference* (Gothenburg, Sweden, May 2009).

104. Pister, K. et al. (2001), "Smart Dust: Autonomous Sensing and Communication in a Cubic Millimeter" (Smart Dust Project Website)[online] Available at http://www-bsac.eecs.berkeley.edu/~pister/SmartDust (Accessed on 05/01/2011)

105. Qureshi, K. et al. (2006), "Public Health Incident Command System: A Guide for the Management of Emergencies or Other Unusual Incidents within Public Health Agencies". Vol. 1. 1st edition. Available at http://www.ualbanycphp.org/pinata/phics/default.cfm [Accessed on 06/08/2010]

106. Rezgui, A. and Eltoweissy, M. (2007), "Service-oriented sensor-actuator networks: Promises, challenges, and the road ahead". *Journal of Computer Communications*, Vol. 30, Iss. 13 (September 2007), pp. 2627-2648.

107. Roberts N. (2004), "Public Deliberation in an Age of Direct Citizen Participation". *The American Review of Administration*, Vol. 34, No. 4, pp. 315-353.

108. Römer, K. (2008), "Discovery of Frequent Distributed Event Patterns in Sensor Networks". In Verdone, R. ed. *Wireless Sensor Networks*, Lecture Notes in Computer Science, Vol. 4913, pp. 106-124. Springer.

109. Rotz, L. D. et al. (2002), "Public health assessment of potential biological terrorism agents". *Emerging Infectious Diseases*, Vol. 8, No. 2, pp. 225-230.

110. Santini, S. and Römer, K. (2006), "An adaptive strategy for quality-based data reduction in wireless sensor networks". In *Proceedings of the 3rd International Conference on Networked Sensing Systems* (INSS 2006, TRF, Chicago, IL), pp. 29-36.

111. Sawyer, S. et al. (2004), "Mobility and First Responder". *Communications of the ACM*, Vol. 47, Iss. 3, pp. 62-65.

112. Silberstein, A. et al. (2007), "Data-Driven Processing in Sensor Networks". In *Proceedings of the 3rd Biennial Conference on Innovative Data Systems Research* (CIDR), 7-10 Jan. 2007, Asilomar, California, USA.

113. Sparchholz, C. et al. (2005), "3D visualization – using high resolution multi sensor data". *2nd Panoramic Photogrammetry Workshop*, Berlin, 24-25 February 2005.

114. Stallings, Robert A. and Quarantelli, E. L. (1985), "Emergent Citizen Groups and Emergency Management". *Public Administration Review*, Vol. 45, Special Iss., pp. 93-100.

115. Stankovic, J. A., et al. (2005), "Wireless Sensor Networks for In-Home Healthcare: Potential and Challenges". In *Proceedings of High Confidence Medical Device Software and Systems (HCMDSS) Workshop* (Philadelphia, PA, 2-3 June 2005).

116. Stanton, N. A. et al. (2006), "Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology". *Ergonomics*, Vol. 49, No. 12-13, pp. 1288-1311.

117. Stanton, N. A. (2006), "Hierarchical Task Analysis: Developments, Applications and Extensions". *Applied Ergonomics*, Vol. 37, No. 1, pp. 55-79.

118. Subramaniam, S. et al. (2006), "Online outlier detection in sensor data using nonparametric models," in VLDB, pp. 187–198.

119. Tan, P. (2006), "Knowledge Discovery from Sensor Data". [online]. Available at: http://www.sensorsmag.com/sensors/article/articleDetail.jsp?id=317466 [Accessed 01/09/2010].

120. Tan, Y. L. et al. (2005), "SensoClean: Handling Noisy and Incomplete Data in Sensor Networks using Modeling". Technical Report, University of Maryland. [online] Available at: http://www.cs.umd.edu/~viveks/finalreport.pdf [Accessed 24/08/2010]

121. Tanenbaum, A.S. (2003), *Computer Networks*, 4th edition, Prentice Hall.

122. The Open Group (2009), *ArchiMate 1.0 Specification*. Available at: http://www.archimate.org [Accessed 07/02/2011]

123. Tseng, Y. C. et al. (2006), "A Distributed Emergency Navigation Algorithm for Wireless Sensor Network". *IEEE Computers*, Vol. 39, No. 7, pp. 55-62, July 2006.

