



Australian Government
Department of Defence
Defence Science and
Technology Organisation

Comparison of an Agent-based Model of Disease Propagation with the Generalised SIR Epidemic Model

Russell Connell, Peter Dawson and Alex Skvortsov**

Air Operations Division
***Human Protection and Performance Division**
Defence Science and Technology Organisation

DSTO-TR-2342

ABSTRACT

Social contacts are an important channel for the propagation of disease through a population and should be considered in conjunction with traditional epidemic diffusion. Such channels should always be taken into account for a realistic estimation of a long-term impact of a disease outbreak (natural or malicious) and for the best response options. This paper describes our recent experience in developing a simple agent-based model to simulate disease propagation through a social network and validating the results of the agent-based simulation by reconciling it with a well-known mathematical model.

RELEASE LIMITATION

Approved for public release

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE AUG 2009		2. REPORT TYPE		3. DATES COVERED	
4. TITLE AND SUBTITLE Comparison of an Agent-based Model of Disease Propagation with the Generalised SIR Epidemic Model				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) DSTO, , , ,				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.					
13. SUPPLEMENTARY NOTES The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 31	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Published by

*Air Operations Division
DSTO Defence Science and Technology Organisation
506 Lorimer St
Fishermans Bend, Victoria 3207 Australia*

*Telephone: (03) 9626 7000
Fax: (03) 9626 7999*

*© Commonwealth of Australia 2009
AR-014-617
August 2009*

APPROVED FOR PUBLIC RELEASE

Comparison of an Agent-based Model of Disease Propagation with the Generalised SIR Epidemic Model

Executive Summary

There are currently three main approaches to epidemiological modeling: mathematical, network theory (when social interactions are modeled based on a network theory approach) and, closely related to this, agent-based computer models. Agent based and network based approaches are complimentary to each other.

The mathematical approach dates back to the SIR model (*S*-susceptible, *I*-Infected, *R*-Recovered) and its further modifications by Kermack & McKendrick in 1927 and Andersen & May in 1979. This approach provides rigorous results and is the simplest to implement, but has an obvious shortcoming in that only simplified scenarios can be treated analytically. Agent-based simulation is the most flexible in terms of realistic scenario evaluation and has become increasingly popular.

With the increasing availability of computer resources agent based methods allow high fidelity modeling of epidemical outbreaks on global, national and community levels. The main disadvantage of an agent based approach is model validation, in particular what is the accuracy of the model output for a given 'what if' scenario (which has never occurred)? Therefore one of the important steps of agent-based model validation is so-called "model alignment" (see Chen *et al* 2004), when the agent-based model output is reconciled with other modeling approaches for realistic (observable) values of model parameters.

This paper describes our recent experience in developing an agent-based disease propagation model to simulate an epidemic outbreak and comparing the results of the agent-based simulation with a mathematical model, SIR, as the first step in validating the agent based model. The guiding principle when designing this model was to create a research tool that would allow quantitative studies (sensitivity analysis, data assimilation, reverse problems) as well as ad-hoc operational scenarios based on a small-scale agent based model (that can run on a PC) but with real census data.

The results of this comparison showed that the epidemiological model developed within the agent based model is capable of representing modelling of disease spread in a realistic social network. By showing an alignment of the output of the agent based model and the SIR model we have shown that it is possible to develop a realistic disease spread model in a complex multi-agent social context (including alignment of model parameters, scenarios and underlying assumptions). This validation in future will be expanded to comparison with real epidemic data.

The results of this work will be used in the DSTO Corporate Enabling Research Program (CERP): Bio-Terrorism Preparedness, A Detection to Treat System for Biological Weapons Attacks on a Community.

Authors

Russell Connell

Air Operations Division

Russell Connell is Simulation Researcher within Air Operations Research Branch specialising in civilian models. Russell holds a Bachelors Degree in Computer Science and Software Engineering(1999) from Swinburne University, Australia

Peter Dawson

Human Protection and Performance Division

Peter Dawson is a Research Scientist in the Mathematical Analysis and Physical Sciences Capability, specialising in Chemical, Biological and Radiological Threats. Peter has a PhD in Theoretical Physics (2003) from Adelaide University, Australia

Alex Skvortov

Human Protection and Performance Division

Alex Skvortov holds a PhD from Moscow University of Technology in Applied Mathematics (Fluid Dynamics). During his early career he was involved in R&D projects on ocean acoustics and vortex noise modelling. Following a few years in academia, Alex worked in industry as an IT Consultant in the areas of Mathematical Modelling, Data Warehousing and Decision Support Systems. Alex joined DSTO in 2005. He is currently a Science Team Leader in the Mathematical, Analysis and Physical Sciences Group, Human Protection and Performance Division.

Contents

1. INTRODUCTION.....	1
1.1 Case Study	1
1.2 Models.....	1
1.2.1 Background to CROWD	2
1.3 Report Structure	3
 2. AGENT-BASED MODELLING: AN EPIDEMIOLOGICAL CASE STUDY	 4
2.1 Modern Approaches to Epidemiological Modelling	4
2.2 SIR Mathematical Model.....	5
2.3 Agent-based Epidemic Case Study.....	6
2.4 Validating Results with the SIR Mathematical Model	8
2.5 Analysis of CROWD Epidemic Data Using Generalised SIR.....	12
2.6 Main Results	13
 3. THE DISEASE MODEL IN DETAIL.....	 14
3.1.1 Disease XML schema and example: Influenza	18
 4. DISCUSSION	 19
4.1 Alternative Agent-based disease models: EpiSims and BIOWar	19
4.1.1 EpiSims	20
4.1.2 BIOWar	20
4.1.3 Comparison with CROWD	20
4.2 Further Work.....	21
 5. CONCLUSION	 21
 6. RECOMMENDATIONS.....	 22
 7. ACKNOWLEDGEMENTS	 22
 8. REFERENCES	 22

1. Introduction

This document reports on the comparison of an Australian agent-based disease propagation model with the generalised SIR [1] (Susceptible – Infectious – Recovered) epidemic model. This case study uses a disease model built upon the agent-based framework provided by the simulation CROWD, developed by DSTO for use in the modelling of civilians within complex urban environments. We provide an overview of CROWD and elaborate on the validation of the agent-based disease model.

1.1 Case Study

Investigations into epidemiology by Human Protection and Performance Division (HPPD) have led to the search for simple tools to explore mitigation strategies to stem or stop the spread of epidemics. A mathematical model of epidemics called SIR (described in section 2.2) was the initial basis of their studies. Agent-based models were considered as a means of exploring the problem space and this case study comparing an agent-based disease model with SIR was initiated.

The agent-based model utilises census data to build a social network, based on family groups and work colleagues. The population generated by the framework then interact in a daily travel rhythm from home to work. This has the effect of changing the epidemic transmission paths over time. A worker could contract a disease from a colleague, give it to their family when they arrive home, and they in turn would spread it onwards the next day.

The aim of the case study is to determine if an agent-based model that is structured to use the same variables that are incorporated within SIR (namely population size, infection probability and probability of recovery per time step) would provide similar results and what measures would be necessary to validate the agent-based model against SIR.

1.2 Models

Once a baseline validation has been undertaken using a variant of the agent-based model that replicates the underlying assumptions of the SIR model, such as a uniformly mixed population, tests would be done to see how the non-uniform mixing of the normal agent-based model affects the results. The major expected difference between the agent-based model and SIR was that the sporadic contact network would impede the speed of the epidemic.

SIR is a dynamical system that models epidemics over time. The acronym SIR stands for Susceptible, Infected and Recovered, which are the three states an individual can be in within the model. The model has been known and used for a long time, going back to the original publication of W.O. Kermack, A.G. McKendrick [8]. The advantage is the clarity of the mathematical framework and simplicity in implementation [2]. The basic *SIR* model is represented by a nonlinear dynamical system of three differential equations whose two main parameters are:

- A. The probability of a specific infectious individual infecting a specific susceptible individual should they meet, multiplied by the chance they will meet per unit time.
- B. The probability an infected person will recover per time unit.

CROWD is an agent-based model built upon a complex urban environment representation. The validation of the disease model within CROWD against the industry standard SIR (see section 2) model, and the inherent extensibility of CROWD, permits the research program to begin investigating more complex social structures and the effect of a larger range of individual behaviours on the behaviour of epidemics. From the effect of individuals either staying at home or flooding hospitals to how diseases spread through crowds of intermingled people in locations such as shopping centres or sporting events. For technical detail of CROWD's disease model the reader is referred to Section 3.

