

**An Investigation of Algorithms to Clean RFID Data
for Activity Monitoring of the Elderly**

Matthew Bai

Master of Computer and Information Sciences

2009

TABLE OF CONTENTS

AUTHORSHIP	i
ACKNOWLEDGEMENTS	ii
ABSTRACT	iii
1 INTRODUCTION	1
1.1 <i>Current Issues Associate with RFID Application</i>	1
1.2 <i>Research Motivation</i>	1
1.3 <i>Research Question</i>	2
1.4 <i>Objective and Structure of the Dissertation</i>	2
2 LITERATURE REVIEW	4
2.1 <i>RFID Applications</i>	6
2.1.1 <i>The Link of RFID Application to Ubiquitous Computing</i>	6
2.1.2 <i>Object Locating</i>	6
2.1.3 <i>Activity Monitoring</i>	7
2.1.4 <i>Proposed Assisted Living Systems for the Elderly</i>	8
2.1.5 <i>Data Cleaning Associated with Sensor-Based Deployments</i>	9
2.2 <i>Processing Method One: Take Non-Overlapping 1 Second (Every Second) Window Method and Multi-Second Window Method</i>	10
2.3 <i>Processing Method Two: The SMURF Windowing Concept</i>	14
2.3.1 <i>Using SMURF to Clean Single Tag Data</i>	15
2.4 <i>Research Methodology and Methods</i>	20
2.5 <i>Experimental Method</i>	21
2.6 <i>Justification of Selected Methodology – Experimental Investigation</i>	21
3 EXPERIMENT - DATA COLLECTION	23
3.1 <i>Experimental Design</i>	23
3.2 <i>Independent and Dependant Variables</i>	23
3.3 <i>Equipment</i>	24
3.3.1 <i>Reader</i>	24
3.3.2 <i>Tags</i>	25
3.4 <i>Experimental Procedure</i>	25
3.5 <i>Pilot</i>	27
3.6 <i>The Amount of Gathered Data for both Experiments</i>	28
3.7 <i>Experiment One</i>	29
3.8 <i>Experiment Two</i>	33
3.9 <i>Result of Experiment Two</i>	Error! Bookmark not defined.
4 EXPERIMENT - DATA PROCESS AND ANALYSIS	35
4.1 <i>Framework of Applied Method in Both Experiments</i>	35
4.2 <i>Differentiations in Data Collecting and Cleaning Methods for Both Experiments:</i> ...	37
4.3 <i>First Experiment</i>	38
4.3.1 <i>Data Process and Analysis – Non-overlapping 1 Second Window Method in Location Verification</i>	38

4.3.1.1	<i>Data Process</i>	38
4.3.1.2	<i>Analysis</i>	42
4.3.2	<i>Data Analysis - Fixed Multi-Second Window Method for Location Verification</i>	42
4.3.2.1	<i>Data Process</i>	42
4.3.2.2	<i>Analysis</i>	45
4.3.3	<i>Data Analysis – Overlapping Variable Window Size (SMURF) for Location Verification</i>	46
4.3.3.1	<i>Data Process</i>	46
4.3.3.2	<i>Analysis</i>	48
4.4	<i>Second Experiment</i>	49
4.4.1	<i>Data Analysis – 1 Second Window Method for Sequence Verification</i>	52
4.4.1.1	<i>Data Process</i>	52
4.4.1.2	<i>Analysis</i>	56
4.4.2	<i>Data Analysis - Fixed Window Method for Location Verification</i>	57
4.4.2.1	<i>Data Process</i>	57
4.4.2.2	<i>Analysis</i>	60
4.4.3	<i>Data Analysis – Sliding Window for Location Verification</i>	61
4.4.3.1	<i>Data Process</i>	61
4.4.3.2	<i>Analysis</i>	63
4.5	<i>Comparison of Processing Methods</i>	64
4.5.1	<i>Issues Relating to Practical and Conceptual Design</i>	67
4.5.2	<i>Hardware and Tag Placement Issues</i>	67
4.5.3	<i>Power and Communication</i>	67
4.5.4	<i>Other Problem Encountered</i>	68
4.5.5	<i>Limitations</i>	68
4.5.6	<i>Future Study</i>	69
5	<i>CONCLUSION</i>	70
6	<i>REFERENCES</i>	71
7	<i>APPENDIX</i>	77
7.1	<i>Appendix A: Original RFID Datasets of Pilots</i>	77

LIST OF ABBREVIATIONS

<i>Abbreviation</i>	<i>Full Description</i>
1. <i>ADL</i>	<i>Activities of Daily Life</i>
2. <i>AUT</i>	<i>Auckland University of Technology, New Zealand</i>
3. <i>LANDMARC</i>	<i>Location Identification based on Dynamic Active RFID Calibration</i>
4. <i>RFID</i>	<i>Radio-Frequency Identification</i>
5. <i>SMURF</i>	<i>Statistical sMoothing for Unreliable RFid data</i>
6. <i>UHF</i>	<i>Ultra High Frequency</i>

LIST OF FIGURES

<i>Figure 1. How Information is transferred through Low-Frequency Passive RFID Tags (Weinstein, 2005).</i>	4
<i>Figure 2. Frequency Characteristics of RFID Systems (Chawla et al., 2007).</i>	5
<i>Figure 3. Potential Landmark Locations (Symonds et al., 2007).</i>	9
<i>Figure 4. Example Reader Tag List.</i>	11
<i>Figure 5. Static Non- Overlapping Window in Tracking RFID Tag.</i>	13
<i>Figure 6. Different Setting for Smoothing-Window Size in Tracking RFID Tag (Jeffery et al., 2006).</i>	14
<i>Figure 7. Attaining the Midpoint of the Windows</i>	18
<i>Figure 8. Reader Configuration Interface.</i>	26
<i>Figure 9. Metronome Timer is used to get Precise Time and Epoch Indications.</i>	27
<i>Figure 10. Tags' Layout and Cycle Movement Sequence in Experiment One.</i>	29
<i>Figure 11. Tracient RFID Reader with Lanyard.</i>	30
<i>Figure 12. Data Processing and Analysis Methods of Indoor Experiment.</i>	35
<i>Figure 13. The Framework of Research Flow for 1 Second Window, Static Window and Sliding Window Methods in Both Experiments.</i>	36
<i>Figure 14. The Result of Using a 1 Second Window Method in Data Processing from 3 Test Groups.</i>	41
<i>Figure 15. The Result of Using 2 Second Window in Data Processing from 3 Groups.</i>	43
<i>Figure 16. The Result of Using 4 Second Window in Data Processing from 3 Groups.</i>	43
<i>Figure 17. The Result of Using 6 Second Window in Data Processing from 3 Groups.</i>	44
<i>Figure 18. Code and Algorithm for SMURF Adaptive Per-Tag Cleaning (Jeffery et al., 2006).</i>	46
<i>Figure 19. The Result of Using Simplified SMURF in Data Processing from 3 Groups.</i>	47
<i>Figure 20. Code for Identifying the Minimum Number of Differences between Two Strings.</i>	51
<i>Figure 21. Comparing Experimental Results with Reference Group by Using VB.net Levenshtein Distance Comparison Algorithm.</i>	54
<i>Figure 22. The Result of Using 1 Second Window Method in Data Processing from 3 Groups.</i>	54
<i>Figure 23. The Result of Using a 2 Second Window in Data Processing from 3 Groups.</i>	58
<i>Figure 24. The Result of Using a 4 Second Window in Data Processing from 3 Groups.</i>	58
<i>Figure 25. The Result of Using a 6 Second Window in Data Processing from 3 Groups.</i>	58
<i>Figure 26. The Result of Using Sliding Window in Data Processing from 3 Groups.</i>	62
<i>Figure 27. First Experiment: The Comparison of the Result from Non-overlapping 1 Second Window, 4 Second Window and Overlapping SMURF of Multiple Tags in One Chart.</i>	65
<i>Figure 28. Second Experiment: The Comparison of the Result from Non-overlapping 1 Second Window, 4 Second Window and SMURF Methods of Multiple Tags in One Chart.</i>	66
<i>Figure 29. Pilot of Single Tags.</i>	79
<i>Figure 30. Pilot of Single Tags by using Non-Overlapping Window Method.</i>	82
<i>Figure 31. Pilot of Single Tags by using Non-Overlapping Window Method.</i>	84
<i>Figure 32 Pilot of 2 Tags by using 1 Second Window Method explain what this is and label the Axes.</i>	86

<i>Figure 33. Pilot 1 of Multiple Tags by using 2 Sec Static Window Method.....</i>	<i>87</i>
<i>Figure 34. Pilot of Multiple Tags by using 4 Sec Static Window Method.....</i>	<i>87</i>
<i>Figure 35. Pilot of Multiple Tags by using 6 Sec Static Window Method.....</i>	<i>88</i>
<i>Figure 36. Pilot of Multiple Tags by using 1 Second Window Method.....</i>	<i>89</i>
<i>Figure 37. Pilot of Multiple Tags by using 2 Sec Static Window Method.....</i>	<i>89</i>
<i>Figure 38. Pilot of Multiple Tags by using 4 Sec Static Window Method.....</i>	<i>90</i>
<i>Figure 39. Pilot of Multiple Tags by using 6 Sec Static Window Method.....</i>	<i>90</i>
<i>Figure 40. Pilot of Multiple Tags by using 1 Second Window Method.....</i>	<i>91</i>
<i>Figure 41. Pilot of Multiple Tags by using 2 Sec Static Window Method.....</i>	<i>91</i>
<i>Figure 42. Pilot of Multiple Tags by using 4 Sec Static Window Method.....</i>	<i>92</i>
<i>Figure 43. Pilot of Multiple Tags by using 6 Sec Static Window Method.....</i>	<i>92</i>
<i>Figure 44. Pilot of Multiple Tags by using 1 Second Window Method.....</i>	<i>93</i>
<i>Figure 45. Pilot of Multiple Tags by using 2 Sec Static Window Method.....</i>	<i>93</i>
<i>Figure 46. Pilot of Multiple Tags by using 4 Sec Static Window Method.....</i>	<i>94</i>
<i>Figure 47. Pilot of Multiple Tags by using 6 Sec Static Window Method.....</i>	<i>94</i>

LIST OF TABLES

<i>Table 1. How 1 Second Window Method Produces Values.....</i>	<i>12</i>
<i>Table 2. How Non-Overlapping Static Window Produces Values from RFID Data Stream.</i>	<i>12</i>
<i>Table 3. How SUMRF Produces Values by using Sliding and Overlapping Window.</i>	<i>16</i>
<i>Table 4. The First Example of Calculating the Probability for Each Epoch.....</i>	<i>16</i>
<i>Table 5. The Second Example of Calculating the Probability for Each Epoch.</i>	<i>17</i>
<i>Table 6. The Third Example of Calculating the Probability for Each Epoch.</i>	<i>17</i>
<i>Table 7. Main Features of Tracient Padl-R UF RFID Reader (Tracient, 2009).</i>	<i>24</i>
<i>Table 8. Use Simple Name to represent tags in Experiment One.</i>	<i>30</i>
<i>Table 9. Example of Data Collection for Reference Group.....</i>	<i>32</i>
<i>Table 10. Use Simple Name to Represent Tags in Experiment Two.</i>	<i>33</i>
<i>Table 11. Location of Results from both Experiments and Various Analysis Methods.....</i>	<i>36</i>
<i>Table 12. Collected Data Source from both Experiments and Their Distribution for Data Processing.....</i>	<i>37</i>
<i>Table 13. An Example of How the Test Data is compared with the Reference Group.</i>	<i>39</i>
<i>Table 14. Calculated Mean for Three Groups.....</i>	<i>42</i>
<i>Table 15. Calculated Mean for Three Groups on Three Different Size of Windows.....</i>	<i>44</i>
<i>Table 16. Calculated Mean for Three Groups.</i>	<i>47</i>
<i>Table 17. Example of Compressed Data in Experiment Two.....</i>	<i>50</i>
<i>Table 18. The Example of How We Compare the Test Data with the Reference Group.</i>	<i>53</i>
<i>Table 19. Calculated Mean for 1 Second Window Method.</i>	<i>55</i>
<i>Table 20. Calculated Mean for Three Groups on Three Different Size of Windows.....</i>	<i>59</i>
<i>Table 21. Calculated Mean for Three Groups.</i>	<i>62</i>

LIST OF EQUATIONS

<i>Equation 1. How to Calculate the Average of Probability.</i>	<i>17</i>
<i>Equation 2. How to Calculate the Window Size from Probability of Detections.</i>	<i>17</i>
<i>Equation 3. How to Calculate the Proportion of Matching between Two Tests.</i>	<i>41</i>
<i>Equation 4. How to Calculate the Proportion of Difference for Two Sequences.</i>	<i>53</i>
<i>Equation 5. The Improvement Made by using SMURF in Comparison with the Best Result of Static Window Method, In Location Alignment.</i>	<i>64</i>
<i>Equation 6. The Improvement Made by using SMURF in Comparison with the Best Result of Static Window Method, in Sequence Detection.</i>	<i>65</i>

LIST OF NAME OF PROGRAMS

1. *Program 1: Metronome Timer*
2. *Program 2: Matlab – 1 (Every) Second Window Data Proceeding*
3. *Program 3: Matlab – Static Window Data Processing Program;*
4. *Program 4: Matlab - Statistical sMoothing for Unreliable RFid data
(SMURF) Cleaning;*
5. *Program 5: VB.net Levenshtein Distance Comparison Algorithm;*

AUTHORSHIP

“I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or institution of higher learning.”

A handwritten signature in black ink on a light yellow background. The signature is stylized, starting with a large 'M' and ending with a long horizontal stroke.

Signature

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor David Parry, who afforded me an excellent opportunity to conduct research in the area of ubiquitous computing and emerging technology, as well as my study background and working experiences. During these studies, he helped me with explicating the research goal, enhanced my field knowledge, provided me with rigorous and valuable feedback, and directed me, a novice researcher, to an exciting academic world.

I also would like to thank my parents for their love and support, and for believing in me, and my girl-friend for supporting me.

Finally, I would like to thank my friends for taking photographs for me during the research, and for their advice.

ABSTRACT

Radio-Frequency Identification (RFID) system functions as a potentially flexible and low cost tool for both object locating and human activity tracking. In this dissertation, a comparative literature review was employed initially, in order to gain wide background and theoretical evidence to answer our research questions. Experimental investigations were carried out, which were focused on examining and evaluating the effectiveness of RFID performance.

Bearing in mind the challenges confronted by the elderly, we developed prototypes in our experiments. In order to improve the data reliability and overall performance of RFID application, this experimental investigation methodology was used within a positivistic paradigm. The research focussed mainly on the development and evaluation of tools to clean the RFID data stream and improve the identification of activity. Based on analysis of experimental results, we examine whether fixed (or static) window cleaning method or Statistical sMoothing for Unreliable RFID data (SMURF) middleware (Jeffery, Garofalakis & Franklin, 2006) is a viable and cost-effective candidate to produce more reliable data for human activity monitoring.

1 INTRODUCTION

RFID is a novel wireless technology to record data transmission and precisely identify objects. It retrieves the unique identification from the tagged item through radio waves (Want, 2004a). The basic premise behind the RFID system is that the RFID reader sends radio frequency signals to tags, these signals contain messages that are readable by tags and reflect back to the RFID reader (Roussos, 2006). RFID functions similarly to a bar code, but it has many advantages in contrast to bar code, including the ability to be read without the line of sight and multiple tags being able to be read at the same time. As ubiquitous computing and information technology evolve, RFID is becoming more and more viable in the areas of inventory control, retail and healthcare.

Recently, there has been a trend for eldercare-focused assistive technology that plays a vital role in most western countries, especially in the home environment and for Activities of Daily Life (ADL) monitoring (Symonds, Ayoade & Parry, 2007). Improving the data reliability based on activity monitoring is the topic that has generated much recent researches. These proposed solutions allow higher location precision, assist in indoor location management, and moreover, they may reduce long-term costs and improve productivity of healthcare providers.

1.1 Current Issues Associate with RFID Application

Theoretically, normal Ultra High Frequency (UHF) tags can offer a long reading distance: 4 – 5 meters (Chawla & Ha, 2007). However, practical study reveals a limited reading range, which can be less than 100mm. According to Jeffery, et al. (2006)'s research, the actual reading rate of RFID data stream is only around 60-70% in a real-world context.

This research is concerned with the comparison of both the static windowing method and SMURF method with raw data stream, to approve the best way of retrieving more accurate and reliable results, in an activity monitoring scenario.

1.2 Research Motivation

RFID activity detection is based on recording the interaction between people holding an RFID detector and RFID tagged objects. However the problem generally encountered is that these data are unreliable. There are a variety of methods that have been proposed to limit and filter error readings including static windowing and SMURF middleware. This dissertation is motivated by interest in support for activity detection for the elderly.

1.3 Research Question

In order to detect activity and location, two approaches to accuracy are used. Firstly, the match between actual location and derived location at any particular time and secondly the match between the actual sequence movement and that derived from the RFID data. Our research question is “How can windowing methods, including SMURF, increase the accuracy of RFID - derived activity and location measures?”

1.4 Objective and Structure of the Dissertation

The objective of this study is to examine and evaluate the effectiveness in a model of a proposed RFID activity detection system. This study simulates a ubiquitous computing system, through the adoption of both non-overlapping static window and overlapping sliding window approaches to find a viable solution in terms of decreasing or limiting dropped readings.

We provide a background study on RFID technology in Chapter 2. It includes a comprehensive survey on existing RFID technology, which provides research background by reviewing prior researches on both RFID applications and activity monitoring.

Chapter 3 discusses our research question and methodology with justifications. It focuses on explaining the selection of research paradigm, and how the research question was formulated. Moreover, it describes the major methods in developing data collection, processing and analysis.

Chapter 4 includes the procedures of data collection of two experimental investigations in terms of using RFID devices to simulate a real living environment for the elderly. It also explains how data were collected, including the explanations of the experimental environment and devices.

In Chapter 5, the data analysis section, it explains how different methods in data processing and analysis affect the results.

In Chapter 6 the discussion provides a summary of reflection on the whole project. Afterwards, we summarise the research findings and results, discuss the limitations in both research design and experimental results, and provide directions for future study.

Chapter 7 is the conclusion chapter which attempts to find the answer to the research questions formulated in this (Introduction) section.

2 LITERATURE REVIEW

The basic principle of RFID technology is that a RFID reader interrogates surrounding tags by emitting radio frequency (RF) signals. These RF signals contain hard coded data that are readable by tags. Nearby tags response to these signals with their unique identification which can represent its attached objects (Roussos, 2006).

RFID technology is widely adopted in different areas, in particular, the fields of shipment tracking and supply chain control. Key RFID applications include livestock ID, supply chains and automated libraries (Chawla & Ha, 2007). On the other hand, the adoption of RFID systems in terms of positioning and assistance of human activities is also emerging. RFID tags are categorised into three types upon operational principle: passive, semi-passive, and active. Active and passive tags are distinguished by their powering modes rather than their accessing capability (Gaetano, 2005). Generally, information transferred between RFID reader and tag will be simplified as four steps, as shown in *Figure 1*:

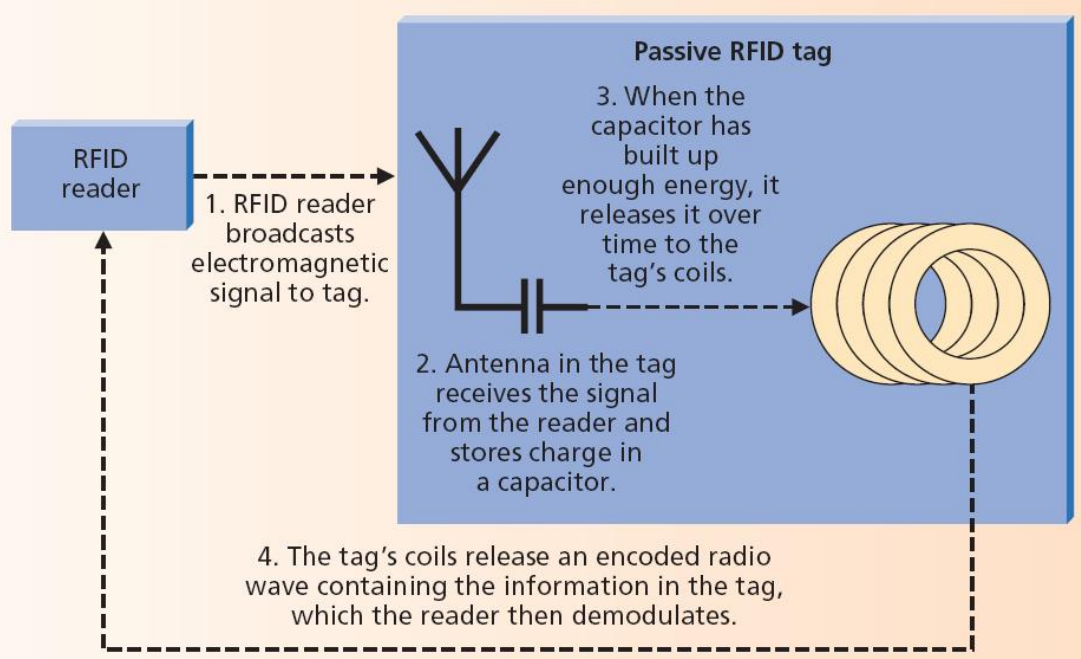


Figure 1. How Information is transferred through Low-Frequency Passive RFID Tags (Weinstein, 2005).

In order to reduce the overall cost and size of the RFID tags, passive tags are designed without the use of batteries (Sarma, Brock & Engels, 2001). They are fully energised by the interrogating readers. To ensure the functionality, these tags have a built-in antenna, which assists in capturing electronic signals from the reader. The embedded capacitor saves energy from the RF signal that is emitted by the reader, and reflects it back once capacitor is fully charged. Weinstein (2005) explains this process as an inductive coupling. Reading cycles consist of a single interrogation cycle, namely a single iteration through the reader's protocol that attempts to determine all tags in the reader's vicinity (Jeffery et al., 2006). Passive RFID tags have the benefits of being cheap, tiny, portable and plentiful.

More sophisticated tags have a wider reading range and higher ability to be detected even close to metal backing or wet surfaces, but may become more costly. Chawla & Ha (2007) summarise the RFID tags by their bands, frequency characteristics, and corresponding standards as shown in *Figure 2*.



