Modelling infection spread using location tracking

Mason, A.M. and Dingle, N.J. and Knottenbelt, W.J. and Bell, D. and Buchanan, W. and Theummler, C.

2010

MIMS EPrint: **2010.104**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

Modelling infection spread using location tracking

Andrew Michael Mason, Nicholas John Dingle and William John Knottenbelt*

Department of Computing,
Imperial College London,
180 Queen's Gate, London SW7 2AZ, UK
E-mail: amm104@doc.ic.ac.uk
E-mail: njd200@doc.ic.ac.uk
E-mail: wjk@doc.ic.ac.uk
*Corresponding author

Derek Bell

Imperial College London,
NIHR CLAHRC for Northwest London,
Chelsea & Westminster Hospital,
369 Fulham Road, London SW10 9NH, UK
E-mail: d.bell@imperial.ac.uk

William Buchanan

Centre for Distributed Computing and Security,
School of Computing, Napier University,
10 Colinton Road, Edinburgh EH10 5DT, UK
E-mail: w.buchanan@napier.ac.uk

Christoph Thuemmler

Centre for Applied Ehealth,
School of Computing, Napier University,
10 Colinton Road, Edinburgh EH10 5DT, UK
E-mail: c.thuemmler@napier.ac.uk

Abstract: The precision of location tracking technology has improved greatly over the last few decades. We aim to show that by tracking the locations of individuals in a closed environment, it is now possible to record the nature and frequency of interactions between them. Further, that it is possible to use such data to predict the way in which an infection will spread throughout such a population, given parameters such as transmission and recovery rates. We accordingly present a software package that is capable of recording and then replaying location data provided by a high-precision location tracking system. The software then employs a combination of SIR modelling and the epidemiological technique of contact tracing in order to predict the spread of an infection. We use this software to conduct a number of experiments using a sample data set, and compare the SIR graphs generated from these to similar graphs generated using the traditional SIR differential equations.

Keywords: location tracking; SIR modelling; contact tracing; infection spread.

Reference to this paper should be made as follows: Mason, A.M., Dingle, N.J., Knottenbelt, W.J., Bell, D., Buchanan, W. and Thuemmler, C. (2010) 'Modelling infection spread using location tracking', *Int. J. Healthcare Technology and Management*, Vol. 11, No. 6, pp.442–461.

Biographical notes: Andrew Mason obtained an MSc Degree with Distinction in Computing Science from Imperial College London in 2008, following a BSc in Mathematics from the same institution in 2007. He researched location tracking technologies as part of the former, under the supervision of William Knottenbelt. After leaving Imperial, Andrew joined the graduate scheme at BT, where he currently works as a Software Engineer.

Nicholas Dingle obtained an MSc Degree with Distinction in Computing Science from Imperial College London in 2001 and a PhD in Computing Science from the same institution in October 2004. He is currently employed as a Research Associate in the Department of Computing investigating the performance optimisation of virtualised data storage systems. His other research interests are techniques for the performance analysis of models of very large concurrent systems, with a particular focus on the use of parallel and distributed approaches for extracting response time densities and quantiles.

William John Knottenbelt is a Reader in Applied Performance Modelling in the Department of Computing at Imperial College London, where he became a Lecturer in 2000. His research work focuses on the performance modelling of systems using high-level formalisms, with applications to real-world systems including healthcare systems and data storage infrastructures.

Derek Bell is the Chair of Acute Medicine at Imperial College London and the Director of the NIHR CLAHRC for Northwest London a large research programme designed to narrow the second translational gap in healthcare research. He qualified in Medicine at Edinburgh University and trained in both Edinburgh and London as a physician. His major areas of research include improving health care systems, patient and public involvement in research and the clinical diagnosis and management of pulmonary embolism and pneumonia.

Bill Buchanan is a Professor in the School of Computing at Edinburgh Napier University. He currently leads the Centre for Distributed Computing, Networking and Security, and works in the areas of security, e-crime, intrusion detection systems, digital forensics, e-health, mobile computing, agent-based systems, and simulation. He has published over 26 academic books, and over 100 academic research papers, along with awards for excellence in knowledge transfer. Presently, he is working with a range of industrial/domain partners, including with the Scottish Police, health care professionals and the FSA.

Christoph Thuemmler studied Medicine, Political Science and Educational Science at Heidelberg University where he earned his Doctoral Degree in Neurology. He is currently the Director of the Centre for Applied Ehealth at Edinburgh Napier University, in Edinburgh, UK and holds a Honorary Consultant post with Chelsea and Westminster Hospital in London. He has an interest in cloud computing, RFID and the internet of things and is involved in research work with the European internet of things knowledge cluster hosted by the European Commission in Brussels.

1 Introduction

The ability to accurately track the location of an individual has improved greatly over the last 30 years. From the launch of the GPS system in 1978 (Strom, 2008), through wireless technologies such as 802.11a/b/g and Bluetooth, and technologies such as assisted GPS, the accuracy of real-time location tracking systems that are widely available today is sub-10 m. While this is useful, and widely utilised, for many location-based applications, the ability to track individuals with an accuracy greater than this has only recently been developed. Increasingly, technologies such as RFID (Choi et al., 2006) and ultrawideband (http://www.ubisense.net/index.php?load=content&pageid=80) are being exploited to provide accurate location tracking of a large number of individuals.

Many uses of this sort of technology are already apparent, for example keeping track of workers in a dangerous environment to protect their safety (Choi et al., 2006), and tracking football players to provide post-match analysis (http://www.ubisense.net/index.php?load=content&pageid=80). However, there are undoubtedly many more to be discovered as the technology matures and is more widely available. One such application is in the healthcare system. Location tracking systems are already being used to keep tabs on at-risk patients, who need constant monitoring to protect their safety (http://www.ekahau.com/?id=1010). Another use may well be in the tracking and prevention of infections, and it is this application that this paper aims to explore.

Many diseases are passed from person to person by physical encounters – from intimate contact in the case of Sexually Transmitted Infections (STI), to simply being in the near vicinity in the case of more virulent airborne infections. While the former type of encounter is relatively easy to keep track of, until recently the latter has been almost impossible to study. With the ability to track individuals' locations to a sub-1 metre accuracy, though, this may change.

If we assume that it is possible to track the spread of an infection at the scale proposed, there are many possible applications of such a system with regard to the prevention and further understanding of epidemics. For example, during a real epidemic, information about contact between individuals could give doctors access to a list of people ordered by the probability that they are infected. This would allow informed targeting of treatment, and quarantine if necessary. If it were possible to alter the model in real-time to reflect the infection status of patients, the model would become more accurate as time went on, and could retro-actively be updated to calculate new lists as required. If recorded, such information would also be invaluable in predicting the spread of future infections.

In a similar vein, given a relatively predictable system (such as a hospital), it should be possible to determine which individuals have the most infection-spreading contact with others. This would again allow doctors to create a list of people for whom immunisation would be a priority in the event of an epidemic. This could dramatically reduce its spread, as those responsible for spreading the infection to most people would no longer do so.

