Gryphon: A Hybrid Agent-Based Modeling and Simulation Platform for Infectious Diseases

Bin Yu, Jijun Wang, Michael McGowan,
Ganesh Vaidyanathan, and Kristofer Younger

Quantum Leap Innovations,
3 Innovation Way, Suite 100,
Newark, DE 19711, USA
{byu,jw,mtm,gv,ky}@quantumleap.us

Abstract. In this paper we present Gryphon, a hybrid agent-based stochastic modeling and simulation platform developed for characterizing the geographic spread of infectious diseases and the effects of interventions. We study both local and non-local transmission dynamics of stochastic simulations based on the published parameters and data for SARS. The results suggest that the expected numbers of infections and the timeline of control strategies predicted by our stochastic model are in reasonably good agreement with previous studies. These preliminary results indicate that Gryphon is able to characterize other future infectious diseases and identify endangered regions in advance.

1 Introduction

Various approaches have been developed to understand and predict the spread of infectious diseases and the impact of treatment and control strategies [1]. These range from compartmental models represented by sets of differential equations [2,3] to highly complex individual-based models which represent daily activities and connections of individuals via transmission networks [4]. Compartmental models can be easily solved, but they cannot model adaptive behaviors of individuals and complex interactions of different groups of populations during disease outbreaks. While individual-based models like EpiSims can capture the spread of diseases with high-fidelity, modeling large populations often resorts to supercomputers and makes it impractical for quick what-if analyses of interventions or treatments under different conditions.

In this paper we present Gryphon, a hybrid agent-based stochastic modeling and simulation platform for characterizing the geographic spread of infectious diseases and the effects of various mitigation strategies in a GIS environment. As a flexible, computationally efficient modeling and simulation platform, Gryphon has been used successfully in several real time exercises such as Cobra Gold 2008 in Thailand (bird flu, pandemic flu, and multi-lateral military exercises); Operation Caring Response to aid the humanitarian response to Cyclone Nargis in Myanmar (hurricane and epidemic disease outbreaks); and most recently as a primary tool to assist the U.S. Northern Command (USNORTHCOM) and the United States Department of Health and Human Services (DHHS) in modeling and managing the impact of the spread of the H1N1 virus.

S.-K. Chai, J.J. Salerno, and P.L. Mabry (Eds.): SBP 2010, LNCS 6007, pp. 199–207, 2010.

Gryphon integrates agent-based modeling with a stochastic structured-population susceptible-exposed-infectious-recovered (SP-SEIR) model. This hybrid approach provides several advantages over each pure method by combining rich modeling capabilities of agent-based modeling and low computational overhead of differential (or difference) equations. Therefore, Gryphon enables multiple rapid what-if analyses to be performed using singular or multiple interventions and allows the users to optimize their pandemic responses. Compared to recent efforts on equation-based infectious disease modeling [5,6], Gryphon can support more complex and intensive user interactions for both modeling and interventions at runtime. These include pausing the simulation and modifying parameters for a specific group during runtime, modifying the probability that a sick individual travels from his home city to any other city, and enforcing temporary travel restrictions.

We study both local and non-local transmission dynamics of stochastic simulations based on the published parameters and data for SARS. SARS (Severe acute respiratory syndrome) is a respiratory disease first identified in Guangdong, China and became a severe health threat to more than 30 countries in 2003. Many studies have been reported to study the local temporal development of the SARS epidemic in one or more countries [7,8], but deficiencies were still lying in those models because of their separate space and time methodology and the lack of stochastic process for local and global disease transmission. The goal of this study is to validate the stochastic modeling and simulation of Gryphon for both local and non-local disease transmission dynamics using the SARS epidemic data. Instead of estimating all parameters in the stochastic SEIR model, we focus on the basic reproductive rate R_0, which is "the average number of secondary cases caused by an infectious individual in a totally susceptible population".

The rest of the paper is organized as follows. In Section 2 we present the design of hybrid agent-based modeling, stochastic disease model and data sets used by Gryphon. The results of validation study are presented in Section 3. Section 4 concludes the paper with some directions for future research.