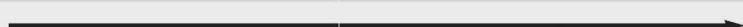
Frequency range	< 135 KHz [LF]	13.56 MHz [HF]	860–960 MHz [UHF]	2.45GHz [Microwave]	
Relevant standards	<ul style="list-style-type: none">• ISO 11784 & 11785• ISO/IEC 18000-2• ISO 14223-1	<ul style="list-style-type: none">• ISO/IEC 18000-3• EPC class-1• ISO 15693• ISO 14443 (A/B)	<ul style="list-style-type: none">• ISO/IEC 18000-6• EPC class-0, class-1	<ul style="list-style-type: none">• ISO/IEC 18000-4	
Typical read range	<0.5 m	~1 m	~4–5 m	~1 m	
Tag type	Passive-inductive coupling	Passive-inductive coupling	Passive or active	Passive or Active	
Typical applications	Access control, animal tagging, vehicle immobilizer	Smart cards, access control, payment ID, item-level tagging, baggage control, biometrics, libraries, transport, apparel	Supply chain pallet- and box-level tagging, baggage handling, electronic toll collection	Electronic toll collection, cold chain management, environment monitoring	
Multiple tag read rate	Slower				Faster
Ability to read near metal or wet surfaces	Better				Worse
Passive tag size	Larger				Smaller

Figure 2. Frequency Characteristics of RFID Systems (Chawla & Ha, 2007).

2.1 RFID Applications

RFID technology has been widely used in the areas of object tracking and identification. In particular, research conducted by Sarma, et al. (2001) indicate that RFID is set to revolutionise industrial control as it holds the potential to simplify and make more robust on the tracking of parts or part carriers through manufacture, storage, distribution and ultimately the supply chain. Alternatively RFID has been proposed for human activity tracing in many recent studies. As Kutiyawala, et al. (2004) state, RFID can be used in a robot-assisted indoor environment, which guides people with amblyopia to avoid indoor barriers.

2.1.1 The Link of RFID Application to Ubiquitous Computing

RFID is sometimes described as the key component for the accomplishment of ubiquitous computing. Researchers define the RFID system as the most pervasive computing technology in history (Michael & McCathie, 2005). Firstly, passive tags can be attached to surfaces for use in indoor environments, such as walls and carpet, (Lin, Su, Chen, Wang, Lin & Chen, 2007). Secondly, RFID readers can form a grid, which can trace the movement of tagged objects based on analysis of the grid's known location information. Thirdly, RFID operates without human involvement, since RFID tags could respond to any readers within a certain range even without human intervention. This is an obvious feature of ubiquitous computing.

2.1.2 Object Locating

One way of looking at the process of finding objects is analogous to ants finding their way to food. An ant is able to mark the returning track by spreading pheromone once food is located. This can also help other ants collaborating and relocating the food (Mamei & Zambonelli, 2007). There are methods, such as the RFID embedded home environment, of using RFID to support object locating in a similar manner to ants finding food. Likewise, RFID tagged environments support people in locating objects. When a tagged individual walks into the room, indoor facilities with RFID readers could record the movement of the individual, and alert them to certain obstacles. Moreover, the system helps users in identifying lost objects (Romer, Schoch & Mattern,

2004). As Romer, et al. (2004) suggest, when an individual carrying a RFID reader walks into the house, the tags attached on different objects are able to respond to the read and provide data for the person's trail. Moreover, the tagged environment stores a trace of the locations of objects such as glasses, therefore, users are able to find objects they lost or which were moved by other people.

Ni, Liu, Lau, & Patil (2004) performed experiments on the Location Identification based on the Dynamic Active RFID Calibration (LANDMARC) system. This system utilises the technology of placing extra location reference tags to help location calibration. Through a low-cost tag network, the system dynamically updates the reference information for lookup based on the detected range from the reference tags in real-time. It is examined from their experiment that LANDMARC system provides more accurate and reliable information on location and object identification. Ni et al. (2004) define the LANDMARC as a flexible and dynamic approach, which can achieve much more accurate and close to real-time location sensing.

2.1.3 Activity Monitoring

There are systems proposed for using ubiquitous computing scenarios to derive people's activity and intentions, such as an activity trace in the home environment, which adopts RFID-driven gloves and bracelets to detect tagged objects (Smith et al. 2005). The system involves the use of active sensor beacons that allow accurate object identification. Rather than direct observation, their scheme provides an indirect approach to inferring people's behaviour through the analysis of their interactions. However they face the issue of batteries, which makes them impractical for long-term dense deployment.

Moreover, Philipose, et al. (2004) illustrate that the use of RFID gloves for activity monitoring in daily living. Wearable sensor gloves are used in their research; the glove interacts with a RFID tagged object, then the system derives the participant's activity based on the gathered information from the tagged objects. Similarly, Philipose, et al. (2004)'s approach infers activity through a data stream gathered by sensor gloves, but the difference is that Philipose, et al. (2004) concentrate on the analysis of a probabilistic

sequence of used objects. It is emphasised that the main aims of both mentioned projects are in activity tracking rather than object detection, with the system designed to track removal and return of items from a specific storage area. Due to technical limitations, users have to operate the RFID device (reader) within an effective range.

2.1.4 Proposed Assisted Living Systems for the Elderly

The key challenge associated with elderly people living alone is falls or other incapacitating events. Moreover, they are unable to summon help once these falls occur. Falls are the major cause of most morbidity and are hard to prevent. However, research is trying to assist elderly people in dealing with these incapacitating events by adopting a behaviour or activity inferring system. Gross activity measurement has been performed by many groups. As a recent paper reveals, this monitoring system is able to be attached out of sight, which makes them particularly well suited for ADL monitoring of the elderly (Philipose et al., 2004).

On the other hand, there are challenges related to effectively and accurately inferring people's behaviour in their indoor environment. This includes inferring what activity users are performing, how they are performing it, and its current stage (Philipose et al., 2004). In fact, some elderly people have early-stage cognitive decline, and their behaviours may be more variable than younger people. Hence, it is especially difficult to detect and infer the behaviour and activities of the elderly group.

Symonds, et al. (2007) propose a system, which involves the placement of RFID tags around the user's home (landmarks). The user is equipped with a reader that has built-in data stream storage. There is also a wireless channel formed between reader and main data storage via Bluetooth, which allows real time communication. During the operation and interaction of reading equipment with tags, recorded detections of objects and landmarks are stored and transferred continuously. Once the user deposits an object, it becomes in-visible in the system. The object's location is recorded in reference to landmarks visited before and after (Symonds et al., 2007). Gathered logs also represent

the sequence or logic of these interactions, which is then mapped to potential activity sequences and behaviours, see *Figure 3*.

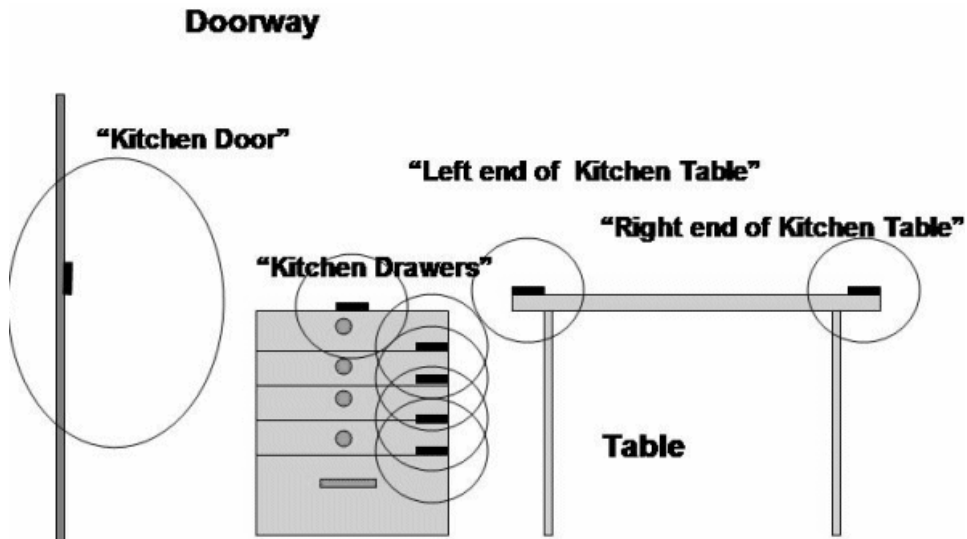


Figure 3. Potential Landmark Locations (Symonds et al., 2007).

The use of landmarks has the benefit of allowing different effective accessing ranges, as well as adapting spatial resolution to meet the requirement of specific environments (Symonds et al., 2007). Their system also reduces the overall costs, and limits the effect on others within the same space. Moreover, this approach is suitable for both less frequently used personal and public area installations (Symonds et al., 2007).

In order to derive activity, we need more sophisticated methods in RFID data cleaning.

2.1.5 Data Cleaning Associated with Sensor-Based Deployments

As motioned in 1.1, the observed efficiency of reading rate of RFID detection is only around 60 -70% in a real-world deployment context (Jeffery et al., 2006). There is a great amount of reading errors in every RFID data stream. The data produced by sensors must be appropriately cleaned to compensate for these failures before they can be used by any application. The detection rates vary significantly for different readers and tags. According to Jeffery et al. (2006)'s experiments, there is a roughly 75% chance of detections within the full detection distance by using an Alien reader with Alien Squiggle tag. However, the proportion of tags detected in this region drops dramatically to 25%

once they use a Sensormatic reader with Alien I2 tag under noisy conditions (Jeffery et al., 2006).

The above detection limitations cause errors during deployment, and high error rates result in raw RFID data being essentially useless in many areas of applications (Jeffery et al., 2008). In order to correct reading errors, and allow data streams to meet the high level information requirements, we deploy the RFID middleware system between reader and applications. This approach is widely developed in the inventory control area. The basic concept behind this middleware is using a non-overlapping static window or an overlapping sliding window approach, which interpolates dropped readings for every tag within the time window on the collected data stream (Jeffery et al., 2006). There is a novel sampling-based approach: SMURF, which is introduced in order to reduce the reading dropping rate on the statistical basis, and incorporating cleaning logic in applications produces more reliable RFID data stream. Through an adaptive smoothing filter for cleaning raw RFID data streams, this greatly increases the complexity of the application. To ensure that all tags in the reader's detection range are read, the smoothing window must be large enough to correct for reader unreliability. Small window sizes will easily omit some tags, causing a false negative. On the other hand, adopting large windows will limit the detection of the tag's movements (or reader's movement), resulting in a false positive (Jeffery et al., 2008).

Due to the differences in reading range for different types of RFID readers and tags, we have to face the reality of being unable to completely detect every single movement, which will cause the RFID data to be unreliable.

2.2 Processing Method One: Take Non-Overlapping 1 Second (Every Second) Window Method and Multi-Second Window Method

In fact, multiple reader interrogation cycles are grouped into time sectors. We define these time sectors as epochs. Generally epochs present units of time or a number of interrogation cycles. Researchers range these epochs from 0.2 to 0.25 seconds (Want, 2004), and the length of epoch can also be extended or reduced upon various requests.

For individual epochs, the RFID reader traces the information about all detected tags. This information also includes multiple interrogation responses from every tag. Normally these data streams are saved into built-in memory. A typical stream list consists of each tag's identification (includes multiple responses) and the time that the tag responded. This information is then periodically transferred into the client's main storage, such as a PC, for further use as shown in *Figure 4*.

<i>Tag's ID</i>	<i>Time of Detection</i>
<i>E2003411B802011371233241</i>	<i>16:56:06 25/May/2009 (LOCAL)</i>
<i>E2003411B802011371233245</i>	<i>16:56:07 25/May/2009 (LOCAL)</i>
<i>E2003411B802011371233239</i>	<i>16:56:07 25/May/2009 (LOCAL)</i>
<i>E2003411B802011371233241</i>	<i>16:56:07 25/May/2009 (LOCAL)</i>

Figure 4. Example Reader Tag List.

In fact, the scheme in terms of using a windowing approach consists of two levels of measurements: epochs and windows. An *epoch* is a basic time element in which one or more reads of a tag is performed; this could be a second in our case. Epochs do not overlap, and windows do not overlap either in this non-overlapping window method.

A *window* defines a finite set of readings (in terms of an interval of time) within a data stream. Within a window, we can aggregate multiple readings into one, or compare readings to detect outliers. The basic idea of using 1 second as a window (every window) and Multi-Second window methods (larger than 1 second) is that when we group data (readings) into static window, a window is a group of epochs. The static (fixed-width) non-overlapping windows of 1, 2 seconds etc are called "1 second window", "2 second window" etc, from now on to avoid complicating the text. For this first 1 second-window method, we treat every epoch as a separate window. Then we select the most frequently appearing tag within each epoch, as shown in *Table 1* (simulated data are used).

Table 1. How 1 Second Window Method Produces Values.

Epoch Number	Tag 1	Tag 2	Tag 3	Most Detected Value for Tag
1	0	1	0	Tag 2
2	2	0	0	Tag 1
3	3	0	0	Tag 1
4	1	0	0	Tag 1
5	0	3	1	Tag 2
6	0	4	0	Tag 2

We select the highest number of detected tags from each epoch, as shown in *Table 1*. For the first epoch we have 1 read for Tag 2 (in first grey cell), and 0 read for both Tag 1 and Tag 3. Obviously, the most frequently appearing tag is Tag 2, thus we use Tag 2 to represent the first epoch. Similarly, we calculate the highest value of the second epoch (Tag 1 = 2, in second grey cell) and third epoch (Tag 1 = 3, in third grey cell), and get a list of values for every separate epoch.

We could also group several epochs into one window as we mentioned above. For instance, if we have a window that includes 3 epochs, we then get a list of results as shown in *Table 2* (Simulated data are used).

Table 2. How Non-Overlapping Static Window Produces Values from RFID Data Stream.

Window Number	Epoch Number	Tag 1	Tag 2	Tag 3	Window Produces Value for Tag
1	1	0	1	0	Tag 1
	2	2	0	0	
	3	3	0	0	
	Total Number of Reads:	5	1	0	
2	4	1	0	0	Tag 2
	5	0	3	1	
	6	0	4	0	
	Total Number of Reads:	1	7	1	
3	7	0	3	0	Tag 2
	8	0	1	0	
	9	2	0	0	
	Total Number of Reads:	2	4	0	
4	10	0	1	4	Tag 3
	11	0	0	2	
	12	0	0	1	
	Total Number of Reads:	0	1	7	

We select the highest value from every window, as shown in *Table 2*. For the first window we have a total number of 5 reads for Tag 1, 1 read for Tag 2 and 0 read for Tag 3 (as shown in first grey line). The most frequently appearing tag is Tag 1, thus we use Tag 1 to represent the first window. Similarly, we calculate the highest value for each window, and get a list of values for every window. Tag 2 represents the second window with the value of 7 (as shown in second grey line), Tag 2: 4 for the third window and Tag 3: 7 for the fourth window.

For example, if the window's size changes from 3 to 5 epochs per window, we then adjust the group size accordingly, and may get a different result. The 1 second window method use 1 second as the basic size of its window, namely we compare readings for every second, (thus 1 window = 1 second). This is different from Multi-Second window method in terms of both window size and processing program, thus we distinct these two methods in data processing section.

There is a limitation with the use of above non-overlapping window method, as small window sizes will easily omit some tags, which causes a false negative. Large window limits the detection of the tag's movements, which causes a false positive, as shown in *Figure 5*.

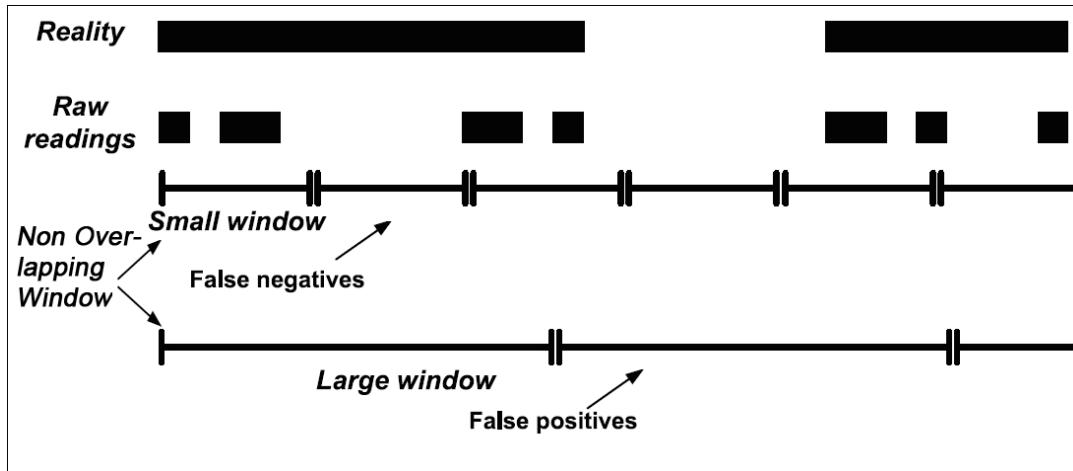


Figure 5. Static Non- Overlapping Window in Tracking RFID Tag.

2.3 Processing Method Two: The SMURF Windowing Concept

As we discussed in 2.1.5, the primary factor limiting the widespread adoption of RFID technology is the unreliability of the data streams produced. Recent work by Jeffery, et al (2008) has identified an RFID middleware system that alleviates issues associated with using RFID data through adaptive techniques based on a novel statistical framework (SMURF). According to Jeffery, et al. (2006), the objective of this middleware is to decrease the dropping reads through allowing every individual tag its opportunity to be accessed within a normalised window. From *Figure 5* and *Figure 6*, for both non-overlapping and overlapping windows, we can find that “small windows fail to fill in dropped readings (False Negatives); large windows fail to capture tag movement (False Positives) by RFID readers” (Jeffery et al., 2006).

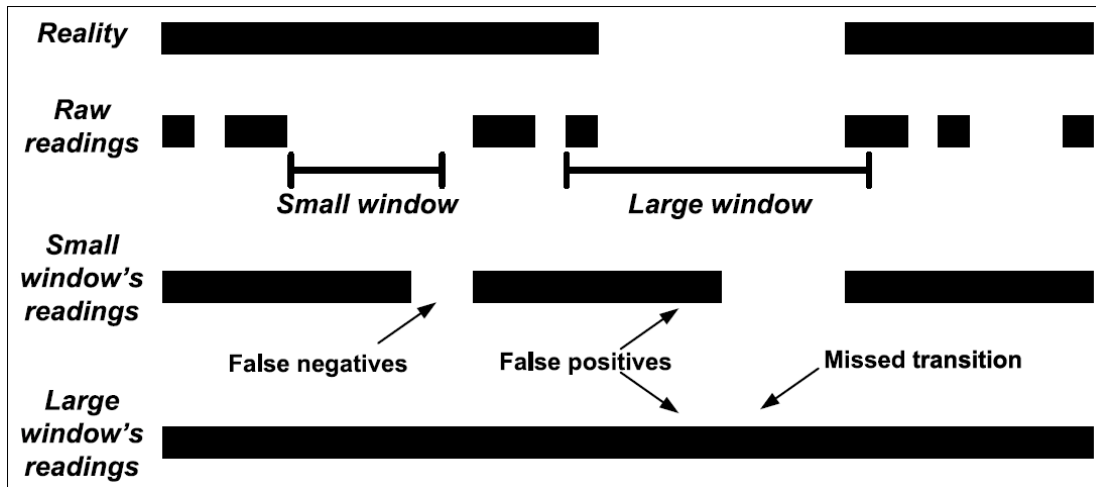


Figure 6. Different Setting for Smoothing-Window Size in Tracking RFID Tag (Jeffery et al., 2006).

The core of Jeffery, et al. (2006)’s smoothing window concept is to ensure the size of the smoothing window is large enough to retrieve remote tags that are normally missed by small windows’ detection, then optimise the unfiltered data (Jeffery et al., 2008). On the other hand, because larger window sizes are sometimes unable to always detect the tag, a small proportion of data streams always drop the tag for no reason, which leads us to mistakenly think this missed tag is out of detection range (results in a False Positive).

Jeffery, et al. (2006) explain that window size must be deeply studied for any RFID filtering scheme deployment. According to their (Jeffery et al., 2006) findings, an efficient RFID filtering solution will substantially reduce the cost of understanding environment characteristics, and configuring devices and software to adopt these factors. Moreover, to guarantee the long term data reliability, the window size is also expected to be reconfigured even for the same filtering middleware deployment from time to time.

2.3.1 Using SMURF to Clean Single Tag Data

SMURF has variants for both detecting single tags and multiple tags. In our case because each location only has one tag, we use the single tag approach. It should be noted that this approach also reverses the model used to develop SMURF which is one where the tags move, and the reader is fixed e.g. a conveyor belt. In our case the tags are fixed and the reader moves.

Generally, we group gathered RFID data into epochs and windows (mentioned in 2.2). SMURF produces a value of ‘detected’ or ‘not detected’ at the midpoint of its window. It consists of two major filtering mechanisms for both single and multiple tag data stream cleaning. Firstly it ensures a precise data stream for a single tag. Secondly, it is able to accurately estimate the total amount of the participating tags through statistical methods.

Although epochs do not overlap, windows do overlap in this approach. Each window starts at one epoch then the next window starts at the next epoch and so on. This is different to the fixed window that motioned in 2.2, which does not involve overlapping method. For instance, if we have epochs of 1 second and a window of five seconds, we could get a list as shown in *Table 3* (simulated data are used):

Table 3. How SUMRF Produces Values by using Sliding and Overlapping Window.

Window Number	Window starts at Epoch	Window ends at Epoch	Window produces value at epoch number
1	1	5	3
2	2	6	4
3	3	7	5
4	4	8	6

From the example above, we find that value 3 (first grey cell) is the middle point between Epoch 1 and Epoch 5, and value 4 (second grey cell) is the middle point between Epoch 1 and Epoch 5, etc.

SMURF applies a data-filtering approach that adopts a sliding window processor for fine-grained RFID data smoothing (Jeffery et al., 2006). There are two stages to get the size of the window. Firstly, we calculate the probability of detection for each tag in every epoch. For instance, assuming we perform 10 scans within a second (a unit of Epoch in this case), and if we only detect 5 times within that second, then the probability of detection in this epoch is $0.5 \left(\frac{5}{10}\right)$. Thus we can get the average “ P_i^{avg} ” = 0.5 over the whole window. Secondly, we have to set the window size, in order to observe each tag with more opportunity to be detected, we provide an example to explain this algorithm (*Table 4*). Afterwards, we need to do a probability calculation per epoch from the large sample. At this stage, we assume that size of window is a constant. For instance, we can get the individual probability for each epoch, as *Table 4* shows:

Table 4. The First Example of Calculating the Probability for Each Epoch.

Epoch Number	Probability of Detection
1	0.6
2	0
3	0.5
4	0.6
5	0.7

Thus we get the P_i^{avg} as:

$$P_i^{avg} = \frac{0.6 + 0 + 0.5 + 0.6 + 0.7}{5} = 0.48,$$

Equation 1. How to Calculate the Average of Probability.

