BioWar: Scalable Multi-Agent Social and Epidemiological Simulation of Bioterrorism Events

Dr. Kathleen Carley[1,2,6], Dr. Douglas Fridsma, MD[4,7], Dr. Elizabeth Casman[3,6], Neal Altman[6], Dr. Jack Chang[5,6], Dr. Boris Kaminsky[6], Demian Nave[5,6], and Alex Yahja[1,2,6]

[1]*Institute for Software Research, International*
[2]*Center for the Computational Analysis of Social and Organizational Systems*
[3]*Engineering and Public Policy Department*
[4]*University of Pittsburgh Medical Center*
[5]*Pittsburgh Supercomputing Center*
[6]*Carnegie Mellon University and* [7]*University of Pittsburgh*
Pittsburgh, Pennsylvania 15213

Abstract

The reality of life is embedded in social networks. Understanding how social networks affect disease propagation and how the consequences of disease change social networks is critical for modeling early manifestation of diseases on diverse human activities. Moreover, disease outbreaks do not happen in vacuum, they are constrained by physical, economical, technological, media, health, and governmental infrastructures. It is critical to be able to systematically reason about the nature of outbreaks, early manifestation of diseases on diverse human activities, the potential of media and inoculation campaigns, and the relative value of various early warning devices. There is simply not enough actual data on bioattacks. What is needed is a cost-effective, ethical system for reasoning about such events. Focusing on sample-based statistical methods, conventional epidemiology ignores the reality of social networks and their dynamics, thus is ill-suited for this. BioWar -- a scalable city-wide multi-agent network model – addresses the above needs. BioWar uses cognitively realistic agents embedded in social, health, and professional networks to describe how people acquire diseases, manifest symptoms, seek information, and seek care. We describe BioWar and present results showing the efficacy of the approach.

Contact:
Dr. Kathleen Carley
Institute for Software Research, International
Center for the Computational Analysis of Social and Organizational Systems
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
Tel: (412) 268-6016
Fax: (412) 268-2338
Email: kathleen.carley@cmu.edu

Keywords: scalable systems, social networks, epidemiology, disease manifestation, early detection, multi-agent systems, bioterrorism, biosurveillance

BioWar: Scalable Multi-Agent Social and Epidemiological Simulation of Bioterrorism Events

Dr. Kathleen Carley, Dr. Douglas Fridsma, MD, Dr. Elizabeth Casman, Neal Altman, Dr. Jack Chang, Dr. Boris Kaminsky, Demian Nave, and Alex Yahja

Introduction

Throughout history, the spread of disease has been linked to social and environmental factors. When people live in close proximity to each other, the risk of disease spread — both infectious and non-infectious — increases. Public health measure and aggressive vaccination policies have reduced the risk of disease spread, and in some cases, have eradicated diseases altogether. Smallpox for example, has been eradicated from the natural population through an aggressive and systematic vaccination program.

However, public health officials remain vigilant to control and prevent disease outbreaks. Recent attention has turned to bio-terrorism and the risk that terrorists might use weapons of mass destruction to infect vulnerable populations. In these scenarios, the social networks in which we live can become the vehicle through which disease is spread. How best to detection and respond to a man-made disease outbreak is not always clear. The goal of our research is to develop tools to systematically think about, model and analyze how diseases would spread through socially connected groups.

In trying to prepare for attacks, policy makers need to be able to think through the consequences of their decisions in various situations. Consider, for example, trying to decide if all US citizens should be vaccinated for smallpox. Speculations abound as to the potential devastation that smallpox could wreak. Medical experts, scientists, and policy makers need a way of thinking through the morass of complex interconnections to understand whether different inoculation or containment strategies will be effective. Unfortunately many existing models are quite limited in that they only apply to a single disease, discount factors such as the urban geography which can influence disease spread, or discount how people use their social networks (who is a friend with whom) to pass information such as when to go to the doctor to be treated. In general, being able to estimating the impacts of large scale biological attacks and the efficacy of containment policies is necessary from intelligence and planning perspective and requires reasoning about social response and disease processes as a complex social system. The recent case of an atypical pneumonia – the Severe Acute Respiratory Syndrome -- which so far claims 54 lives among the 1,550 infected emphasizes the importance of and close linkage between social networks, transportation networks, and early detection.