1.2.1 Background to CROWD

In 1995 DSTO acquired the rights to use of the Close Action Environment (CAEN) simulator from DERA (UK). CAEN is both a wargame and a simulator that allows multi-sided multiplayer wargaming. CAEN's primary focus is close combat infantry battles and as such it models up to 500 individual entities, vehicles and infantry, on a raster terrain map. When first acquired, CAEN's urban model was rudimentary utilising 10 metre square buildings with the outcome of in-building combat resolved using a modified Lanchester equation. This limited the usefulness of CAEN in modelling the MOUT¹ environment and scenarios were limited to low-resolution inter-building scenarios, as demonstrated in studies supporting Restructuring The Army (RTA) Phase 1. [3]

At the same time as the acquisition of CAEN Land Operations Division (LOD) began experimenting with Intelligent Agent Technologies to deliver a wider variety of tactics to the entities within their wargames. The first technology to be evaluated was dMARS [4], which was employed to explore tactical, environmentally-aware movement in CAEN. This was followed by JACK [5] agents that autonomously planned and conducted a company sized attack on an enemy platoon within the OTBSaf² wargame [6].

Towards the end of the 1990's, with activities such as the USMC Urban Warrior experiments and the enunciation of the Three Block War [7] concept, it became obvious that Urban Combat simulation was an important requirement for modelling future operations. Tools available to DSTO at the time were unable to model urban environments at a level of fidelity high enough to permit detailed urban tactical studies. This deficiency led to the development of the Advanced Urban Environment (AUE). AUE is a vector based model of the physical structures that make up an urban environment. It includes high detail buildings incorporating internal and external features, terrain, trees and other vegetation, roads and infrastructure. Urban classification studies were done using AUE as the environment representation to explore the effects of the limited line of sight within cities on the tactical awareness of troops [8]. As well as providing a representation of the physical environment AUE was designed to provide a symbolic representation of the environment that could be used by agents to intelligently move around the landscape in a tactically aware manner.

¹ MOUT : Military Operations on Urban Terrain

² OTBSaf: OneSAF Test Bed- Semi Automated Forces, a US brigade level wargame/simulation

As complex as urban battle-scapes are they are merely more confined versions of other types of complex terrain such as jungles and mountains. The missing element for true realism is the indigenous populations of the cities and the level of chaos they bring to urban combat.

CROWD was begun as an answer to both the above observations and the Australian Army's "Future Land Warfighting Concept – Complex Warfare" [9] document that highlighted three inter-related complex terrains: physical, human and information. CROWD utilises the complex physical terrain provided by AUE and layers the human and informational environments over the top. This provides us with an urban model that is capable of representing the levels of complexity and chaos inherent in a populated urban environment.

1.3 Report Structure

This report begins with a description of the current state of Epidemiological modelling and then describes the SIR epidemic model in particular. The case study is outlined noting in particular how the two models, SIR and CROWD, are aligned. The results of the case study and further research are discussed.

We describe in technical detail how a disease and its propagation are modelled within CROWD. An XML Schema used to define new diseases is also presented.

In the next section we draw comparisons between CROWD and two other well known epidemiological models, BIOWar and EpiSims, and go on to highlight some of the important differences between the models.

We conclude with a description of the current state of our modelling, some thoughts for future directions as well as conclusions on the utility of the CROWD disease model.

2. Agent-based Modelling: An Epidemiological Case Study

2.1 Modern Approaches to Epidemiological Modelling

There are a number of approaches to Epidemiological Modelling (i.e. to the modelling of a disease spread through a population). The traditional well-known approach is to model the population and disease spread via a dynamical system, where the dynamics of the infected population is described with a system of coupled non-linear differential equations. Such a system can provide a simplistic model of social interaction and mixing of individuals and, as a result, an evolution of an infectious state of the whole population. The parameters of these dynamical systems ('population models') are usually provided from historical observations (data fitting) or from physics-based reasoning (population fluxes, rate of contacts etc). This approach and the SIR model have been known for a long time, going back to the original publication of W.O. Kermack, A.G. McKendrick [10] in 1927. Its sound advantage is the clarity of its mathematical framework and simplicity in implementation [11].

The second approach to Epidemiological Modelling is to model an underlying social network and population interaction with the modern methods of Network Theory [12]. Although Network Models still require some fitting parameters, they generally provide broader models of epidemic outbreaks than simple equation-based models (i.e. they have richer and more flexible capability for scenario representations) [12]. For example, from the Network Theory point of view an epidemic outbreak simply corresponds to a percolation threshold on an underlying social network. Such an approach to Epidemiological Modelling provides a new theoretical foundation for consistent data analysis, parameter estimation, modelling and forecasting. Finally, with the increasing availability of computer resources agent-based simulation has become a practical method for conducting Epidemiological Modelling. In the agent-based approach the whole township can be modelled as a system of software agents interacting in time and space in according with the prescribed rules. Agent technologies can provide individual citizens with motivations, opinions and biases or work on aggregations of population. It becomes possible to model the entire population in a plausible manner, providing a model of community that acts as if it inhabits the city or the whole country [13,14,15]. The development of such an infrastructure is a critical step in the realistic simulation of complex social network dynamics and its response to epidemiological events.

It is important to recognise that disease spread through population is controlled by two groups of parameters: the disease properties (infection and incubation period, symptom timing, infection mechanisms) and the structure of the social network of the population (contact rate, social clustering, and migration). These groups of parameters are completely unrelated and can be studied independently. For instance, the structure of the social network of the community can be studied long before influenza outbreak occurs within it (if ever). Only by combining together a given disease and a given social network might an epidemic be created, but this depends on threshold conditions in the relationship between the network connectivity and infection and recovery rates (see the SIR model: Section 2.2), otherwise the disease dies out.

The critical factor for a high fidelity Epidemiological Model is the ability to independently validate its predictive results. It is often very difficult due to a lack of reliable field data, as the simulated event has never occurred, or based on ethical grounds to conduct such a validation. The logical choice of validation techniques in such situations is to use cross-validation, to run a validated model for some simplified scenarios (where the result is known or obvious) or to compare its output with other available models that have been validated (so called model alignment [13, 16, 17, 18]). The model alignment method provides a rather simple yet effective way to ensure model plausibility (no noticeable flaws in functionality), before it can be moved to the parameter tuning stage.

In this report we demonstrate an application of such a technique to the validation of the agent-based disease propagation model CROWD against the dynamical system model SIR.

2.2 SIR Mathematical Model

The basic *SIR* model can be represented by a nonlinear dynamical system of three differential equations [2]:

S = Number of Susceptible Individuals
 I = Number of Infected/Infectious Individuals
 R = Number of Recovered Individuals

α = probability of a specific member of I infecting a specific member of S should they meet, multiplied by the chance they will meet per unit time
 β = the probability an infected person will recover per time unit

$$\frac{d S}{d t} = -\alpha S I \quad (1)$$

$$\frac{d I}{d t} = \alpha S I - \beta I \quad (2)$$

$$\frac{d R}{d t} = \beta I \quad (3)$$

Figure 1: The three nonlinear equations that comprise the SIR model

These equations describe the time evolution of a population between disease states *Susceptible*, *Infected* and *Recovered*. The parameter α is the probability of a specific member of *I* infecting a specific member of *S* should they meet, multiplied by the chance they will meet per unit time. The parameter β is the probability an infected person will recover per time unit. For the extreme scenario $S = 0$ (no susceptible people left) we would have a monotonic exponential decay of $I = I_0 e^{-\beta t}$.

The threshold condition for epidemics can be derived from the second equation of the SIR system (positive rate of infected people) and can be written as $S(t = 0) > \beta / \alpha$.

The SIR system was initially developed based on a very simplified model of social interaction. For instance an assumption of uniform population mixing was implicitly used to deduce the nonlinear terms at the right hand side of the SIR system. It has been a well- recognised weakness of the SIR model and has been a base for its ongoing criticisms. This issue was recently addressed in a paper [19] where an important generalisation was made to extend the standard SIR model to the non-uniform mixing case:

$$\frac{d S}{d t} = - \alpha S^{\delta} I \quad (4)$$

$$\frac{d I}{d t} = \alpha S^{\delta} I - \beta I \quad (5)$$

$$\frac{d R}{d t} = \beta I \quad (6)$$

where $\delta \approx 2$ (estimated value (5)) is a new parameter describing an efficiency of social mixing. The case $\delta \approx 1$ corresponds to a uniform mixing, as in the standard SIR model. Since the extended SIR model has an extra fitting parameter (δ), it can be better reconciled with field observations and can then be aligned with the two major observable values: time of epidemics and its amplitude.[19, 20]

In section 3.5 the results of applying equations (4)-(6) to CROWD's output are discussed.

2.3 Agent-based Epidemic Case Study

In previous case studies [16, 17, 18] we modelled a rather simple scenario with CROWD to demonstrate its modelling capabilities in the epidemiological domain so only the standard SIR model was used to cross-validate CROWD output. More complicated scenarios and corresponding CROWD alignment with the extended SIR model will be published elsewhere.