Then, we use the equation of:

$$w_i \geq \left\lceil \frac{\ln(1/\delta)}{P_i^{avg}} \right\rceil$$

Equation 2. How to Calculate the Window Size from Probability of Detections.

to set the number of epochs within the smoothing window (w_i), in order to guarantee that tag i is observed within w_i with probability $> 1 - \delta$ ($\delta = \text{Required Completeness Confidence}$). We use 95% confidence, then get the result of ≥ 6.2 for w_i from:

$$w_i \geq \left\lceil \frac{\ln(1/\delta)}{P_i^{avg}} \right\rceil \geq \left\lceil \frac{\ln(1/0.05)}{0.48} \right\rceil \geq 6.2$$

Since the window size has to be a constant, we need at least 7 epochs per window to ensure better reading.

Table 5. The Second Example of Calculating the Probability for Each Epoch.

Epoch Number	Probability of Detection
1	0.9
2	0.9
3	0.9
4	0.9
5	0.9

In this second example, as *Table 5* indicates: if the probability of detection in each epoch is relatively high and:

$$w_i \geq \frac{2.99}{0.9} \geq 3.3$$

So we need at least 4 epochs per window.

Table 6. The Third Example of Calculating the Probability for Each Epoch.

Epoch Number	Probability of Detection
1	0.1
2	0.1
3	0.1
4	0.1
5	0.1

Similarly, if the probability of each epoch is relatively low, in the third example as illustrated in *Table 6*, and is calculated as:

$$w_i \geq \frac{2.99}{0.1} \geq 29.9$$

Then we need at least 30 epochs for each window. SMURF randomly select the number of epochs to calculate the probability for its first window, it could be the first 5 epochs in this case. Then it adjusts the next window size according to the calculated probability. For instance, when the probability decrease to $P_i^{avg} = 0.1$, SMURF increases the window size to 30 (≥ 29.9) epochs for the next window.

As mentioned in Section 2.3, an epoch is a basic time element in which one or more reads of a tag is performed. For instance, we can use a second as a basic epoch. Epochs do not overlap. A window is a group of epochs. SMURF produces a value of ‘detected’ or ‘not detected’ at the midpoint of its window. SMURF recognises a tag reading for certain windows if at least one reading exists for the tag within that window. As Jeffery, et al. (2006) explain, there are two fundamental differences between SMURF’s sliding-window processor and conventional RFID filters: (1) SMURF implements partitioned RFID smoothing, and (2) SMURF adopts an epoch-based mid-window slide. As *Figure 7* illustrates, SMURF gathers the data of the midpoint from one window under normal window processing. As the window slips, it gathers the next midpoint of the windows, which overlap.

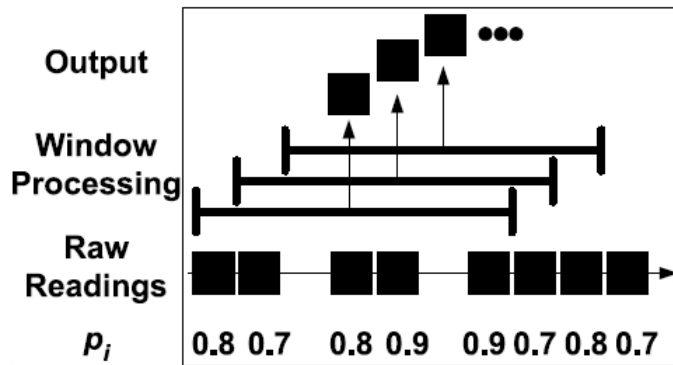


Figure 7. Attaining the Midpoint of the Windows

In order to ensure completeness, once there are low probabilities of detection occurring in a certain window, SMURF has to increase the window size to cover more epochs such that the tag has a high probability of being detected. Because of the simplicity of the test, the windows size w_i was not reduced during this work. Thus SMURF refers to this method that we have modified from the original in this dissertation.

METHODOLOGY

2.4 Research Methodology and Methods

A comparative literature review was adopted as the first stage of this study, systematically illustrating the emerging technology of RFID with its contemporary applications. Literature was gathered from a wide range of online libraries and databases, including: ACM, IEEE, ScienceDirect and Springer, along with RFID vendors' websites, and a number of standards organisations.

The methodology of the second part of this study involves a quantitative approach, by using the empirical investigation study within a design science paradigm. Most previous studies have employed a positivist paradigm as their primary research tool in the areas of RFID with assisted living research. According to Dobbins, Lane and Steiner (1998), experimental investigations are valuable and even studies using simulated data as subjects have validity. Moreover, Sridhara, et al. (2005) claim that a simulation approach is suitable for this type of research, large-scale deployments and evaluations of RFID performance in particular.

The design science paradigm plays a central role for problem solving within the IT discipline. Since IT research deals with artificial phenomena rather than natural phenomena, design science (or design research) seeks to create innovations that define the ideas, practices, technical capabilities, and products through which the analysis, design, implementation, and use of information systems can be effectively and efficiently accomplished. Rather than apply a method to examine and better understand reality, design science seeks an effective technique (Varshnavi & Kuechler, 2004). Design research methodology consists of five general stages: awareness of problems, suggestion, development, evaluation, and conclusion.

Design science is technology driven, and aims to create things to serve human needs. It adopts field experiments, which allow us retrieve and examine the data within real and specific environments in order to ensure the reliability of test. Design science creates the

methodological tools that natural scientists use. It has twofold research activities: to build and evaluate (March & Smith, 1995).

We use the key word “Build” to indicate the demonstration stage of the creation of the artefact, and to explain the technical principle behind this artefact. “Evaluate” represents the assessment stage. This process examines and evaluates the performance of constructed artefacts against certain criteria of value. It always prompts the questions: does it work? Is it improved? (March & Smith, 1995). As we assess the performance of RFID systems to improve their reliability by using various methods of data cleaning, our experiment examines the RFID data stream against criteria of utility, and provides a substantive test and practical proof of its actual performance. Our proposed experiment will examine Hypothesis 1: How the data stream effectively cleaned to support daily activity identification.

2.5 Experimental Method

Experimental studies aim to manipulate the independent variables to find the effect on the dependent variable. Field experiments in particular offer the advantage that they are conducted in a real situation, and thus avoid many of the drawbacks of the laboratory experiment. The only shortcoming for experimental studies is that laboratory settings are unable to completely reflect the actual environment (Collis & Hussey, 2003). However, in this study we conduct simulations of real home environments, carefully controlling movement to simulate a real situation.

2.6 Justification of Selected Methodology – Experimental Investigation

As Collis and Hussey (2003) explain, experimental research is a positivistic methodology. Experimental investigation is normally carried out in a systematic manner in a laboratory or natural setting.

Nevertheless, my research goal is to objectively examine and evaluate a real situation in a systematic manner. Thus adopting a positivistic paradigm with experimental investigation methodology is suitable for my project.

Moreover, Dobbins, Lane & Steiner (1998) recommend that the choice of research method should be based on the purpose of the research and the investigator's orientation. In their view, laboratory experiments are useful to examine behaviour at the individual level. There are also advantages of experimental study regarding the level of control and replication.

A simulation method has been widely employed, in many RFID based studies, whereby researchers combine both real experiments and simulation methods to evaluate the RFID performance, reading rate and algorithms in terms of positioning (Lin et al., 2007 and Breaha & Jons, 2006). In fact, real experiments allow researchers to gather more precise data, and guarantee the reliability of reference groups. Thus, we adopt this method for reference group data collection in the early stage of the experiment, and then employ the simulation approach to gather data that better represents the simulated circumstance.

3 EXPERIMENT - DATA COLLECTION

3.1 Experimental Design

The experiments described measure the performance of RFID equipment in a simulated activity tracking environment. In these experiments, we place tags in a real home situation in order to collect real data in a real-world environment, which underpins the validity of processing and analysis.

The basic equipment of our experiment consists of a mobile reader and few stationary tags. These tags are attached to objects or furniture, such as desks and doors.

Data were collected from monitoring simulated personal activity. To achieve this, a simulation experiment was setup in an indoor environment. The personal activity will be represented by a series of tags detected using various approaches such as SMURF, which is an RFID middleware platform, organised as a pipeline of processing stages (Jeffery et al., 2008). After data collection, the sequence generated from the RFID system will be processed and analysed. The known movement sequence or set of locations at various times will then be compared to the derived sequence or time/location list.

3.2 Independent and Dependent Variables

In order to control the validity of the simulation and determine the most effective method for data stream cleaning, we set the independent variable as the measured time at a location and sequence of these locations for the first experiment (actual movement). The dependant variable is the RFID tag detection time from the gathered data stream. In the second experiment, the sequence of locations identified is the independent variable (location sequence). Similarly, the dependant variable is the cleaned RFID tag detection sequence from the gathered data stream. Variable controls were the same RFID reader with the same configurations, and the same processing approach was used for all data, remaining constant during the experiment.

3.3 Equipment

3.3.1 Reader

Due to the high cost and configuration complexity for a RFID system with multiple stationary readers and tags, we believe a single mobile reader plus UHF tags will be more cost efficient and easier to use. A Tracient Padl_R UF reader is used in our experiments, it produces 855 – 960mHz Radio Frequency output. Unlike traditional and parallel driven readers, the reader used in our research has both USB and blue-tooth connections. Due to these characteristics, it has the benefit of being portable and having a long access range. The basic features of our experimental reader are displayed in *Table 7*.

Table 7. Main Features of Tracient Padl-R UF RFID Reader (Tracient, 2009).

Size	210mm(h) x 70mm (w) x 15mm (d) (approx)
Weight	135 grams (approx)
Operating temperature/Storage temperature	-10° to +60° C
Ingress Protection	IEC 529 IP64
Approvals	CE, C-Tick, FCC, EN302 208, EN301 489, EN 60950
Drop Specification	1m to plywood on concrete, all faces and corners.
User Interface	Single button for RFID reading, Multi-colour Leds, Audio beeper (multi tone). Configurable use case scenarios via USB or Bluetooth.
UHF Transponder	ISO 18000-6A, ISO 18000-68, ISO 18000-6C, EPC Class 0, EPC Class 1 (GEN2), and UHF Proprietary
Approximate read range	1.5m (typically, depends on transponder)
Date rate	~40kbps (EPC C1G2), ~80kbps (ISO 18000-6C)
O/P Power	0.5W
Bluetooth Interface	Class 2(typically 10m operation).
Antenna	Internal only
Additional Data Communication Interface	USB-mini-B connector for wired applications
Power Supply	Via USB for high-capacity internal Li-Ion battery recharge
Firmware Support	Field upgradeable via USB. Some configuration settings also available via Bluetooth interface.

3.3.2 Tags

Active tags have the benefit of longer detection distances. However, active tags are not adopted in our experiment due to two reasons. The first reason is active tags require an embedded battery, and this restricts the duration of use of active tags. Since passive tags function without an internal power source, these tags are more economical for large-scale deployment than active tags. Secondly, since the fundamental requirement of the investigation is to place and attach tags in different areas or allocate them to objects, passive tags provide enough functionality for indoor environment. A shorter range is an advantage as only the tag's ID is used to represent the landmark and object, so any detection within range will indicate the same location. In other words, the employment of large scale and highly sensitive devices, such as active tags is unnecessary, as affordable and smaller passive RFID tags are capable enough for all requirements of the experiment.

Due to the trade-off between costs, access range and power source, we employ Ultra High Frequency (UHF) passive tags in our experiment. The UHF has a wider detection range than normal passive tags, it operates between 860 – 960mHz, is 40mm (L) x 40mm (W) in size, and has the same frequency as our experimental reader. These tags are Alien UHF- “SquiggleTM”, which have been successfully deployed in the fields of inventory control and library management across the world.

3.4 Experimental Procedure

There are certain issues that can make simple mapping from tag to location difficult, such as the interference from other Radio Frequency sources or manufacturing differences, as tags may fail to be detected within the expected range (Symonds et al., 2007). This section describes how data were collected in order to test various data cleaning methods.

We simulate reader's movement and tag placement typical of a home scenario, each experiment performed over the course of 3 groups, each group includes 10 tests, and each test consists of 90 seconds of collection.

RFID tag detection has been used for recognising the relevant human motion, (as mentioned in Section 2.1.3). To conduct these experiments, a number of RFID tags are placed in my home. As longer intervals between reads have the potential to cause readers to fail to catch tags' responses while moving, the status of the reader was set to continuous mode, and its reading interval was adjusted to 0.05 seconds (Minimum interval), and pause time was set to 0.00 seconds. This reader will collect 20 interrogations within 1 second (See *Figure 8*).

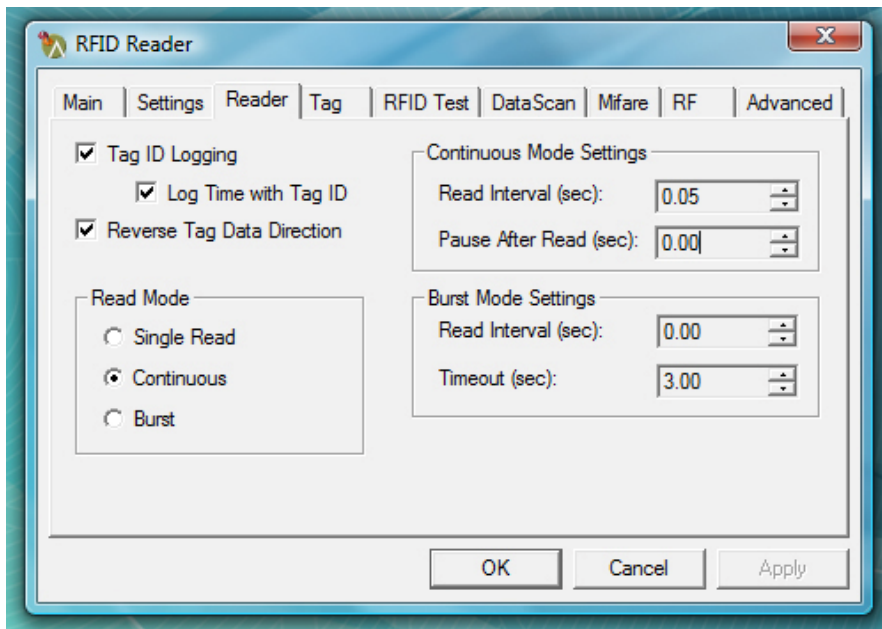


Figure 8. Reader Configuration Interface.

In order to compare the RFID data to the physical location we used the Metronome Timer to manually record the actual position of the subject at each time point. Metronome Timer is a customisable timer program, it allows user counts up or down for maximum of 1 hour with tick, as shown in *Figure 9*. During our experiments, we adjust Metronome Timer to tick every second, and emphasises beat every two seconds. This was used to indicate to the experimental subject when they should move between locations when collecting the data in experiment one.



Figure 9. *Metronome Timer* is used to get *Precise Time* and *Epoch Indications*.

There are two major ways of assessing the accuracy of the data collected:

- a) By comparing the calculated location with the actual location at each time point;
- b) By comparing the sequence of locations actually visited with the sequence derived from the RFID data;

The first assessment is particularly valuable if the person's location is important – for example if the system is being used to support context aware devices that use location as an input. The second case is important if sequences of location are being used to infer activity – where the context depends on what the person has just done and what they are about to do. Obviously a good score on both measures is desirable, and a sequence can be derived from a set of locations, but if a relatively sparse array for tags is used, it is likely that there will be a high degree of uncertainty in location for much of the time, which may not be important if the sequence is all that is required.

3.5 Pilot

Before we conducted our real life experiments, we carried out pilots just on examining the correctness of all our algorithms, data processing methods and processing programs. There were 5 tags used for this pilot, as with the full experiments.

The experiment pilot: We stayed close to each tag for exactly 2 seconds, and took 4 seconds to move to the next recognised location, then stayed for 2 seconds again. We

turned on the reader detection indication (“beep” mode), which could allow us to receive confirmation for every detection that had occurred. The reading data were collected manually, we could find out the actual movement and interval between movements from gathered reading data. These data were transferred to an Excel file, we then created scatter and x-y chart based on gathered data, generated diagrams that provided a virtual guide reflecting the reader’s movement. Thus in this pilot, there were 10 streams (2 streams for each tag) recorded, and the interval between each tag was 4 seconds. To confirm the accuracy of gathered data we turn the reader back to mute mode after the pilot, in order to ensure the randomness and validity of the later experiment.

3.6 The Amount of Gathered Data for both Experiments

For both experiments, we define our basic movement as a circuit. A circuit represents one movement around all tags, as explained in *Figure 10*; a test represents 3 consecutive circuits; a group represents 10 tests; and 3 groups were performed for each experiment.

3.7 Experiment One

The first experiment is designed to examine the location alignment (where am I now?) of an indoor RFID data stream. Namely, we compare the results for collected data from the reader with the actual position (where are you) to see the difference in every specific time period. We use Metronome Timer for exact timing indication.

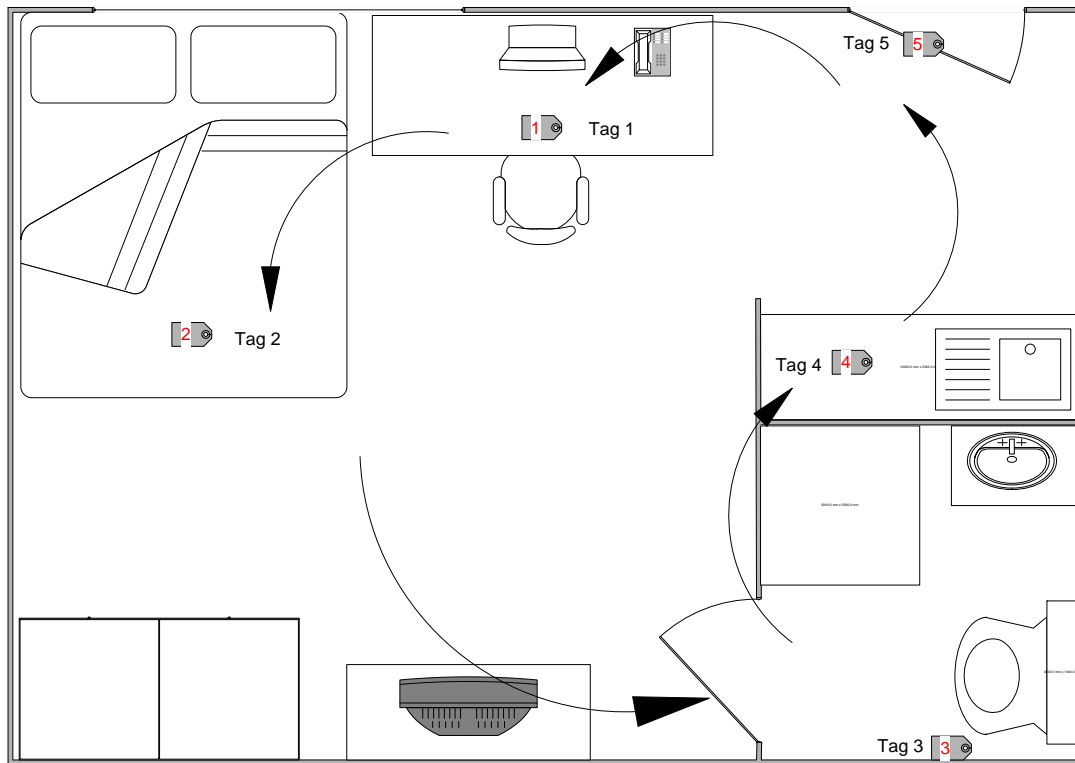


Figure 10. Tags' Layout and Cycle Movement Sequence in Experiment One.

In this experiment, there were five tags placed in my house, which represented five locations or objects. Each tag had its individual unique identification. For instance, Tag 3 was attached by the toilet, it was designed to record the time when people attend the toilet. *Figure 10* illustrates the sequence and direction of movement for the first experiment.

Table 8. Use Simple Name to represent tags in Experiment One.

<i>Tag ID</i>	<i>Caption in First Experiment</i>
<i>E2003411B802011371233244</i>	<i>Tag 1</i>
<i>E2003411B802011371233239</i>	<i>Tag 2</i>
<i>E2003411B802011371233242</i>	<i>Tag 3</i>
<i>E2003411B802011371233245</i>	<i>Tag 4</i>
<i>E2003411B802011371233241</i>	<i>Tag 5</i>

To collect the RFID data we used a Tracient reader attached on a lanyard, with a coil aerial resting around the lower end of the sternum as shown in *Figure 11*.



Figure 11. Tracient RFID Reader with Lanyard.

UHF tags (Alien) were placed as shown in *Figure 10*, and tags were at the height of the aerial during normal operation. Normal walking movement induced movement in the tag reader, much as a pendant around the neck will swing, and this effect could assist in detection when the aerial angle of detection was quite small – around 30 degrees (Parry & Symonds, 2009). Firstly we set our RFID reader to emit a beep when it detected a tag in order to get precise indication of when tags were detected. Since we had to record the actual position as our reference group, and this reference group was only performed once for first experiment,

Table 9 shows the example of readings for reference group. During the first experiment, we stayed close to each tag for exactly 2 seconds, and spent 4 seconds to move to the next recognised location, then stayed for 2 seconds again. The time was collected manually by using the Metronome Timer programme, then transferred to an Excel file as reference for later comparison. Thus in this test, there were 90 seconds of recording, and the interval between each tags was 4 seconds. This data were used to produce a file that had tags recorded as accurately as possible – i.e. at each stop the appropriate tag was recorded. This part of the experiment also confirmed that tags could definitely be detected at each location.

For the actual experiment 1, the detector beep was turned off so the subject was not aware as to whether a tag had been detected or not. Once Metronome Timer started beeping, we started to move and pass by the attached objects while wearing the reader. We stayed close to each tag for exactly 2 seconds, and spent 4 seconds moving to the next recognised location, then stayed for 2 seconds again. In order to collect the data as accurately as possible, we divided the data collection into 3 different and separate occasion groups. Each occasion was called a group. During each group we performed 10 complete tests (3 circuits for each test), and each test consists of 90 second. There is an example of readings for one single test, shown in *Table 11*.

Table 9. Example of Data Collection for Reference Group.

Location/activity	Cumulative time (seconds)
Tag 1	2
Moving to tag 2	6
Tag 2	8
Moving to tag 3	12
Tag 3	14
Moving to tag 4	18
Tag 4	20
Moving to tag 5	24
At Tag 5	26
Moving to tag 1	30
Tag 1	32
Moving to tag 2	36
Tag 2	38
Moving to tag 3	42
Tag 3	44
Moving to tag 4	48
Tag 4	50
Moving to tag 5	54
At Tag 5	56
Moving to tag 1	60
Tag 1	62
Moving to tag 2	66
Tag 2	68
Moving to tag 3	72
Tag 3	74
Moving to tag 4	78
Tag 4	80
Moving to tag 5	84
At Tag 5	86
Moving to tag 1	90

The reader was set up to collect data at up to 10 times a second; if no tag was within range then no data were collected. The total collection time of this experiment is 2,790 sec (3 groups \times 10 \times 90 seconds = 2,700 sec plus the one we collected for reference group: 1 x 90 sec).