While conventional epidemiology has achieved significant successes in managing diseases and epidemics, the approach is inadequate in dealing with the high noise to signal ratio in case of bioattacks where the focus is early detection. Part of the reason for this is that conventional epidemiology does not inherently account for spatial, geographical, and social dimensions in its mathematical modeling of diseases. There is a recent progress in spatial epidemiology which takes into account the spatial and geographical dimensions directly [Lawson 2001][Gimblett 2002], however it does not consider the social dimension nor the intricate interplay between different dimensions. The interplay would affect the shape of signal fed into detection systems. Moreover, the organizational, media, government/institutional action dimensions also play significant roles in the event of bioattacks. There is also work in Stanford Medical Center of BioSTORM (Biological Spatio-Temporal Outbreak Reasoning Module) which uses Bayesian correlation in its Bayesian network reasoning, in addition to Kalman filter, of its RASTA engine [Buckeridge et. al. 2002]. While BioSTORM is a step closer to real-time disease surveillance system, it ignores social networks, physical constraints, and the dynamics of diverse human activities. Furthermore, the Bayesian networks it uses have inherent difficulties modeling abnormal and rare signal not anticipated in the prior distribution. This is critical as rare, minuscule, & abnormal signals are what bioattacks most likely produce, especially during the first critical 48 hours of early manifestation.

To aid the analyst and decision maker we are developing BioWar. In BioWar we are combining state-of-the-art computational models of social networks, communication media, disease models, demographically accurate agent models, wind dispersion models, and a diagnostic error model into a single integrated model of the impact of an attack on a city. Unlike traditional models that look at hypothetical cities, in BioWar the analyst can examine real cities using census, school track, and other publicly available information. Moreover, rather than just providing information on the number of infections, BioWar models the agents as they go about their lives – both the healthy and the infected. This enables the analyst to observe the repercussions of various attacks and containment policies on factors such as absenteeism, medical web hits, medical phone calls, insurance claims, death rate, and over the counter pharmacy purchases. Analysts could use BioWar to ask and answer "what if questions" of the form "what would happen to this city if three people returned from vacation with smallpox?" Moreover, these questions could

be answered from the person's perspective, the perspective of the organization where he or she works, and the city's perspective.

Social Networks

The world we live in is a complex socio-technical system [Carley and Gasser 1999]. Although, social, organizational and policy analysts have long recognized that groups, organizations, institutions and the societies in which they are embedded are complex systems; it is only recently that we have had the tools for systematically thinking about, representing, modeling and analyzing these systems. These tools include multi-agent computer models and the body of statistical tools and measures that have arisen in social networks. Using these tools we can begin to think about these complex socio-technical systems in which people live, and in which diseases can propagate. Such an illustration is particularly salient in lieu of the tragic events of October 2001 when anthrax was distributed through the US postal system.

Over the past few years, major advances have been made in the ability to model and analyze social networks and graphs of agent interactions. Techniques including the P* family of tools, and the graph level metric tools (such as that for clustering networks) can be used to locate patterns. Among the data that can be represented as graphs are interaction or communication networks, monetary networks, inter-organizational alliances, mental models, texts, web pages, who was present at what event, story lines. The combination of these techniques with machine learning is likely to be especially powerful for locating anomalies, assessing coherence, and locating underlying fundamental patterns.

Multi-agent network models, if based on known information about general or specific characteristics of groups, can suggest general or specific guidance about how to affect or protect the underlying group, organization or society. Exactly what these models can address depends on the model. For example, food poisoning is often discovered through analysis of social links and shared experiences. More worrisome, infectious bioagents spread through the social networks of people who are infected. While non-contagious bioagents such as anthrax do not spread through social networks, it affects the behavior of social networks and more importantly social networks would facilitate piecing up together the puzzle of non-contagious attack (e.g., if anthrax is released at a stadium, sport-fan networks would show a malignant pattern). Since BioWar has imbedded within it agents that are linked through social, work, and knowledge networks, it allows us to recognize patterns with the network that may suggest a disease outbreak or bioterrorism event has occurred.

BioWar is an effort to develop a scaleable and precise simulation tool to examine disease propagation and agent behavior in response to disease and illness. We believe it will serve to help researchers understand, predict, and analyze weaponized biological attacks at the city level and engage in "what-if" analyses to help inform decision-making in this complex socio-technical policy domain. For example, it can be used in a "what-if" mode to examine the impact of and response to various weaponized attacks for contagious and non-contagious diseases under high-alert and no-alert conditions.