124. Tufekci, S. and Wallace, W. A. (1998), "The Emerging Area of Emergency Management and Engineering". *IEEE Transactions on Engineering Management*, Vol. 45, No. 2, pp. 103–105.

125. Turoff, M. (2002), "Past and future emergency response information systems". *Communications of the ACM*, Vol. 45, No. 2, pp. 29-32.

126. Turoff, M. et al. (2006), "The Design of a Dynamic Emergency Response Management Information System (DERMIS)". In International Engineering Consortium ed. *Annual Review of Network Management and Security*, Vol. 1, pp. 101-124.

127. U.S. National Response Team (2000), "Incident Command System/Unified Command (ICS/UC) Technical Assistance Document". Available at: http://www.nrt.org [Accessed 06 August 2010]

128. Van de Walle, B. and Turoff, M. (2007), "Emergency response information systems: emerging trends and technology". *Communications of the ACM*, Vol. 50, No. 3, pp. 29-31.

129. Vasardani, M. and Flewelling, D. (2005), "Cascading web map services and GIS web servers for Emergency Response Management". *Geomatica*, Vol. 59, No. 4, pp. 435-444.

130. Waugh, W. L. Jr. and Hy, R. J. (1990), "Handbook of Emergency Management: Programmes and Policies Dealing with Major Hazards and Disasters". Greenwood Press.

131. Waugh, W. L. Jr. and Streib, G. (2006), "Collaboration and Leadership for Effective Emergency Management". *Public Administration Review*, Vol. 66, No. s1, pp. 131-140.

132. Welsh et al. (2004), "Wireless Sensor Networks for Emergency Medical Care". [online] Available from: www.eecs.harvard.edu/~mdw/talks/ge-codeblue.pdf [cited 09 December 2007]

133. West Yorkshire Fire and Rescue Service (2007). Operational Procedure No.37: Incident Ground Risk Assessment.

134. Wickler, G. and Potter, S. (2009), "Information-Gathering: From Sensor Data to Decision Support in Three Simple Steps". *Special Issue on Intelligent Decision Making in Dynamic Environments: Methods, Architectures and Applications of the Intelligent Decision Technologies (IDT): An International Journal*, Vol. 3, No. 1, 2009.

135. Xue, W. et al. (2006), "Contour map matching for event detection in sensor networks". In *Proceedings of the ACM SIGMOD international Conference on Management of Data* (SIGMOD '06, Chicago, IL, USA, 27 – 29 June 2006), pp.145-156.

136. Yang, L. (2007), "On-site Information Sharing for Emergency Response Management". *Journal of Emergency Management*, Vol. 15, Iss. 5, pp. 55-64.

137. Yang, L. et al. (2009), "On-Site Information Systems Design for Emergency First Responders". *Journal of Information Technology Theory and Application*, Vol. 10, Iss. 1, pp. 5-27, April 2009.

138. Yang, S.H. and Frederick, P. (2006), "SafetyNET – a wireless sensor network for fire protection and emergency responses". *Measurement and Control*, Vol. 39, Iss. 7, pp. 218-219.

139. Yang, Y. et al. (2007), "Emergency Information Management System Architecture". In *Proceedings of 14th TIEMS Annual Conference 2007* (Split, Croatia, 5-8 June 2007).

140. Yang, Y. et al. (2009), "Making Sensor Data Work for Emergency Response". In *Proceedings of 16th TIEMS Annual Conference 2009*, Istanbul, Turkey, June 9-11, 2009, pp. 231-240.