A small Australian town, population just over 3000, was built from Australian Census Bureau data (see [16, 17, 18]). The data used included the age/gender break-down and family-household-workplace makeup. The model generates a population based on this data and then builds families, households and work places based upon the census data. An initialisation file is used to determine the types of businesses within the town as well as the number of employees and hours worked. Town cadastral town data is used to match businesses and residences to brick and mortar buildings. The generated families and workforce is then randomly distributed amongst the residential and business buildings. For the purposes of this study such a distribution was adequate, in future studies other methods will be explored. The population then moves around town according to the basic daily rhythm of traveling to and from home and the workplace. At each location a different set of people come into contact, and this set changes as different shifts start and finish work, and make their way home to their families. The contacts derived from these rhythms are used to drive the disease model. The disease model is derived as a Finite State Machine with a probability of moving from one disease state to another as a result of the social contacts (S-susceptible, I-infected, R-recovered). The structure of social network generated by CROWD is presented in Figure 2.

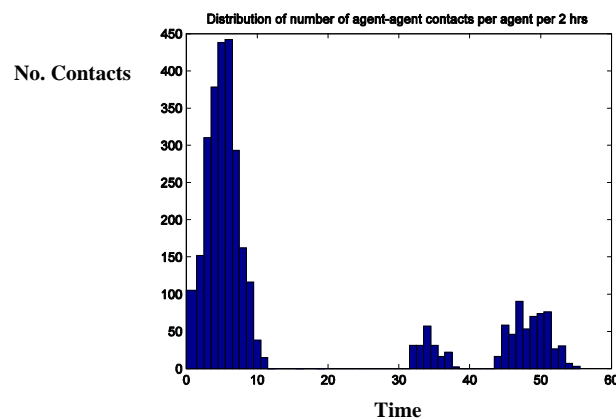


Figure 2: Plot of distribution of the number of contacts per 2 hr period per agent

The peaks to the right correspond to the occupants, students and teachers, within the three schools of the town, which have significantly greater numbers of co-located individuals than other locations.

An artificial epidemic spread was created in this township within the CROWD model, where each infected agent has a probability $P_1 = 7.1 \times 10^{-4}$ of infecting uninfected agents it meets, and a probability $P_2 = 0.9959$ of staying infected per two hour period (giving a half life of 24 hours). The agents then spread the infection through the simulated town. The resulting graph of number of people in SIR states is given in Figure 3. As can be seen from this graph, at first the disease spreads quickly through the susceptible populace, however, as more people become infected, the availability of susceptible people drops, making it less likely for those infected to pass the disease on. Thus after a week, the epidemic dies out.

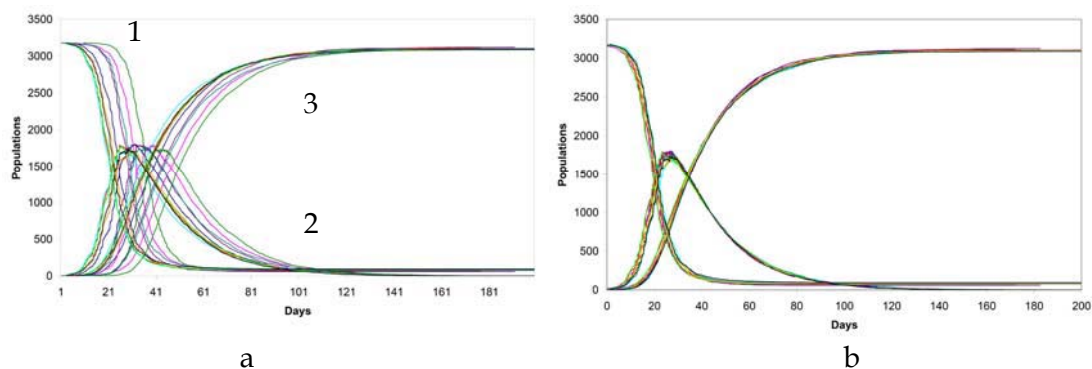


Figure 3: CROWD simulation epidemic data: a) results of 10 runs. Group 1, 2 and 3 lines represent the susceptible, infected and recovered populations respectively. Peak time differing due to amplification of effect of statistical variation when I is small, b) same data compensated for statistical variation in time to show shape is conserved.

Of particular interest is the variation in timing of the epidemics in each of the 10 runs. The epidemic is a slowly spreading one. For a significant time, I stays small as there are few people to pass the disease, and each person recovering at this early stage has a large proportional

effect on the size of I . It is not until $I \gg 1$ that the epidemic takes off quickly, with many people to spread it. At this stage statistical variation in infections and recoveries makes little difference and the course of the epidemic is constant for each run to within a few percent (see Figure 3b). The time to peak I is 32 ± 6 days, with the peak percentage of population infected $55 \pm 1\%$. The final number of uninfected people is 80 ± 10 .

In Figure 4, the CROWD output data has been used to create a map of where infections occur within the virtual town. Here the radii of the red circles are proportional to the number of people infected at a site. The three large infection zones are the three schools.

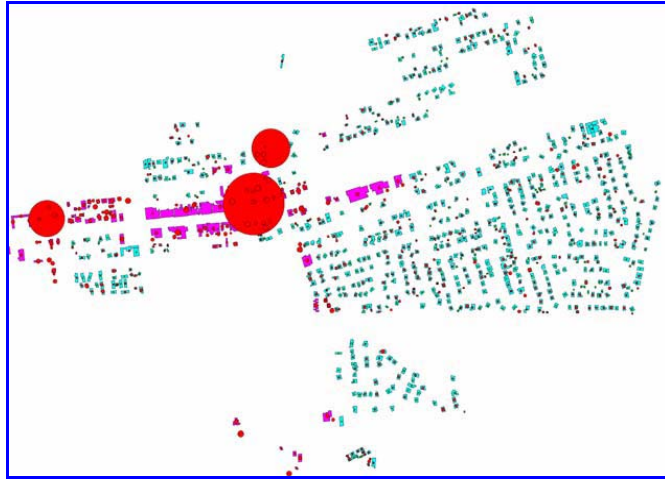


Figure 4: Map of where infections occur. Larger circles correspond to more infections. Light blue buildings are residential, pink buildings are businesses. Major infection sites correspond to the 3 schools.

2.4 Validating Results with the SIR Mathematical Model

As discussed earlier, it is critical for a high fidelity epidemiological model to be able to independently validate its predictive results. Here the CROWD epidemiological model is compared to the SIR mathematical model of section 2.2.

In order to adequately compare CROWD results to the SIR model, β can be set to reflect the same half life time of 14 days used to generate the CROWD data in Fig. 2.

$$(7) \quad \beta = -\frac{\log(.5)}{14 \times 24} = .0021 / hr$$

Then by using the following function derived from (1) and (2) (see [2]):

$$I(t) - I(0) = \frac{\beta}{\alpha} \log\left(\frac{S(t)}{S(0)}\right) - (S(t) - S(0))$$

and using the data depicted by Fig. 13 to assign the start ($t = 0$) and end ($t = \infty$) values of S and I , we can find the ratio β / α . This provides a value $\alpha = 2.46 \times 10^{-6}$ per head per hr.

Evaluating the SIR model for these values of parameters α, β (done in Mathematical) leads to the plot in Fig. 15. As can be seen, the general behaviour of the two models is the same, with an epidemic lasting multiple months, a sharp drop in S followed by a levelling out as I dies out, unable to sustain itself with the reduced S , and an increase in R as the sick recover. The same general conclusion holds for different values of parameters α, β and this is a strong indication of CROWD fidelity.

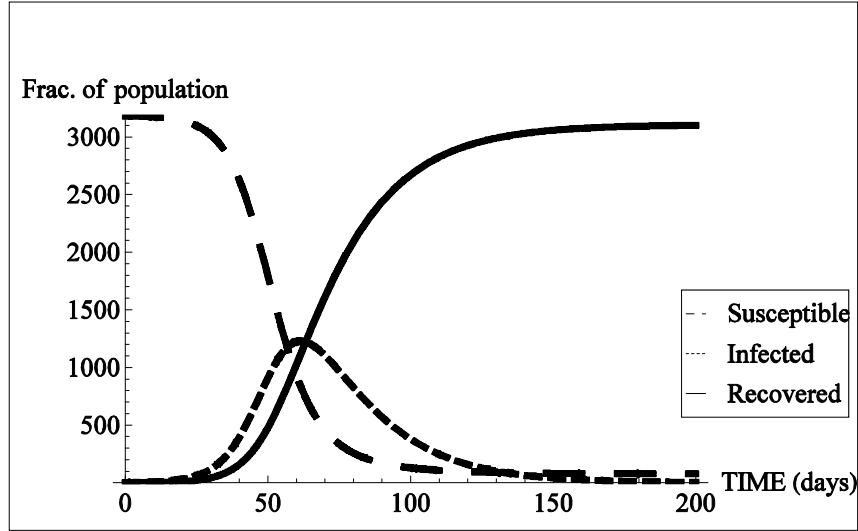


Figure 5: SIR model epidemic state evolution, matching recovery time and initial and final S, I and R with the CROWD simulation

Although the plots are not perfect matches, that difference is the expected outcome. The timing and magnitude of the *Infected* peaks are different. CROWD predicts I reaches a peak of 55% of the population on the 32nd day, whereas the SIR model predicts a peak of 39% on the 61st day. This is due to the differences in the population's contact topology. As mentioned earlier, the SIR model is derived assuming uniform mixing among all members of the population, whereas a real population has people with a wide range of contact rates [12, 21]. In such systems the part of the population with higher than average contact rates (i.e. the students in see Fig. 12) spread infection fast, more than compensating for those with lower than average rates. This causes the infected population to peak earlier and higher (see also Rahmandad *et al* 2004)[22].