BioWar

The following sections describe the components of BioWar and the underlying rationale of the disease, network, and agent models. Graphically, these components are illustrated in Figure 1.
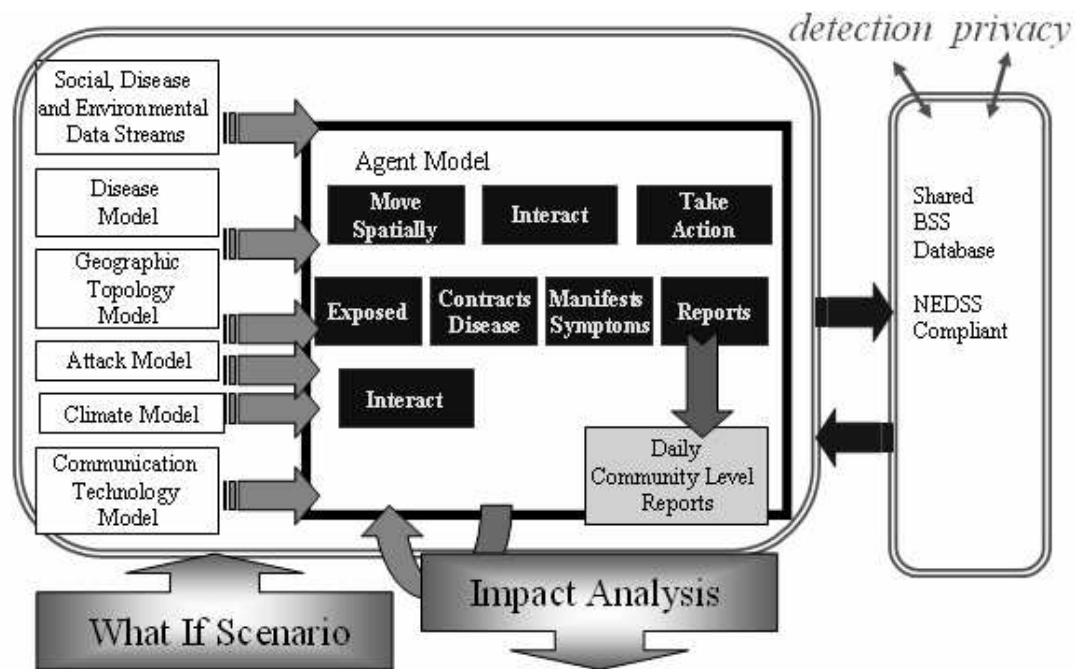
Figure 1. The design of BioWar. Input characteristics of the agents, diseases, geography, and communication technology are used by the agent model to change behavior. The raw output can be used to test early detection algorithms, and hypothetical what-if scenarios used to analyze policy impact.

Disease Model

In constructing our disease model, we used historical accounts of known anthrax releases, documents from the October, 2001 bioterrorism attacks, and known disease knowledge bases. The disease model considers risk factors (to identify subpopulations at more risk of developing a disease), symptoms, diagnostic tests, and disease progression to describe how a bioagent release would affect an individual, and focuses on how these individual characteristics affect behavior and the data that might be collected from non-routine sources. We have drawn on the experience of other medical expert systems developed to assist in diagnosis to ground our disease model in well-founded medical knowledge representations [Miller 1982].