141. Yang, Y. et al. (2010a), "Sensor data processing for emergency response". *International Journal of Emergency Management*, Vol. 7, Nos. 3/4, pp. 233-248.

142. Yang, Y. et al. (2010b), "Opportunities for WSN for Facilitating Fire Emergency Response". In *Proceedings of IEEE International Conference on Information and Automation for Sustainability (ICIAfS) 2010* , pp. 81-86.

143. Yao, Y. and Gehrke, J. E. (2002), "The cougar approach to in-network query processing in sensor networks". *SIGMOD Record*, Vol. 31, No. 3, September 2002, pp. 9-18.

144. Yoon, S. and Shahabi, C. (2007), "The Clustered AGgregation (CAG) Technique Leveraging Spatial and Temporal Correlations in Wireless Sensor Networks", *ACM Transactions on Sensor Networks (TOSN)*, Vol. 3, Issue 1, Article 3, March 2007.

145. Zlatanova, S. (2005), "A proposed system architecture for emergency response in urban areas". *Directions Magazine*, pp. 1-5.

146. Zhuang, Y. and Chen, L. (2006), "In-network outlier cleaning for data collection in sensor networks," *CleanDB Workshop*, pp. 41-48, 2006.

147. Zhuang, Y. et al. (2007), "A Weighted Moving Average-based Approach for Cleaning Sensor Data". In *Proceedings of the 27th International Conference on Distributed Computing Systems* (ICDCS'07, Toronto, 25-27 June 2007), pp. 38-45.

Appendix

I. Interviews of Incident Commanders

1. Brief introduction about the research project, purpose of the interview and reason.

The research aims to investigate how to make use of the sensor readings collected to provide useful information for Incident Commanders. The research is technical, but it also focuses on what real incident commanders' needs are, it aims to provide technical solutions to meet Incident Commanders' information requirements.

So far, based on the documented literature, hypotheses have been extracted about the incident commanders' goals, and the required information and required format that the system should provide derived from the user goals. The interview aims to verify the hypotheses and find out what an Incident Commander's real goals are, what tasks are required to achieve the goals, what information is required to make decisions and how the required information are gathered at the moment.

2. Interview Questions

1. As an incident commander, what are your main goals when you arrive on site? What kind of site/incident? Are the goals different according to the specific type or situation of incident?

Hypothesis: Protect fire-fighters, rescue people, fire fighting, etc.

2. What are the priorities of these goals? Are the goals/priorities different for different incidents/building/situation at different incident stages?

Hypothesis: Protecting people is more important

3. What would you do to make sure that you achieve the priorities of goals?

Hypothesis: Do not send the crew into the building when it is not necessary. If it is necessary to send them in, then make sure the safe exiting route is clear.

4. What information do you need to know?

5. Where do you get that information at the moment?

6. What is missing at the moment?

Hypothesis: incident seriousness distribution, safe escaping route

7. What role does escape routes play?

Continue with question 8-14 if safe escaping routes are important, otherwise find out what is important, and propose other potential technical solutions and verify them.

8. How do you work out safe escape routes?

9. How do you tell these to the fire-fighters?

10. What is the best way of telling fire-fighters what/where the safe escape routes are?

11. How do you do these at the moment?

12. Does these work well at the moment?

13. At what types of fires would safe escape routes be particularly important?

14. Describe a typical situation where the potential technical solution based on sensor data (e.g. safe escape routes) can make a difference.

15. Describe a story of what you do at an incident, from start to finish. What do you need to know and what do you have to do at each stage?

Stages: S1- alarm call

S2- on your way (mobilisation phase)

S3- arrive on scene

16. Where is your position on scene? Stay in vehicle? Portable device? Go into building?

17. How do you communicate information to the fire-fighters?

18. Do fire-fighters have a map? Do fire-fighters get navigation instructions? If you get shown fire fighter moving, how do you give directions etc.

19. What level of detail do you want? e.g. a detailed plan of the building or simple plan, just showing exits and water points or something else?