3.8 Experiment Two

As mentioned in Section 2.1.3, the RFID-based location identification sequence has implications for recognising relevant human motion. The second experiment is designed and based on this factor. Moreover, to clean the gathered data, the second experiment is designed to examine the accuracy of the reading sequence match.

For the reference group data collection, the subject approached and stopped at each tag in sequence at a normal walking, but uncontrolled speed. In this step we turned on the reader detection indication (tick mode), thus we were alerted for every detection which occurs. Then we executed the collected result by using our *Matlab – 1 second window data proceeding program* and then compared the result with an ordinal experimental design. After we confirmed the accuracy of gathered data we turned the reader back to mute mode, in order to ensure the randomness and validity of the experiment.

Table 10. Use Simple Name to Represent Tags in Experiment Two.

<i>Tag ID</i>	<i>Caption in Second Experiment</i>
<i>E2003411B802011371233244</i>	<i>Tag A</i>
<i>E2003411B802011371233239</i>	<i>Tag B</i>
<i>E2003411B802011371233242</i>	<i>Tag C</i>
<i>E2003411B802011371233245</i>	<i>Tag D</i>
<i>E2003411B802011371233241</i>	<i>Tag E</i>

Similarly, in this experiment, the subject did not move at a fixed speed so the Metronome Timer was only used to indicate when the test was completed (we set Metronome Timer to give tick for the last 5 second). Instead the subject moved in a natural manner on a fixed sequence (A – B – C – D – E).

The idea of conducting the second experiment is to perform a known sequential movement with random speed. Furthermore, this experiment also involves cyclic movement around every tag within the environment. By adopting three different data processing methods, we compare the moving pattern in terms of sequence, to find the most precise result. In order to collect the data as accurately as possible, we also divided the data collection into 3 different occasion groups. During each group we performed 10

complete tests (3 circuits for each test), and each test consists of 90 second. Because this experiment was conducted within a random speed, we obtained a slightly difference in the amount of collected data. A total collection time of for this experiment is 2,700 sec ($3 \text{ groups} \times 10 \times 90 \text{ seconds} = 2,700$).

4 EXPERIMENT - DATA PROCESS AND ANALYSIS

4.1 Framework of Applied Method in Both Experiments

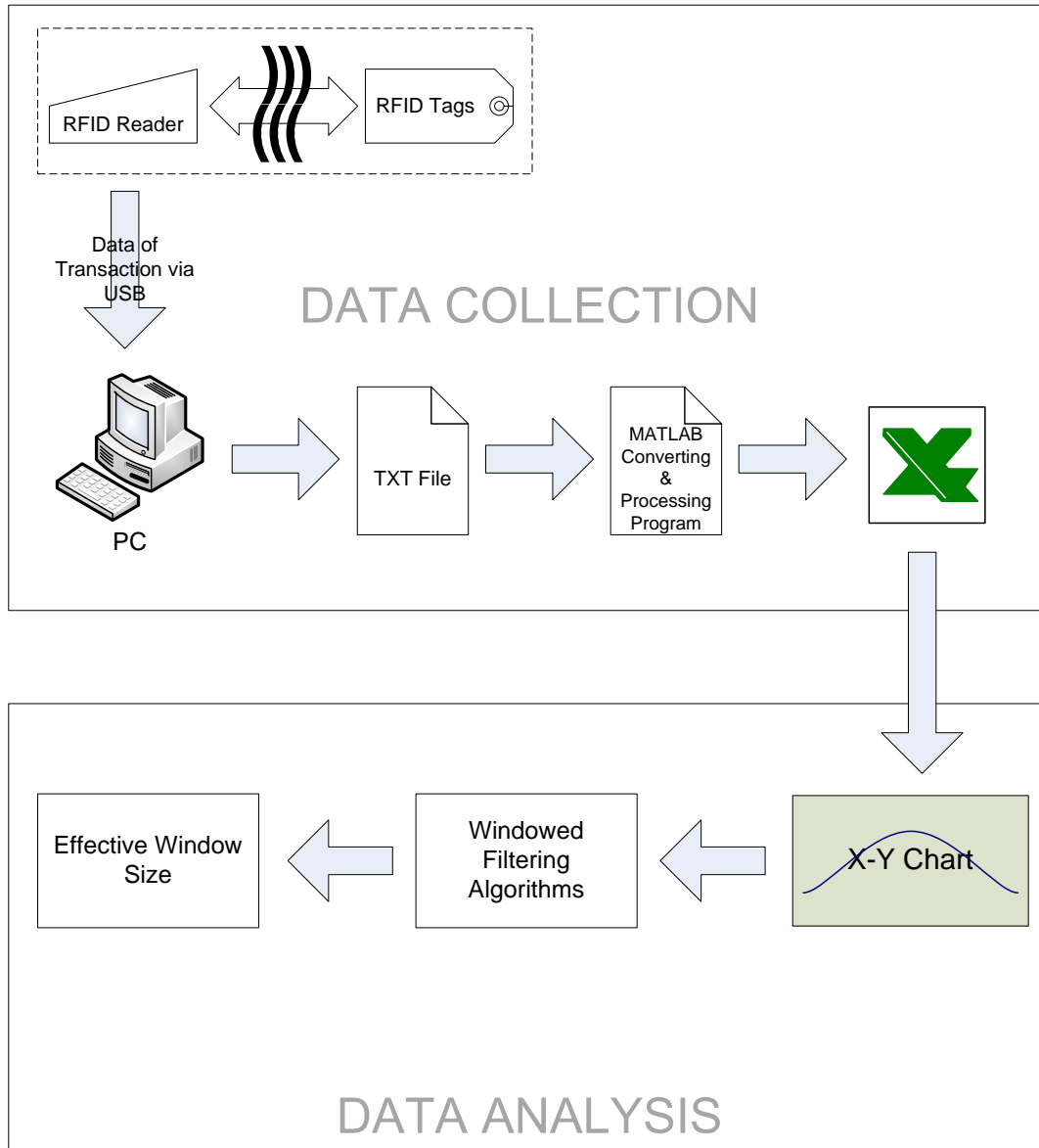


Figure 12. Data Processing and Analysis Methods of Indoor Experiment.

Standard Location Time Model and Accurate Sequence

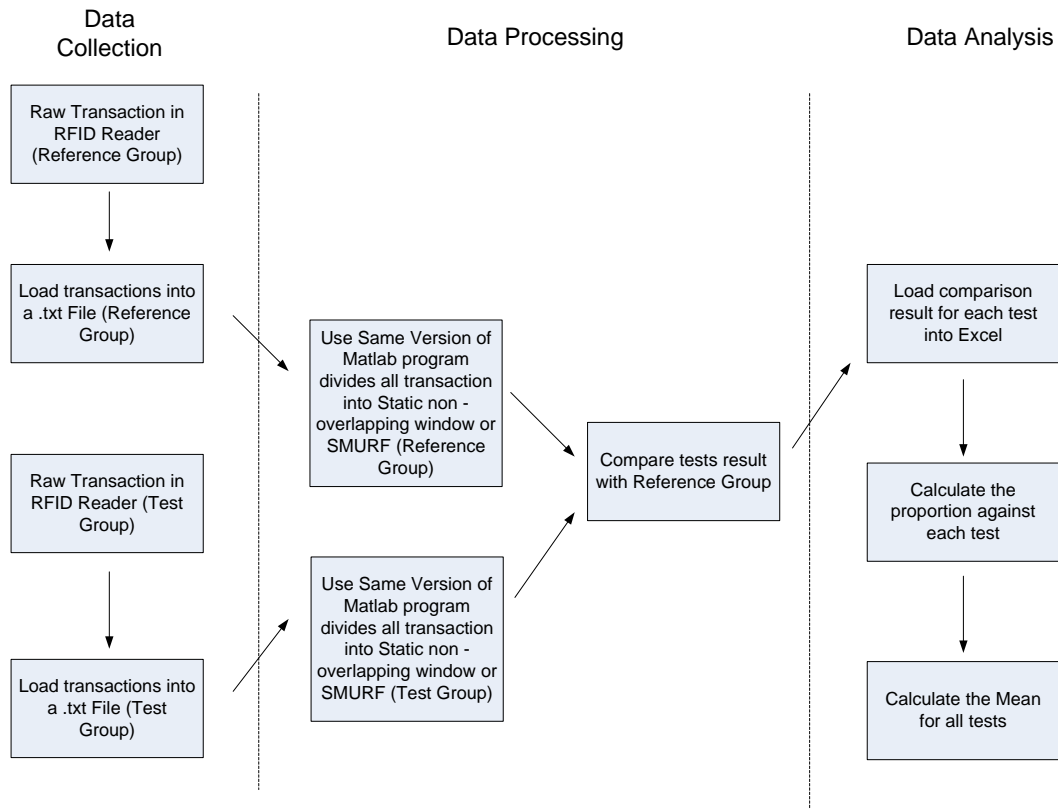


Figure 13. The Framework of Research Flow for 1 Second Window, Static Window and Sliding Window Methods in Both Experiments.

Table 11. Location of Results from both Experiments and Various Analysis Methods

Experiment 1 Data Collected 2700 sec	1 Second Window (Figure 14)	2 Second Window (Figure 15)	4 Second Window (Figure 16)	6 Second Window (Figure 17)	SMURF Window (Figure 19)
Experiment 2 The Amount Data Collected 2700 sec	1 Second Window (Figure 22)	2 Second Window (Figure 23)	4 Second Window (Figure 24)	6 Second Window (Figure 25)	SMURF Window (Figure 26)

4.2 Differentiations in Data Collecting and Cleaning Methods for Both Experiments

- We used two methods of data collecting as illustrated in *Table 12*: 1) Collected by using controlled timing (Experiment One); 2) Collected without using controlled timing (Experiment Two). In the first experiment we aimed to examine the location accuracy within specific time points, thus we must control the timing in order to ensure we provide accurate time points for both reference and test groups. However, timing control will not affect the moving sequence, thus we need not control time in the second experiment.
- There are two methods of data cleaning as illustrated in *Table 12*: 1) Static non-overlapping window (includes both 1 second and Multi-Second window); 2) SMURF.
- There are also two ways of testing quality of data as illustrated in *Table 12*: 1) the accuracy at each time point; 2) the sequence difference (Levenshtein Distance) algorithm. In second data processing, we adopt a Levenshtein Distance algorithm to compare between the reference group and test group, then clearly reveal the difference between them in terms of the difference in location sequence.

Table 12. Collected Data Source from both Experiments and Their Distribution for Data Processing.

	Timed	Static Window	SMURF	Location Accuracy	Location Sequence
Experiment One	Yes	Yes	Yes	Yes	No
Experiment Two	No	Yes	Yes	No	Yes

4.3 First Experiment

4.3.1 Data Process and Analysis – Non-overlapping 1 Second Window Method in Location Verification

4.3.1.1 Data Process

We employ the 1 second window method for data processing in this step. Firstly, we load collected data into our *Matlab – 1 second window data proceeding program*. The basic function of this program is to convert .txt into .m format, which can be recognised by Matlab. We conduct individual tests for this method, hereby there are 30 .txt files ready for processing. Secondly, we convert streams' time into an absolute format. For instance, "11:05:20 22/Jun/2009 (LOCAL)" was converted to "00:00:01", then "11:05:21 22/Jun/2009 (LOCAL)" was converted to "00:00:02", this approach will allow us group and count multiple streams within the same second. Thirdly, the program selects the most frequently appearing tag within every second, in order to avoid unwanted multiple detections. The program will write filtered data streams into Excel spreadsheet lists, thus we compare streams' tag ID in each cell list with the reference group gathered from experiment one, and get a proportion. Each proportion represents the percentage of similarity between each single test and reference group, as shown and compared in Table 13. Finally we calculate the average from 30 tests, since single test results sometimes vary widely.

Table 13. An Example of How the Test Data is compared with the Reference Group.

	Time	Reference Group	Test Group	Match
First Circuit	1	Tag 1	Tag 1	1
	2	Tag 1	Tag 1	1
	3	N/A	N/A	1
	4	N/A	N/A	1
	5	N/A	Tag 2	0
	6	N/A	Tag 2	0
	7	Tag 2	Tag 2	1
	8	Tag 2	N/A	0
	9	N/A	N/A	1
	10	N/A	N/A	1
	11	N/A	N/A	1
	12	N/A	N/A	1
	13	Tag 3	N/A	0
	14	Tag 3	Tag 3	1
	15	N/A	Tag 3	0
	16	N/A	Tag 3	0
	17	N/A	N/A	1
	18	N/A	Tag 4	0
	19	Tag 4	Tag 4	1
	20	Tag 4	N/A	0
	21	N/A	N/A	1
	22	N/A	N/A	1
	23	N/A	Tag 5	0
	24	N/A	N/A	1
	25	Tag 5	Tag 5	1
	26	Tag 5	Tag 5	1
	27	N/A	Tag 5	0
	28	N/A	N/A	1
	29	N/A	N/A	1
	30	N/A	N/A	1
Second Circuit	31	Tag 1	Tag 1	1
	32	Tag 1	Tag 1	1
	33	N/A	N/A	1
	34	N/A	N/A	1
	35	N/A	N/A	1
	36	N/A	Tag 2	0
	37	Tag 2	N/A	0
	38	Tag 2	Tag 2	1
	39	N/A	Tag 2	0
	40	N/A	N/A	1
	41	N/A	N/A	1
	42	N/A	N/A	1
	43	Tag 3	N/A	0
	44	Tag 3	Tag 3	1
	45	N/A	Tag 3	0

Table 14 Continued.

	Time	Reference Group	Test Group	Match
Second Circuit	46	N/A	Tag 3	0
	47	N/A	N/A	1
	48	N/A	N/A	1
	49	Tag 4	N/A	0
	50	Tag 4	Tag 4	1
	51	N/A	Tag 4	0
	55	N/A	Tag 4	0
	53	N/A	N/A	1
	54	N/A	N/A	1
	55	Tag 5	Tag 5	1
	56	Tag 5	N/A	0
	57	N/A	N/A	1
	58	N/A	N/A	1
	59	N/A	N/A	1
	60	N/A	Tag 1	0
Third Circuit	61	Tag 1	Tag 1	1
	62	Tag 1	Tag 1	1
	63	N/A	N/A	1
	64	N/A	N/A	1
	65	N/A	Tag 2	0
	66	N/A	Tag 2	0
	67	Tag 2	Tag 2	1
	68	Tag 2	N/A	0
	69	N/A	N/A	1
	70	N/A	N/A	1
	71	N/A	Tag 4	0
	72	N/A	N/A	1
	73	Tag 3	N/A	0
	74	Tag 3	Tag 3	1
	75	N/A	Tag 3	0
	76	N/A	N/A	1
	77	N/A	N/A	1
	78	N/A	N/A	1
	79	Tag 4	Tag 4	1
	80	Tag 4	Tag 4	1
	81	N/A	N/A	1
	82	N/A	N/A	1
	83	N/A	Tag 5	0
	84	N/A	N/A	1
	85	Tag 5	Tag 5	1
	86	Tag 5	N/A	0
	87	N/A	N/A	1
	88	N/A	N/A	1
	89	N/A	N/A	1
	90	N/A	N/A	1
Total		90		61 (Total of Matching)

From the example above, we calculate the proportion for this test by using:

$$\frac{\text{TheNumberOfMatchingOutputs}}{\text{TotalNumberOfWindows}} = \frac{61}{90} = 0.6778$$

Equation 3. How to Calculate the Proportion of Matching between Two Tests.

So the accuracy for this test is 0.6778 compared with our reference group. Then we calculate the accuracy for each test, and get the mean for 3 groups.

Each test represents 3 circuits of the living area; each group represents 10 tests. Thus each group represents 30 circuits of the room. The first experiment consists of 3 groups; each group is performed in a single session. The data is presented for each group separately as there may have been changes that affected efficiency of detection between sessions.

In this non-overlapping window method, we used reference group data and produce values for each window, and we did the same for individual test.

Accuracy was calculated based on the overall mean of 3 groups, with the mean being 57.37% for 1 second window method (*Figure 14*).

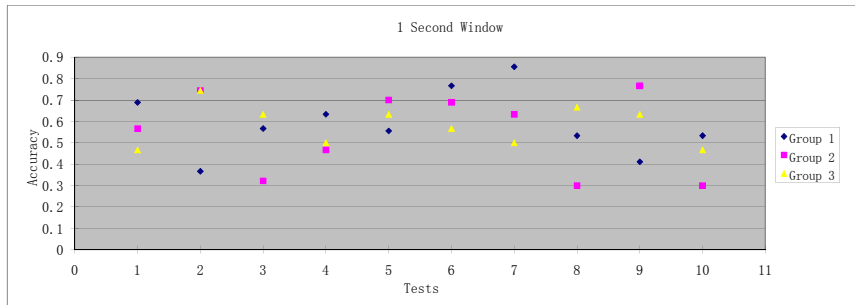


Figure 14. The Result of Using a 1 Second Window Method in Data Processing from 3 Test Groups.

4.3.1.2 Analysis

Accuracy was calculated based on the overall mean of 3 groups.

Table 15. Calculated Mean for Three Groups

Group 1	Group 2	Group 3	The Overall Mean of This Method
59.11%	54.89%	58.11%	57.37%

Based on data processing result, the mean (57.37%) represents the chance of correct reading. Because we filter multiple detections before similarity comparison, the analysis for this method of first experiment indicates that if we apply a 1 second window approach on location verification at a certain time, we can only get 57.37% of correct reading, which is the correct tag will be detected when the person is located at the measurement point near the tag.

4.3.2 Data Analysis - Fixed Multi-Second Window Method for Location Verification

4.3.2.1 Data Process

In this stage, we adopt a fixed non-overlapping window approach as our analysis method. We load collected data into another *Matlab – static window Data* processing program, the basic function of this program is to convert .txt into .m format, which can be recognised by Matlab. We also conducted 3×10 tests for this analysis method, resulting in 30 .txt files ready for processing. Secondly, we convert data stream time into an absolute format as this approach will allow us to group and count multiple streams within the same second. We were allowed to select the size of window in this stage. The window size was adjusted to 2 seconds, 4 seconds and 6 seconds, for instance we enter “>> [result] = get_result(6)” if the size of epoch is 6 seconds. Thirdly, *Matlab – static window data processing program* selects the most frequently appearing tag within every window, in order to allow the accuracy to be calculated. The program will write filtered data streams into a excel spreadsheet list. The program compares streams’ tag ID in each cell list with the reference group data gathered from experiment one, and calculates the proportion for the single test. Each proportion represents the percentage of similarity between each

single test and reference group just for a certain window size. Finally we calculate the average from 30 tests, since single test results sometimes vary widely.

The mean for each window is shown in *Table 16*.

For each method (different window size), accuracy was calculated based on the overall mean of 3 group , with the mean being 62.67% for this two second window method (*Figure 15*); 68.19% for this two second window method (*Figure 16*); 61.56% for this two second window method (*Figure 17*).

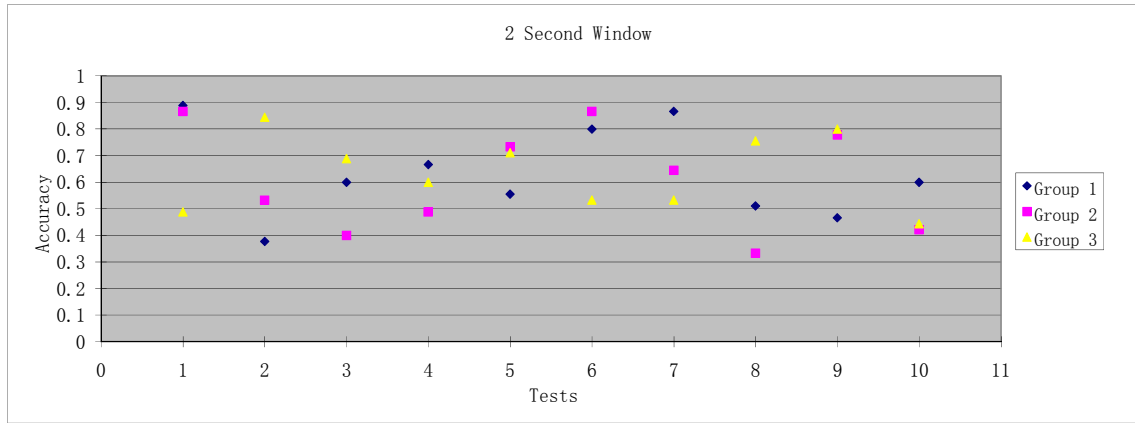


Figure 15. The Result of Using 2 Second Window in Data Processing from 3 Groups.

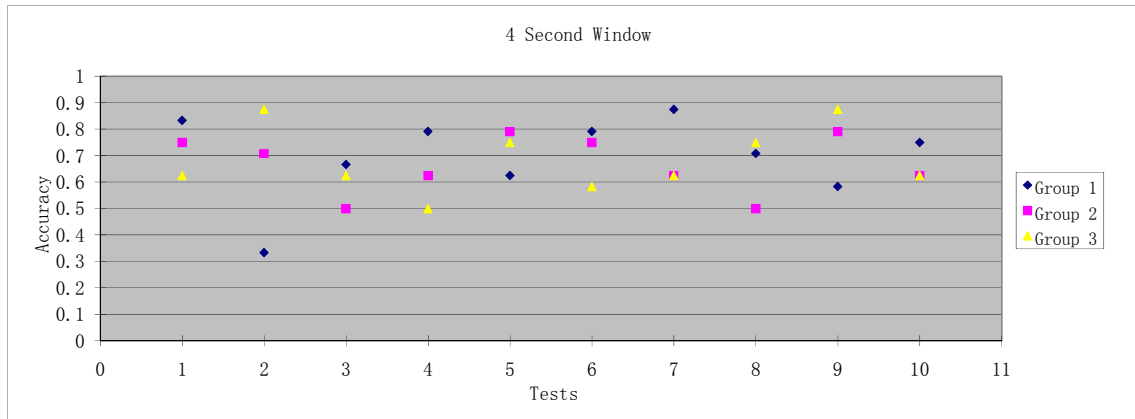


Figure 16. The Result of Using 4 Second Window in Data Processing from 3 Groups.

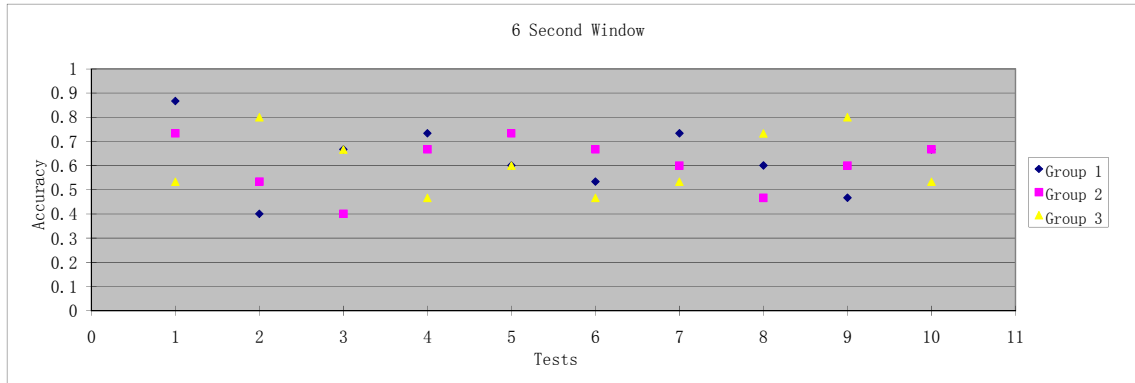


Figure 17. The Result of Using 6 Second Window in Data Processing from 3 Groups.