2 Methods

2.1 Hybrid Agent-Based Modeling

A group of individuals associated with a geographic location (e.g., a country) is modeled as a primary group agent. A primary group agent can be decomposed into several secondary group agents. Each of the secondary group agents can be further decomposed into multiple tertiary group agents. Translocation is the process of decomposing each primary group into various secondary groups and populating locations with the corresponding secondary groups. The mixing of secondary groups at a location can be localized mixing or non-localizing mixing. Localized mixing refers to the manner in which members of all secondary groups at a location interact with one another. Non-localized mixing is the manner in which members of secondary groups at different locations indirectly interact with one another or with environments to spread disease such as indirect transmission of cholera via water. In this paper only localized mixing is considered.

Different from equation-based models such as SP-SEIR, the hybrid agent-based model does not have a migration matrix to determine the mixing rates among

different groups. Instead, the mixing process is naturally driven by the behaviors of different groups. The behaviors of an agent include two parts: active and reactive. Active behaviors of an agent are modeled by a set of decision rules such as movement patterns, condition-based behaviors caused by interventions and environmental changes. The reactive behaviors of an agent in the context of infectious diseases refer to localized and non-localized mixing for a location, where the numbers of individuals at different disease states change constantly due to the interaction with other agents at the location.

Each simulation time step consists of three steps in the order of pre-step, step, and post-step. In pre-step, a secondary group agent may change its behaviors in response to either interventions or environmental changes. In step, secondary group agents at a location mix with each other based on a given disease model. In post-step, the system will update the state of each secondary group agent based on the calculation of the disease model. Subsequently, each secondary group agent notifies its primary group agent of the state changes. At the end of post step, all secondary groups at each location are cleared and the translocation process of each primary group agent is executed to prepare for next simulation time step.

2.2 Disease Model

We use a discrete-time stochastic susceptible-exposed-infectious-recovered (SEIR) model to simulate the localized mixing of all secondary group agents at a location, where S(t), E(t), I(t) and R(t) represent the number of susceptible, exposed, infectious, and recovered individuals, respectively, at a location at time t. The total population at the location $N(t) = S(t) + E(t) + I(t) + R(t)$ is assumed to be a constant (birth and death are ignored). Specifically, the stochastic SEIR model is specified by the following difference equations.

$$S(t+h) = S(t) - B(t) \tag{1}$$

$$E(t+h) = E(t) + B(t) - C(t) \tag{2}$$

$$I(t+h) = I(t) + C(t) - D(t) \tag{3}$$

$$R(t+h) = R(t) + D(t) \tag{4}$$

where h represents the time interval between two continuous simulation steps and h is set to 1 day.

$B(t)$ is the estimated total number of infections resulting from individuals in the $I(t)$ state. For a given infectious person, the number of new infections from $N(t)$ is sampled from a binomial distribution as $M(t) = Binomial(Binomial(Poisson(c), S(t)/N(t)), p)$, where c is the mean number of daily contacts per person and p is the probability that a contact produces infection. Given $M(t)$, $B(t)$ can be calculated as the sum of $M(t)$ for all individuals at I state. Since a Poisson distribution is a special case of a Binomial distribution, the compound distribution $Binomial(Binomial(Poisson(c), S(t)/N(t)), p)$ can be reduced to $Poisson(\beta IS(t)/N(t))$, where β is the transmission rate and $\beta = c * p$ [9]. The number of individuals becoming infectious C(t) in a day can be represented by a binomial distribution $Binomial(E, \alpha)$, where $1/\alpha$ is the length of the mean latent period. Similarly, the number of recoveries $D(t)$ can be represented by $Binomial(I, \gamma)$, where $1/\gamma$ is the length of the mean infectious period.

The values of mean latent period and mean infectious period are 0.85 day and 2.95 days, respectively. The daily transmission rate β is estimated from the basic reproductive rate R_0 as $\beta = R_0 * \gamma$.

2.3 Data Sets

Important events in the timeline of the 2003 SARS epidemic in Hong Kong and other Asian countries are as follows

- February 15, 2003: Official report of a 33-year male and a 9 year old son in Hong Kong with Avian influenza (H5N1).
- March 12, 2003: First global alert about atypical pneumonia in Vietnam and Hong Kong was issued by World Health Organization (WHO).
- March 15, 2003: Second global alert about name of SARS and case definition was issued by WHO.

One simple way to model the transmission dynamics and control strategies is to change the basic reproductive rate R_0. Therefore, instead of using one value for R_0, we have a pairwise value (R_H, R_L) for R_0 to reflect the level of effectiveness of control strategies after WHO warnings. The value of R_0 is switched from R_H to R_L in the experiments based upon one of the WHO global alerts. The range of R_H is $\{2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5\}$ and the range of R_L is $\{0.6, 0.7, 0.8\}$.