20. Sketch how you want the information to be presented, e.g. the type of map/plan.

20. What information do you not want?

21. A real example of fire-fighters get trapped or in the danger of getting trapped because of the lack of information.

3. Summary

The hypotheses are:

- Main user goals are to protect fire-fighters, rescue people, fire fighting.
- Protecting people has higher priority than save properties.
- Incident Commanders always make sure they are clear about the safe exit route if they have to send fire-fighters into a building.
- Incident Commanders need better situation awareness about seriousness and seriousness distribution of incident in the building.

Therefore, the initial plan is to produce incident seriousness distribution in the building for situation awareness and fire fighting goals, and produce safe exit routes calculation for people deployed in the building for the goals of protecting fire-fighters and rescuing people.

II. Architecture Models for Each Captured System Action (in ArchiMate language)

Before the Incident:

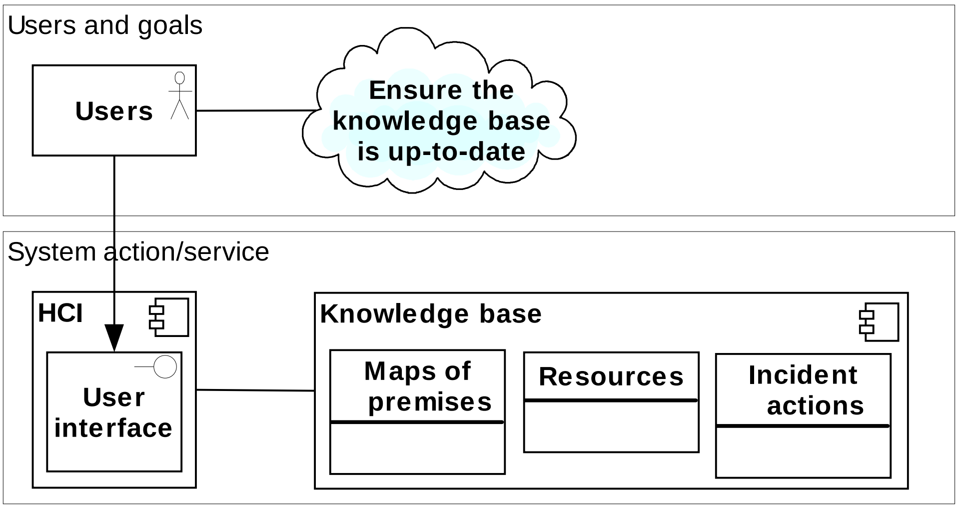


Figure AII-0-1: Architecture model for the system action: Prompt users to update the information and store the most up-to-date information

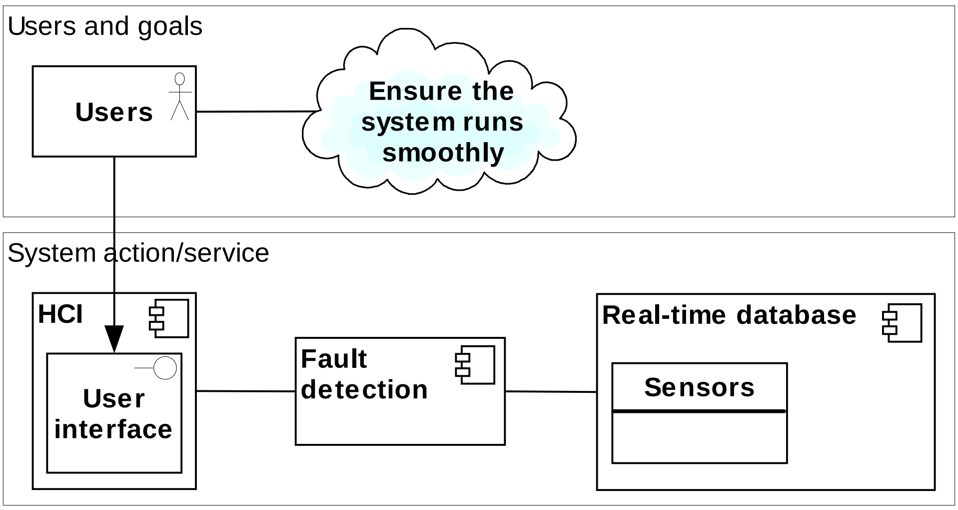


Figure AII-0-2: Architecture model for the system action: Run regular test and generate report on faulty parts