Representation of Disease Course

We divide the disease symptoms into three groups, divided by time course: those symptoms and risk factors that predispose a person to development of a disease, those symptoms that might prompt a visit to the doctor or pharmacist, and those symptoms or tests that would help with diagnosis or may affect mortality and morbidity. Pre-existing risk factors affect who has a higher probability of getting a disease, and will use information about agent demographics during simulation to afflict agents with disease. For example, an elderly or immunosupressed patient may be at a higher risk of contracting anthrax if exposed. Once infected, the patient then begins to acquire symptoms as their disease progresses. This affects patient behavior in the simulation. For example, a patient with multiple, non-specific symptoms might be more likely to go to a physician or visit a pharmacy. When a patient has developed enough symptoms to warrant further investigation, a portion of the patients who are ill will seek medical care. It is at that point a patient may have additional tests performed, use more healthcare resources, and likely have their information recorded in databases that serve as a source for early detection. This progress is shown in Figure 2.
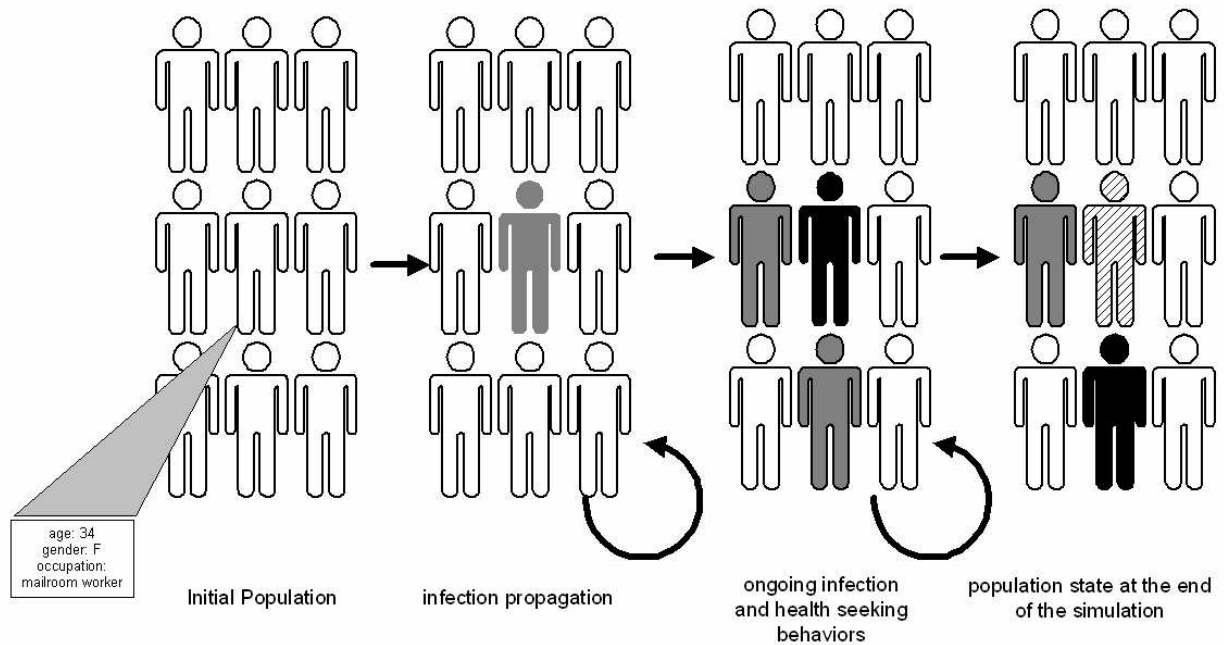
Figure 2. Example population changes for a simulated contagious disease. (From the left.) The population is initialized to match the demographics and social characteristics of the population under study. Once the simulation begins, individuals become infected, and can propagate the infection to people nearby in their social network. As patients manifest symptoms (black color), they change their behavior, and may seek medical care, or stay home from work (reducing the risk of spreading the infection). The final population state may include patients who have died, who currently have symptoms or who have changed their behaviors in response to their disease symptoms.

Risk factors

As we discovered in the anthrax releases in October 2001, certain demographic groups are more likely to be susceptible to particular diseases than other. These risk factors increase a person's susceptibility to diseases either through host factors, or environmental factors to which that person is exposed. For example, individuals who have contact with animals (sheep shearers, for example), are more likely to contract cutaneous anthrax than other occupations. Risk factors are distributed to individuals in the population, according to demographic characteristics, and we can determine these risks based on age, sex, and disease prevalence.

Symptoms

Symptoms are important in our model in that they motivate behavior changes. People with symptoms will stay home from work, will visit their doctor or pharmacist, and will change their patterns of interacting with others. This change in interaction patterns leaves a "footprint" on the data streams used to detect bioterrorism events. It is this manifest pattern of symptoms that enables early detection of bioattacks.

The symptoms have two different measures that influence which symptoms a person gets and how that changes their behavior. The frequency measure is a qualitative measure of how frequently people with a particular disease will manifest this symptom. The evoking strength is a qualitative measure of how frequently someone with a particular symptom will think they have a particular disease. For example, the symptom of cough may have a very high frequency among patients with anthrax—however cough would have a low evoking strength because people with a cough do not immediately think they have anthrax.

Frequency

Frequency is a number between 1 and 5 which answers the question: "In patients with disease x, how often do you see symptom y?". For example, patients with the diagnosis of anthrax will have a fever frequency of 5 – nearly all patients with anthrax will have fevers at some point in the course of their disease.

Evoking strength

Evoking strength is a number between 0 and 5 which answers the question: "When you see symptom y, how often do you think the cause is disease x?" For example, fever symptoms are not specific to any one disease – in our disease profile of anthrax, fever is given a ES of 1. However, widened mediastinum is a more specific manifestation of anthrax – in patients who have a widened mediastinum, the diagnosis of anthrax should be considered. Evoking strength is similar to specificity.

How symptoms relate to agent behavior

The frequency and evoking strength associated with individual symptoms are used in different ways within BioWar. Although individuals will get symptoms based on the symptom frequency for the disease that they have, they will actually alter their behavior based on the evoking strength of that symptom for a serious illness. For example, if a person contracts an anthrax infection and displays fevers and chills, in the absence of a known attack, may consider their symptoms to be an influenza infection or simple cold and not significantly alter their behavior. However, if they began heaving shortness of breath, chest pains, or other symptoms more suggestive of a serious infection, they would likely stay home from work, seek medical attention or go to an emergency room.