The travel data sets are generated from the International Air Transport Association (IATA) (http://www.iata.org) database, which contains the number of available sets between any two given countries. The country data sets, including population, latitude and longitude for each country, are generated from the website (http://www.geonames.org).

3 Results

In general, there are two ways to approximate R_0: by assuming an exponential increase in the number of cases over time (e.g., [10]) or by fitting a specific model that summarizes

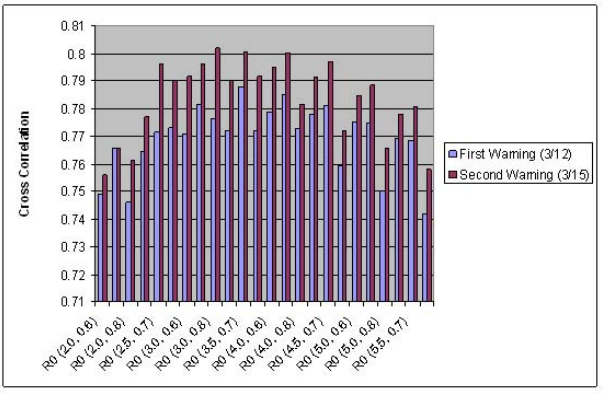

Fig. 1. The cross-correlation coefficient for different pairwise R_0, where the value of R_0 is switched on 3/12 and 3/15, respectively

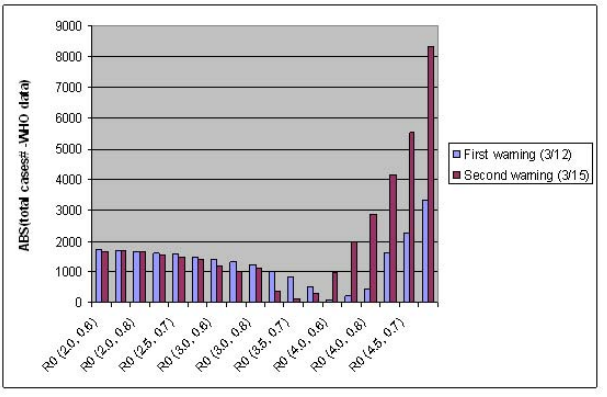

Fig. 2. The difference of simulated accumulative $case\#$ and actual accumulative $case\#$ for pairwise R_0, where the value of R_0 is switched on 3/12 and 3/15, respectively

assumptions about the epidemiology of an infectious disease (e.g., [7]). In this paper we follow the second approach to estimate the spread of a disease based on R_0. Moreover, we choose the metric of cross correlation to quantify the similarity between simulation data and WHO data.

3.1 Parameters for Local Transmission Dynamics

The first experiment studies the basic reproductive rate for local transmission dynamics in Hong Kong. Figure 1 describes the cross-correlation coefficient between the mean of the simulated data for 100 rounds and the WHO data for different pairwise R_0 values, where we seed two infected individuals in each simulation. According to the WHO data, there are only two infections in Hong Kong on February 15, 2003. Note that the two data series in Figure 1 are aligned to calculate the maximal cross-correlation coefficient between the simulated data and the WHO data. We can find that, for Figure 1, the cross-correlation coefficient is consistently higher when R_0 is switched on March 15, 2003. This indicates that those serious control measures such as quarantine and isolation are implemented in Hong Kong only after WHO issued the second global warning on March 15, 2003.

However, it is hard to find the proper value of R_0 only from Figure 1. The reason is that cross-correlation coefficient only models the shape of two data series. As we can see from Figure 1, it is difficult to tell whether pair (3.0, 0.7) is better than pair (3.5, 0.7) on modeling the spread of SARS in Hong Kong. One idea is to use accumulative case numbers as the second measurement to model the scale of the curves. Figure 2 describes the difference of simulated accumulative $case\#$ and actual accumulative $case\#$ for pairwise R_0, where the value of R_0 is switched on 3/12 and 3/15, respectively. From Figure 2, we find that the accumulative case number for (3.5, 0.7) is much closer to the WHO data based on the second global warning. The experimental results show that the combined two metrics, cross-correlation coefficient and cumulative case number, can effectively estimate R_0 values from temporal patterns in an observed epidemic.