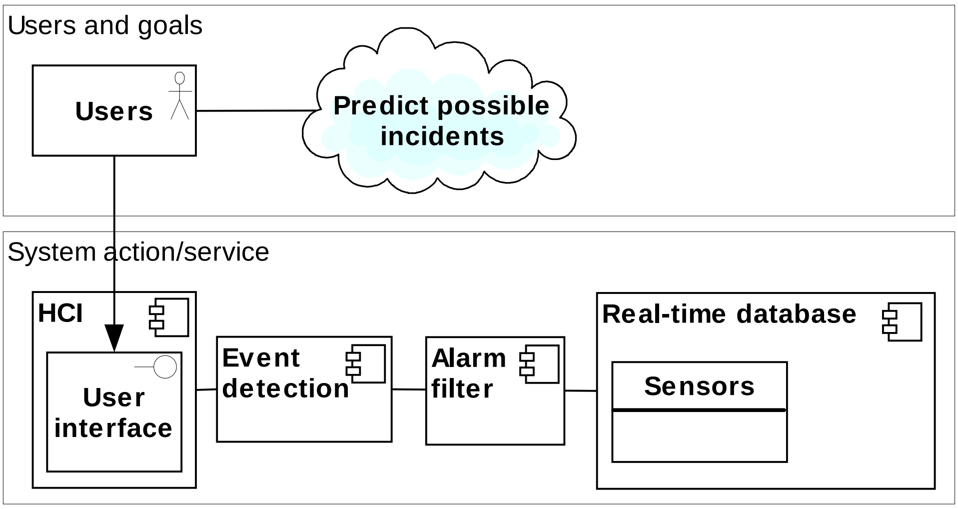


Figure AII-0-3: Architecture model for the system action: Generate warning of abnormal phenomenon detected

During the incident:

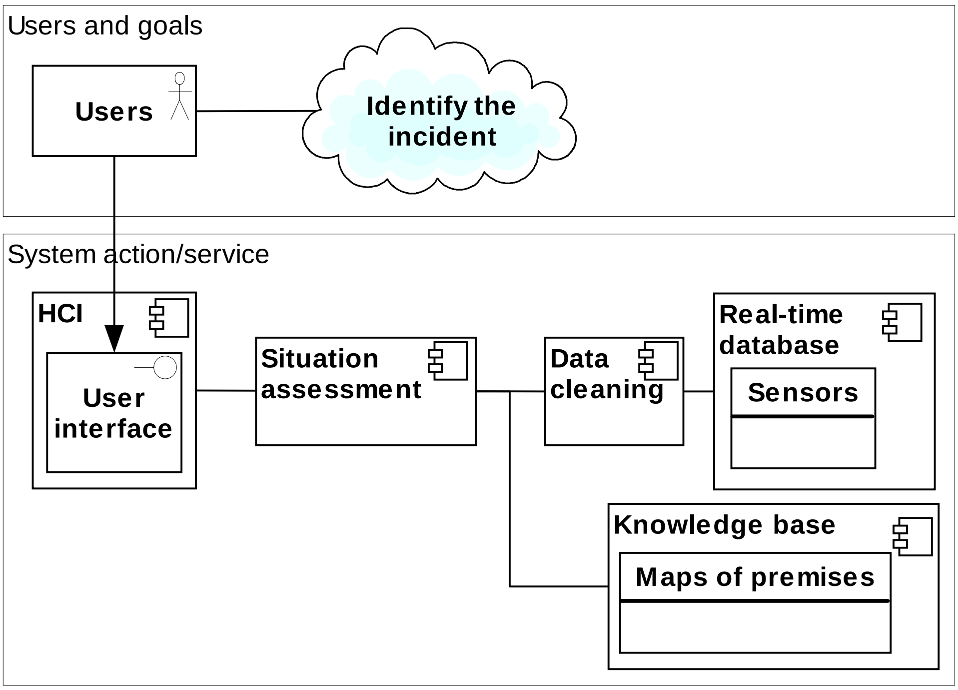


Figure AII-0-4: Architecture model for the system actions: Generate real-time monitoring of incident, integrate location map/floor plan with real-time incident development

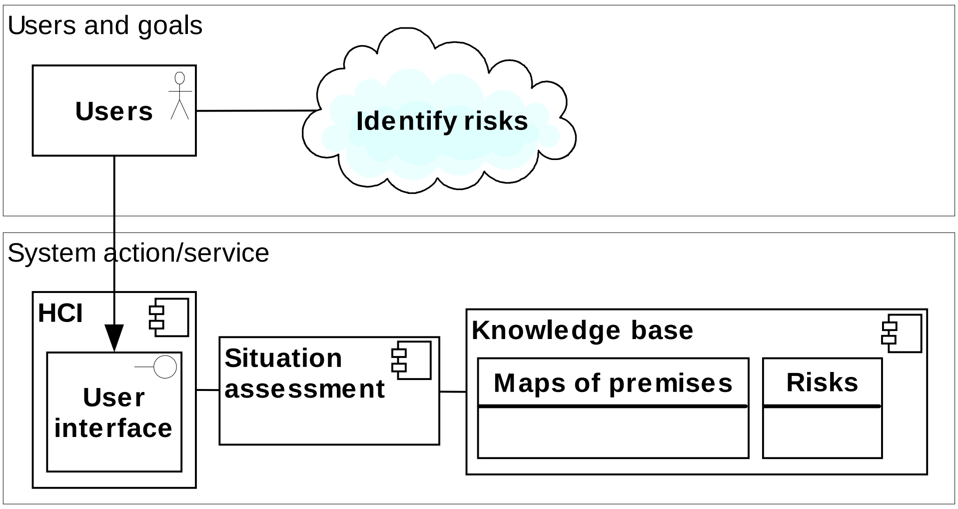


Figure AII-0-5: Architecture model for the system action: Integrate dynamic risks with floor plan

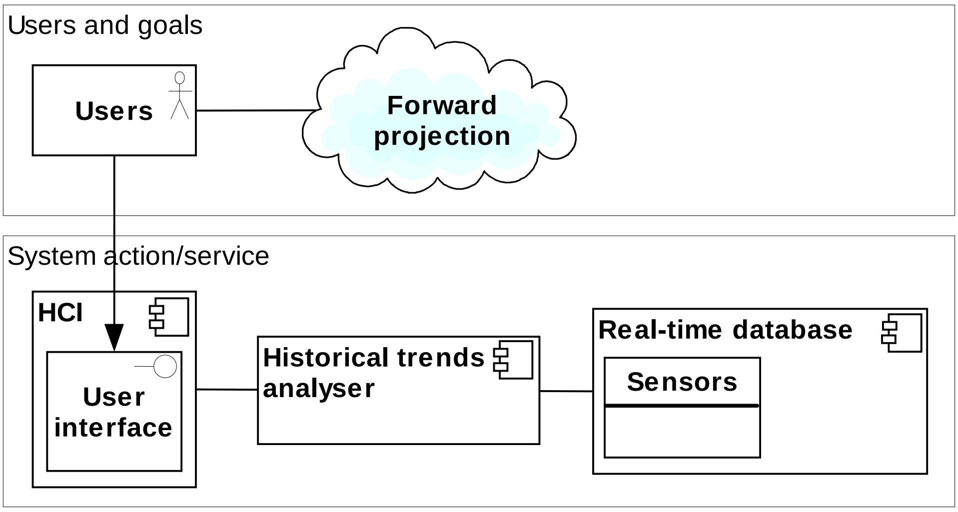


Figure AII-0-6: Architecture model for the system action: Calculate historical trends and forward projection, and display them on request

