Agent-Based Stochastic Simulations of Shipboard Disease Outbreaks

Bin Yu, Jijun Wang, Michael McGowan, and Ganesh Vaidyanathan
Quantum Leap Innovations
3 Innovation Way, Suite 100
Newark, DE 19711, USA
{byu, jw, mtm, gv}@quantumleap.us

Abstract
Infectious diseases aboard naval ships may rapidly spread within shipboard populations and severely disrupt operational activities. In this paper we present Gryphon, an agent-based stochastic modeling and simulation platform for characterizing the spread of shipboard infectious diseases. We discuss the stochastic process of disease transmission, features of the Gryphon system for decision support and the emergent dynamics of observed epidemics. We focus on the sensitivity analysis of stochastic simulations for shipboard disease outbreaks and document the results across various population sizes and seeded infections. Our results show that the dynamics of a disease outbreak can be successfully predicted with a reasonable variance when the number of seeded infections and the population size become relatively high. We discuss the implications of various behaviors exhibited by the stochastic simulation engine and conclude with several possible improvements to the development of Gryphon platform.

Keywords: Agent-based stochastic simulations, infectious diseases, epidemiology, shipboard

1. INTRODUCTION
Infectious diseases such as influenza and norovirus may rapidly spread within shipboard populations, causing considerable morbidity and disrupting operational activities [5] [12]. In this paper we describe Gryphon, a Shipboard Disease Decision Support Toolkit (SDDT) developed at Quantum Leap Innovations, Inc. (QLI). The goal of Gryphon SDDT is to construct a computational model of the spread and impact of shipboard diseases as well as related intervention strategies on personnel and their shipboard operational capability. Gryphon enables Medical Officers (MOs) of a vessel to estimate the benefit of specific interventions or treatments, determine resource requirements for treatment of personnel, and assess the potential impact of the disease on the operational capability of the vessel.

Deterministic models for infectious diseases have been widely studied in epidemiology, such as SIR (Susceptible, Infectious, and Recovered) and SEIR (Susceptible, Exposed, Infectious, and Recovered) models [1]. A SEIR model has four compartments S (for susceptible), E (for exposed), I (for infectious) and R (for recovered). The stochastic SEIR model simulates the dynamics of the spread of an infectious disease outbreak in the sense that individuals are initially susceptible, then may or may not acquire the infection (move into the exposed and infectious compartments) and finally recover (move into the recovered compartment). Thus each member of the population typically progresses from susceptible to infectious to recovered. The diagram shown in Figure 1 illustrates the SEIR model, in which the boxes represent different compartments with the arrows representing the transition between compartments.

![Figure 1. An example of SEIR diagram in epidemiology](image)

However, by aggregating groups of people and modeling the spread of a disease in a deterministic way, a significant level of details of different human behaviors and environmental variables is missing. In practice, the likelihood that a particular individual will catch a given infectious disease depends on where he/she works, whom he/she interacts with, as well as his/her habits of personal hygiene. In order to introduce stochasticity into disease modeling, a discrete-time stochastic SEIR model (standing for Susceptible, Exposed, Infectious, and Recovered) has been developed to simulate the transmission dynamics of shipboard disease outbreaks.

A stochastic SEIR model is suitable for modeling diseases such as (pandemic) influenza, pneumonia and norovirus. In contrast with a deterministic model, the transition between any two compartments in a stochastic SEIR model is sampled from a probability distribution per time unit. In the context of shipboard disease modeling, each time unit represents 1 hour. Because sailors belong to different departments and shifts may have significantly different social behaviors, Gryphon integrates agent-based modeling with stochastic structured-population SEIR (SP-SEIR) modeling to simulate the heterogeneous behaviors among groups and the stochastic behaviors within each group. Particularly, the agent-based modeling addresses the spatial/temporal variations of the interaction in different locations, and the probable locations of officers and enlists within an hour. While the stochastic model, associated with an agent (sailor group), simulates the stochastic-
ity of disease spread within the group. This hybrid approach provides several advantages over each individual method by combining the rich modeling capabilities of agent-based modeling with the low computational overhead of equation-based modeling [13]. Therefore, Gryphon enables multiple rapid what-if analyses to be performed using singular or multiple interventions and allows the users to optimize their guidance and actions for effective course of action planning.