Diagnostic tests

Diagnosis is simulated based on the apparent symptoms. It produces certain accuracy in the form of ROC curve similar to the real diagnosis. The result of diagnosis would determine whether a person is treated properly and recovered. To determine which disease a person has, the evoking strengths among potential diseases are compared and the highest one is chosen as the diagnosed disease.

Representation of Social Networks

Each agent is linked to other agents in the simulation population by a set of links (relationships) modeling a social network. Each link is a pair that identifies an agent and the relationship type. Agents may be linked unidirectionally or reflectively. Relationships currently implemented are:

```
SPOUSE,              /* Person is spouse. */
PARENT,              /* Person is parent. */
SIBLING,             /* Person is sibling. */
CHILD,               /* Person is child. */
OTHERFAMILY,         /* Person is other family member. */
COWORKER,            /* Person is coworker. */
GROUPMEMBER,         /* Person is member of group. */
NEIGHBOR,            /* Person is neighbor. */
FRIEND,              /* Person is friend. */
ADVISOR,             /* Person is advisor. */
SCHOOLMATE,          /* Person attends same school. */
OTHER                /* Person has other relationship. */
```

The relationship types were drawn from the GSS survey data: http://www.icpsr.umich.edu:8080/GSS/homepage.htm (Codebook Indices -> S -> Social Networks) with one additional category (SCHOOLMATE) for younger agents. The overall network size and distribution was drawn from Klovdahl, Alden S. "Social Networks in Contemporary Societies" along with some target numbers for individual relationship counts.

Representation of Agent Interaction

Agents interact with each other based on CONSTRUCT model and on the spatial and social network proximities. In the core of CONSTRUCT [Carley 1991] lies the co-evolution and emergence of communication and knowledge networks, driven by homophily and expertise seeking. The principle of homophile states that people are more likely to communicate with people that are similar to them. Similarity in CONSTRUCT is assessed by attributes such as age, sex, race, prestige, occupation, educational level, social class, belief, culture, interests, or attitudes. The principle of expertise seeking states that the information poor are more likely to initiate communication with the information rich to fulfill their information needs. Moreover, the interaction is mediated by social networks, professional networks, and other networks. One would more likely to interact with family members, friends, and acquaintances than to complete strangers. However due to the nature of the physical world, people would interact to others nearby and also by random chance. Referrals (who knows what and who knows who) also affect the interaction. Figure 3 shows a nice illustration of the complexity of agent interaction. BioWar implemented the interaction based on common knowledge and knowledge difference, and also the interaction based on social network and random chance, as follows:

```
Algorithm Interaction_by_Social_Network
Let A be an agent
Throws a random number R
If R is less than a threshold T
    Choose an interaction candidate C randomly, weighted by location proximity
    Let agent A interact with agent C based on common knowledge and knowledge difference
Else
    Look into agent's ego-network
    Find all agent's ego-network immediate neighbors
    Let N be the total number of neighbors
    Choose an interaction candidate C among the N neighbors randomly
    Let agent A interact with agent C based on common knowledge and knowledge difference
End-of-If
```
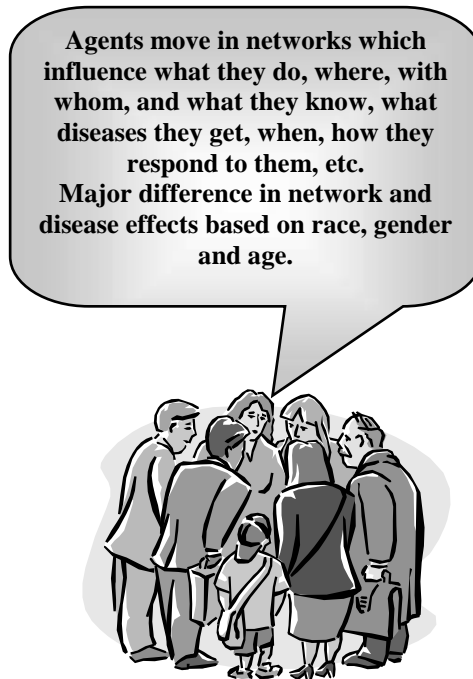
Agents move in networks which influence what they do, where, with whom, and what they know, what diseases they get, when, how they respond to them, etc.
Major difference in network and disease effects based on race, gender and age.

Figure 3. The complexity of agent interaction and disease interference

Representation of Recreation

The recreation code simulates some of the activities people engage in beyond normal routine (time spent at home, work or school) and health related concerns (time spent with doctors, at pharmacies or in hospitals). In the current BioWar model, an agent's activity is mapped 1 to 1 to a location (thus, agents work in workplaces and sleep at home).