Table 17. Calculated Mean for Three Groups on Three Different Size of Windows

Window Size	Group 1	Group 2	Group 3	Overall Mean
2 Sec	63.33%	60.67%	64%	62.67%
4 Sec	69.58%	66.67%	68.33%	68.19%
6 Sec	62.67%	60.67%	61.33%	61.56%

4.3.2.2 Analysis

Based on data processed in 4.3.2.1, the mean of the 4 second windows is the highest value of this static Multi-Second window method. If we apply 4 seconds as the window size, 68.19% of readings are effective (*Table 17: Mean of 4 Sec*). This figure 68.19% is substantially higher than the mean of 2 second windows; 62.67% (*Table 17: Mean of 2 Sec*). Accordingly, we find that small windows fail to fill in dropped readings which results in false negatives. However, we find the means from large window size 6 Sec (64.0%; *Table 17: Mean of 6 Sec*) are also lower than that of the 4 second windows. This result tells us that a 6 second window is too wide for this detection, as Jeffery, et al. (2006) suggest that large windows fail to capture tags' movement, which causes false positives by RFID readers. Based on these processing results, we conclude that a 4 second window size is the best option for both dropped readings avoidance and movement capture in terms of static window methods in this experiment.

4.3.3 Data Analysis – Overlapping Variable Window Size (SMURF) for Location Verification

4.3.3.1 Data Process

In the third data analysis stage, we aim to adopt the Overlapping Variable Window Size (we refer to this method as SMURF or simplified SMURF to avoid complicating the text) filtering technique (explained in 2.3) to continue investigating the gathered results. We still use the data from the converted .m file. Firstly, we calculate the probability for each epoch by using *Matlab - Statistical sMoothing for Unreliable RFid data (SMURF) Cleaning*, a program that can get the probability for each tag over the whole streams lists, the code for this program is shown in *Figure 18*. And we had explained this algorithm and given examples in section 2.3.1.

Require: $T = \text{set of all observed tag IDs}$
 $\delta = \text{required completeness confidence}$
 $\forall i \in T, w_i \leftarrow 1$
while (getNextEpoch()) **do**
 for (i in T) **do**
 processWindow(W_i)
 $w_i^* \leftarrow \text{completeSize}(p_i^{avg}, \delta)$ // Lemma 4.1
 end for
end while

Figure 18. Code and Algorithm for SMURF Adaptive Per-Tag Cleaning (Jeffery et al., 2006).

Using these probabilities, the SMURF Program then calculates the average over the time period (P_i^{avg}) for each tag. This is an example of the P_i^{avg} result for every tag. Secondly, we employ the equation from the literature (discussed in 2.3):

$$w_i \geq \left\lceil \frac{\text{Ln}(1/\delta)}{P_i^{avg}} \right\rceil$$

Assuming delta is 0.05 (we want 95% confidence), to calculate the epochs per window (window size) for each tag. Thirdly, we use these window sizes to reprocess the data (.m file). As seen in the literature review in 2.3, SMURF only emits a reading with an epoch value corresponding to the midpoint of the window, thus the program will select the tag with the highest proportion to represent the midpoint, resulting in only one tag being

detected within a given epoch. Finally, we save these processing results into an Excel spreadsheet. We carry out 3 groups for this analysis method, and there are 10 tests for each group. By using Excel's built-in average formula, we compare each test result with the reference group and calculate the similarity proportion for each.

In this overlapping sliding window method, we compare each test with reference group for corresponding window. The windows size w_i varies from 4 to 18 seconds based on P_i^{avg} for each tag. Accuracy was calculated based on the overall mean of 3 groups, with the mean being 70.71% for this SMURF method (*Figure 19*).

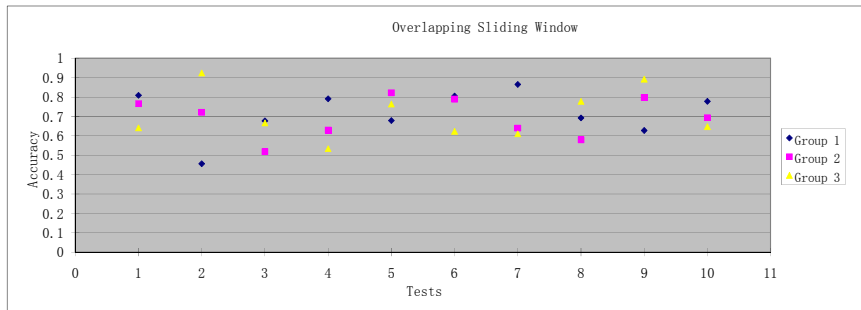


Figure 19. The Result of Using Simplified SMURF in Data Processing from 3 Groups.

Table 18. Calculated Mean for Three Groups.

Group 1	Group 2	Group 3	The Overall Mean
71.79%	69.54%	70.79%	70.71%

4.3.3.2 Analysis

Based on data processed in 4.3.3.1, we understand that if we apply a simplified SMURF approach to location verification at a known time, we can get 70.71% (Table 18) of similarity from our reference group.

In this case the result is (70.71%) closer to the reality in comparison with using a 1, 2, 4 or 6 second static non-overlapping window, if we apply this method in an environment that is close to our experiment, as we verify and examine the reference group data to represent reality.

4.4 Second Experiment

In the follow data analysis procedure, we employ three different windowing methods to analyse the collected data based on the second experiment – Sequence Verification.

In this experiment, each test represents 3 circuits of the room; each group represents 10 tests. Thus each group represents 30 circuits of the room. The second experiment consists of 3 groups performed on different occasions. Each group was performed in a single session. Then we compare the results for collected data from a reader with the actual sequence by adopting three methods. These three methods consist of 1 second windowed comparison, fixed size window comparison and simplified SMURF algorithm comparison. Based on gathered data, however, we can find that multiple streams for a single tag repeat (e.g.: AAABBBCCCDDD), and this factor causes confusion and miscalculation. Thus we have to compress the raw data before algorithm comparison, by converting “AAABBBCCCDDD” into “ABCD”. We use *Matlab – 1 second window data proceeding program* to accomplish this function, shown in *Table 19*. And we can find the difference for this conversion from *Table 20*. Similarly, before we conduct the second experiment, we carry out a pilot to examine both the rationality and correctness of all our algorithms, data processing methods and processing programs.

Table 19. Example of Compressed Data in Experiment Two

Tag's Id	Time	Processed and Converted Tag's Id
E2003411B802011371233244	11:05:20 22/Jun/2009 (LOCAL)	E2003411B802011371233244
E2003411B802011371233244	11:05:21 22/Jun/2009 (LOCAL)	
E2003411B802011371233244	11:05:21 22/Jun/2009 (LOCAL)	
E2003411B802011371233239	11:05:24 22/Jun/2009 (LOCAL)	E2003411B802011371233239
E2003411B802011371233239	11:05:25 22/Jun/2009 (LOCAL)	
E2003411B802011371233239	11:05:25 22/Jun/2009 (LOCAL)	
E2003411B802011371233239	11:05:26 22/Jun/2009 (LOCAL)	
E2003411B802011371233239	11:05:26 22/Jun/2009 (LOCAL)	E2003411B802011371233242
E2003411B802011371233242	11:05:32 22/Jun/2009 (LOCAL)	
E2003411B802011371233242	11:05:32 22/Jun/2009 (LOCAL)	
E2003411B802011371233242	11:05:32 22/Jun/2009 (LOCAL)	
E2003411B802011371233242	11:05:32 22/Jun/2009 (LOCAL)	
E2003411B802011371233242	11:05:33 22/Jun/2009 (LOCAL)	E2003411B802011371233245
E2003411B802011371233245	11:05:38 22/Jun/2009 (LOCAL)	
E2003411B802011371233245	11:05:38 22/Jun/2009 (LOCAL)	
E2003411B802011371233245	11:05:39 22/Jun/2009 (LOCAL)	
E2003411B802011371233245	11:05:39 22/Jun/2009 (LOCAL)	
E2003411B802011371233245	11:05:39 22/Jun/2009 (LOCAL)	
E2003411B802011371233245	11:05:39 22/Jun/2009 (LOCAL)	E2003411B802011371233241
E2003411B802011371233241	11:05:44 22/Jun/2009 (LOCAL)	
E2003411B802011371233241	11:05:44 22/Jun/2009 (LOCAL)	
E2003411B802011371233241	11:05:45 22/Jun/2009 (LOCAL)	
E2003411B802011371233241	11:05:45 22/Jun/2009 (LOCAL)	
E2003411B802011371233241	11:05:45 22/Jun/2009 (LOCAL)	

After data stream compression, the next step of this experiment is to compare the sequence of between results. For instance, AAABBBCCDDDD is the same as ABCD but different to ABACD. We also write a program to conduct the comparison. The program can compare the collected and loaded transitions result by using Edit Distance or Levenshtein Distance Algorithm, we call the program as *Levenshtein Distance Program*, both algorithm and example of this method is explained in *Table 20*, and the code shown in *Figure 20*. The Levenshtein Distance measures the amount of difference between two sequences by using a metric comparison approach (Navarro, 2001). Originally, the Levenshtein Distance was used to compare the distance between two strings, identifying the minimum number of different operations needed to transform one string into the other (Navarro, 2001).

```
int LevenshteinDistance(char s[1..m], char t[1..n])
// d is a table with m+1 rows and n+1 columns
declare int d[0..m, 0..n]

for i from 0 to m
  d[i, 0] := i
for j from 0 to n
  d[0, j] := j

for j from 1 to n
  for i from 1 to m {
    if s[i] = t[j] then
      cost := 0
    else
      cost := 1
    d[i, j] := minimum(
      d[i-1, j] + 1,    // insertion
      d[i, j-1] + 1,    // deletion
      d[i-1, j-1] + cost // substitution
    )
  }
return d[m, n]
```

Figure 20. Code for Identifying the Minimum Number of Differences between Two Strings.

4.4.1 Data Analysis – 1 Second Window Method for Sequence Verification

4.4.1.1 Data Process

The 1 second window method uses the data gathered from experiment two. Firstly, we still use the same approach to load data into *Matlab – 1 second window data proceeding program*, however the difference is that we employ *Matlab – static window data processing program* for data processing, thus the Matlab program was modified to fulfil sequence analysis instead of locations. The program then converts .txt into .m format, which can be recognised by Matlab. There are also 3×10 tests for this analysis method, and gathered stream's time is also converted to an absolute format, in order to allow us to group and count multiple streams within the same second. Secondly, the program selects the most frequently appearing tag within 1 second, in order to avoid unwanted multiple detections. The program will write filtered streams into a spreadsheet list, and compress the repeating location into a non-repeating string (e.g.: "AAABBBCCCDDD" is compressed into "ABCD"), since the repeating location means the RFID reader stays at the same location. The reason for compressing these repeating IDs is to standardise raw data, in order to be processed by Levenshtein Distance program, as shown in *Figure 21*. Results from each group are compared with the reference group data gathered from experiment one using a proportion calculated by Excel. Each proportion represents the percentage of similarity between a single test and the reference group, as shown in *Figure 18*. Finally we calculate the average from 30 tests, since single test results sometimes vary widely.

Table 20. The Example of How We Compare the Test Data with the Reference Group.

Actual Sequence for Reference Group (tag's Id)	Calculated Sequence for Test (tag's Id)	Difference
A	A	
B	B	
C	D	1 (Should not be a "D")
D	C	
E	E	1 (Should be a "D" before "E")
A	A	
B	B	
C	C	
D	D	
E	C	1 (Should not be a "C")
A	E	
B	A	
C	B	
D	C	
E	D	
	E	
15 (Total)		3

From the example in Table 21, we calculate the proportion for this test by using:

$$1 - \frac{\text{TheNumberOfDifference}}{\text{TotalNumberOfDetectedTags}} = 1 - \frac{3}{15} = 1 - 0.2 = 0.80$$

Equation 4. How to Calculate the Proportion of Difference for Two Sequences.

So the accuracy for this test is 0.8 compared with our reference group. Then we calculate the accuracy for each test, and get the mean for 3 groups. The reason for using "A" or "B" instead of using "Tag 1" or "Tag 2" to represent the tags is to meet the requirement of Levenshtein Distance program.

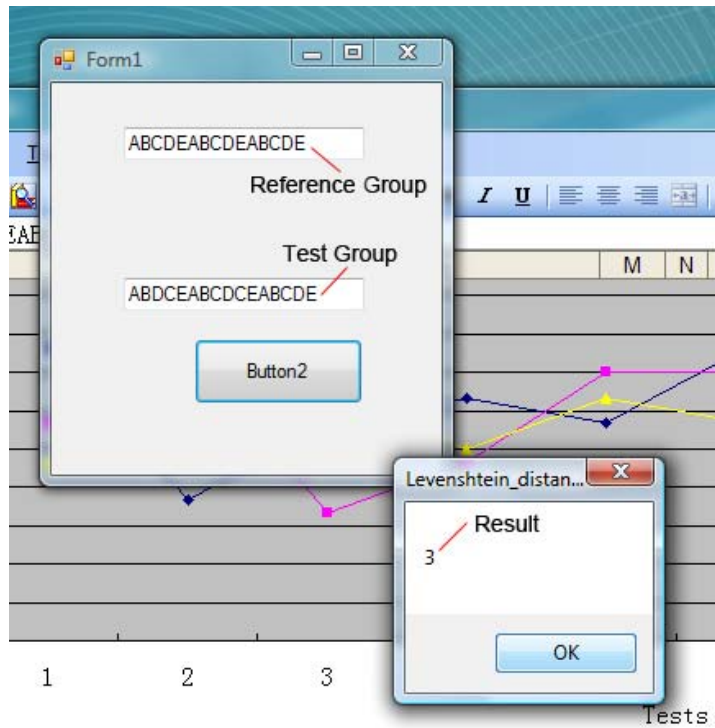


Figure 21. Comparing Experimental Results with Reference Group by Using VB.net Levenshtein Distance Comparison Algorithm.

Accuracy was calculated based on the overall mean of 3 groups, with the mean being 78.9% for this 1 second window method (Figure 22).

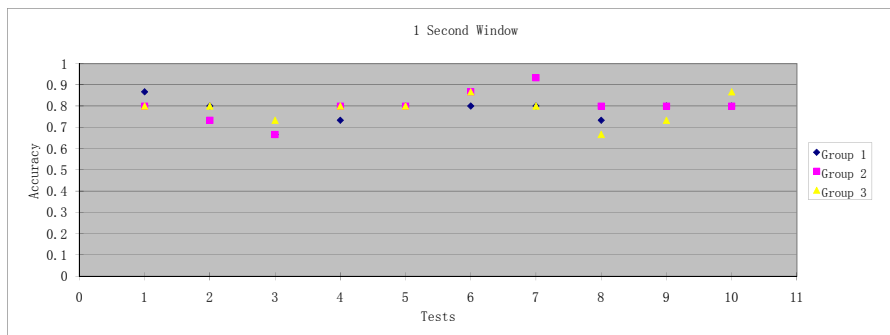


Figure 22. The Result of Using 1 Second Window Method in Data Processing from 3 Groups.

Table 22. Calculated Mean for 1 Second Window Method.

Group 1	Group 2	Group 3	Overall Mean
78%	80%	78.9%	78.9%

The average from overall mean of three groups is 78.9% (*Table 22*) for this 1 second window method.

4.4.1.2 Analysis

Based on processed data in 4.4.1.1, we can find that the average 78.9% (*Table 22*) of three means of 78%, 80% and 78.9% represents the chance of effective reading using the 1 second window method. Because we filter multiple detections before similarity comparison, analysis for this method of experiment two indicates that if we apply a non-epoch approach on location verification at a certain time, we only get 78.9% reliable streams. More than 20% of the streams could be dropped, interfered or blocked.

4.4.2 Data Analysis - Fixed Window Method for Location Verification

4.4.2.1 Data Process

In this stage, we adopt a static window approach as our analysis method. Similar to the method in 4.2 we firstly load collected data into another *Matlab – static window data processing program*. The basic function of this program is to convert .txt into .m format, which can be recognised by Matlab. We also conducted 30 tests for this analysis method, resulting in 30 .txt files ready for processing. Secondly, we convert stream's time into an absolute format as this approach will allow us to group and count multiple streams within the same second. We were allowed to select the window size in this stage. The epoch was adjusted to 2 seconds, 4 seconds, and 6 seconds, for instance we enter ">> [result] = get result(6)" if the window size is 6 seconds. Thirdly, we select the most frequently appearing tag within every epoch, in order to avoid unwanted multiple detections. The program will write filtered streams into a cell list. There are 4 separate lists against each test. The program compares a stream's tag ID in each cell list with the reference group data gathered from experiment one, and calculates the proportion for single test. Each proportion represents the percentage of similarity between each single test and reference group just for a given epoch size. Finally we calculate the average from 30 tests, since single test results sometimes vary widely.

For each method (different window size), accuracy was calculated based on the overall mean of 3 groups, with the mean being 80.2% for the 2 second window method (*Figure 23*); 82% for this 4 second window method (*Figure 24*); 80.4% for this 6 second window method (*Figure 25*).

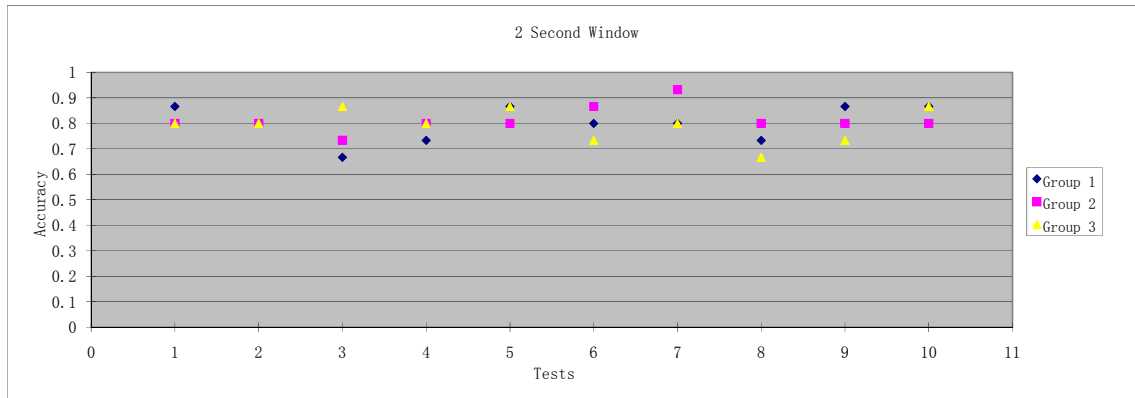


Figure 23. The Result of Using a 2 Second Window in Data Processing from 3 Groups.

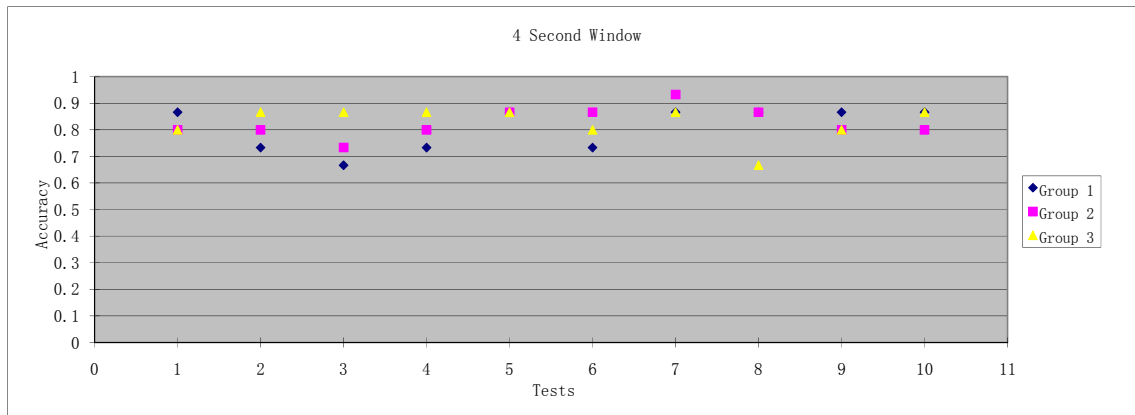


Figure 24. The Result of Using a 4 Second Window in Data Processing from 3 Groups.

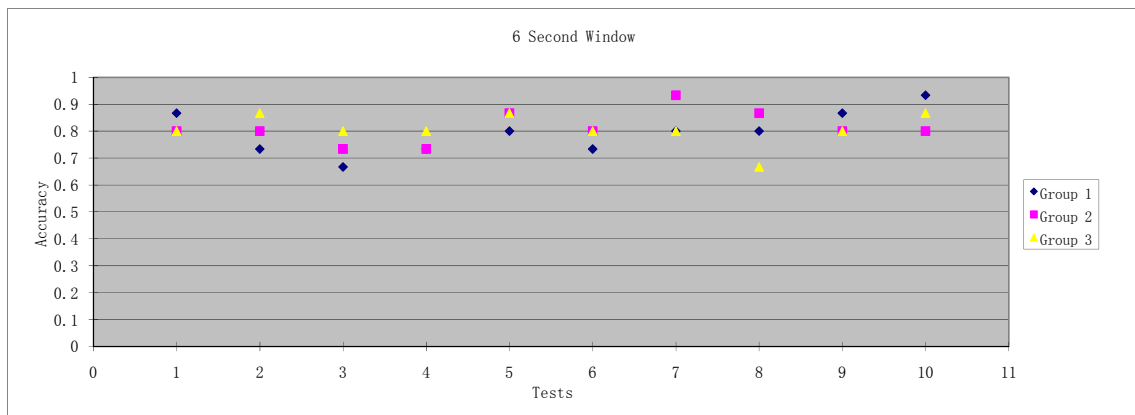


Figure 25. The Result of Using a 6 Second Window in Data Processing from 3 Groups.

The mean for each method is shown in *Table 23*.