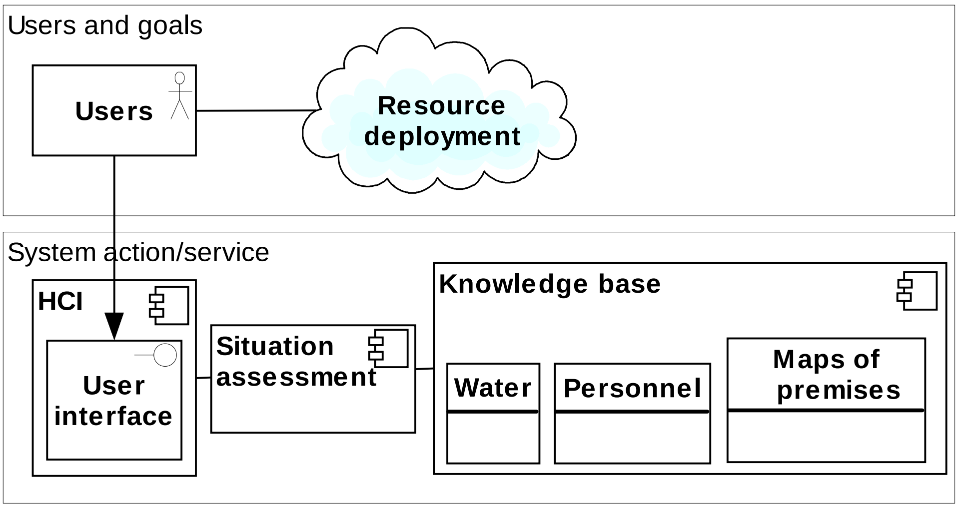


Figure AII-0-7: Architecture model for the system action: Store resources information, integrate them with location map/floor plans

After incident

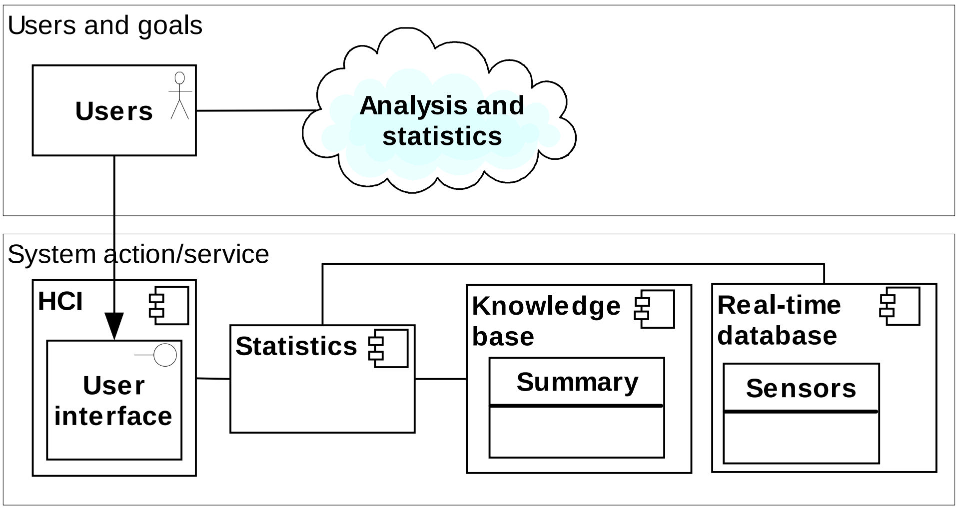


Figure AII-0-8: Architecture model for the system action: Store the statistical information of the incident