The recreation code adds several recreation venues where agents gather and potentially interact:

```
STADIUM,            /* Open air events. */
THEATER,            /* Indoor events. */
STORE,              /* Shopping locations (excludes pharmacies). */
RESTAURANT          /* Eating locations. */
```

Recreation rates were derived from the 1994 EPA Time Use Survey (http://sda.berkeley.edu:7502/cgi-bin/hsda?harcUMD+time) by grouping activity categories to determine the percentage of the day normally spent in recreation and using the time at location (where) data to make a determination as to how much time is spent in each recreation location. (Note: the time use data is not ideally suited to BioWar requirements and the derived rates for recreation locations are only approximations). The data confirms that recreation occurs in many contexts, so much recreation is assumed to occur at the agent's home.

The dataset is sufficiently large to allow some seasonal, weekly and demographic factors to be modeled. For current implementation, tables were constructed for the four seasons (conventional definition with winter starting on December 21, spring on March 21, etc.). Each seasonal table contained separate entries for each day of the week, for young versus old agents and for male versus female. Paired tables were employed: one holding the overall recreational probability and the second determining recreational location (if any since recreation may not occur at a

defined location). In addition to the seasonal data, tables for minor and major holidays were constructed. Since the time use database excluded holidays, the information was scaled up from annual averages for Saturday.

Since the simulator divides each day into smaller time units (four hour "ticks"), an adjustment table was also introduced. The adjustment table allows recreational rates to be adjusted to match the normal day cycle: little recreation occurs during hours of sleep or work and more when the agent would logically have free time (for example in the evenings during the week). In order to reflect the empirical data, these tables are set to preserve overall rates: recreation that is deferred is taken later and recreation taken early reduces the chance of recreation later in the day. These tables were derived for stereotypical patterns and are not empirically based.

The recreation code's role is advisory: it indicates that an agent "wants" to recreate, but does not place the agent at a recreational location.

Representation of Wind

Wind Model is a part of the current implementation. It relates to the generation of the wind information (wind speed and direction) for the whole period of simulation. Generated wind pattern distributions by speed and direction closely match the empirical data for the simulated regions - San Diego, CA, Pittsburgh, PA, and Norfolk, VA published on the www.epa.gov and averaged for the 5 years.

In fact we are interested in wind model mostly at the moment of the attack especially when the attack is performed outdoors and the dispersion of biomaterial is performed though the wind puff movement. We use Gaussian model of the wind dispersion that is well defined and the most popular one. The assumptions of the model are:

- The dispersed material is chemically stable and is not deposited to the ground.
- The lateral and vertical variations of the material concentration can both be described by Gaussian distributions, which are functions of x – downwind distance only.
- The wind direction is constant with height.
- Although in the simplest Gaussian model the wind speed is constant with height our Wind Model calculates the dependence of wind speed with height that may be important when the hazardous material release is not at the ground.

The essential part of the Wind Model is assessing the Pasquill atmosphere stability category for the period of the attack. In the absence of the detailed meteorological data we assigned Pasquill atmosphere stability category based on the wind speed and time of the attack but not the sky condition that may be considered as a reasonable approximation [Barrat 2001].

Atmospheric dispersion modeling is usually performed in the local coordinate system with the origin of the system at the ground level at the point of emission (for the ground releases) or directly beneath the point of emission (for elevated releases). Since the BioWar itself employs geographical coordinate system the Wind Model includes methods for the transformation between these two coordinate systems. The matter was even more complicated when the multiple point releases or airborne release was simulated. In such case the origin of the local coordinate system was put at the location of the agent and the total effect on the agent (the summary dosage) was calculated as a sum of dosages from individual releases.

The dosage inhaled by the agent was calculated using equation below [Turner 1994][Meselson 2001]:

$$Dose = [QB][\pi u \sigma_y \sigma_z]\text{-}1 exp[-(1/2)(y/\sigma_y)^2] exp[-(1/2)(H/\sigma_z)^2]$$

where Q- source strength (number of spores), B – breathing rate (usually for light work $B = 5 *10^{-4}$ m^3/sec), u – wind speed in m/sec, σ_y, σ_z – dispersion parameters that are functions of downwind distance x, H – height of the release in meters.

The Wind Model also includes methods that allow determining whether the agent is located in the downwind zone and how far it is from the point of the release.

Validation of the Wind Model was performed by the comparing historical averaged wind data with the generated data. One of the methods in Wind Model allows to calculate frequency distribution of the wind speed by direction for the cases of wind speed less than 1.5 m/sec, between 1.5 m/sec and 2.6 m/sec, between 2.6 m/sec and 3.8 m/sec, and greater than 3.8 m/sec, and also for the total of all wind speed values.