Window size	Group 1	Group 2	Group 3	Overall Mean
2 Sec	80%	81.33%	79.33%	80.2%
4 Sec	86.67%	82.67%	82.67%	82%
6 Sec	79.33%	81.33%	80.67%	80.4%

Table 23. Calculated Mean for Three Groups on Three Different Size of Windows

4.4.2.2 Analysis

Based on processed data in 6.2.1, we find that the mean of 4 seconds is the highest value for this fixed size window method. Namely, if we apply 4 seconds as the window size, we can get 81.8% of readings to be effective. This figure 81.8% (*Table 23: Mean of 4 Sec*) is substantially higher than the mean of 2 seconds (*Table 23: Mean of 2 Sec*). This is consistent with findings by Jeffery, et al. (2006) that small windows fail to fill in dropped readings which results in false negatives. However, we find the mean of 80.4% from the large 6 Sec window size (*Table 23: Mean of 6 Sec*) is also lower than the 4 seconds window mean of 81.8% (*Table 23: Mean of 4 Sec*). This result tells us that a 6 second window is too wide for this detection. This is also consistent with the reviewed literature, which indicates that large windows fail to capture tag movement, causing false positives by RFID readers (Jeffery et al, 2006). Then we can get a final result based on the above analysis: the 4 second window size is the best option for both dropped readings avoidance and movement capture.

4.4.3 Data Analysis – Sliding Window for Location Verification

4.4.3.1 Data Process

In this third data analysis stage, we continue to examine the results collected from experiment two using the SMURF cleaning approach. The experimental results consist of 3 test groups, 10 tests within each group. We aim to compare each test result with the known sequence (the reference group), then calculate the average of each group. Afterwards, we compare the result both statistically and graphically. To gain the vital result, firstly we have to use the data from .m file, and calculate the probability for each epoch by using SMURF method. This program can determine the probability for individual tags by randomly selecting a sampling section from the streams lists. Secondly, the probability leads us to calculate the average over the time period (P_i^{avg}) for each tag. This is an example of the P_i^{avg} result for every tag. Afterwards, we use the equation from the literature (discussed in 2.3.1):

$$w_i \geq \left\lceil \frac{\ln(1/\delta)}{P_i^{avg}} \right\rceil$$

With 0.05 of delta (we want 95% confidence), to calculate the epochs per window (window size) for each tag. Thirdly, we use these window sizes to reprocess the data (.m file) by adapting our SMURF program in order to get the sliding mid-point for each tag. The size of the following window is enlarged once low probability appears for this tag. If there are two tags selected from the same epoch, we select the tag with higher probability to represent the midpoint. Fourthly, we incorporate the previous *Matlab – 1 second window data proceeding program* to compress gathered results into a non repeating string (e.g.: “AAABBBCCCCDDD” is compressed into “ABCD”). We conduct 3 groups in the experiment, thus there are 30 results in total. Finally we apply the Levenshtein Distance program to compare each result with the reference group, and get the accuracy rate for each. Accuracy rate was calculated based on the overall mean of 3 groups, the window size w_i varies from 4 to 16 seconds, based on P_i^{avg} , with the mean being 83.6% for this overlapping sliding window method (*Figure 26*).

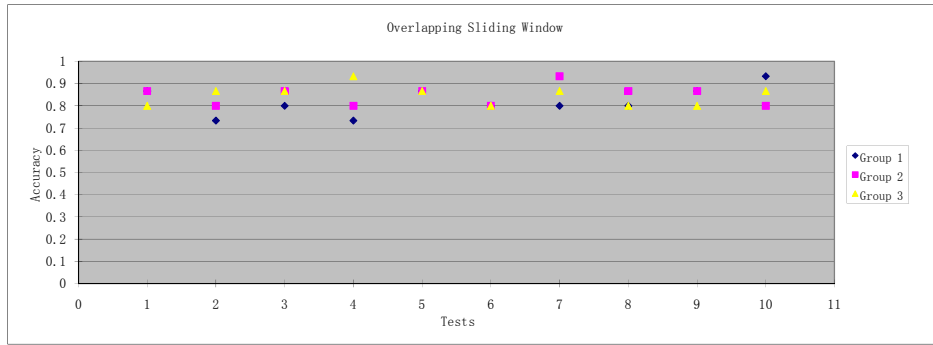


Figure 26. The Result of Using Sliding Window in Data Processing from 3 Groups.

Group 1	Group 2	Group 3	Overall Mean
82%	84.67%	84.67%	83.6%

Table 24. Calculated Mean for Three Groups.

4.4.3.2 Analysis

Based on data processed in 4.4.3.1, we understand that if we apply the SMURF approach on location verification at a certain time, we can get 83.6% (*Table 24*) of similarity from our reference group.

In this case the result is (83.6%) which is closer to reality in comparison with using static non-overlapping window.

DISCUSSION

4.5 Comparison of Processing Methods

Based on the analysis of data processing for the first experiment, we understand that the adoption of SMURF provides the best result of (70.71%, *Table 18*) in comparison with 1 second window method (57.37%, Accuracy was calculated based on the overall mean of 3 groups. *Table 15*) and static window methods (68.19%, *Table 17*: 4 second window). The data processing was conducted by three different methods based on data gathered from experiment one. From the comparison of results shown on *Equation 5*, SMURF performs better than a 4 second window approach in a home environment. SMURF filtering offers the lowest error rate across all data processing methods of first experiment. The result improved by 3.70% (*Equation 5*) once we apply the SMURF method compared with the best result offered from a static non-overlapping window approach:

$$\left(\frac{SMURF'sRatio - StaticWindowMethod'sRatio}{StaticWindowMethod'sRatio} \right) = \left(\frac{70.71\% - 68.19\%}{68.19\%} \right) = 3.70\%$$

Equation 5. The Improvement Made by using SMURF in Comparison with the Best Result of Static Window Method, In Location Alignment.

This partitioned smoothing scheme provides particularly good results, because it adjusts the window size based on each tag's performance. In comparison, a static windowing approach merely uses a single window size for all tags, which is unable to catch the variation. We graphically compare data processing results from these three selected methods in *Figure 27*.

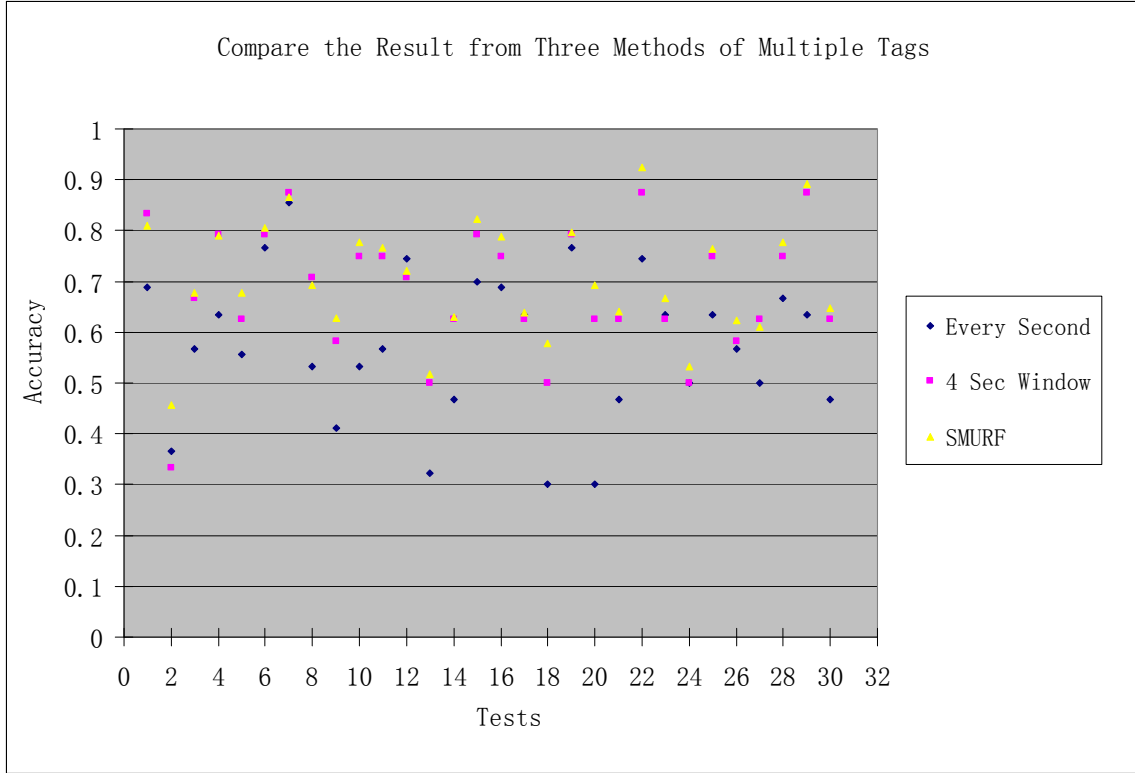


Figure 27. First Experiment: The Comparison of the Result from Non-overlapping 1 Second Window, 4 Second Window and Overlapping SMURF of Multiple Tags in One Chart.

The result of the second experiment also illustrates the use of SMURF method can substantially increase the reliability of collected data. As Figure 28 illustrates, SMURF provides the accuracy rate of (83.6%, Table 24) in the second experiment. However, compared with the best result from other static window methods, SMURF only improved by 1.95%.

$$\frac{SMURF'sRatio - StaticWindowMethod'sRatio}{StaticWindowMethod'sRatio} = \left(\frac{83.6\% - 82\%}{82\%} \right) = 1.95\%$$

Equation 6. The Improvement Made by using SMURF in Comparison with the Best Result of Static Window Method, in Sequence Detection.

This is almost improved by 2% (1.95%, Equation 6), and substantially lower than the result from the first experiment (3.67%, Equation 5). However this is not unexpected because of the better overall performance when looking at sequences rather than absolute

locations and times. For example, the result of a 4 second window in the second experiment provides a higher value of 82.67% (*Table 23: Mean of 4 Sec*) than the 68.19% obtained by the 4 second window from the first experiment (*Table 17: Mean of 4 Sec*).

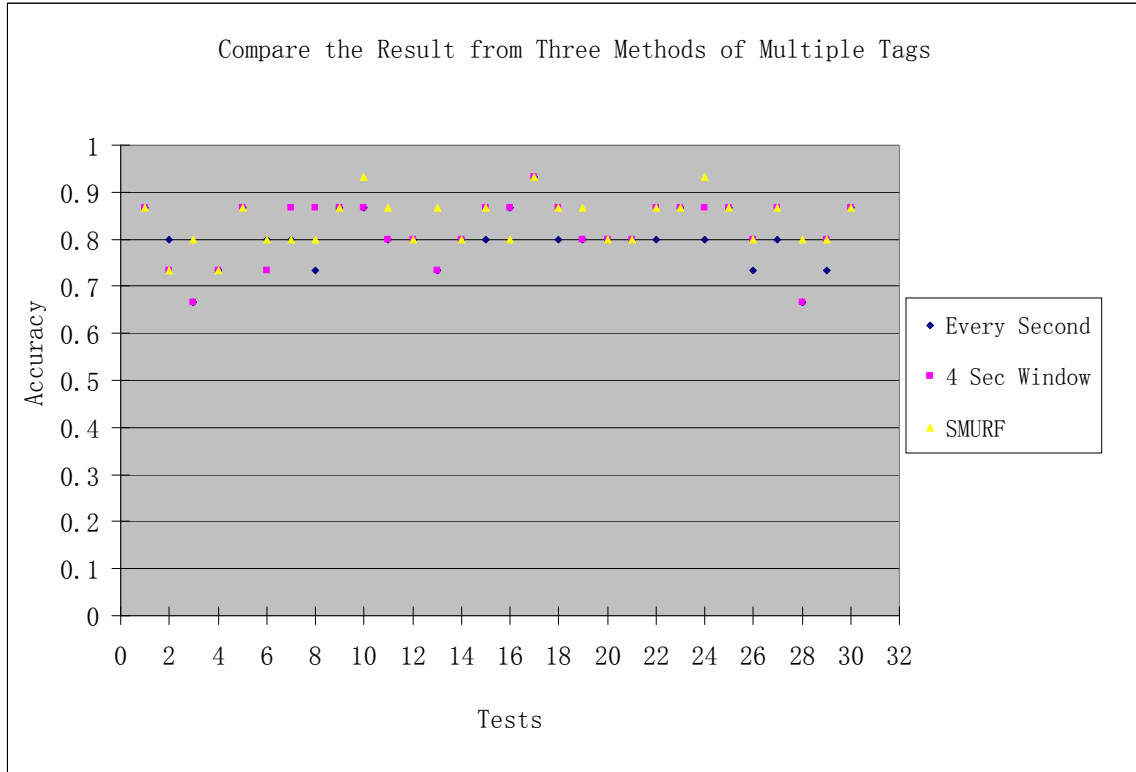


Figure 28. Second Experiment: The Comparison of the Result from Non-overlapping 1 Second Window, 4 Second Window and SMURF Methods of Multiple Tags in One Chart.

To bridge the gap between theory and practice, we use three different processing and analysis methods for both experiments, each experiment involves 3 different groups, and each group consists of 10 tests with 90 seconds for each separate test. According to our findings from Chapter 5, we find that the SMURF method provides more reliable results and improved outcomes. Based on two results of the SMURF methods (70.71% - First experiment: *Table 18*) and (83.6% - Second experiment: *Table 24*), it appears that the simplified SMURF approach is the best performing approach in both situations. In our experiment these performance accuracies are increased by 3.70% (*Equation 5*. from first experiment) and 1.95% (*Equation 6*. from second experiment).

4.5.1 Issues Relating to Practical and Conceptual Design

During the experiment, we encountered a number of issues relating to both practical and conceptual design. The main issues arose from the size of the area, which is dependent on the detecting range of the interrogator, both size and layout of the tags. In this study, we adjust the number, size and alignment of tags to suit the environment in order to optimise the personal interaction between user, reader and tags.

4.5.2 Hardware and Tag Placement Issues

One problem encountered is that the real environment has the potential for false readings in a multi-room environment. For instance, the tags can be detected through some building materials, which are generally used in the home environment, such as wood wall or partitions. We carried out a number of pilots before the experiment in terms of reading effectiveness. Therefore, in both experiments we placed tags at least 70cm from walls, in order to limit the potential for false readings. Nevertheless, there are still some drawbacks relating to the tags' placement, such as the distance between tags. If there is only a tiny space between two tags, it will confuse both static, non-overlapping window method and simplified SMURF, as both our non-overlapping and overlapping programs do not decrease their window's size to capture these tiny movements, thus significantly affecting the processing result. It is possible to create rules on tag placement to decrease the false reading rate. On the other hand, this could become quite complex and require expert assistance for real environment installation. This may significantly increase overall costs and restrict the performance of the object location system (Parry & Symonds, 2009).

4.5.3 Power and Communication

We found the reader's battery is a major issue for longstanding detection, especially for continuously reading status in our experiment. In fact a longer detecting range requires wider antennae tags and higher energy consumption. These facts result in the reduction of battery life (Parry & Symonds, 2009). However actual duration for both reader and tags actually depends on user defined scanning frequency and data size transmitted (Lin et al., 2007). Thus, both life and duration of battery are treated as major issues for most wireless

devices. Moreover, in reality there is a high demand for power, since when the elderly are alone in a private living space, the wearable sensor must be turned on.

4.5.4 Other Problems Encountered

Through our experiment, we find metal is a major barrier for RFID detection, which is quite common in an indoor environment. Philipose, et al. (2004) demonstrate that metal and water absorb the radio waves among RFID communication. We placed some tags into our toilet, and we found the effective reading rates were extremely low for these tags in both experiments. Initially we believed this was caused by the high humidity of that area. However, after we removed these tags from the washing machine, the effective reading rates returned to average even though it was still placed in that area (under the same humidity level). We also conducted several tests against this phenomenon by varying the size and material of attached metal material. Finally we found the size of metal material backing affects the rate and chance of detection, especially for large pieces of aluminium. In order to increase the reliability and accuracy of results, we suggest avoiding the use of metal backing for both experiments as well as real life. As a result, these materials will lower the nearby detection rate, and make RFID data streams unreliable (Philipose et al., 2004).

4.5.5 Limitations

There are a number of limitations to our study. Firstly, from the experimental environment aspect: temperature and humidity factors are not considered. Nevertheless, these factors affect the performance of the RFID system (Chen et al., 2005). Secondly, only passive tags are used in our experiment, we did not use active tags. Thirdly, from the design and methodology aspect, laboratory settings are unable to completely reflect the actual environment (Collis & Hussey, 2003). This gives rise to other issues in establishing and carrying out this research. In particular, we may not have such strong control over confounding and extraneous variables. Therefore, to minimise above issues, it is important to carefully set variables, and conduct iterative pre-test and post-test in a rigorous manner.

Furthermore, the objective of this study is to improve the reliability of inferring human behaviour based on the analysis of tracked movements. Thus the style of these movements is designed as linear or circular with no sophisticated movements involved in either experiment. Real world human activities are more complex in comparison with our experiments and laboratory setting, and may therefore avoid various data collection and reduce the precision of inference. Moreover, we place tags with a large space between two contiguous tags in order to avoid the tiny movements and the insufficiency of our processing tools. All RFID tags are placed within the same dimension and at a certain distance from each other. All above factors have the potential to avoid noise and reduce the chances of multiple detections. In addition, we only used one combination of readers and tags, the results may vary a lot in both reading ranges and detection rates if other combinations of readers and tags are applied. Differences in tag layout and spacing may result different experiment results. Thus all above factors urge us to adopt various types of tags in various environments for our future research.

4.5.6 Future Study

As we mentioned in Section 2.1.3, a group of passive, low-cost, UHF tags were placed under ground by Willis and Helal (2005). This RFID grid provides precise location and detailed attributes about the surrounding areas (Willis & Helal, 2005). It also offers the foundation of precise indoor/outdoor location for the disabled users. Future work will be focused on following their scenario for the guide feedback and sensing distance from obstacles to avoid collisions. This may involve the study of a different application area.

According to Varshney's (2007) study, around 40% of senior citizens in the US have to face more than one kind of physical disability. This figure will increase significantly with the growth of aging population in the near future. Previous research illustrates that adopting RFID systems can assist us with monitoring care. Thus our long-term goal will focus on applying RFID technology to effectively reduce suffering and chronic medical problems of the elderly (Varshney, 2007). The direction of potential study includes the development of RFID monitoring systems for senior or dementia patients with preventing and cognitive disability recovery.

5 *CONCLUSION*

This dissertation has given some examples of use of RFID in particular, in reference to living assistance for the elderly. The research report is aimed to determine how effective RFID support is for tracking every day activities of the elderly. Through the study of previous literature, we gain theoretical background and ideas of contemporary study efforts on RFID systems.

Specifically, RFID readers suffer from low detecting rates and frequent failure in capturing tags. Additionally, unexpected noises that cause the unreliable data in the final dataset are concluded as an issue of false reading. These false readings are classified into False Positive and False Negative readings. These factors and concerns become the significant challenge in large scale implementation (Parry & Symonds, 2009). In order to save the costs in both configuration and maintenance in terms of large scale implementation, RFID middleware has to provide more accurate data on reflecting the real world.

The answer to the our research question is that both the fixed size windowing approach and SMURF middleware can slightly increase the accuracy of gathered RFID data streams, thus increasing the reliability of the RFID system in activity detection for the elderly. Simplified SMURF provides a more reliable result. More complex environments where multiple tags are detected simultaneously may require new approaches but relatively simple data cleaning methods appear to be a promising start. However, due to complexity of human activity in real life (discussed in research limitations Section 4.5.5), and different standards of RFID systems (in both literature Section 2.1 and research limitations Section 4.5.5), these methods are too limited to sufficiently reflect physical reality by themselves, and further work is required in this area.

6 REFERENCES

- Alsos, O. A., & Svan, D. (2006). Interaction techniques for using handhelds and PCs together in a clinical setting. Paper Presented as the Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles, Oslo, Norway.
- Breahna, T., & Jonhns, D. (2006). Simulation spices RFID read rates. *Microwaves & RF*, 45(3), 66-76.
- Chawla, V., & Ha, D. S. (2007). An overview of passive RFID. *IEEE Applications & Practice*, 5(9), 11-17.
- Chen, Y. C., Chiang, J. R., Chu, H. H., Huang, P., & Tsui, A. W. (2005). Sensor assisted Wi-Fi indoor location system for adapting to environmental dynamics. Paper Presented as the Proceeding of MSWiM 05, Montreal, Quebec, Canada.
- Collis, J., & Hussey, R. (2003). *Business research: A practical guide for undergraduate and postgraduate students*, (Second Ed.). New York: Palgrave Macmillan.
- Coroama, V. (2006). Experiences from the design of a ubiquitous computing system for the blind. Paper Presented as the Proceeding of CHI '06 Extended Abstracts on Human Factors in Computing Systems, 664-669.
- Dobbins, G. H., Lane, I. M. & Steiner, D. D. (1998). A note on the role of laboratory methodology in applied behavioural research: Don't throw out the baby with the bath water. *Journal of Organisational Behavioural*, 9, 281-286.
- Ding, B., Yuan, H., Zang, X., & Jiang, L. (2007). The research on blind navigation system based on RFID. *Wireless Communications, Networking and Mobile Computing*, 9(21-25), 2058-2061.

- Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532-550.
- Fishkin, K.P., Jiang, B., Philipose, M., Roy, S. (2004). I sense a disturbance in the force: Unobtrusive detection of interactions with RFID-tagged objects. *UbiComp*.
- Gaetano, B. (2005). RFID: Tagging the world. *Communications of the ACM*. 48(9), 34-37.
- Garfinkel, S. L., Juels, A., & Pappu, R. (2005). RFID privacy: An overview of problems and proposed solutions. *Security & Privacy Magazine*, 3(3), 34-43.
- Hansmann, U., Merk, L., Nicklous, M. S., & Stober, T. (2003). *Pervasive computing*. Germany, Berlin: Springer.
- Jeffery, S. R., Franklin, M., & Garofalakis, M. (2008). An adaptive RFID middleware for supporting metaphysical data independence. *The VLDB Journal*, Berlin, Heidelberg: Springer, 17 (2), 265-289.
- Jeffery, S. R., Garofalakis, M., & Franklin, M. J. (2006). Adaptive cleaning for RFID data streams. Paper presented as the proceeding of the 32nd International Conference on Very Large data bases, Seoul, Korea.
- Juels, A. (2006). RFID Security and privacy: A research survey. *IEEE Journal on Selected Areas in Communications*, 24(2), 381-394.
- Kutiyanawala, A., Kulyukin, V., & LoPresti, E. (2006). A rollator-mounted wayfinding system for the elderly: A smart world perspective. Paper Presented as the Proceeding of the 8th International ACM SIGACCESS conference on Computers and Accessibility, 245-246.