Further applications are also possible. Given information about the people who are infected, and when they showed symptoms, it might be possible to determine the likely source of an infection within a closed system (again, such as a hospital). From this point, a more traditional contact tracing study could be performed to identify those at risk in the outside world.

There is a wide body of prior work on the estimation of epidemiological model parameters from incomplete sets of real-life observations, including Expectation Maximisation (EM) (Becker, 1993), Markovian (Gibson, 1997; Gibson and Renshaw, 1998; O'Neill and Roberts, 1999) and non-Markovian approaches (Streftaris and Gibson, 2002, 2004). In contrast, the aim of the current paper is to investigate the feasibility of using high-precision real-time location tracking as a basis to conduct a low-level emulation of the evolution of the spread of infection under the assumptions of epidemiological models.

We have designed a software package that can record location tracking data and allows the user to analyse the data in a variety of ways in order to evaluate its effectiveness in the analysis of infection. This involved being able to create a scenario in which a given set of individuals is infected, and viewing the results of this as contacts between individuals occur in the monitored environment. We then analysed the results from the location tracking system and compared them with existing epidemiology models.

The remainder of this paper is organised as follows: Section 2 briefly introduces the background theory of the SIR model of infection spread. Section 3 then describes the software package that we have developed to track and record the movements of entities exposed to infection, which is then used in Section 4 to obtain quantitative data regarding the spread of infection throughout a given scenario and provide a qualitative comparison with an existing infection model. Finally, Section 5 concludes and discusses opportunities for future work.

2 SIR modelling

Epidemiology is concerned with the spread of disease and its effect on people. This in itself encompasses a range of disciplines, from biology to sociology and philosophy, all of which are utilised to better understand and contain the spread of infection.

One of the most basic models that is used in epidemiology is the SIR model. This was introduced as far back as 1927 by Kermack and McKendrick (1927), and although simple, is a good model for many infectious diseases. The SIR model assumes that a given fixed population whose members are at risk of a given infection can be split into three groups:

- $S(t)$ susceptible (those at risk of infection)
- $I(t)$ infectious (those who have been infected by the disease, and can infect others.
- $R(t)$ recovered (those who have recovered from the disease, and therefore are no longer at risk)

As denoted above, S, I and R are all functions of time, and vary as an epidemic progresses. Assuming the total population remains constant (and is denoted by N), it is clear that the following holds:

$$S(t) + I(t) + R(t) = N.$$

Using this equation as a base, we then consider the transition rates between the three states, i.e., the rate at which susceptibles become infected, and at which those who are

infected recover. These are denoted by βI and ν respectively, where β is the transmission rate (the rate at which susceptible individuals become infected during an encounter with an infections individual), and ν is the rate at which those who are infected recover. Thus is it possible to derive a set of differential equations to describe the system:

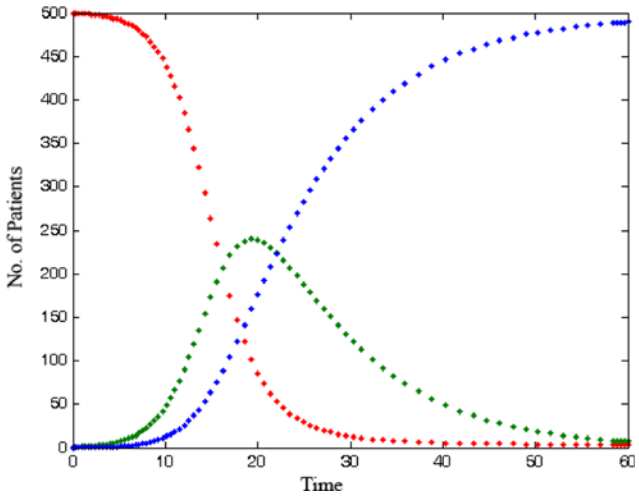
$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta IS$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta IS - \nu I$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \nu I.$$

While this system cannot be solved analytically, plotting it on a graph yields a typical timeline for an epidemic, for example as shown in Figure 1.[1]

Figure 1 A typical SIR model. Red represents susceptible individuals, green infectious and blue recovered (see online version for colours)



This basic model can be extended to include many more factors, for example an incubation period. However, at a basic level the SIR model provides a reasonable approximation for the spread of a disease.

3 The infection tracking software

A disease spread model is a good analytical tool, but by their very nature models are only an approximation of real life. In order to evaluate a model, we need to collect empirical data regarding the transmission of a disease. Contact tracing is a method used by epidemiologists that involves creating a list of infection-spreading contact that one individual has had with another (Kretzschmar et al., 2000).

For example, assume individual A has an infection that is spread by physical contact. Person A recalls that he has recently come into contact with persons B and C. B recalls that since coming into contact with A, he has come into contact with person D, and so on.

Using this method, it is possible to build up a graph (for example Figure 2)[2] of those who may have become infected, complete with the probability of infection if β, the transmission rate, is known.

Figure 2 An example of the type of graph produced by contact tracing. Each node represents an infected individual, and each line connects two individuals who have been in contact (see online version for colours)

In certain cases, this method works well. It is particularly useful in the study of STIs (Huerta and Tsimring, 2002), when infection-spreading encounters are likely to be remembered, and the timing pinpointed relatively accurately. However, if person B can become infected by person A simply by being within a given distance of A for a given time, it is obvious that the method breaks down. Short of requiring every individual in a given environment to record every such encounter, it is impossible to create such a graph.

However, with recent advances in location tracking technology, it has become possible to record the location of any number of individuals reliably to a high degree of accuracy (sub-25 cm). This in turn makes it possible to create such a graph when even the briefest of encounters can be the cause of an infection being transmitted. Using the location tracking data, therefore, it should be possible to construct an SIR model for a given infection whose parameters are known by performing a contact tracing study.

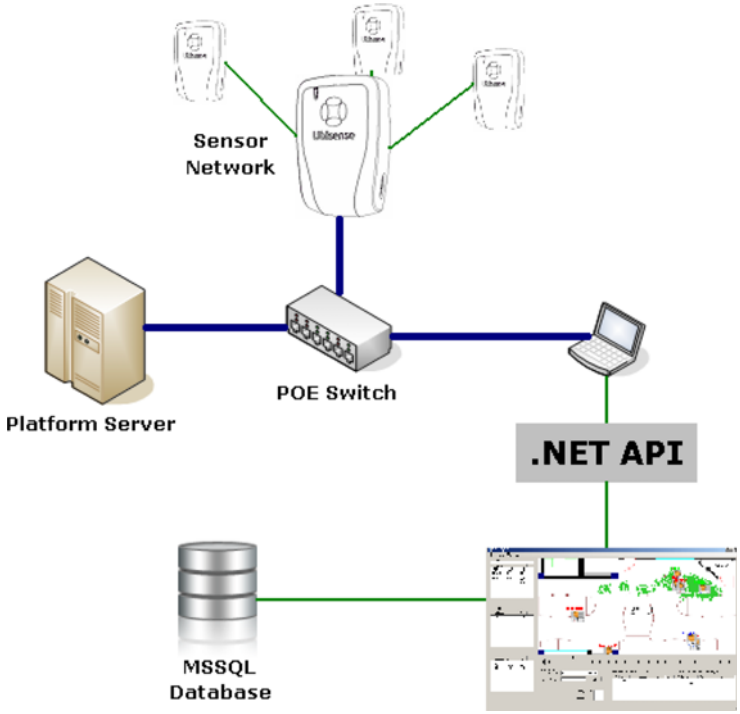
In order to evaluate the feasibility of this approach, we have implemented a prototype piece of software that takes data from a high-precision location tracking system and applies basic infection and recovery rules between the tracked entities. Our intention is to use this to demonstrate that it is feasible to infer information about infection from such location tracking data.