The comparison between the test data for San Diego and historical 1990 – 1992 average is shown below at Figure 4. Relative difference between test and average historical frequency distribution values is less than 30%. Similar results are for Pittsburgh, PA and Norfolk, VA.
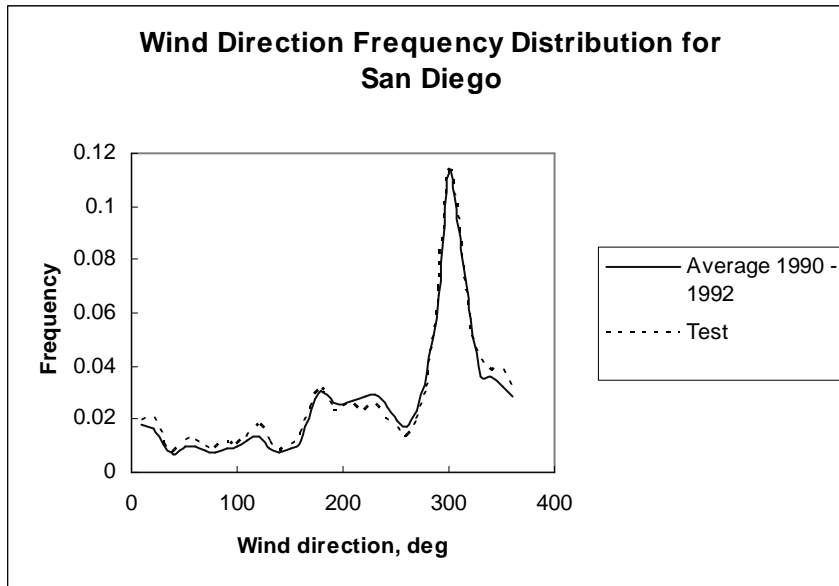
Wind Direction Frequency Distribution for San Diego

Figure 4. Comparison between test wind direction frequency distribution
and average historical data 1990 – 1992 for San Diego.

Representation of Climate

Climate Model is a part of the current implementation. It relates to the generation of the climate information (temperature, pressure, precipitation) for the whole period of simulation. Generated temperature, pressure, and precipitation yearly distributions closely match the historical data for the simulated regions - San Diego, CA, Pittsburgh, PA, and Norfolk, VA published on the www.epa.gov.

Other Data Sources

BioWar uses various other data sources for validating its model. Among the data sources are:
- General Social Survey data: this data set describes the general and comprehensive social characteristics such as work status
- Census 2000 data: this data describes the demographic characteristics of a population, income level, area profiles, birth & death rates, migration rate, estimate of population growth, among others. It also provides census tract cartographic data.
- ArcGIS data set describes the characteristics of geographical area. It provides maps, charts, views, and graphical user interfaces for geographical information systems.
- BigFred data set: this decribes the drug purchase behaviors of a population in a particular area.
- QMR data set: this describes the interrelationship between diseases, symptoms, diagnosis, and treatments, among others. We have 62 diseases with their complete symptoms implemented.

System Implementation

BioWar is designed as a modular system -- a component, or module, in the system interacts with other modules via a simple, published interface of methods. A module conceptually corresponds to a BioWar simulation capability, like disease progression and diagnosis (the "disease module"). This correspondence is not necessarily one-to-one. The agent model, for example, consists of both an agent behavior module and an agent social network module. This flexibility, which is not possible with a similar monolithic design, is crucial for the rapid development of reliable and more realistic simulation models.

BioWar is designed to be reasonably portable, and currently runs under Linux, Windows 2000 and XP, and Tru64 UNIX. For our experiments, we have chosen to use the Alpha processor-based computational resources of the Pittsburgh Supercomputing Center (PSC). Most experiments were performed using the PSC TCS1 system, which comprises 64 4-way Alpha SMP processing elements (PEs), each with 4GB of RAM and 4 667MHz Alpha 21264A (EV6.7) processors. A simulation with about 300,000 agents can take 5 or more hours to complete -- this would be prohibitive if run on a single-node system (e.g. a workstation). The ability to run many concurrent large

simulations was invaluable both to help study the effects of various parameters on the simulation, and to produce a large number of datasets for distribution to the challenge groups.

Results

We have simulated and validated runs for several geographical areas (San Diego MSA, Pittsburgh MSA, Norfolk MSA, a part of Norfolk MSA, and the city of Hampton). The scenarios simulated are for no attack, anthrax attack, smallpox attack, inside and outside buildings. The runs are for 52,000 agents, except for the part of Norfolk, in which we ran 260,000 agents. For the city of Hampton, we simulated all the 140,000 inhabitants without scaling. The length of the runs is 2 years, with each tick of the simulation being 4 hours (a day is 6 ticks of simulation). There are 62 diseases, including a seasonal flu. The simulation starts on September 15, 2001.