- Landt, J. (2001). Shrouds of time: The history of RFID (Tech. Rep). Pittsburgh, PA, USA: AIM Global.
- Lin, Y. J., Su, M. J., Chen, S. J., Wang, S. C., Lin, C. I., & Chen, H. S. (2007). A study of ubiquitous monitor with RFID in an elderly nursing home. Paper Presented as the Proceeding of MUE 2007 International Conference on the Multimedia and Ubiquitous Engineering.
- Mamei, M., & Zambonelli, F. (2007). Pervasive pheromone-based interaction with RFID tags. *ACM Transactions on Autonomous and Adaptive Systems*, 2(2), 4.
- March, S., T., & Smith, G., F. (1995). Design and natural science research on information technology. *Decision Support Systems Archive*, 15(4) 251-266.
- Michael, K., & McCathie, L. (2005). The pros and cons of RFID in supply chain management. Paper Presented as Proceedings of the International Conference on Mobile Business, IEEE.
- Navarro, G. (2001) A guided tour to approximate string matching. *ACM Computing Surveys*, 33(1):31-88.
- Ni, M. L., Liu, Y., Lau, Y. C., & Patil, A. P. (2004). LANDMARC: indoor location sensing using active RFID. *Wireless Networks of Kluwer Academic*, 10(6), 701-710.
- Niederman, F., Mathieu, R., G., Morley, R., & Kwon, I., W., (2007). Examining RFID applications in supply chain management. *ACM Communications*, 50(7), 92-101.
- Ohkubo, M., Suzuki, K., Kinoshita, S. (2005). RFID privacy issues and technical challenges. *ACM Communications*, 48(9), 66-71.

- Parry, D., & Symonds, J. (2009). RFID and assisted living for the elderly, 119 – 136. In Symonds, J., Ayoade, J., & Parry, D. (2009). Auto-identification and ubiquitous computing applications: RFID and smart technologies for information convergence. New York: Information Science Reference.
- Phillips, T., Karygiannis, T., & Kuhn, R. (2005). Security standards for the RFID market. IEEE Security & Privacy Magazine, 3(6), 85 – 89.
- Philipose, M., Fishkin, K.P., Perkowitz, M., Patterson, D.J., Fox, D., Kautz, H. & Hahnel, D. (2004). Inferring activities from interactions with objects. IEEE Pervasive Computing, 3(4), 50-57.
- Rao, J., Doraiswamy, S., Thakkar, H., & Colby, L. (2006). A deferred cleansing approach for RFID data analytics. Paper Presented as the Proceeding of the 32nd International Conference on Very Large Data Bases, Seoul, Korea.
- Romer, K., Schoch, T., & Mattern, F. (2004). Smart identification frameworks for ubiquitous computing Applications. Wireless Networks, 10(6), 689-700.
- Roussos, G. (2006). Enabling RFID in retail. IEEE Computer Society, 39(3), 25-30.
- Sarma, S., Brock, D., & Engels, D. (2001). Radio frequency identification and the electronic product Code. IEEE Micro, 21(6), 50-54.
- Song, B., Qin, P., Wang, H., Xuan, W., & Yu, G. (2009). “bSpace: A Data Cleaning Approach for RFID Data Streams Based on Virtual Spatial Granularity, “Hybird Intelligent Systms. Paper Presented as the Proceeding of the Ninth International Conference on Hybrid Intelligent Systems, 3, 252-256.

- [illegible]

- Weinstein, R. (2005). RFID: A technical overview and its application to the enterprise. *IT Professional*, 7(3), 27-33.
- Willis, S., & Helal, S. (2005). RFID information grid for blind navigation and wayfinding. Paper Presented as the Proceeding of the 9th IEEE Symposium on Wearable Computers.

7 APPENDIX

7.1 Appendix A: Original RFID Datasets of Pilots

[illegible]

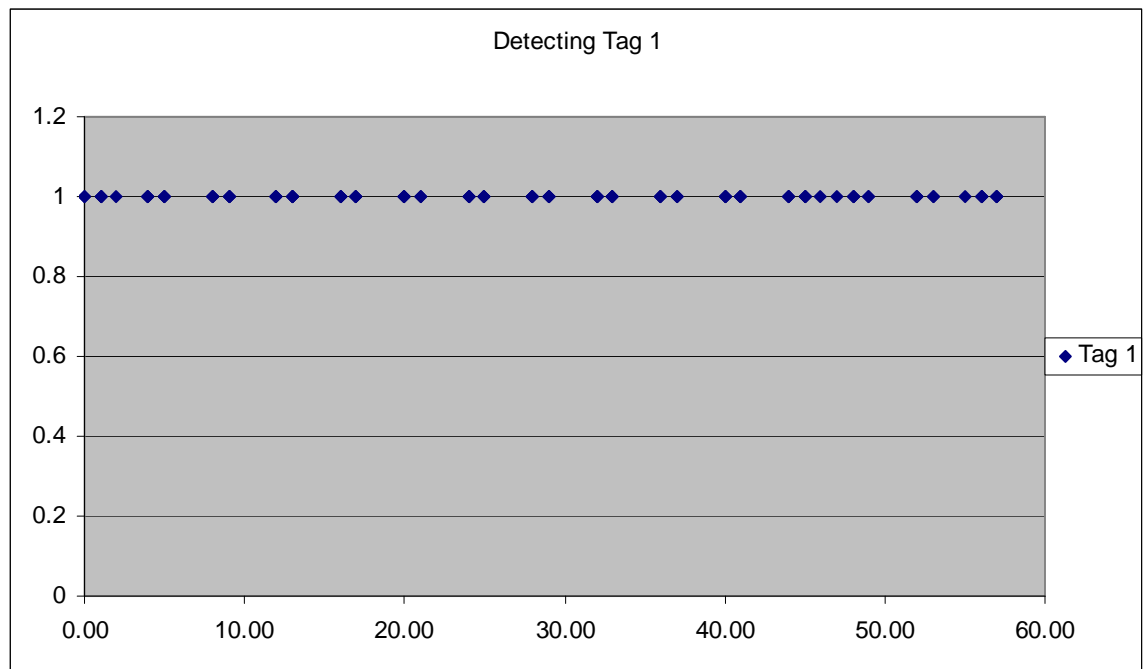


Figure 29. Pilot of Single Tags.

[illegible]

[illegible]

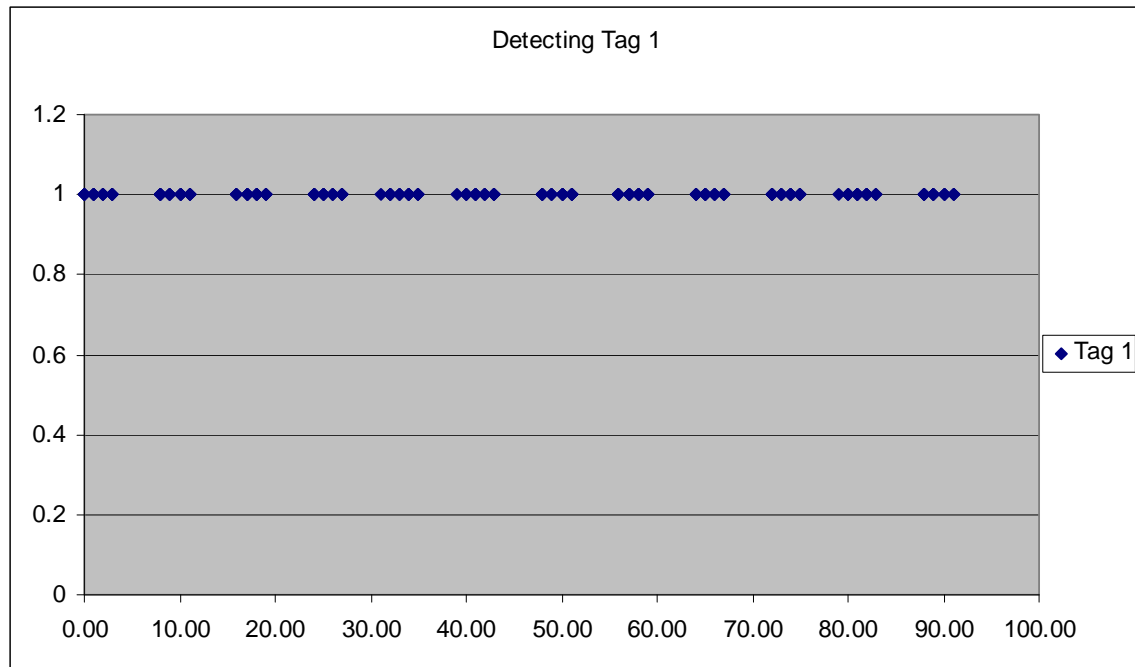


Figure 30. Pilot of Single Tags by using Non-Overlapping Window Method.

- 83

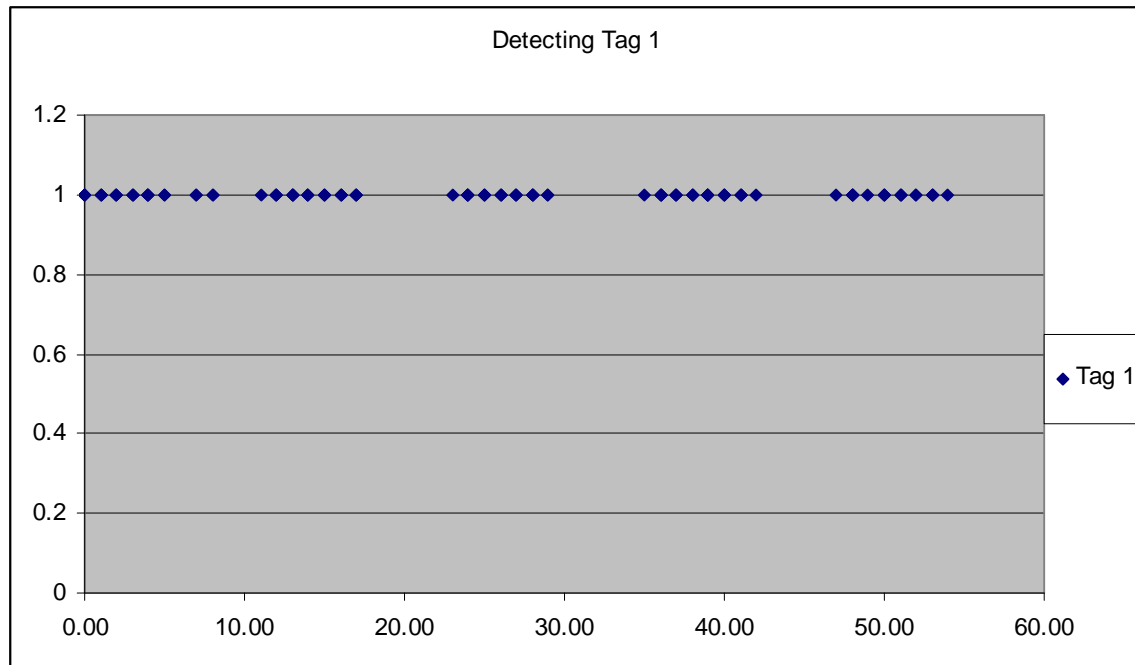


Figure 31. Pilot of Single Tags by using Non-Overlapping Window Method.

- 85 -

E2003411B802011371233242	11:27:52 04 May 2009 (Local)	11	27	52	41272.00	19.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:27:53 04 May 2009 (Local)	11	27	53	41273.00	20.00	E2003411B802011371233242	0	1
E2003411B802011371233244	11:27:53 04 May 2009 (Local)	11	27	53	41273.00	20.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:27:54 04 May 2009 (Local)	11	27	54	41274.00	21.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:27:54 04 May 2009 (Local)	11	27	54	41274.00	21.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:27:55 04 May 2009 (Local)	11	27	55	41275.00	22.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:27:55 04 May 2009 (Local)	11	27	55	41275.00	22.00	E2003411B802011371233244	1	0
E2003411B802011371233242	11:27:55 04 May 2009 (Local)	11	27	55	41275.00	22.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:27:56 04 May 2009 (Local)	11	27	56	41276.00	23.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:27:56 04 May 2009 (Local)	11	27	56	41276.00	23.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:27:56 04 May 2009 (Local)	11	27	56	41276.00	23.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:27:57 04 May 2009 (Local)	11	27	57	41277.00	24.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:27:57 04 May 2009 (Local)	11	27	57	41277.00	24.00	E2003411B802011371233242	0	1
E2003411B802011371233244	11:27:57 04 May 2009 (Local)	11	27	57	41277.00	24.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:27:57 04 May 2009 (Local)	11	27	57	41277.00	24.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:27:58 04 May 2009 (Local)	11	27	58	41278.00	25.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:27:58 04 May 2009 (Local)	11	27	58	41278.00	25.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:27:58 04 May 2009 (Local)	11	27	58	41278.00	25.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:27:59 04 May 2009 (Local)	11	27	59	41279.00	26.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:27:59 04 May 2009 (Local)	11	27	59	41279.00	26.00	E2003411B802011371233244	1	0
E2003411B802011371233242	11:28:00 04 May 2009 (Local)	11	28	00	41280.00	27.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:00 04 May 2009 (Local)	11	28	00	41280.00	27.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:00 04 May 2009 (Local)	11	28	00	41280.00	27.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:00 04 May 2009 (Local)	11	28	00	41280.00	27.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:01 04 May 2009 (Local)	11	28	01	41281.00	28.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:01 04 May 2009 (Local)	11	28	01	41281.00	28.00	E2003411B802011371233242	0	1
E2003411B802011371233244	11:28:01 04 May 2009 (Local)	11	28	01	41281.00	28.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:02 04 May 2009 (Local)	11	28	02	41282.00	29.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:02 04 May 2009 (Local)	11	28	02	41282.00	29.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:02 04 May 2009 (Local)	11	28	02	41282.00	29.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:02 04 May 2009 (Local)	11	28	02	41282.00	29.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:03 04 May 2009 (Local)	11	28	03	41283.00	30.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:03 04 May 2009 (Local)	11	28	03	41283.00	30.00	E2003411B802011371233244	1	0
E2003411B802011371233242	11:28:03 04 May 2009 (Local)	11	28	03	41283.00	30.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:04 04 May 2009 (Local)	11	28	04	41284.00	31.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:04 04 May 2009 (Local)	11	28	04	41284.00	31.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:04 04 May 2009 (Local)	11	28	04	41284.00	31.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:05 04 May 2009 (Local)	11	28	05	41285.00	32.00	E2003411B802011371233242	0	1
E2003411B802011371233244	11:28:05 04 May 2009 (Local)	11	28	05	41285.00	32.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:05 04 May 2009 (Local)	11	28	05	41285.00	32.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:06 04 May 2009 (Local)	11	28	06	41286.00	33.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:06 04 May 2009 (Local)	11	28	06	41286.00	33.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:06 04 May 2009 (Local)	11	28	06	41286.00	33.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:07 04 May 2009 (Local)	11	28	07	41287.00	34.00	E2003411B802011371233244	1	0
E2003411B802011371233242	11:28:07 04 May 2009 (Local)	11	28	07	41287.00	34.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:08 04 May 2009 (Local)	11	28	08	41288.00	35.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:08 04 May 2009 (Local)	11	28	08	41288.00	35.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:08 04 May 2009 (Local)	11	28	08	41288.00	35.00	E2003411B802011371233242	0	1
E2003411B802011371233242	11:28:09 04 May 2009 (Local)	11	28	09	41289.00	36.00	E2003411B802011371233242	0	1
E2003411B802011371233244	11:28:09 04 May 2009 (Local)	11	28	09	41289.00	36.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:09 04 May 2009 (Local)	11	28	09	41289.00	36.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:10 04 May 2009 (Local)	11	28	10	41290.00	37.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:10 04 May 2009 (Local)	11	28	10	41290.00	37.00	E2003411B802011371233244	1	0
E2003411B802011371233244	11:28:11 04 May 2009 (Local)	11	28	11	41291.00	38.00	E2003411B802011371233244	1	0

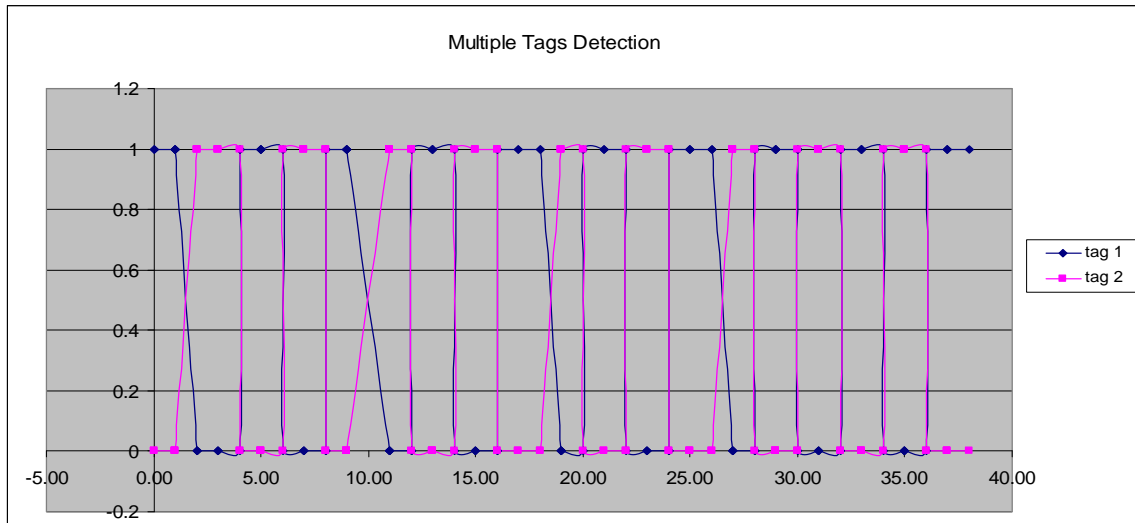


Figure 32 Pilot of 2 Tags by using 1 Second Window Method explain what this is and label the Axes.

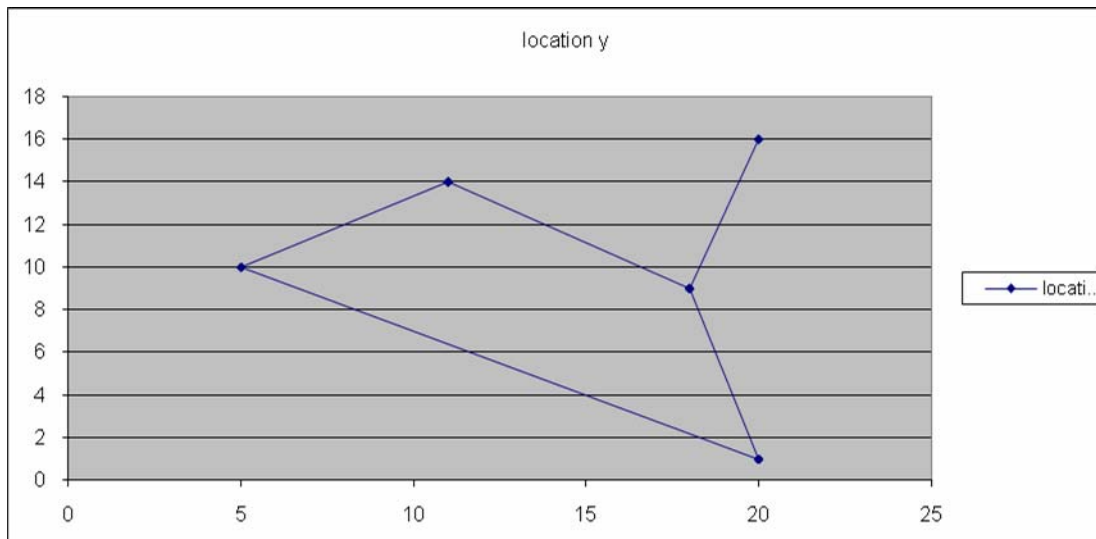


Figure 33. Pilot 1 of Multiple Tags by using 2 Sec Static Window Method

These figures show the calculated movement around the space using the actual x-y locations of the tags and linking the detected tag sequence. Although crude these graphs give an indication of where tags have been detected and the sequence they were detected in. Any backtracking or bypassing indicates a double reading or failure to read a tag.

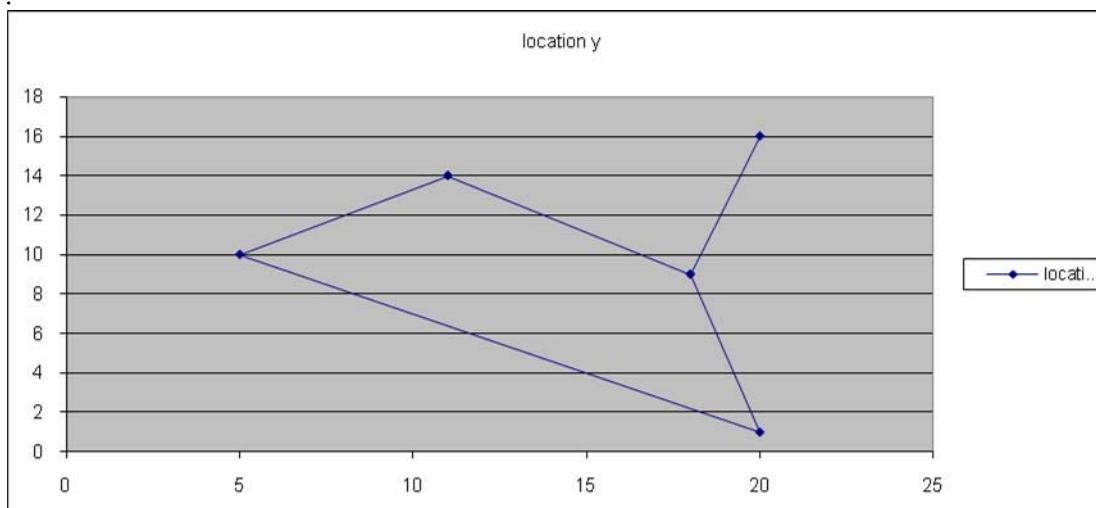


Figure 34. Pilot of Multiple Tags by using 4 Sec Static Window Method

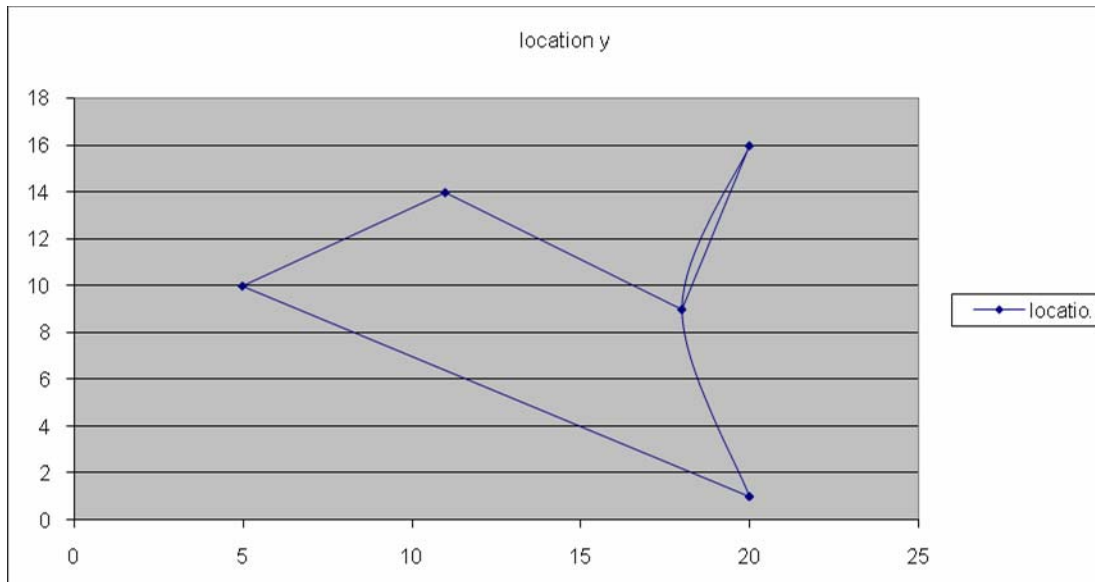


Figure 35. Pilot of Multiple Tags by using 6 Sec Static Window Method

The double line indicates that the RFID data shows movement from 20,16 to 18, 9 and back again which is not correct.