The goals of this paper are twofold. First, we present the methods and the key features for simulating shipboard infectious diseases in Gryphon. Second, we study the sensitivity of stochastic simulations under various initial conditions. Different mathematical deterministic models have been analyzed for varying population sizes [6] [10] [11]. Usually a threshold is identified to determine the dynamics of an epidemic model with a varying population size, where the disease dies out or becomes an endemic or epidemic. Stochastic simulations for shipboard disease outbreaks have been studied by [9] in terms of the average number of infections over multiple runs. However, the analysis of stochastic simulations for varying population sizes and seeded infections remains to be addressed in epidemiology.

The transmission dynamics of a shipboard disease outbreak may rely on several factors. At the population level, these include but are not limited to the population size and the initial number of infection cases. To better understand the stochastic simulation engine in Gryphon, we have collected the results for the sensitivity study of our stochastic simulation engine for shipboard disease outbreaks and analyze the metrics used to quantify the randomness of the stochastic simulation engine. Moreover, we discuss the implications of different behaviors exhibited by the stochastic simulation engine.

2. METHODOLOGY

Figure 2 describes the ship model, population model, and the stochastic disease model integrated by the simulation engine for shipboard disease modeling. Next we discuss the details of each model, the design principle of agent-based modeling and simulation, and the metrics for sensitivity analysis of stochastic simulations.

2.1. Ship model

The SDDT model is based on the LHD-1 class of amphibious assault ships with 12 compartments (distinct physical locations aboard a ship) and 1065 sailors (officers and enlists). These sailors are organized into the following departments: Air, CSI, Deck, Dental, Engineering, Executive, Medical, Navigation, Operations, Shop, Enlisted Mess Supply, Officer Mess Supply, Enlisted Berthing Supply, Officer Berthing Supply, and Stowage.

Figure 2. Integrated disease, personnel (population) and vessel models in Gryphon

2.2. Population model

In order to model the personnel behavior aboard a naval ship in a realistic manner, we simulate both on-duty and off-duty activities. There are three shifts for shipboard populations: red, white, and blue. Each officer and enlisted crew member in a shift will have a different routine and interact with a number of crew members on a daily basis. In addition to the on-duty activities, a probability is associated with a sailor or officer for his/her off-duty locations in any of the 12 compartments during a period on an hourly basis.

2.3. Hybrid Agent-based modeling

In shipboard disease modeling, a group of individuals associated with the same on-duty schedule (e.g., deck) is modeled as a primary group agent. A primary group agent can be decomposed into several secondary group agents based on the behaviors of the primary group agent. A primary group agent can be decomposed into several secondary 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 (e.g., a compartment in a ship) 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 flu via air or shared pathways. 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 [14] [15]. 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 pat-
terms, 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 step.

### 2.4. Stochastic 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.

\[
S(t + h) = S(t) - B(t) \quad (1)
\]
\[
E(t + h) = E(t) + B(t) - C(t) \quad (2)
\]
\[
I(t + h) = I(t) + C(t) - D(t) \quad (3)
\]
\[
R(t + h) = R(t) + D(t) \quad (4)
\]

where \( h \) represents the time interval between two continuous simulation steps and \( h \) is set to 1 hour.

\( B(t) \) 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 is sampled from a binomial distribution as

\[
M(t) = \text{Binomial}(\text{Binomial}(\text{Poisson}(c), S(t)/N(t)), p),
\]

where \( c \) is the mean number of hourly contacts per person and \( p \) is the probability that a contact produces infection. Given \( M(t) \), \( B(t) \) is the sum of \( M(t) \) for all individuals at \( I \) state. Since Poisson distribution is a special case of a Binomial distribution, the compound distribution \( \text{Binomial}(\text{Binomial}(\text{Poisson}(c), S(t)/N(t)), p) \) can be reduced to \( \text{Poisson}(\beta S(t)/N(t)) \), where where \( \beta \) is the transmission rate and \( \beta = c \cdot p [7] \).

The number of individuals becoming infectious \( C(t) \) in a hour can be represented by \( \text{Binomial}(E, \alpha) \), where \( 1/\alpha \) is the length of the mean latent period. Similarly, the number of recoveries \( D(t) \) can be represented by \( \text{Binomial}(I, \gamma) \), where \( 1/\gamma \) is the length of the mean infectious period.