To further illustrate this, we could have chosen to align the values α and β with CROWD using equations 4 and 5 and then using the CROWD S, I, R initial conditions to model the SIR model, giving Figure 6.

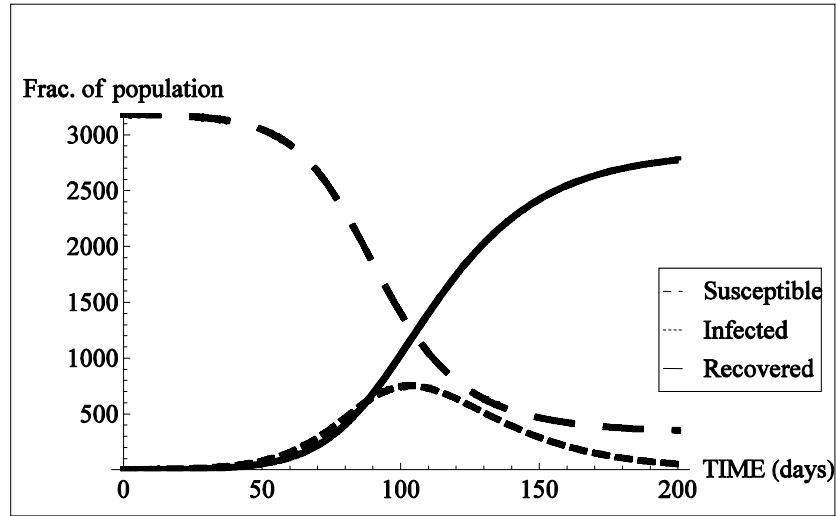


Figure 6: SIR model epidemic state evolution, matching initial S , I and R , recovery time and chance to infect with the CROWD simulation

For an epidemic with matching properties (same β , P_1 and n) we see that the SIR model predicts a much milder epidemic. Again, the difference is due to the more efficient passing of infection through CROWD's more realistic social network than a uniform network assumed by the SIR model.

It is also true, however, that the contact network of CROWD requires improvement. In particular the connection networks within schools and work places are currently uniform, rather than more realistic scale free networks, resulting in overly high contact rates and clustering coefficients, which in turn lead to overly efficient disease propagation. Changing this will result in (amongst other effects) bringing the student peaks in Figure 12 back towards the main population, but they should still be higher than the general populace.

In general, we found that when trying to simulate more complex scenarios (spatially inhomogeneous populations, special events, etc.) the SIR-like models become rougher approximations and the agent-based approach becomes more appropriate (see also Toroczkai *et al* 2007[26]).

As stated earlier, the match between the standard SIR model and that of CROWD is because the SIR model uses a simplistic uniform social network. It is possible to get CROWD to use an identical network by forcing all agents to co-locate. To study this, CROWD was rerun for 1055 people are locked in a room, where every agent comes into contact with every other each time period. A weakly infectious disease is introduced and allowed to spread (the chance of passing the infection per contact is 0.00009, and the chance to recover per time period is 0.04). The corresponding SIR model is then run, from the times when CROWD's I first exceeds 1, 5, 10 and 20. The level of fit between the CROWD model I population and that of the SIR model is then found in each case, using the R^2 score ($R^2 = 1$ is a perfect match, $R^2 = 0$ indicates no relationship between the predictions). Figure 7 contains an example of this output.

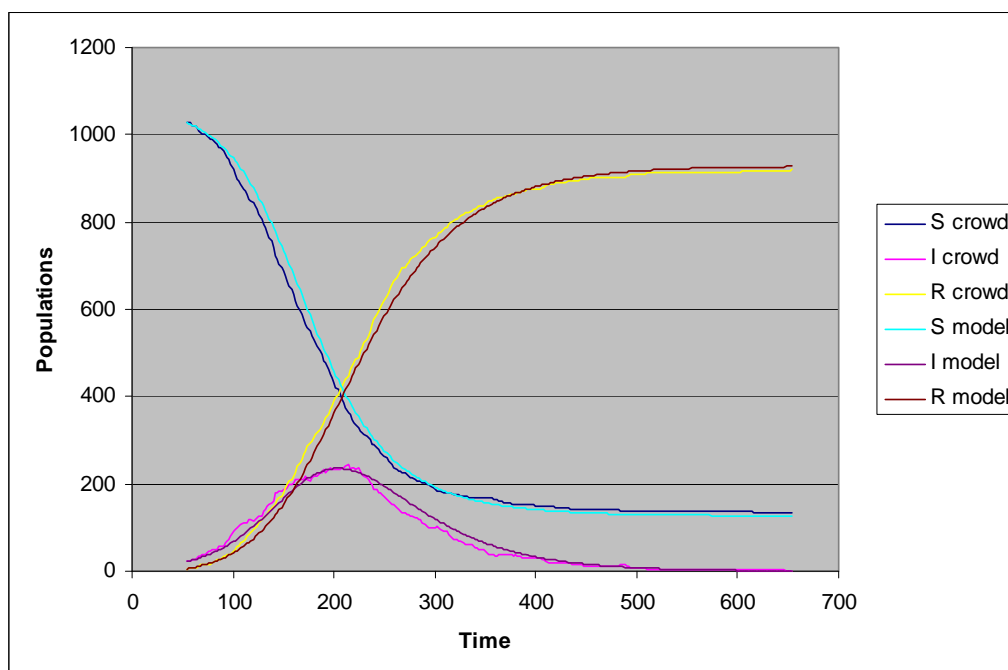


Figure 7: SIR matches to CROWD (agents in same room), starting at $I=20$

The R^2 score results, over 6 runs of CROWD, are summarised in the following table. The purpose of higher starting thresholds of I are to get past the initial stage of the epidemic where small statistical fluctuations have large time impacts. As can be seen, the R^2 score dramatically improves between a threshold of $I=1$ and $I=10$.

Table 1: R^2 score comparison of I predictions of CROWD and SIT models, for different threshold starting values of I

Starting I	$I=1$	$I=5$	$I=10$	$I=20$
	0.964	0.966	0.939	0.927
	0.363	0.833	0.901	0.915
	0.922	0.923	0.975	0.906
	0.956	0.969	0.972	0.972
	0.226	0.867	0.915	0.972
	0.944	0.938	0.952	0.941
Average	0.729	0.916	0.942	0.939

An R^2 of over .9 indicates a good match between the two models. Varying the model parameters to create a more severe epidemic produces matches in excess of .99.

There is a caveat involved here. In CROWD, for the weak epidemic used here, there are two possibilities: in almost 40% of cases, the disease died out before infecting more than a few people; in the other 60%, large epidemics occurred. This is due to statistical fluctuation being very significant at the start of an epidemic, a feature that the more simplistic SIR model cannot capture. The CROWD runs used in the above table are those that do bloom into a large epidemic.

2.5 Analysis of CROWD Epidemic Data Using Generalised SIR

In this section the results of applying the generalised SIR equations (4)-(6) revealed new and informative science about epidemics.

An analysis of the predicted value of δ of equation (5) revealed that δ is not a constant of the system. Rather it varies with the number of agents recovered and immune to the disease. The affect can be seen in Figure 8 as the cause of the effective hysteresis loop. If δ was constant, the number of new cases per hour should be proportional to I^δ , and the dark blue data points would all lie on a single line. Instead, δ varies with R . The physical interpretation of this is that at the start of the epidemic, $\delta < 1$, meaning infection spreads through the agents more quickly than for a uniform social network. However as the number of recovered agents grows, paths of possible transmission through the social network are blocked, meaning fewer new cases, equivalent to a raising δ .