3.1 Overall architecture

The overall architecture of our approach can be seen in Figure 3. We use a high-precision sensor network (provided by Ubisense (http://www.ubisense.net)) to track the location of tagged individuals within a fixed environment. This network communicates location data to the central platform server over a Power-Over-Ethernet (POE) switch. Any number of client machines can be connected to the network, and receive updates from the server via

multicast IP packets. The client machines receives the location updates, which are passed through the C# .NET API of the location tracking system into our visualisation and analysis software. The information is then displayed in the application's GUI, or can be saved to a database.

Figure 3 Overall architecture of the system (see online version for colours)

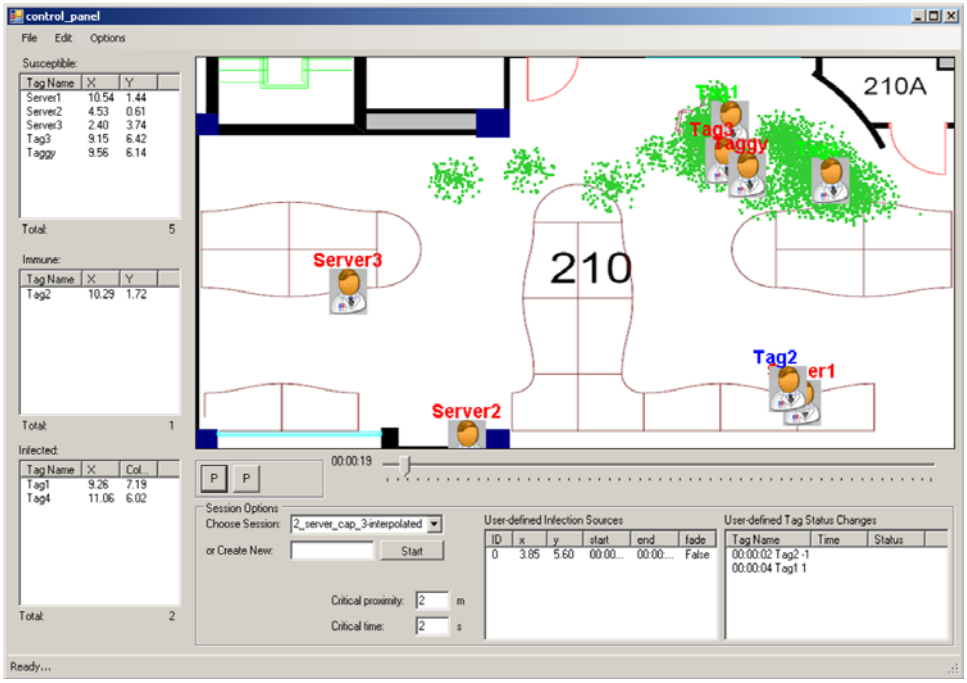


The database contains one table that contains details of sessions (a unique session id and an identifying session name), and one table that contains all of the session events. Each session event row contains the session id it relates to, a unique tag name, the x, y and z coordinates of the tag, and the standard error and time of the reading. When a session is being recorded, each location received by the program is translated into a row of this format that can later be saved to the database. This allows for easy playback of the session, as each row is read in turn and translated into a visual representation on the map.

3.2 GUI layout

As shown in Figure 4, a map sits at the top right hand corner of the GUI. Each tag is displayed on the map in the correct position for the current time, as an icon depicting a person, with its name above. The text of the name is colour-coded according to its infection status: red for susceptible, green for infectious and blue for recovered. An area in which a tag might become infected (for example the area around an infectious tag), is depicted by a cloud of green dots. The map is capable of displaying multiple tags and infection sources.

Figure 4 A screenshot of the application in use (see online version for colours)

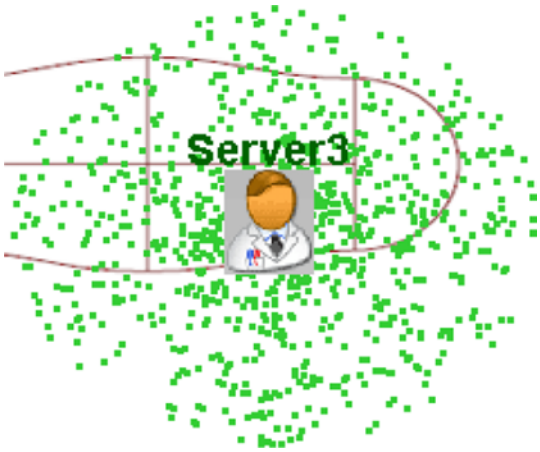


To the left of the map lie the tag status windows. Between them these contain the names of all of the tags, and also display their constantly-updated *x* and *y* coordinates. From top to bottom, the three status windows correspond to susceptible, infectious and recovered tags. As a tag's status changes, it moves to the correct window, and the window's total (displayed just below each one) is updated accordingly. This allows the user to see at a glance both the location and infection status of any given tag.

A time bar is displayed just below the map and allows the user to move both forwards and backwards freely through the session. The current time is also displayed in text form just to the left of the bar, to give a more precise view. Below the time bar are the Infection Source window, which lists each infection source added by the user, and the Tag Status Change window, which lists each change in status of a tag that is actioned by the user. Finally, there are various controls to start and stop playback and recording, and to create new, open and save sessions. These are located to the left of the time bar and in the File menu respectively.

While a session is being recorded, it is played in real time on the map. This allows the user to monitor the session, and correct any problems that might occur in the data capture such as tags switching to 'standby mode' or excessive noise in location readings caused by phenomena such as signal reflections. Infection radius is represented by a cloud of dots, as shown in Figure 5.

Figure 5 Green dots representing an infectious area centred on Server 3, whose name is coloured green to indicate that he is infectious (see online version for colours)



3.3 Representing spread of infection

The user can set four parameters when playing back a recorded location data-set, which will be used to determine the way in which infection will spread between the tracked entities:

- *Infection radius*: How close two entities have to be before there is a chance of passing an infection between them.
- *Infection persistence*: How long the area that an infected entity has been in remains infected after they have left it. The radius of a persistent infection decreases over this time until it is zero.
- *Transmission rate*: The rate at which an infectious entity will infect a susceptible entity if the latter is within the infection radius. This is different depending on the infection being studied.
- *Recovery time*: The time before an entity's immune system deals with an infection and they become 'recovered'.