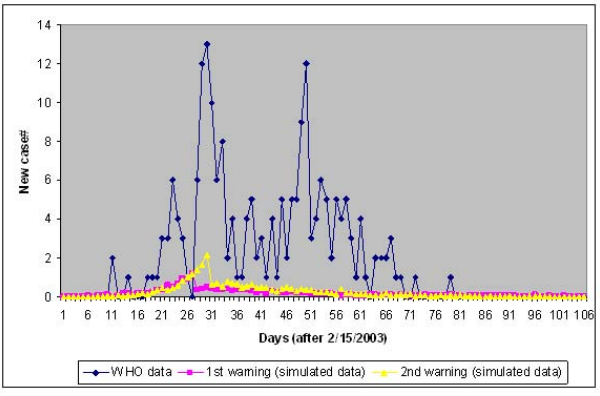

Fig. 3. The new $case\#$ of SARS in Singapore from WHO data and the mean new $case\#$ of SARS from the simulations (100 rounds) for the pairwise R_0 at (3.5, 0.7)

3.2 Parameters for Non-local Transmission Dynamics

The non-local transmission dynamics for SARS at the country level may depend on two factors: the airline travel and the probability that a sick individual travels from his home city to other cities. In this experiment we seed two infections in Hong Kong on February 15, 2003 and we examine the disease outbreak in two Asian countries: Singapore and Japan. We assume that the probability that a sick individual with SARS travels is 0.5. Figure 3 shows the new $case\#$ of SARS in Singapore from the WHO data and the mean new $case\#$ of SARS from simulations. We can see that, based on the airline travel data and the probability that a sick individual travels, the stochastic simulation engine significantly underestimates the SARS outbreak in Singapore. The peak of the mean new case number is one for the first warning and two for the second warning.

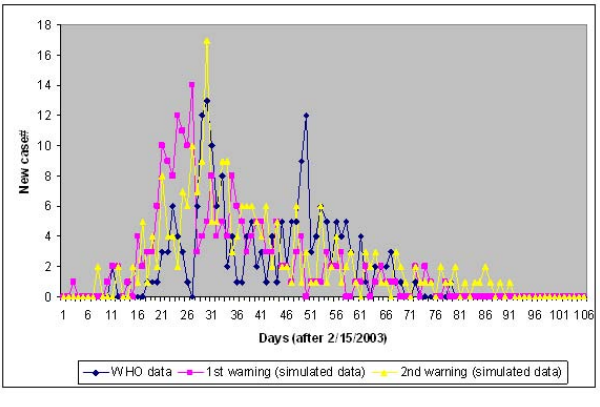

Fig. 4. The new $case\#$ of SARS in Singapore from WHO data and the best match new $case\#$ of SARS from the simulations for the pairwise R_0 at (3.5, 0.7)

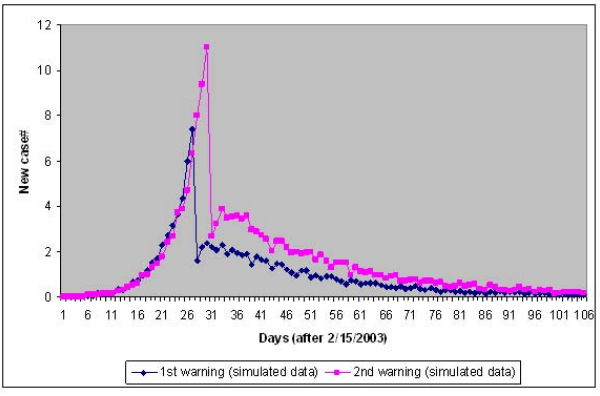

Fig. 5. The mean new $case\#$ of SARS in Japan from the simulations (100 rounds) with the pairwise R_0 at (3.5, 0.7)

However, some specific simulations did capture the dynamics of the SARS outbreak in Singapore. Figure 4 describes the best matched epidemic curve in Singapore from stochastic simulations. The cross correlation coefficient between the best matched simulation data and the WHO data is around 0.6. This indicates that the spread of the SARS epidemic in Singapore is probably the worse case scenario predicted by stochastic simulations. Most likely the first few infections are super-spreaders (a super-spreader is a person having many contacts) and they transmit the disease to large numbers of people.