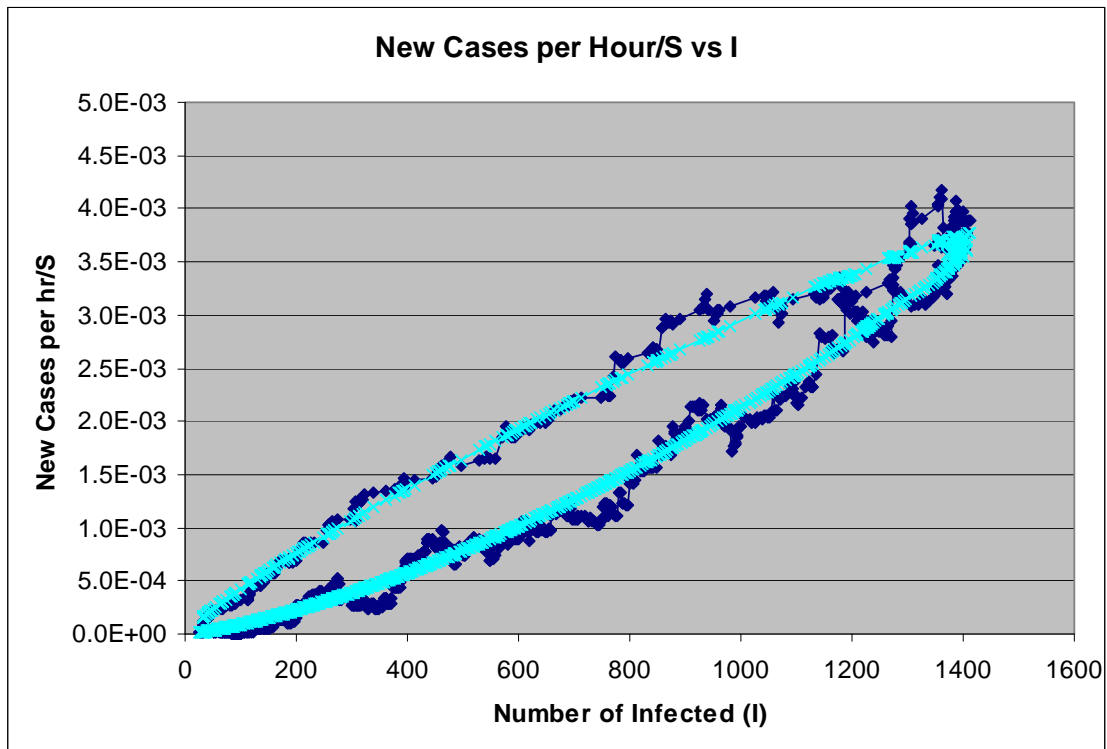


Figure 8: Dark blue is partially time averaged CROWD data. Pale blue line results from a model for $\delta(R)$. The top edge of the loop describes the rate of new cases per hour as I grows, and the bottom edge describes the number of new cases as the epidemic dies out. That the graph forms a loop rather than a line demonstrates that δ varies with the number of recovered.

By studying the effective initial and final values of δ , which can be read from the gradient of Figure 8 in Log-Log scale, a model for $\delta(R)$ was constructed. As seen in Figure 8, it well reproduces the behaviour seen here ($r = R/\text{Total Population}$).

$$\delta(r) = ((.85 + 1.3) + (.85 - 1.3) \cos(\pi r)) / 2 \quad (8)$$

Preliminary investigations also show that this model better predicts the precise course of the epidemic than the standard or constant δ SIR models.

This improved form of the SIR equations can be used as a foundation for the development of simplified operational models with a capability to produce a realistic evaluation of numerous "what-if" scenarios within a limited timeframe by using ordinary computer resources (PC).

2.6 Main Results

The epidemiological model within CROWD is capable of high fidelity simulation for modelling of disease spread in a realistic social network. By careful alignment of the output of CROWD and the SIR model we have obtained a sense of validity that is needed to develop a realistic disease spread model in a complex multi-agent social context (including alignment of model parameters, scenarios and underlying assumptions). This validation in future will be expanded to comparison with real epidemic data.

We believe that our new agent-based model for disease outbreaks provides a cost-effective tool for reasoning about such events and for the simulation of the typical "what-if" scenarios, as well as for the evaluation of various response options. Such a model can be used by civilian health officials for formulating health management policy, as well as by military commanders wishing to assess the impact of disease (naturally occurring or through deliberate biological warfare attack) on their operational capability.

Moreover, the complexity and high fidelity of the model is such that it is revealing new science about the effect epidemics have on population networks.

Shortcomings within the contact graph generated within CROWD, in particular within the classroom environment of schools, have been identified and will be addressed in later versions with further research into generating realistic scale free networks for such environments.

3. The Disease Model in Detail

The disease model is a plug-in model for CROWD that hooks into the Health module which stores each individual's physiological state. To propagate diseases it utilises mechanisms within CROWD that track the population's locations and movements. With the disease module a *Disease List* keeps track of all diseases an individual has contacted; disease progress is monitored via the *Immunity* class and follows the process outlined in Figure 9.

A requirement for the disease model is the ability to track how long people are in contact with each other. The contact model developed for the study is a proximity model, with people being counted as being in contact if they are in the same building. This is reasonable for diseases that are spread via airborne means but a more detailed contact model would be required for other disease transmission vectors.

The disease Life Cycle is linked into the *TimeTicker*, with the *TimeInContactMatrix* and each individuals *Immunity* linked into the *TimeTicker* as *TickListeners*.

- **TimeInContactMatrix:** This maintains a list of each individual's current constant time in contact with each other individual in the environment. This time is updated each Tick.
- **DiseaseModel:** As well as maintaining time in contact the *TimeInContactMatrix* also kicks off the checks for disease transmission by telling the *Model* the amount of time each pair of individuals has been in contact. The *DiseaseModel* checks each person's *DiseaseState* and conducts a disease transmission test if they have been in contact long enough and one of them is infectious.
- **Immunity:** This holds the current *DiseaseState* of a single *Disease* that an Individual has contacted. The *Immunity* is held within a *DiseaseList* which maintains all the diseases that a particular individual has contacted. Each time tick the *Immunity* checks the probability of the current disease state transitioning to a new state.

The following page outlines the timing of events within the model. UML class diagrams for the disease module are shown in Figure 10 and Figure 11.

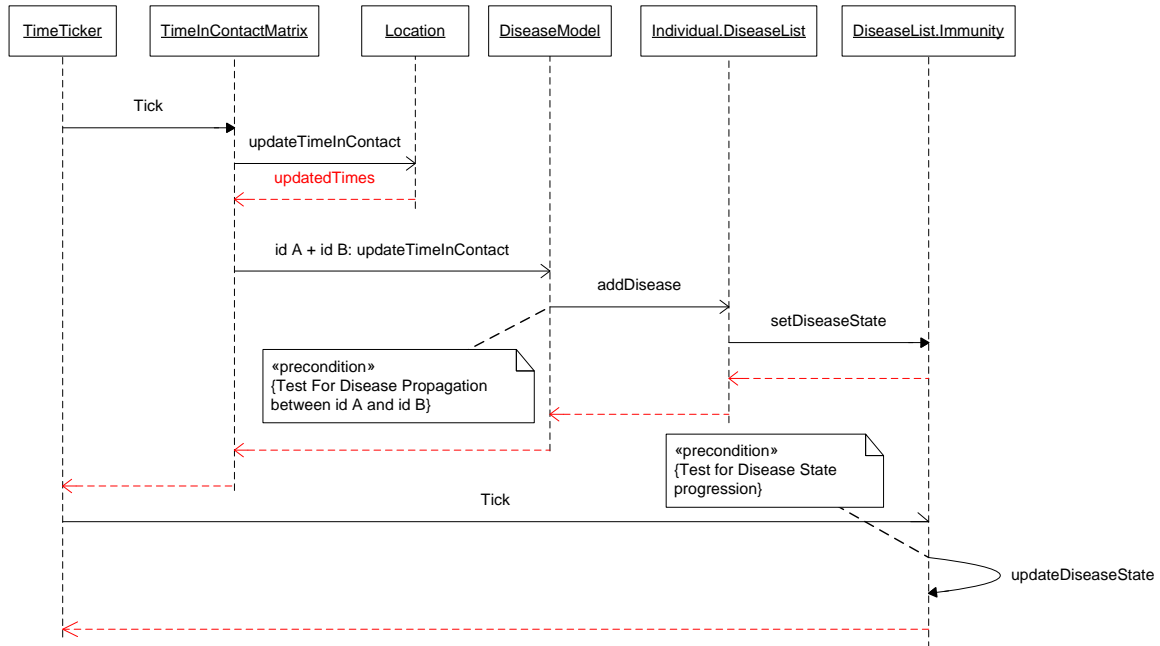


Figure 9: Function State Diagram for the disease model

Function State Diagram Description:

- The *TimeTicker* ticks the *TimeInContactMatrix* which in turn updates the times in contact for each individual in each location.
- Individuals within a location and the updated contact time are sent to the *DiseaseModel*.
- Test for Disease Propagation: *DiseaseModel*
 - a. Each pair of individuals who have a positive contact time have their current Disease States checked. If they have a disease and are infectious and the contact time is > recheck time for that disease state then a check is made to see if the Disease is transmitted to the other party. If the disease is propagated it is added to the individual with the first stage Disease State of that disease.
- Once all these tests have been completed the *TimeTicker* then ticks each Individuals *DiseaseLists*.
- Test for Disease State progression: *DiseaseList.Immunity*
 - a. Each tick every Individuals Immunities are checked to see whether any disease they are currently carrying moves to a new Disease State, i.e. from *Infectious* to *Recovered*. This transition is probability based and dependant on the disease, its current state and the recheck period of the disease.