Whenever the time is advanced to a point beyond the time currently being shown on the map, the program runs through each database row in turn, and performs an algorithm (detailed in Figure 6), to determine its status at that point. If the tag is deemed to be susceptible, and has been exposed to an infection source for long enough to constitute a contact, the program uses C#'s pseudo-random number generator to generate a double between 0 and 1. If this is less than the user-defined transmission rate, the tag status is set to infected.

Once a user has defined the parameters to his satisfaction, and has advanced the time bar to the desired point, he is able to plot an SIR-style graph of the session so far (as shown in Figure 7). This allows him to compare sessions where different parameters have been set easily, and also to compare the software with the results from a standard infection model.

Figure 6 Accuracy analysis – red dots denote survey positions (see online version for colours)

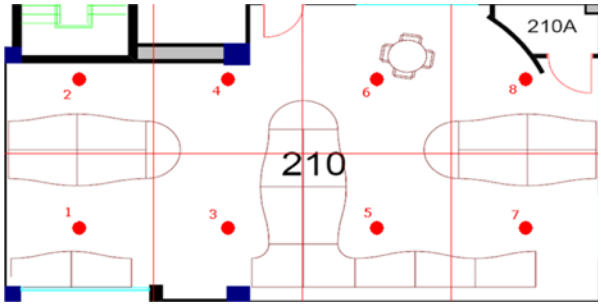
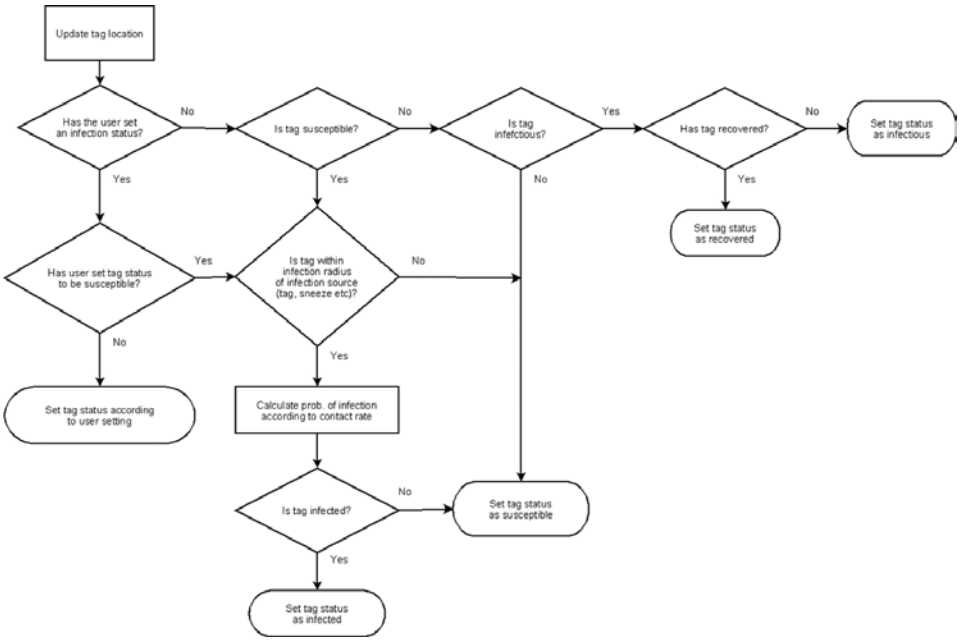


Figure 7 Algorithm to determine infection status



4 Evaluation

4.1 Accuracy of location tracking hardware

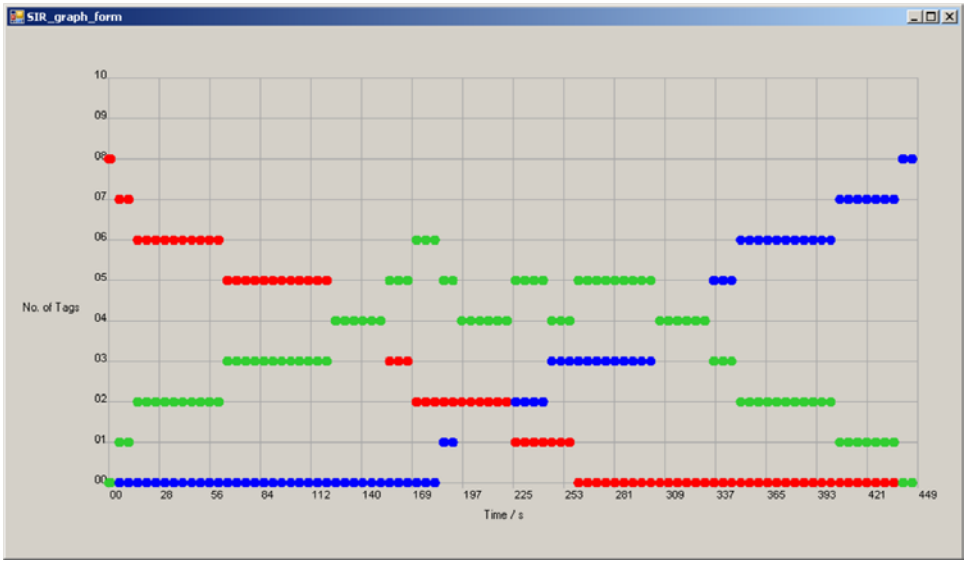
In order to obtain a quantitative analysis of the Ubisense hardware's accuracy, we set up an experiment, splitting the experimental area into eight equal sections, and placing the same tag at the same height at the centre of each section (using a laser distance measuring device to ensure accurate positioning), as shown in Figure 8.

We then recorded 40 location measurements for each location, and averaged the 40 to obtain a location reading. We calculated the difference in each reading from the expected location, and averaged the errors. The averaged errors are shown in Table 1, and from these we calculated the total average error to be 0.25 m, or 25 cm, with an estimated experimental error of 3 cm.

Table 1 Average measurement errors for the eight survey locations

Point	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Error (m)	0.077	0.608	0.111	0.184	0.138	0.253	0.331	0.287

Figure 8 An SIR-style graph generated by the software. Red: susceptible, green: infectious, blue: recovered (see online version for colours)



4.2 Effect of infection model parameters

In order to test the effect of changing the four infection model variables (infection radius, infection persistence, transmission rate and recovery time), we first collected a sample set of location data. We defined our environment generally in terms of customers and servers, but this could easily represent a specific situation such as healthcare (e.g., customers represent patients and servers represent doctors).

In our environment, there were four customers and three servers. The inter-arrival times between customers entering were sampled from an exponential distribution with $\lambda = 1$. They were then served by Server 1 for a time also determined by an exponential distribution with $\lambda = 1$, before moving on to either Server 2 or Server 3 with probability 0.7 and 0.3 respectively, and waiting there for a number of minutes determined by exponential distributions with parameters $\lambda = 1.5$ and 1.2, respectively. A typical situation during the experiment is depicted in Figure 9, where Customer *C* is being served by Server 1, and Customer *B* is being served by Server 2.