In the sensitivity study presented in this paper we use pandemic flu as an example. The values for mean latent period and mean infectious period are 0.85 ± 24 hours and 2.95 ± 24 hours, respectively. The mean hourly transmission rate is estimated from the basic reproductive number \( R_0 \), where \( R_0 = 1.8 \) [3] [4]. Intuitively \( R_0 \) is known as “the average number of secondary cases caused by an infectious individual in a totally susceptible population”.

### 2.5. Metrics

The following three metrics are relevant in the public health domain to measure the sensitivity of the stochastic Gryphon simulation engine,

- Peak value: the maximal infectious population during a disease outbreak.
- Peak time: the time from initial exposure of a disease to the maximum of the infected population.
- Total case number: the accumulative case number during a shipboard disease outbreak.

To study the randomness or dispersion of stochastic simulation results, we calculate the coefficient of variation across multiple runs for each metric defined above. Coefficient of variation (CV) is a normalized measure of dispersion of a probability distribution [7]. It is defined as the ratio of the standard deviation \( \sigma \) to the mean \( \mu \).

\[
CV = \frac{\sigma}{\mu}
\]

The coefficient of variation measures the degree of randomness compared to the mean. We do not use standard deviation, since standard deviation by itself cannot capture wildly different means for data sets with different conditions.

### 3. Prototype System

The goal of the Shipboard Disease Decision-Support Tool (SDDT) program is to utilize recent advances in agent-based models (ABM) to construct computational models of the effects that shipboard disease and related intervention and treatment procedures have on personnel and their shipboard operational capability. The SDDT will enable the MOs to assess the benefits of specific health interventions or treatments, determine resource requirements for treatment of personnel, and advise their chain of command about the potential impact of the disease on the operational capability of the vessel.

Figure 3 shows the SDDT status window during the simulation, where the user can switch between a department.
view and a compartment view. We see from Figure 4 that the compartment view displays two types of information for a group agent: disease dynamics (e.g., percentage of personnel in each disease state) and operation levels (optimal, reduced, and compromised). When a location is fully operational the cell is green indicating an optimal capacity. When the location is between 85% – 99% of it’s fully functional status, the cell is amber indicating a reduced capacity; and when the location has less than 85% of the required staff, the cell is red indicating a compromised capacity.

Figure 3. Snapshot of the user interface of Gryphon SDDT

In addition to the spread of infectious diseases, SDDT also models several medical (e.g., antiviral, vaccines) and non-medical interventions. The non-medical interventions include

- **Mess to Helipad** - In this intervention we send some percentage of the people who would have eaten in the mess hall to the helipad. The helipad has a much larger capacity than the mess hall, so people are not packed as close together. In our experiments we send one half of the people to eat at the helipad.

- **Social Distancing** - In this intervention we reduce the contact rate of people at all locations in the ship by restricting direct contact. This can mean limits on physical contact, ensuring that crowds do not gather in close contact (i.e. at a movie night or football game), or any other protective measure. Since we do not have direct statistics on the reduction, we are simple claiming a percentage of reduction (reduced to 80% in the experiments).

- **Meal spacing** - This intervention reduces the peaks of congregation in the mess hall around lunch and dinner by spacing out when people come to eat. This will keep the mess hall well under capacity, reducing the contact rate.

The analysis window provided by Gryphon allows users to examine how the shipboard population is affected by a pandemic across simulations (e.g., different seeded infections, different intervention strategies). These include the number of individuals at disease states S, E, I and R, the daily/accumulative morbidity rate and the daily/accumulative

Figure 4. The visualization of the propulsion compartment in SDDT
mortality rate. Figure 5 shows a snapshot of an analysis window in Gryphon, where the number of infected individuals changes over time.

**Figure 5.** The analysis window for the number of infected individuals over time

4. EXPERIMENTAL RESULTS

We simulate a disease outbreak for varying population sizes at \( n \times 10^6 \), where \( n \in \{0.25, 0.33, 0.5, 1.0, 2.0, 3.0, 4.0\} \). We choose four as a maximal value of \( n \), as usually there are about 4000 sailors and marines onboard a U.S. aircraft carrier. For each experiment we seed 1 to 10 infections in the department of NCO mess supply to initialize the simulation. We will not change the number of compartments when we scale the shipboard population size up or down. Finally, we assume that all Navy ships we simulate have a consistent number of physical locations for sailors.