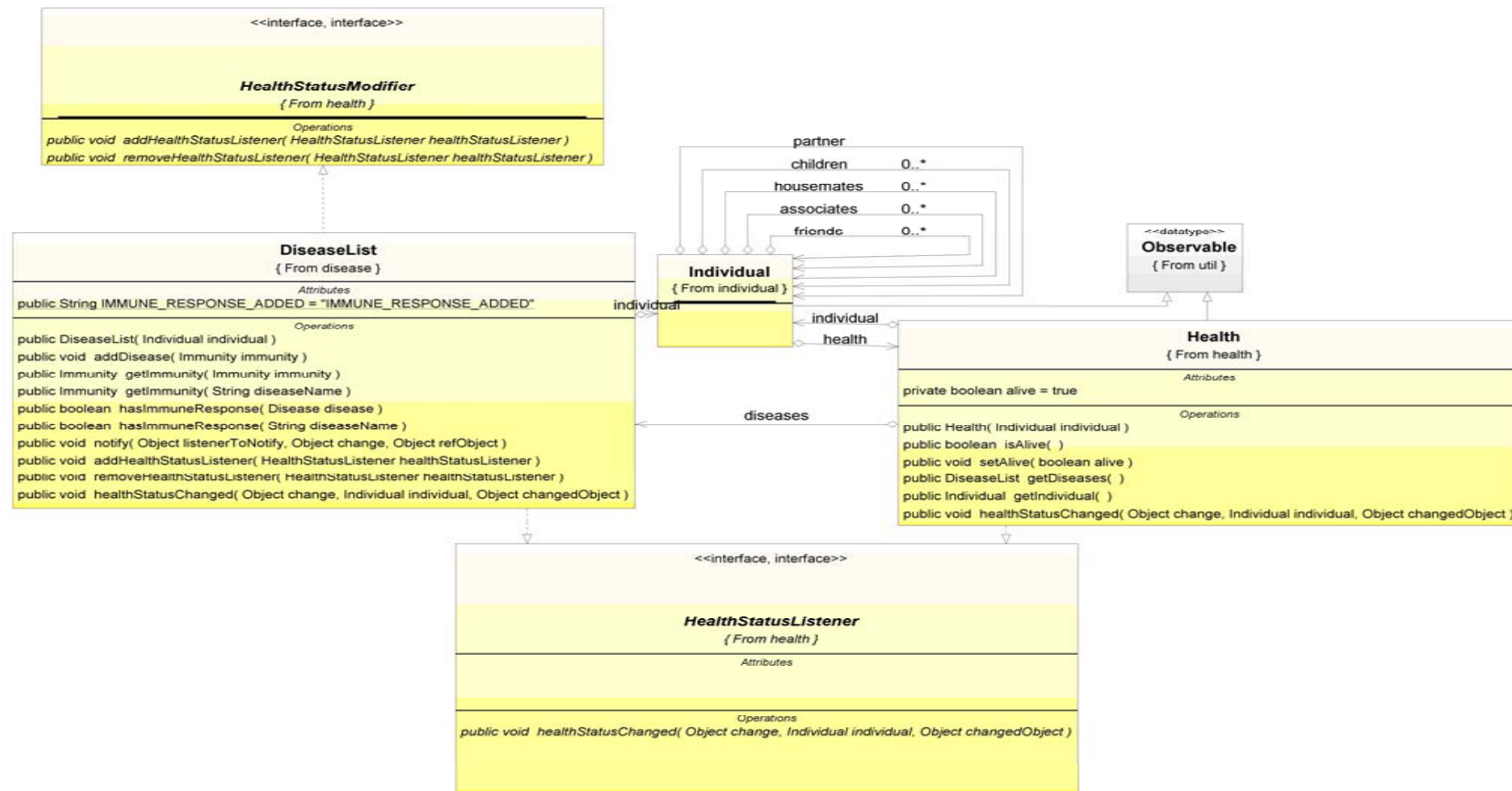


Figure 10: Health Dependencies

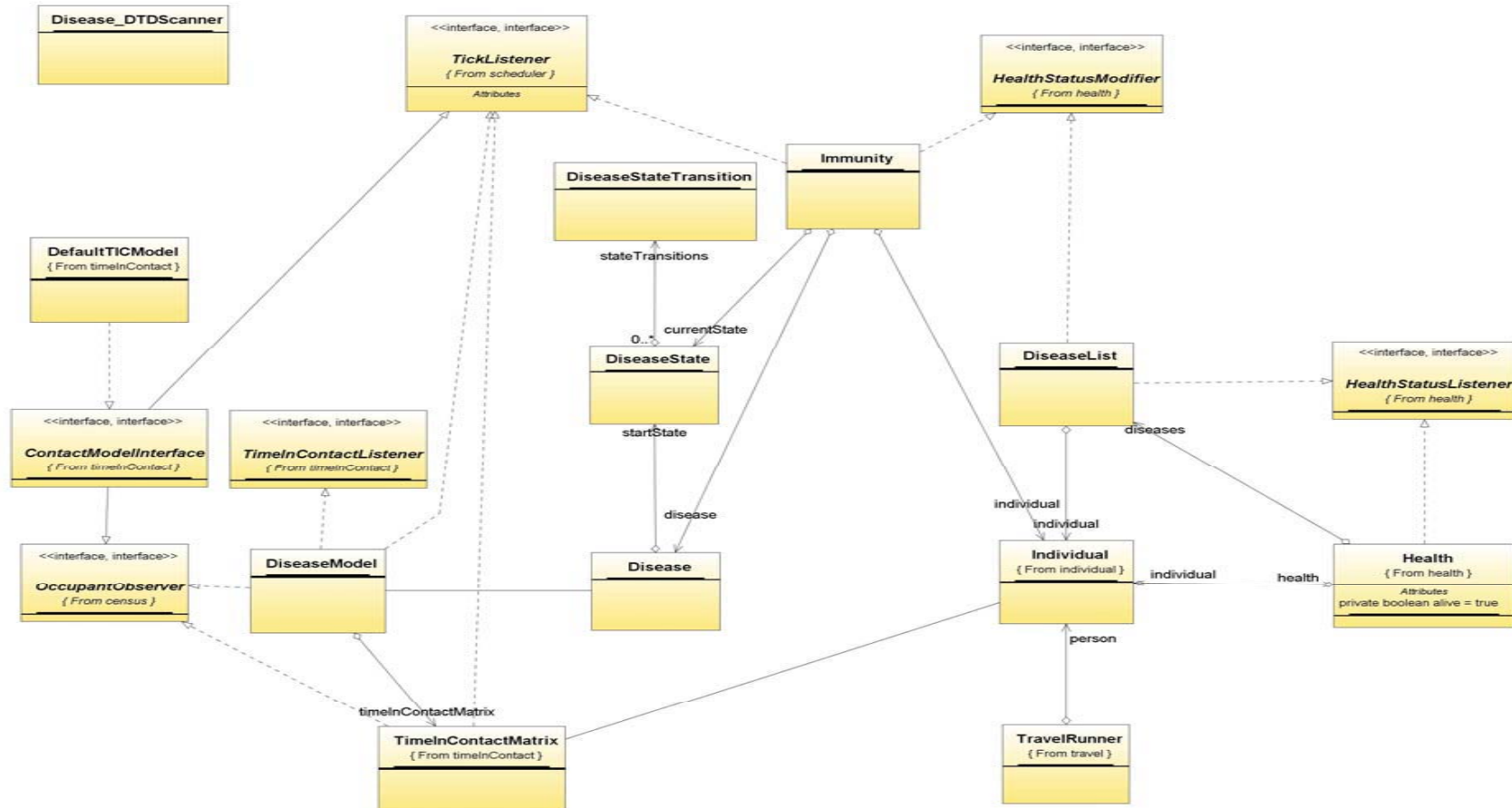


Figure 11: Disease Model Components

3.1.1 Disease XML schema and example: Influenza

Diseases are stored in an XML format that captures the requisite information regarding all aspects of a diseases life cycle. This enables users to produce new diseases and modify old ones by creating new XML documents accorded to the Disease_DTD schema.