Using this data set, we ran a control experiment with the following parameters:

- the first customer was infected at time 20 s
- the infection radius was 1 m
- the infection persistence was 0 s

- the transmission rate was 0.5 infections per second
- the recovery time was 180 s.

We then generated an SIR graph for this experiment, shown in Figure 10, which shows that in total five entities (customers and servers) were infected with all recovering within 390 s.

Figure 9 A screen capture showing locations from the sample data set (see online version for colours)

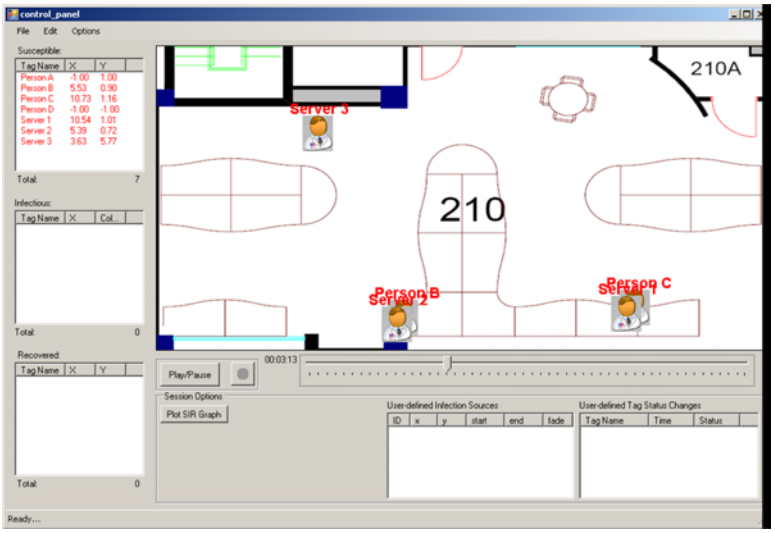
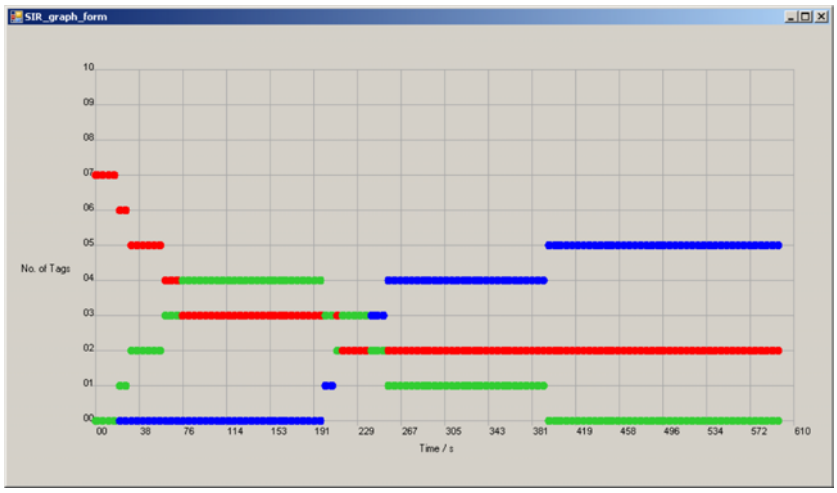


Figure 10 SIR graph for the control experiment (red: susceptible, green: infectious, blue: recovered) (see online version for colours)



We then carried out four sets of experiments, in which we varied one parameter in each set, and obtained an SIR graph for each. Varying one parameter at a time ensured that any change was due to that parameter alone, and not to a combination of changes.

However, some non-determinism was inevitably introduced by the probabilistic effect of the transmission rate.

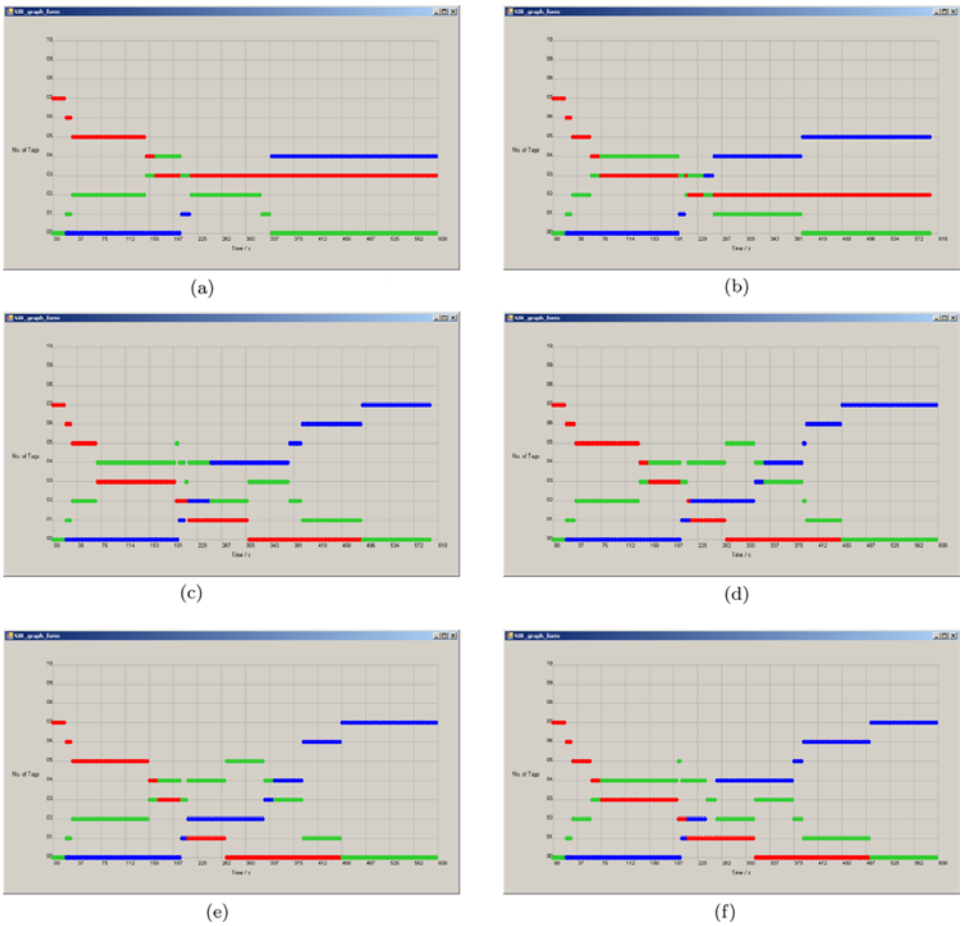
4.3 Effect of varying parameters in our simulation

We first sought to determine if the effect of varying the infection radius and persistence parameters in our simulation produced intuitive results.

4.3.1 Infection radius

We ran simulations using values of 0.5, 1.5, 2, 2.5 and 3 m for the infection radius (the value of 1 m was not used as it had already been used for the control). The SIR graphs generated from these are shown in Figure 11.