The following two conditions are changed for each set of stochastic simulations,

- Population size: The number of total sailors \( N \) is selected from \( n/4, n/3, n/2, n, 2n, 3n, 4n \), where \( n \) is equal to 1065.
- Initial infected population: ranging from 1 to 10.

We calculate the coefficient of variation (CV) for each set of stochastic simulations with 100 runs. For simplicity, we only describe simulation results for total case numbers. The other two metrics, peak time and peak value, demonstrate similar behaviors as exhibited by total case numbers.

4.1. Effects of the number of seeded infections

Figure 6 shows the value of CV for total case numbers across various population sizes. We find that the possibility of disease outbreaks is highly sensitive to the number of the seeded infections. From Figure 6, we can see that the CV for a given population size is high when the number of seeded infections is small. For example, the minimal CV for all population sizes is around one. We also can find that the value of CV tends to decrease greatly with varying population sizes if we seed with more than two infections.

**Figure 6.** Coefficient of variation for the total case number across different population sizes

Our results indicate that the randomness of stochastic simulations can be significantly reduced if there are more infections seeded in the population. The reason is that a disease caused by a seeded infection can either die out or become an outbreak depending on the randomness in transmission and mixing dynamics. Figure 7 shows the percentage of the runs where a seeded disease dies out among 100 runs. Per [12], we define that a disease dies out in a simulation if the case number does not exceed 1% of the total population in 48 hours. From Figure 7 we can find that the probability of diseases dying out is relatively high in stochastic simulations if there is only one seeded infection. The probability of the disease dying out is close to zero if we seed three or more infections in a ship.

**Figure 7.** The percentage of disease outbreaks dying out across different seeded infections
4.2. Effects of simulation conditions

In the second experiment we study whether the simulation results are consistent across different experimental conditions. One hypothesis is that the randomness caused by small seeded infections can be removed after we run more simulations. Figure 7 shows the changes of CV for total case number across different rounds of simulations, where we seed three infections in the department of NCO mess supply.

Figure 8 shows that the CV for total case number oscillates between 0.6 and 1 if the number of sailors in a ship is less than 1065. Moreover, we have discovered that there are some “jumps” for the CV value of the total case number when the number of sailors goes beyond 1065. The scale of these jumps becomes smaller and smaller as the number of simulations increases. These jumps in terms of CV are caused by those simulation runs where a seeded disease dies out. The average effects of these runs with diseases dying out become smaller and smaller when the number of simulations increases.

4.3. Effects of population size

In a stochastic simulation, the probability of a disease outbreak spreading depends on the number of people being contacted, and the probability of the disease being transmitted. In a larger population an infected individual tends to contact and mix with more people in the main gathering locations such as quarters and messes of a ship. If we seed more infections in a large ship, crowded populations in both quarters and messes appears to contribute to the wide spread of a shipboard disease. In the stochastic SEIR model, this corresponds to a high mixing rate in the locations [2]. Our experimental results show that three or more infections will likely cause the emergence of a disease outbreak if none of intervention strategies are applied.

In Figure 9 we examine the value of CV for total case number from the perspective of different population sizes. We can see that the value of CV is high if the number of sailors in a ship is small. This indicates that the actual disease transmission for a small population size is highly unpredictable at the start of the simulation when we seed with only one or two infections. From Figure 7, we know that about 42% of simulation runs in a population with 250 sailors will not lead to an outbreak if we only seed one infection. However, with the progression of a disease outbreak, our system can rapidly predict the best and worst scenarios of a disease outbreak over time with an increasing confidence even for small population sizes.

One hypothesis is that the value of CV for small population sizes will be close to the one for large population sizes if we keep increasing the number of seeded infections. From Figure 10, we find that the value of CV for small population sizes does not drop as much as we expect when we increase the seeded infections from five to ten. There is a clear margin area for CV between the population size of 500 and the population size of 1000. The irregular patterns indicate that, in addition to seeded infections, the shipboard population size does affect the system dynamics of shipboard disease outbreaks [10]. This phenomenon may be related to the chaotic behaviors of the non-linear stochastic simulation systems for infectious diseases and will be addressed in our future work.