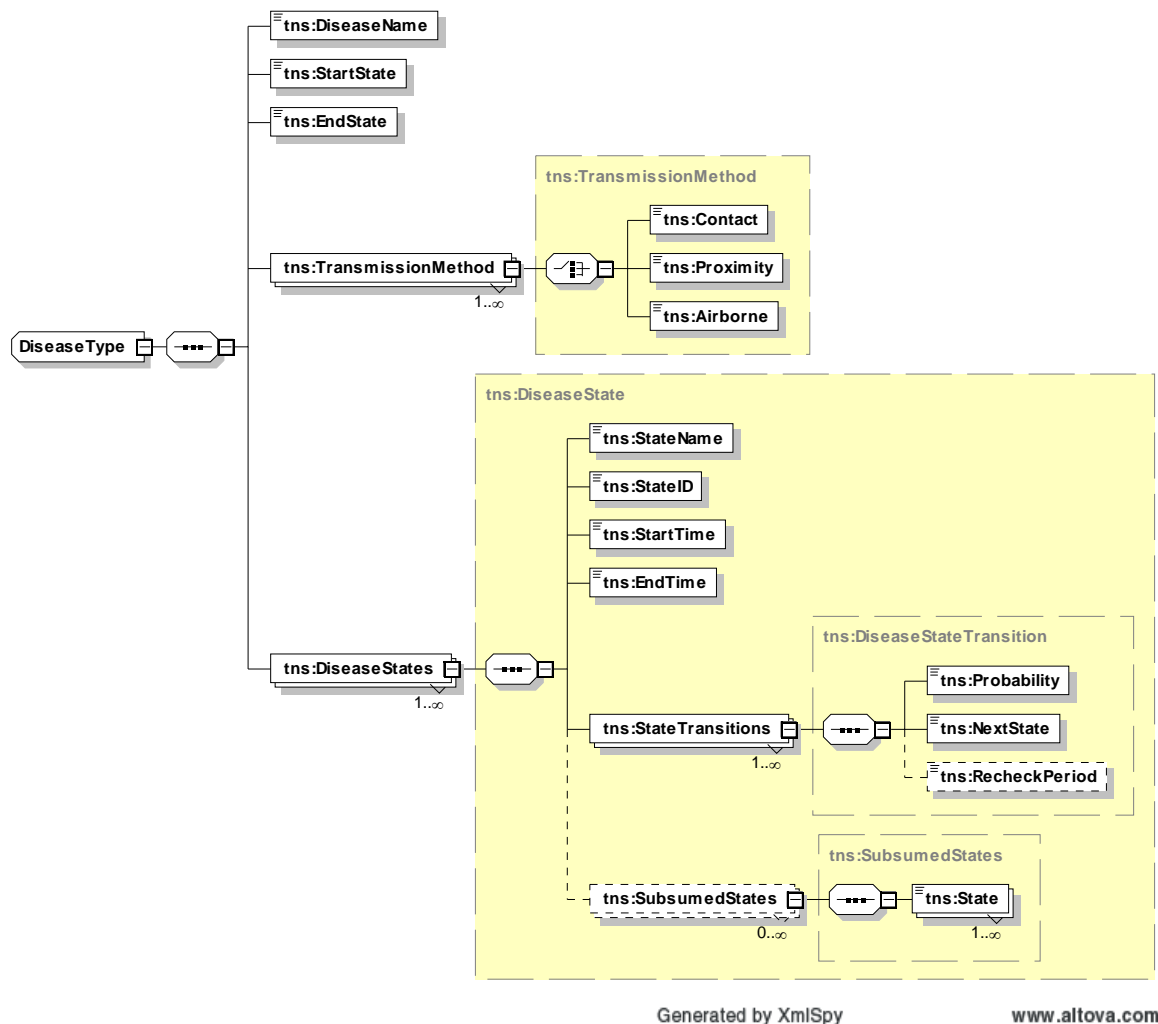


Figure 12: Disease Type XML DTD

While the *DiseaseType* contains the information about the disease it is the *DiseaseState* that drives the Disease Model. The disease states capture the progress of disease as it runs its course. For each *DiseaseState* there a number of *DiseaseStateTransitions* that describe which *DiseaseState* the disease will next pass on to. These transitions are probability based but more advanced version of the disease model could modify these probabilities based on the interventions that have taken place with a particular patient. The presence of *DiseaseStates* allows Individuals to modify their behaviour based on the state they are currently in. For example someone with severe flu symptoms might stay home instead of going to work. These kinds of actions naturally will change the way the disease performs in the wider population.

Table 2: Sample Flu XML file

Disease	Disease Name: Influenza			
	Disease State	State Name: Healthy		
		State Transitions	Probability	0.00714286
			Next State	Infectious
			Recheck Period	2 Hours
	Disease State	State Name: Infectious		
		State Transitions	Probability	0.004117
			Next State	Recovered
			Recheck Period	2 Hours
	Disease State	State Name: Recovered		
		State Transitions	Probability	1
			Next State	Recovered
			Recheck Period	-

The example disease *Influenza* only has 3 states, Health -> Infectious -> Recovered, and was used as a comparison to the mathematical model SIR. The flexibility that this design gives you is that you can add extra disease states as required for the complexity and fidelity of the model you are studying. For example adding in states for “*Pre-Symptoms*” and “*Post-Symptoms*” which the model might utilise to modify the behaviour of the entities based on the states they are in. In the “*Pre-Symptoms*” state the individual would continue to go to work potentially still spreading the disease while the “*Post-Symptoms*” state the individual might stay home from work and wait to recover.

4. Discussion

The scenario for the case study was a flu outbreak within an established social network. This simple scenario, or variations thereof, has been widely studied with both mathematical and agent-based approaches. One of the benefits of agent-based models lies in the ability to extend to cover other scenarios such as

- Non homogeneous population mixing.
- Disease spread amongst time variant social networks, such as seasonal workers.
- A lone infectious agent moving amongst a random contact network, such as an infectious individual in a large shopping complex or business travellers flying in and out of cities/countries.
- Interventions in the epidemic life cycle.
- Tracing the path of a single actor in an epidemic

4.1 Alternative Agent-based disease models: EpiSims and BIOWar

There are a number of available agent-based disease models in the literature. Two of the better developed models available are EpiSims [23] and BIOWar [25].

4.1.1 EpiSims

EpiSims [23], is a highly resolved agent-based simulation tool which combines realistic estimates of population mobility, based on census and land-use data, with parameterised models for simulating the progress of a disease within a host and of transmission between hosts is based upon Transportation Analysis and Simulation System (TRANSIMS) [24], developed at Los Alamos. TRANSIMS provides a way to generate synthetic realistic social contact networks in a large urban region. It is based on the assumption that the transportation infrastructure constrains people's choices about what activities to perform and where to perform them, the nine activities included within EpiSims are: *work, shop, visit, social, other, serve, school, college*.

4.1.2 BIoWar

BIOWar [11, 25] is a computer simulation that combines computational models of social networks, communication media, and disease transmission with demographically resolved agent models, urban spatial models, weather models, and a diagnostic error model to produce a single integrated model of the impact of a bioterrorist attack on an urban area. BIoWar is configured to represent real cities by using census data, school district boundaries, and other publicly available information. Moreover, rather than just providing information on the number of infections, BIoWar models the population of individual agents as they go about their lives – both the healthy and the infected. This allows the analyst to observe the repercussions of various attacks and containment policies as revealed through indicators such as absenteeism, medical web hits, medical phone calls, insurance claims, death rates, over-the-counter pharmacy purchases, and hospital visit rates, among others.

4.1.3 Comparison with CROWD

Where CROWD differs from the above models is the fidelity of the traffic modelling and the activity model. For the study individual movement from place of work to home was assumed to take a certain amount of time, but was not otherwise modelled, no traffic was generated. The activity model used within our study encompassed only *work, school, and home*, which contained enough complexity for our aim of validation against the SIR model. The disease model has been designed to include the level of detail exhibited in the disease model of BIoWar (such as symptoms, immunity and variant transmission rates) however for our purposes a simple implementation was required that matched the variables within the SIR equations.

What CROWD offers that is not within either of EpiSims or BIoWar is the high fidelity physical environment that it inherited from its roots in military simulation. The social network model is generated from the same sets of data as used in BIoWar and Transims, although the transport model is not as sophisticated. The ability to produce results similar to those of SIR informs us that the model has enough complexity to provide a plausible alternative to SIR while providing us with greater insight into the disease transport paths and mechanisms for intervention.

4.2 Further Work

As the model is extended we will be better placed to address the scenarios posited at the start of this section. New studies will provide the impetus for adding and upgrading plug-ins within CROWD. A short list of possible extensions is in Table 3.

Table 3: CROWD Model Extensions

Future Requirements	Description
Extended Individual Activities	Impacts of the fidelity of various outputs, including traffic modelling, disease spread, and social network structures.
Social Networking	Improve the social networking model to better represent how networks form and evolve.
Communications and Media	A mechanism for information, rumour and propaganda to propagate through the population.
Multiple Districts and Regions	Enabling larger populations to be modelled
Climate Models	Incorporating seasonal influences and weather patterns

5. Conclusion

An agent-based model of disease propagation was compared with a validated mathematical model of epidemics, the SIR model. The comparison validated the agent-based disease model, CROWD, in its simple form and providing a foundation for further studies. The benefits of agent-based models for studying factors that impact upon disease propagation through scale free social networks were also briefly discussed. Amongst these is the ability to investigate targeted disease mitigation strategies on select members of the population.

Using available census data to generate our populations gives us a measure of confidence in the demographics of the simulated population. The data provided within a census is broad enough to derive a useful picture of an individual within that society. Information such as: incomes, household outgoings, professional qualifications are all generalised within ABS census data packs and thus are available to clothe the simulated individuals.