Figure 11 SIR graphs for increasing infection radius. Red: susceptible, green: infectious, blue: recovered: (a) infection radius = 0.5 m; (b) infection radius = 1m (control experiment); (c) infection radius = 1.5 m; (d) infection radius = 2 m; (e) infection radius = 2.5 m and (f) infection radius = 3 m (see online version for colours)



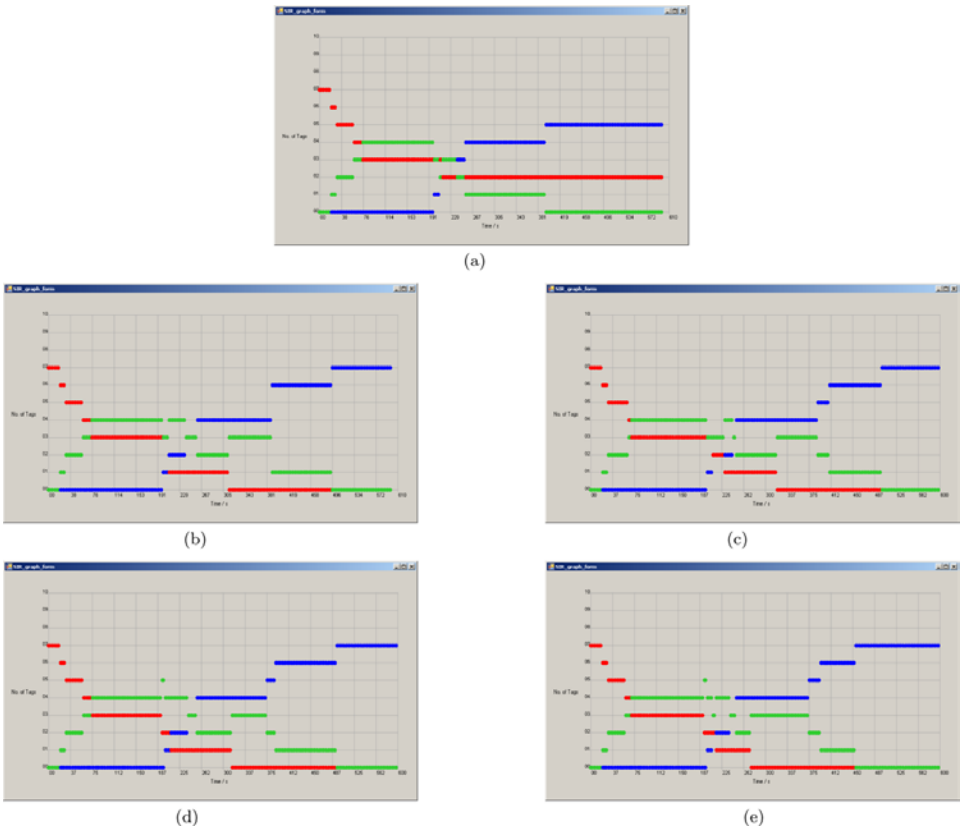
From these graphs, it is obvious that as the infection radius increases, the overall number of tags infected increases. This is particularly marked between 1 m (Figure 11(b)) and 1.5 m (Figure 11(c)), where there is a jump from five to seven tags infected. This number remains constant from 1.5 m (Figure 11(d)) to 3 m (Figure 11(f)), although the peak of the infection (the point at which the number of infectious tags is highest) occurs progressively earlier.

Although the infection radius is not a standard parameter that can be altered in the SIR equations, it is logical that as the infection radius increases, so the number of people infected will increase. In addition, it is likely that people will be infected sooner, as more fleeting contacts between individuals will become infection spreading contacts.

4.3.2 Infection persistence

To test the effect of varying the infection persistence (the amount of time that an area remains infectious after an infectious person has moved through it), we ran simulations using values of 15, 30, 45 and 60 s for the infection persistence. The resulting SIR graphs are shown in Figure 12.

Figure 12 SIR graphs for increasing infection persistence. Red: susceptible, green: infectious, blue: recovered: (a) infection persistence = 0 s (control experiment); (b) infection persistence = 15 s; (c) infection persistence = 30 s; (d) infection persistence = 45 s and (e) infection persistence = 60 s (see online version for colours)



(a)

(b)

(c)

(d)

(e)

As with the increase in infection radius, it is clear from the graphs that the total number of tags infected increases from five to seven when the infection persistence is increased from zero (comparing Figure 12(a) to (b)). However, there is little difference when the infection persistence is increased above this point. This may be attributed to the fact that there is a relatively low number of individuals in my experiment, meaning that nobody crosses the path that an infected person has taken that he would not otherwise come into contact with. This is impossible to evaluate without further study with a larger sample size.

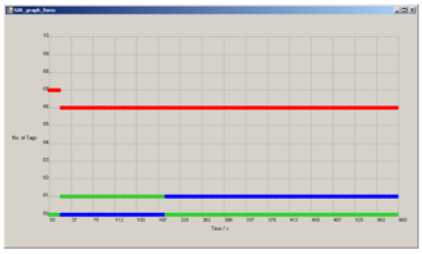
4.4 Comparison with SIR model

The final two parameters (transmission rate and recovery rate) are included in the standard SIR model. This means that the experiments where we varied these factors serve two purposes: as above, they can be used directly to observe the results of varying the parameters, but they can also be used to compare the effects against a standard SIR graph that has been generated using the system of differential equations stated in Section 2.

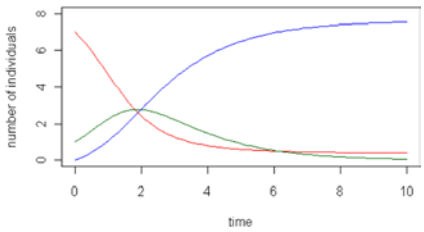
4.4.1 Transmission rate

We ran three simulations in which the transmission rate was increased in increments of 0.25 infections per second. The results are shown in Figure 13, where they are compared to plots of the corresponding differential equation systems. We generated these plots using the statistical package R (R Development Core Team, 2008) and an SIR modelling module (Debarre, 2010).