5. DISCUSSION

In this section we discuss the implications of our sensitivity analysis for the stochastic simulations and several possible improvements of our simulation engine to accurately model shipboard disease outbreaks.

Multiple modes of stochastic systems: Nonlinear dynamics and chaotic behaviors have been studied for determinis-
important role for infectious disease transmission dynamics within a ship. Currently, Gryphon uses a uniform mixing rate to model the spread of a disease among individuals in a specific location, but in practice the rate is highly dependent on the specific disease and needs to be customized for each disease. The mixing rate may rely on the size of the location and ventilation for air-borne diseases and contact-based diseases.

Disease-sensitive behaviors of sick individuals: Gryphon currently uses exactly the same patterns to model sick individuals for all diseases. The behaviors of sick individuals can vary from one disease to another. In order to accurately predict the operational level for a shift of sailors in a given period, disease-specific behaviors of individuals need to be carefully studied and implemented in Gryphon. For example, a sick individual may not be able to work for a severe disease such as cholera, but he may still work after being infected by a less severe disease.

6. CONCLUSION

The sensitivity studies of Gryphon indicate that the probability of a disease outbreak in large military ships can be high if three or more sailors get infections during a port visit. The situation can become even worse since early detection of a disease outbreak might not be available aboard a ship. Even when the disease is detected by medical officers, the effectiveness of many non-medical interventions such as social distancing may be very limited given that the shipboard personnel cannot be dispersed in the closed environment with a limited number of compartments [12]. [5] suggests that rapid mass vaccination may be best solution to disease outbreaks in confined settings like a navy ship.

One caveat is that we do not reduce the number of departments when we decrease the total number of sailors. Therefore, the actual value of the coefficient of variation for total case number in a small ship may be lower than that reported here due to few number of locations for mixing. However, we do not anticipate any changes of our conclusion regarding the transmission dynamics regarding the total population and seeded infections. The reason is that crowded locations such as quarters and messes exist for any ship and the likelihood of mixing at these locations will dominate the transmission dynamics in stochastic simulations.

7. 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 James Patrey at the Office of Naval Research for his support and valuable comments. This research has been sponsored by the Office of Naval Research (ONR) under Contract No. NOOO14-07-C-0014.
REFERENCES


BIOGRAPHY

Dr. Bin Yu, a Senior Research Scientist at Quantum Leap Innovations, has more than ten years of research experience in modeling and simulating complex adaptive systems from robotic sensor systems to biological systems. Currently, he is leading the development of Gryphon, a flexible and scalable modeling and simulation platform for infectious diseases. Before joining Quantum Leap, Dr. Yu was a Postdoctoral Fellow at Carnegie Mellon University, Pittsburgh, PA. Dr. Yu received his Ph.D. in Computer Science from North Carolina State University in 2002.

Dr. Jijun Wang is a senior research scientist and the leading software architect at Quantum Leap Innovations, Inc. He received his Ph.D. from the School of Information Sciences at the University of Pittsburgh in 2008. His current research involves complex dynamic system modeling and simulation, as well as human-computer interaction. He has B.S. and M.S. degrees in Electo-Mechanical Engineering from Tsinghua University of China.

Michael McGowan is a Mathematician/Software Developer at Quantum Leap Innovations. Since joining Quantum Leap he has been heavily involved in both the design and implementation of software and algorithms related to the Gryphon modeling and simulation platform. Recently he has begun work related to flexible and scalable knowledge discovery and hypothesis generation. He earned his M.A. in mathematics from the Johns Hopkins University in 2007 and a B.S. in mathematics and computer science from John Carroll University in 2006.

Dr. Ganesh Vaidyanathan, the Chief Scientist at Quantum Leap Innovations, has a successful track record and significant experience in delivering innovative solutions to solve complex business problems. Formerly a Senior Research Associate with DuPont, he led their empirical data mining efforts and is the inventor of the proprietary InfoEvolve suite of data mining tools based on the marriage of information theory with genetic algorithms. At Quantum Leap, he is leading modeling and simulation efforts to enable scalable knowledge discovery and planning around infectious/chronic disease.