As shown in Figure 3 and Figure 4, the mean new case number predicted from the stochastic simulation engine can overestimate the SARS epidemic. On the other hand, the stochastic engine can also underestimate the SARS epidemic in some countries like Japan in 2003. The daily airline traffic (in 2005) between Hong Kong and Singapore is 38000 seats/per day, while the daily airline traffic between Hong Kong and Japan is 133000 seats/per day. Based on the traffic data as well as the potential of disease transmission from mainland China, the estimated SARS epidemic should be more severe in Japan than that in Singapore.

Figure 5 shows the mean new $case\#$ of SARS in Japan generated by 100 rounds of simulations. The stochastic simulation engine predicts that there will be about 98 SARS cases for the first warning and 160 SARS cases for the second warning in Japan. This value is close to the mean total case number of SARS predicted by the continuous-time stochastic model published in [6]. However, there is no reported SARS case in Japan during the 2003 SARS outbreak. The actual scenarios in Singapore and Japan motivate us to rethink more complex processes of the spatial and temporal transmission as well as different modes of transmission. These include but not limited to the role of the super-spreader and the heterogeneous mixing among different social groups. For example, Japanese tourists may not well mix with the Chinese community in both Hong Kong and mainland China. The high standard of hygiene conditions in Japan may also prevent the spread of SARS.

4 Conclusion

Gryphon provides a scalable, flexible and interactive disease modeling capability that combines agent-based modeling and mathematical modeling to perform rapid, reasonable fidelity simulations and what-if analyses. We studied the effectiveness of Gryphon, an agent-based stochastic simulation engine for infectious diseases using the historic SARS data. The estimated pairwise value of R_0 for Hong Kong is consistent with [10] by assuming an exponential increase in the number of cases over time, while the predicated total case number for non-local disease transmission is close to the one given in [6], in which Hufnagel et al. used a continuous-time stochastic SEIR model. The experimental results suggest that the expected numbers of infections as well as the timeline of enforced control strategies predicted by our stochastic engine are in reasonably good agreement with previous approaches.

In this validation study we simply use a pairwise R_0 to capture the control strategies deployed upon the first and second WHO warnings. We can see that the peak of the simulated data from Gryphon in Figure 5 drops very fast. This motivates us to develop the next generation of Gryphon technology for data-driven stochastic simulations, where the basic reproductive rate R_0 is dynamically changing based on the available data during a disease outbreak. The data-driven Gryphon will serve as a real-time epidemiological environment for pandemic preparedness and response planning.

Acknowledgements

The authors would like to thank Jay Askren, Albert Boehmler, Julie Cowart, David Hample, Eric Jean and Dr. Steve Prior for their contribution to the system development. We would also like to thank Dr. Amy Kircher at the U.S. Northern Command for her assistance and valuable comments. This research has been sponsored by the Office of Naval Research (ONR) under Contract No. NOOO14-07-C-OO14.

References

1. Anderson, R.M., May, R.M.: Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford (1992)
2. Rvachev, L.A., Longini, I.M.: A mathematical model for the global spread of influenza. Mathematical BioSciences 75, 3–22 (1985)
3. Sattenspiel, L., Simon, C.P.: The spread and persistence of infectious diseases in structured populations. Mathematical BioSciences 90, 341–366 (1988)
4. Eubank, S., Guclu, H., Kumar, V.S.A., Marathe, M., Srinivasan, A., Toroczkai, Z., Wang, N.: Modelling disease outbreaks in realistic urban social networks. Nature 429, 180–184 (2004)
5. Colizza, V., Barrat, A., Barthelemy, M., Valleron, A.J., Vespignani, A.: Modeling the worldwide spread of pandemic influenza: Baseline case and containment intervention. PLOS Medicine (2007)
6. Hufnagel, L., Brockmann, D., Geisel, T.: Forecast and control of epidemics in a globalized world. PNAS 101(42), 14124–15129 (2004)
7. Chowell, G., Fenimore, P.W., Castillo-Garsow, M.A., Castillo-Chavez, C.: SARS outbreaks in ontario, hong kong and singapore: the role of diagnosis and isolation as a control mechanism. Journal of Theorectical Biology 224, 1–8 (2003)

8. Chowell, G., Castillo-Chavez, C., Fenimore, P.W., Kribs-Zaleta, C.M., Arriola, L., Hyman, J.M.: Model parameters and outbreak control for SARS. Emerging Infectious Diseases 10(7), 1258–1263 (2004)
9. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 1. Wiley, Chichester (1968)
10. Wallinga, J., Teunis, P.: Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. American Journal of Epidemiology 160(6), 509–516 (2004)