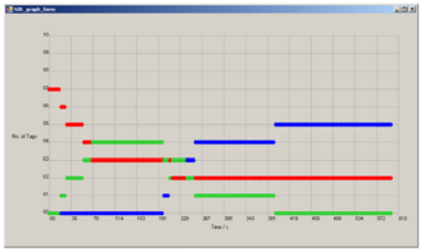
Figure 13 SIR graphs for increasing transmission rate: (a) transmission rate = 0.25 – test data; (b) transmission rate = 0.25 – calculated values; (c) transmission rate = 0.5 (control experiment) – test data; (d) transmission rate = 0.5 – calculated values; (e) transmission rate = 0.75 – test data; (f) transmission rate = 0.75 – calculated values; (g) transmission rate = 1.0 – test data and (h) transmission rate = 1.0 – calculated values (see online version for colours)



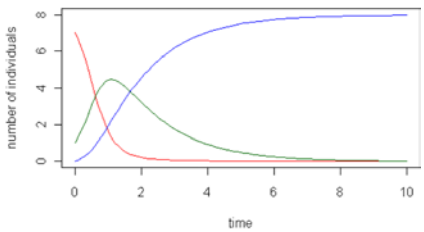
(a)



(b)

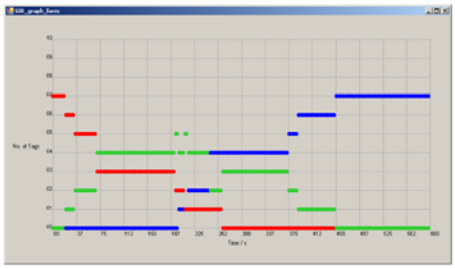


(c)

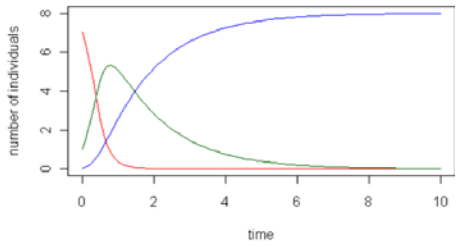


(d)

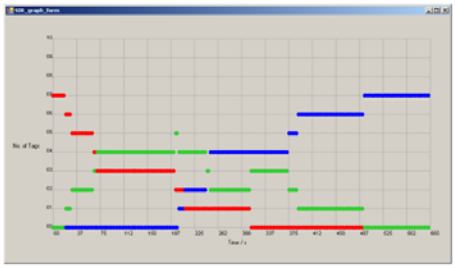
Figure 13 SIR graphs for increasing transmission rate: (a) transmission rate = 0.25 – test data; (b) transmission rate = 0.25 – calculated values; (c) transmission rate = 0.5 (control experiment) – test data; (d) transmission rate = 0.5 – calculated values; (e) transmission rate = 0.75 – test data; (f) transmission rate = 0.75 – calculated values; (g) transmission rate = 1.0 – test data and (h) transmission rate = 1.0 – calculated values (see online version for colours) (continued)



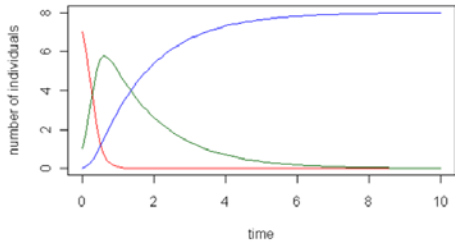
(e)



(f)



(g)



(h)

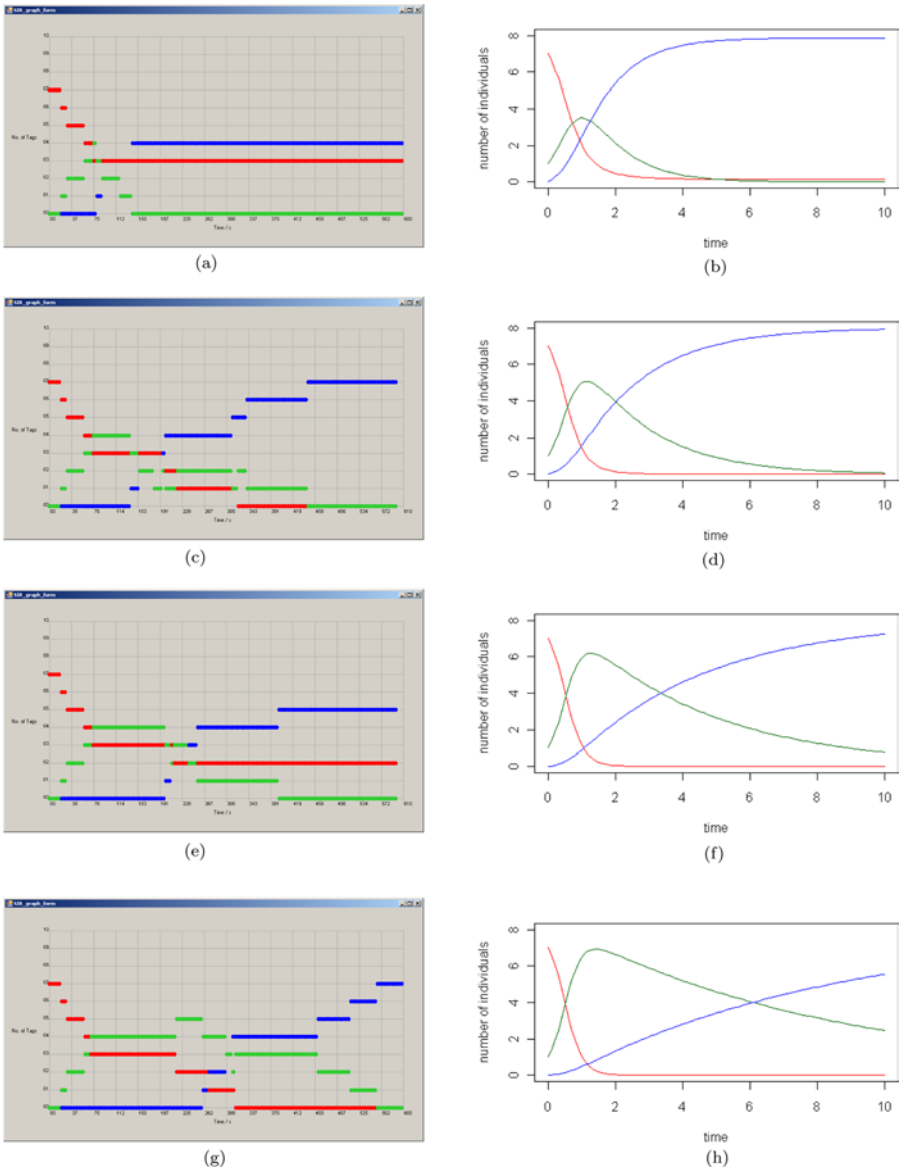
Looking at the graphs generated from the SIR equations, two effects are apparent when the transmission rate is increased. The first is that the peak number of infectious individuals is higher. The second is that this peak is reached sooner the higher the transmission rate. The first of these effects can be observed in the experiment data, although not as clearly as in the calculated graphs. Between Figure 13(a) and (c), the peak number of infectious tags increases from one to five. However, it appears that there is little difference in the time that it takes for these peak values to be reached. This discrepancy from the calculated graphs is likely to be due to the necessarily small sample size, and discrete nature, of the experiment data. However, it could also be indicative of a problem with the design of the software. As with the infection persistence experiment, further study is needed to determine the cause.