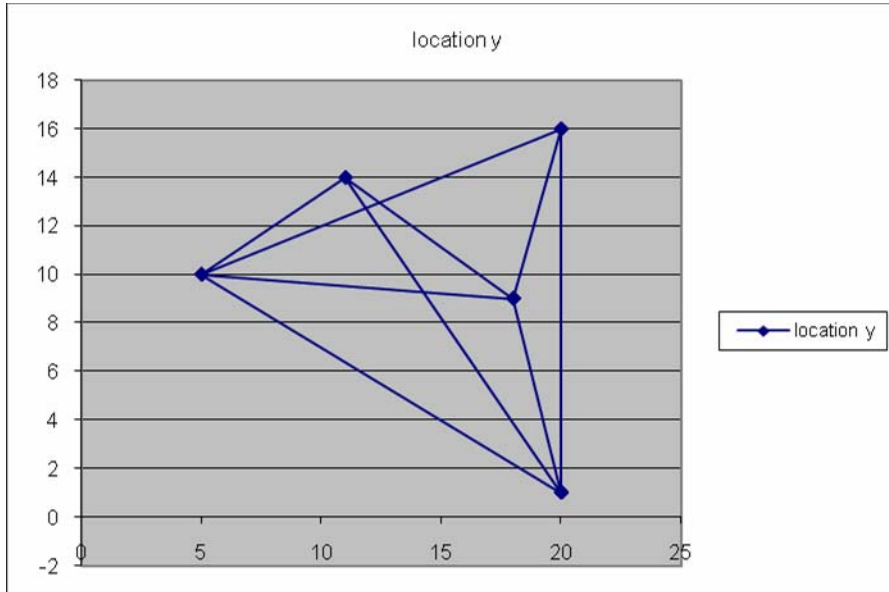


Figure 36. Pilot of Multiple Tags by using 1 Second Window Method

Here is the X-Y scatter plot with multiple circuits, which is used in the pilot for measure and moving trail of the reader, and the actual position of the subject at each time point.

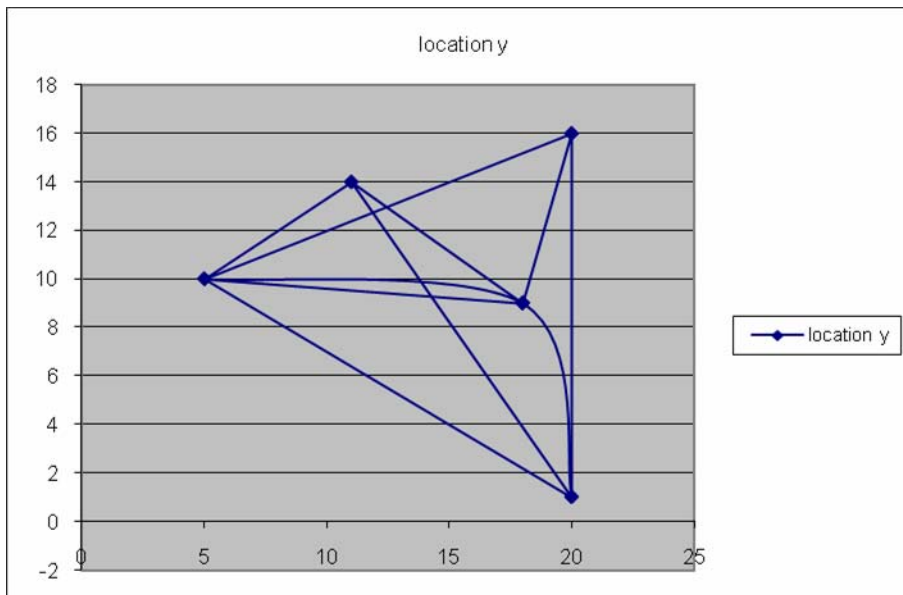


Figure 37. Pilot of Multiple Tags by using 2 Sec Static Window Method

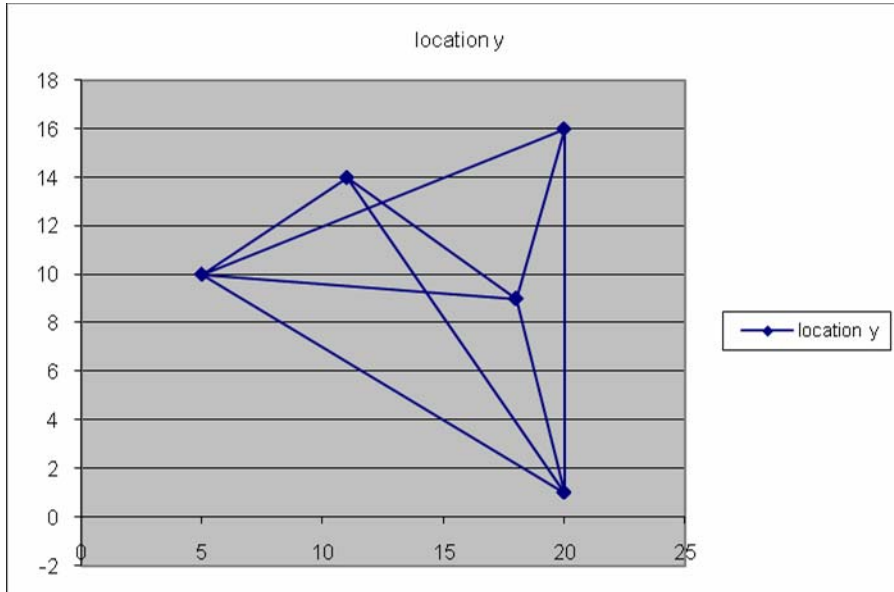


Figure 38. Pilot of Multiple Tags by using 4 Sec Static Window Method

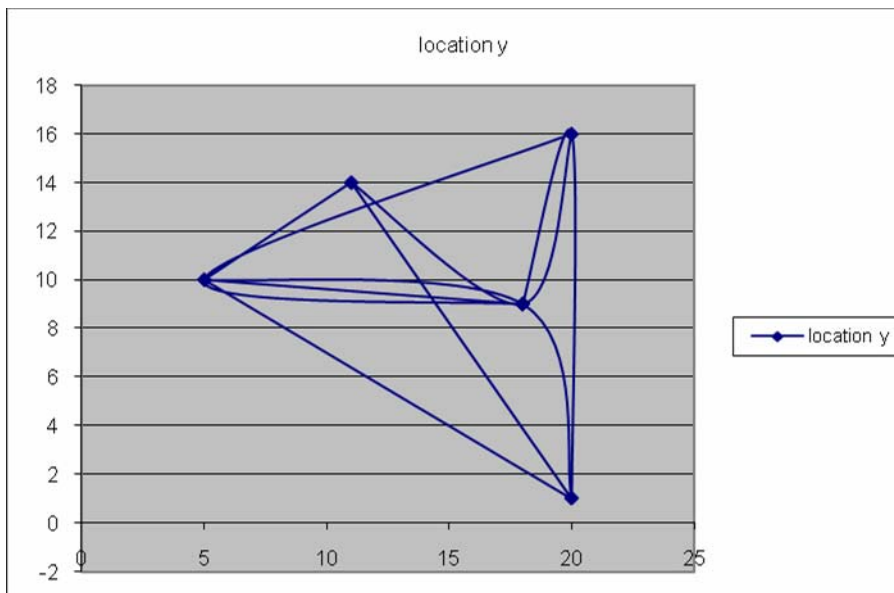


Figure 39. Pilot of Multiple Tags by using 6 Sec Static Window Method

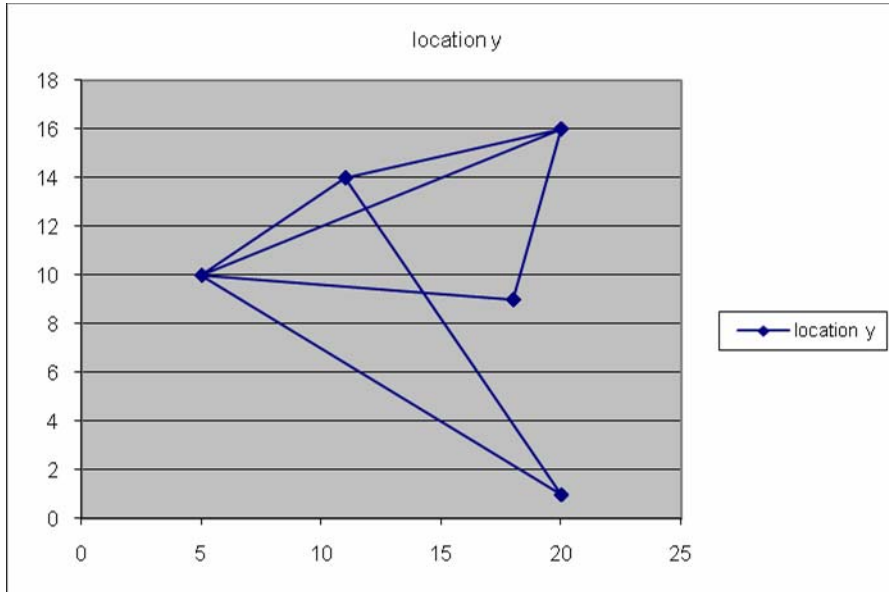


Figure 40. Pilot of Multiple Tags by using 1 Second Window Method

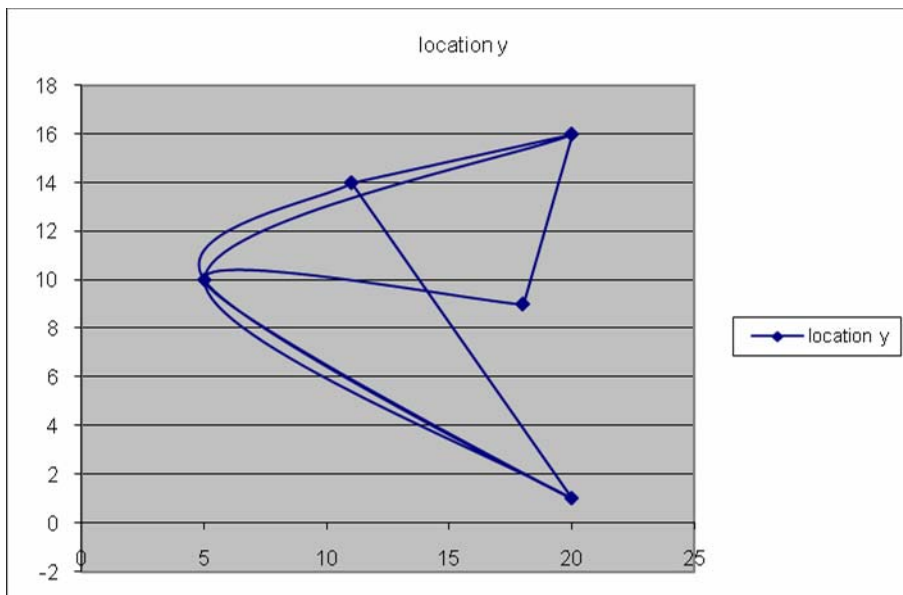


Figure 41. Pilot of Multiple Tags by using 2 Sec Static Window Method

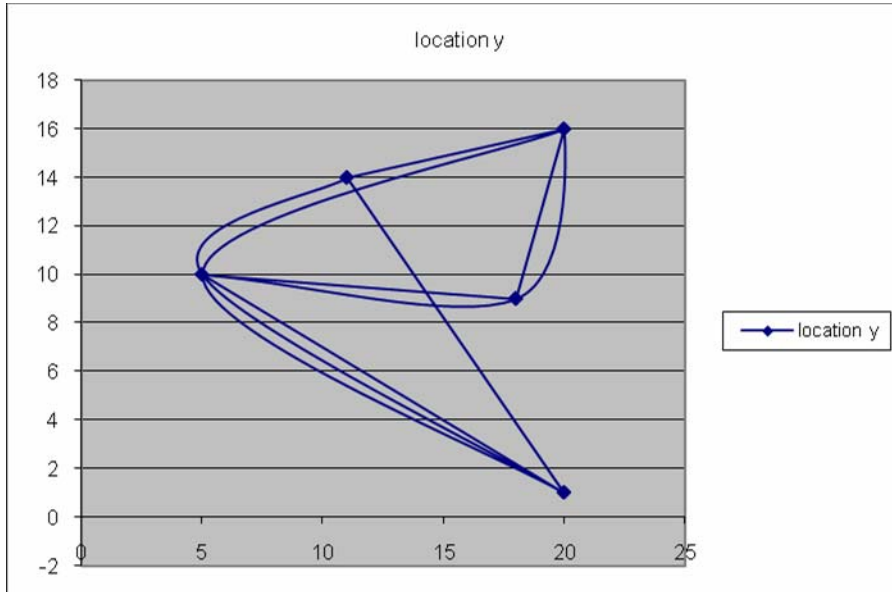


Figure 42. Pilot of Multiple Tags by using 4 Sec Static Window Method

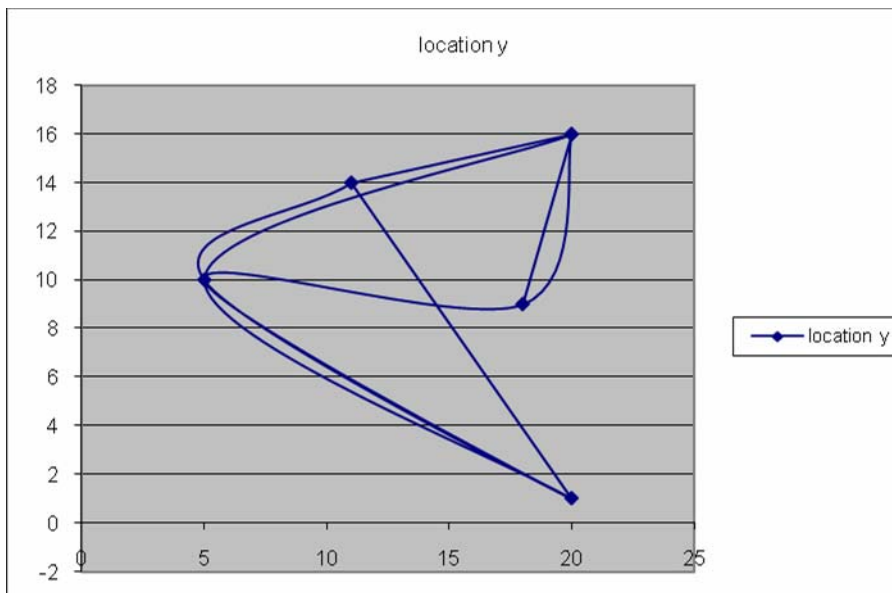


Figure 43. Pilot of Multiple Tags by using 6 Sec Static Window Method

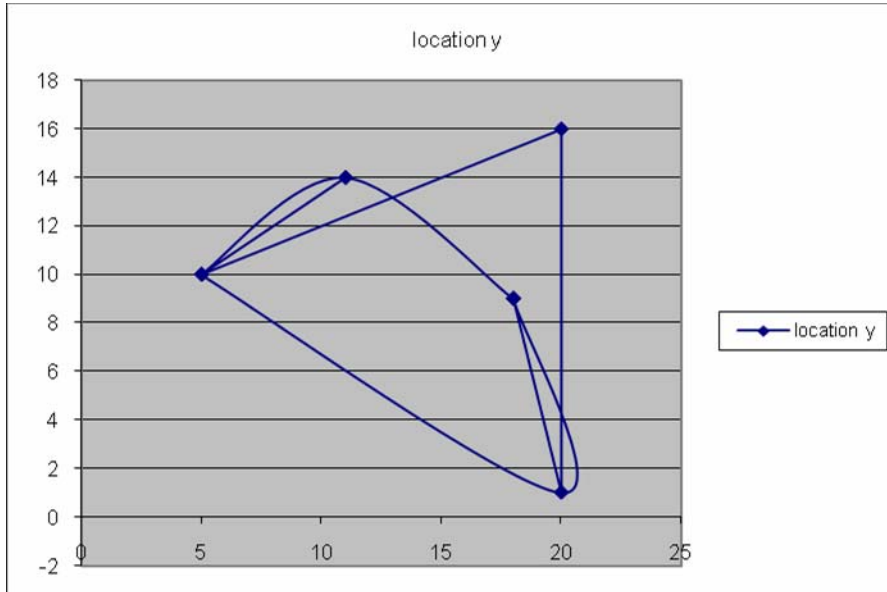


Figure 44. Pilot of Multiple Tags by using 1 Second Window Method

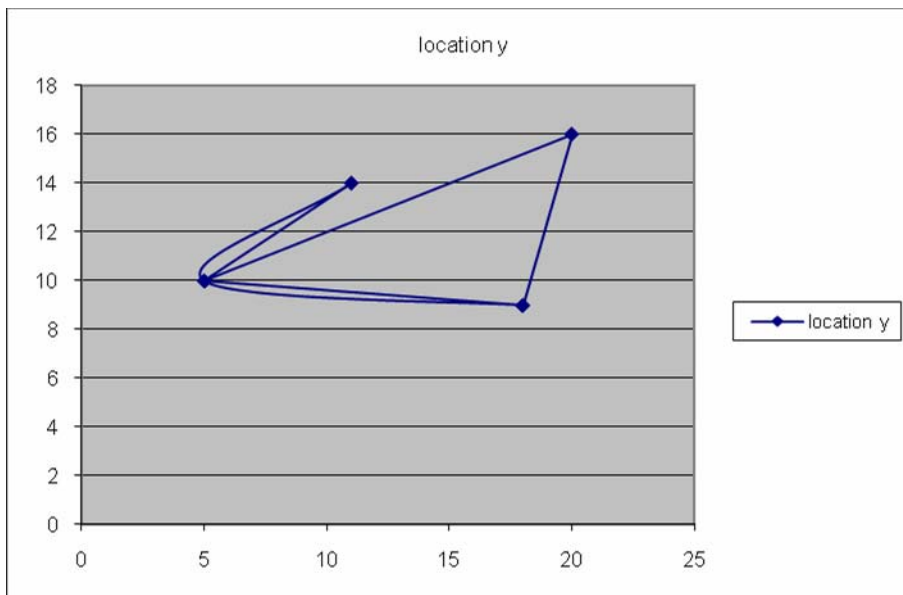


Figure 45. Pilot of Multiple Tags by using 2 Sec Static Window Method

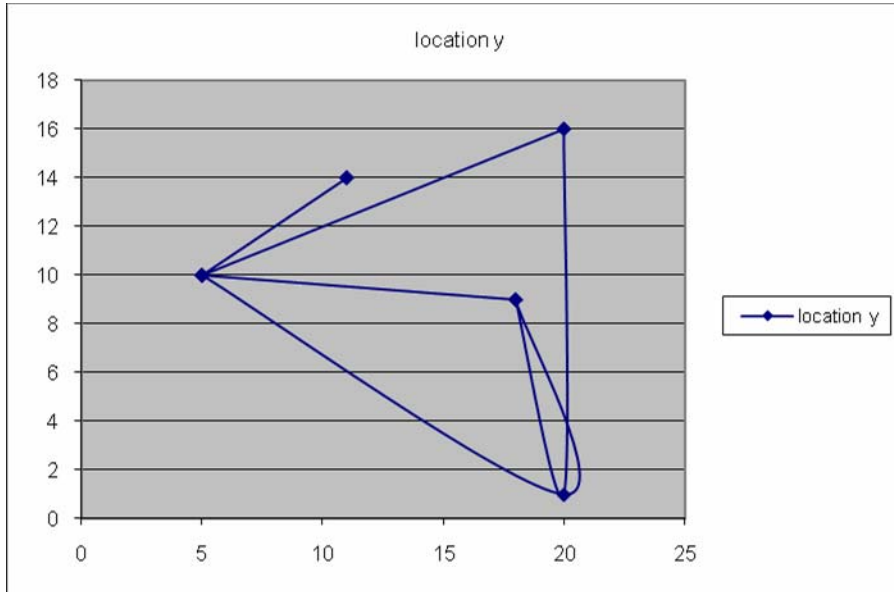


Figure 46. Pilot of Multiple Tags by using 4 Sec Static Window Method

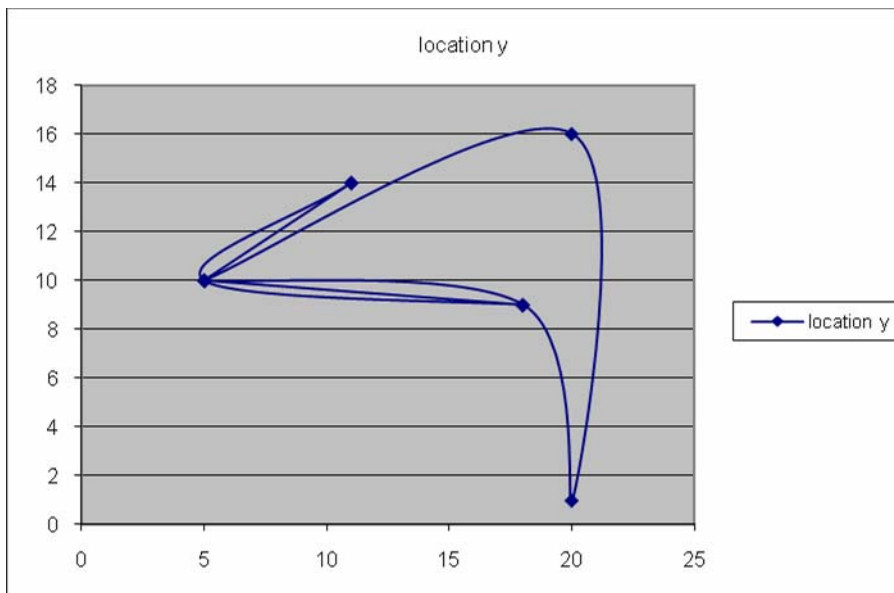


Figure 47. Pilot of Multiple Tags by using 6 Sec Static Window Method