Coupled to a modular and extendible model of a disease lifecycle we have the ability to increase the fidelity of the model in areas such as an individual's response to disease symptoms, medical interventions and the social impact of disease spread. This capability and the method whereby disease and social models are combined with complex urban models expands our capability to explore disease propagation a society.

6. Recommendations

This method of disease analysis lends itself to exploring current issues, such as swine flu or SARS, where a highly mobile community can rapidly spread a disease outside of its initial boundaries. Other interesting cases such as a single infectious traveller moving through the population would be difficult to validate however it is a scenario that concerns health officials.

Both HPPD and LOD have groups investigating civilian modelling and a shared model will permit cross-divisional collaboration at a fundamental level. The results of this work will be used in the DSTO Corporate Enabling Research Program (CERP): Bio-Terrorism Preparedness, A Detection to Treat System for Biological Weapons Attacks on a Community.

7. Acknowledgements

This report is based on the papers presented and published in the Proceedings of

- MODSIM 2007 International Congress on Modelling and Simulation (Modelling and Simulation Society of Australia and New Zealand, December 2007);
- Land Warfare Conference 2007 (Australia, October 2007);
- COMPLEX SYSTEM 2007 (8th Asia-Pacific Complex Systems Conference , Australia, July 2007); and
- BIOCOM2007 (2007 International Conference on Bioinformatics and Computational Biology, November 2007, USA).

8. References

1. W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. Royal Soc. London, Series A CXV* (1927) 700.
2. J.D. Murray, *Mathematical Biology II*, 3rd ed. Springer, 2004.
3. Restructuring The Australian Army (DPUBS:24432/96)
4. d'Inverno, M. and Luck, M. and Georgeff, M. and Kinny, D. and Wooldridge, M. (2004), The dMARS Architecture: A Specification of the Distributed Multi-Agent Reasoning System. *Autonomous Agents and Multi-Agent Systems*, 9 (1-2). pp. 5-53.
5. Busetta, P and Rönquist, R and Hodgson, A and Lucas, A (1999), JACK Intelligent Agents - Components for Intelligent Agents in Java. *AgentLink News* (2). pp. 2-5.
6. R Connell, F Lui, D Jarvis and M Watson, "The Mapping of Courses of Action Derived from Cognitive Work Analysis to Agent Behaviours", *Proceedings of Second International Joint Conference on Autonomous Agents and MultiAgent Systems*, Melbourne, Australia, 11th - 14th July 2003.

7. The Three Block War is a concept devised by Gen Charles Krulak in the late 1990s to describe the complex spectrum of challenges likely to be faced by soldiers on the modern battlefield. In three contiguous city blocks soldiers may be required to conduct full scale military action, peacekeeping operations and humanitarian relief. "The Strategic Corporal: Leadership in the Three Block War" Marines Magazine, January 1999
8. Ashley Stephens, Russell Connell and Phil Davies DSTO-TR-1541 AR 013-008 "Urban intelligence preparation of the battlespace: methods for determining spatial vulnerability in urban environments" Technical Report Unclassified, Limited Release E9505-25-112
9. "Future Land Operating Concept: *Complex Warfighting*", *Australian Army*.
10. W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. Royal Soc. London, Series A CXV* (1927) 700.
11. J.D. Murray, *Mathematical Biology II*, 3rd ed. Springer, 2004.
12. M. E. J. Newman. Spread of epidemic disease on networks. *Phys Rev E*, 66(2002), pp 016128.
13. K.M. Carley, D. Fridsma, E. Casman, A. Yahja, N. Altman, L.C. Chen, B. Kaminsky and D. Nave. BIOWar: Scalable Agent-Based Model of Bioattacks, *IEEE Trans. on Sys. Man and Cyber.-Part A*, .36 (2006), .2, pp 252-265.
14. P.D. Stroud, S. J. Sydoriak, J. M. Riese, J. P. Smith, S.M. Mniszewski, P. R. Romero . Semi-empirical power-law scaling of new infection rate to model epidemic dynamics with inhomogeneous. *Mathematical Biosciences*, 203 (2006). pp 301-318
15. P.D. Stroud, S.D.Valle , S. J. Sydoriak, J. M. Riese, J. P. Smith, S.M. Mniszewski, P. R. Romero . Spatial Dynamics of Pandemic Influenza in a Massive Artificial Society. *Journal of Artificial Societies and Social Simulation*, 10 (2007), 4, pp 9-18 (<http://jasss.soc.surrey.ac.uk/10/4/9.html>)
16. A. Skvortsov, R..B. Connell, P. Dawson, R. Gailis. Epidemic Modelling: Validation of Agent-based Simulation by Using Simple Mathematical Models. *Proceedings of Land Warfare Conference 2007*, ed V.Purui, D.Fillipidis, S.Dunn. (2007), pp 221-227.
17. A. Skvortsov, R. Connell, P. Dawson, R. Gailis. Alignment of Agent-based Simulation with a Simple Mathematical Model. *Proceedings of the 2007 International Conference on Bioinformatics & Computational Biology (WOLDCOM'07), Vol.II*, ed H.R.Arabnia, M.Qu Yang, J.Y.Yang , CSREA Press, (2007), pp 487-490.
18. A. Skvortsov, R. Connell, P. Dawson, R. Gailis, Epidemic Modelling: Validation of Agent-based Simulation by Using Simple Mathematical Models, *Proceedings of MODSIM07*, pp 657-662, 2007
19. Keeling, MJ. & Grenfell, BT (2000) Individual Based Perspectives on $R_{sub}(0)$. *Journal of Theoretical Biology* Vol 203, no 1, pp 51-61 7 Mar 2000
20. Barthelemya, M. Barrat, A, Pastor-Satorrasc, R. Vespignani, A (2005), Dynamical patterns of epidemic outbreaks in complex heterogeneous networks, *Journal of Theoretical Biology* Vol 235 pp 275-288

21. Dekker, A.H., (2007), Realistic Social Networks for Simulation using Network Rewiring, *Proceedings MODSIM 2007*.
22. Rahmandad, H. & Sterman, J. (2004), Heterogeneity and network structure in the dynamics of diffusion: comparing agent based and differential equation models, *MIT ESD-WP-2004-5*.
23. Eubank, S et al (2004), Modelling disease outbreaks in realistic urban social networks, *Nature* 2004 Nov; 429:180 – 184.
24. TRANSIMS: <http://transims.tsasa.lanl.gov>
25. Kathleen M. Carley, Neal Altman, Boris Kaminsky, Démian Nave and Alex Yahja
BIOWar: A City-Scale Multi-Agent Network Model of Weaponized Biological Attacks,
CASOS Technical Report, January 2004 , CMU-ISRI-04-101
26. Z. Toroczkai, H. Guclu, Proximity Networks and Epidemics , *Physica A* 378, 68 (2007)

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA						
				1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)		
2. TITLE Comparison of an Agent-based Model of Disease Propagation with the Generalised SIR Epidemic Model			3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) <div style="display: flex; justify-content: space-between;"> Document (U) </div> <div style="display: flex; justify-content: space-between;"> Title (U) </div> <div style="display: flex; justify-content: space-between;"> Abstract (U) </div>			
4. AUTHOR(S) Russell Connell, Peter Dawson and Alex Skvortsov			5. CORPORATE AUTHOR DSTO Defence Science and Technology Organisation 506 Lorimer St Fishermans Bend Victoria 3207 Australia			
6a. DSTO NUMBER DSTO-TR-2342		6b. AR NUMBER AR-014-617		6c. TYPE OF REPORT Technical Report		7. DOCUMENT DATE August 2009
8. FILE NUMBER 2008/1064796	9. TASK NUMBER DS 07/245	10. TASK SPONSOR CAOD		11. NO. OF PAGES 24	12. NO. OF REFERENCES 26	
13. URL OF ELECTRONIC VERSION http://dsto.defence.gov.au/corporate/reports/DSTO-TR-2342.pdf				14. RELEASE AUTHORITY Chief, Air Operations Division		
15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT <div style="text-align: center;"><i>Approved for public release</i></div>						
OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111						
16. DELIBERATE ANNOUNCEMENT No Limitations						
17. CITATION IN OTHER DOCUMENTS Yes						
18. DSTO RESEARCH LIBRARY THESAURUS http://web-vic.dsto.defence.gov.au/workareas/library/resources/dsto_thesaurus.shtml Demography, Epidemiology, Intelligent Agents, Mathematics, Validation						
19. ABSTRACT Social contacts are an important channel for the propagation of disease through a population and should be considered in conjunction with traditional epidemic diffusion. Such channels should always be taken into account for a realistic estimation of a long-term impact of a disease outbreak (natural or malicious) and for the best response options. This paper describes our recent experience in developing a simple agent-based model to simulate disease propagation through a social network and validating the results of the agent-based simulation by reconciling it with a well-known mathematical model.						