4.4.2 Recovery rate

Due to the nature of the way the location data are recorded, the recovery rate set in the program does not correspond with that used in the actual SIR equations. The former is defined as a number of seconds, while the latter is expressed as a probability. The reason for this discrepancy is that the experimental data are discrete (i.e., they are recorded at discrete time intervals, the periodicity of which varies due to the location tracking hardware), which makes it extremely difficult to calculate a continuous probability. However, broadly speaking, an increase in recovery time should correspond to a decrease in the standard SIR model recovery rate. Therefore, in order to compare this set of experiments to a calculated SIR graph, we estimated the recovery rate and

attempted to look for broad trends rather than a precise comparison. We ran simulations for recovery times of 60, 120, 180 and 240 s, the results of which can be seen in Figure 14.

Figure 14 SIR graphs for increasing recovery time: (a) recovery time = 60 – test data; (b) recovery rate = 1 – calculated values; (c) recovery time = 120 – test data; (d) recovery rate = 0.5 – calculated values; (e) recovery time = 180 (control experiment) – test data; (f) recovery rate = 0.25 – calculated values; (g) recovery time = 240 – test data and (h) recovery rate = 0.125 – calculated values (see online version for colours)



Two features of the calculated graphs are apparent. The first is that the gradient of the recovered line decreases as the recovery rate decreases (and therefore the recovery

time increases). The second is that the peak number of infectious individuals increases as the recovery rate decreases, and after it has peaked, its gradient is less. Both of these trends can be seen in the experimental data, although the second more clearly than the first. The graphs show that as the recovery time increases, the initial gradient of the recovered line decreases. However, once tags have started to recover, the gradient increases, and is eventually very similar in all of the graphs. The peak number of infectious tags rises from four to six, and the gradient of the infectious line after its peak clearly decreases with increasing recovery time. The discrepancy between the gradient of the recovered line in the test data and calculated values could again be attributed to the small sample size, but again needs further study. However, the decrease in gradient of the infectious line shows a good correlation with the calculated data.

5 Conclusion

As a proof of concept, the combination of the high precision location tracking hardware and the software that we have designed is a success. Our experiments show that, in a test environment at least, the hardware is capable of producing location readings that are sufficiently accurate to monitor the movement of individuals, to the extent that a contact tracing study that provides meaningful results can be performed. Given the small-scale nature of the test environment, the fact that any results that mimic real-life models of infections can be obtained at all is very promising.

Given the relative youth of the technology currently used, it is also likely that the accuracy of location tracking information will increase as more progress is made in the field. This can only help to improve the reliability of such models in the future. Likewise, as the concept of using location tracking data as a basis for contact tracing is relatively new, and we have designed the algorithms that determine the infection status of individuals from scratch, these can undoubtedly be improved and expanded upon, with the result that they will more closely match the known behaviour of infections. It is hoped that such a program could even help epidemiologists to learn more about the nature of infections, and how they spread.

There is a large amount of further work to be done regarding the use of location tracking in infection modelling. Our software is largely a proof of concept, and requires much more extensive testing than we have been able to perform in the lab with a limited number of tags. Ideally, this would involve covering a hospital with a multi-cell high-precision location tracking network, and capturing data over a long period. This could then be used as a basis for improving the software and the algorithms used in it. However, the logistics and cost of doing this at present are considerable, and it may be some time before the hardware is mature enough to report accurately the amount of data that would be required for a large-scale study.

It would be interesting to attempt to estimate the parameters for an epidemic model from interactions observed in location tracking data. It should also be possible to extend the existing SIR model to include more variables. The Susceptible, Exposed, Infectious, Recovered (SEIR) model, for example, includes an incubation period, where an individual has been infected but is not yet infectious. It should also be possible to include births and deaths, which would of course lead to a variable population size. These extra variables should be easy to add to the program, but each requires testing to ensure that it is implemented correctly.

It would also be useful to use the location tracking data to produce an ordered list of the individuals in a given environment according to how many contacts they have with others. This could provide invaluable information for doctors, allowing them to target immunisation and treatment at those who interact with a large number of people. It should also be possible to produce a contact graph, such as shown in Figure 2, listing every contact between every monitored individual in a certain period of time.

Aside from the obvious applications in the healthcare sector, there are undoubtedly other uses to which software of this kind could be put. It essentially tracks social interactions, and as such could well have implications in the study of psychology. For example, the exchange of information between people can be analogised to the spread of an infection.

References

Becker, N.G. (1993) 'Parametric inference for epidemic models', *Mathematical Biosciences*, Vol. 117, Nos. 1–2, pp.239–251.

Choi, S., Yun, K. and Kim, D. (2006) 'A robust location tracking using ubiquitous RFID wireless network', *Ubiquitous Intelligence and Computing, Lecture Notes in Computer Science*, Vol. 4159, pp.113–124.

Debarre, F. (2010) *SIR Models of Epidemics*, Website: http://www.tb.ethz.ch/education/model/SIR, Retrieved on 10/09/2008.

Gibson, G.J. (1997) 'Markov Chain Monte Carlo methods for fitting spatiotemporal stochastic models in plant epidemiology', *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Vol. 46, No. 2, pp.215–233.

Gibson, G.J. and Renshaw, E. (1998) 'Estimating parameters in stochastic compartmental models using Markov chain methods', *IMA Journals of Mathematics Applied in Medicine and Biology*, Vol. 15, No. 1, pp.19–40.

Huerta, R. and Tsimring, L.S. (2002) 'Contact tracing and epidemics control in social networks', *Physical Review E*, Vol. 66, No. 5, pp.056115.1–056115.4.

Kermack, W.O. and McKendrick, A.G. (1927) 'A contribution to the mathematical theory of epidemics', *Proceedings of the Royal Society of London, Series A*, Vol. 115, pp.700–721.

Kretzschmar, M., Muller, J. and Dietz, K. (2000) 'Contact tracing in stochastic and deterministic epidemic models', *Math. Biosci*., Vol. 164, pp.39–64.

O'Neill, P.D. and Roberts, G.O. (1999) 'Bayesian inference for partially observed stochastic epidemics', *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, Vol. 162, No. 1, pp.121–129.

R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Streftaris, G. and Gibson, G.J. (2002) 'Statistical inference for stochastic epidemic models', *Proc. 17th International Workshop on Statistical Modelling*, Chania, Crete, pp.609–616.

Streftaris, G. and Gibson, G.J. (2004) 'Bayesian inference for stochastic epidemics in closed populations', *Statistical Modelling*, Vol. 4, No. 1, pp.63–75.

Strom, S.R. (2008) *Charting a Course Toward Global Navigation. The Aerospace Corporation*: http://www.aero.org/publications/crosslink/summer2002/01.html, Retrieved on 02/09/2008.

Notes

[1]http://en.wikipedia.org/wiki/Image:Sirsys-p9.png

[2]http://www.visualcomplexity.com/vc/images/28big01.jpg

Websites

Demonstrations of Ubisense Solutions in Action, Website: http://www.ubisense.net/index.php?load=content&pageid=80, Retrieved on 14/07/2008.

Ekahau – Wi-Fi Based Real-Time Tracking and Site Survey Solutions, Website: http://www.ekahau.com/?id=1010, Retrieved on 14/07/2008.

Ubisense, Website: http://www.ubisense.net, Retrieved on 14/07/2008.