For Emergency Department visits, the outputs of all the simulation fall inside the empirical range of 0.056-0.23 visits per person per year, while for doctor office visits, all the outputs fall inside the empirical range of 0.415-1.61 visits per person per year. The empirical ranges are based on CDC Advance Data, Vital and Health Statistics 2002, no. 326 and no. 328, respectively. The appendix A contains the calculations.

Due to the paper's page limit, we only show a representative output curve of the runs.
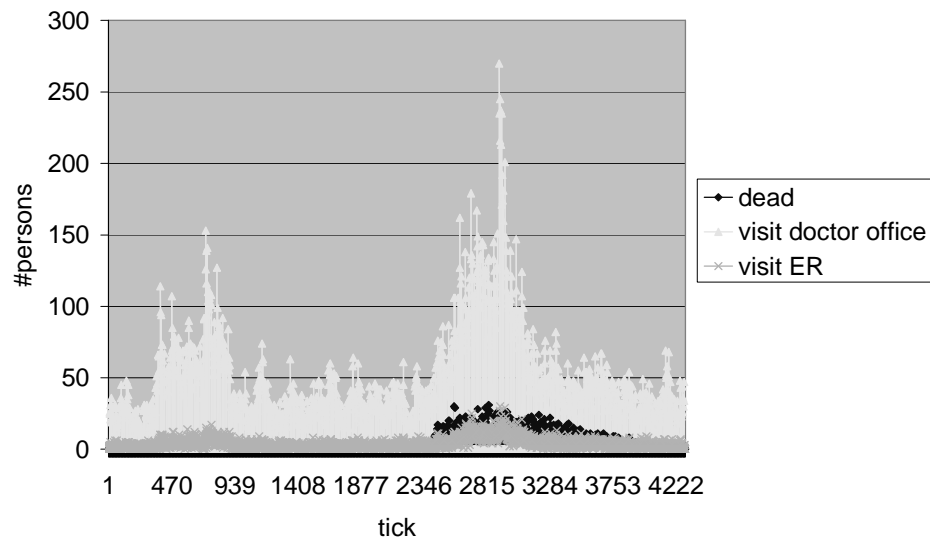

Figure 5. The number of deaths, doctor visits, and ER visits when a medium smallpox attack happens in Norfolk MSA

As shown, soon after a smallpox attack around tick 2346, the number of deaths starts increasing, tailing off significantly long afterwards, due to the infectious nature of smallpox. The number of ER visits follow the same trend as the number of deaths near and after the smallpox attack. During 2 flu seasons in the 2 years of the simulation, the number of doctor visits and ER visits reflects the seasons.

Conclusion

The results show that social multi-agent -- grounded in physical model of the world -- approach is effective in dealing with heterogeneity, dynamic nature, and the complexity of diseases spreading through a population.

Extensions

We are extending BioWar on many fronts, as shown below (listed not according to priority).

- Adding road and transportation networks, especially major streets and subways
- Enhancing the models of hospitals, pharmacies, and doctor offices with queue models.
- Enhancing agent interaction with transactive memory model of CONSTRUCT-TM [Carley and Hill 2001][Carley and Schreiber 2002]
- Examining and verifying each and every causal link in BioWar
- Enhancing the wind & climate model using HPAC, especially precipitation and heat

- Enhancing graphics display and user interface
- Increasing simulation speed by multithreading
- Parallelizing BioWar beyond multithreading
- Getting more empirical data and doing more validation
- Emerging social networks from agent interactions
- Validating the density and structure of social networks
- Adding an automated validation & experiment analyzer engine WIZER
- Adding new data streams such as water usage
- Simulating more cities
- Improving agent & disease seasonal behaviors in addition to flu seasons (e.g., shopping seasons, allergy seasons, etc.)
- Utilizing a synthetic population
- Scaling up of everything to become closer to reality: 120 diseases, 520,000 agents, 20 drugs, etc., or more
- Connecting with the Protégé real-time ontology and data streams

Future Directions

Automatic data collection is an important issue facing BioWar. As we go increasing the complexity of the models in BioWar, the demand for better quantity and quality (timeliness, precision, granularity, etc.) of data also increases. Automatic multi-agent software meta-model, model, and code construction is also critical, as we struggle to increase the fidelity of BioWar. How to combine the opinions of multiple subject-matter experts into BioWar and how to use BioWar multi-agent model to facilitate this is an open question. One pressing policy issue is how to strike a balance between transparency and privacy in the pursuit of realistic BioWar simulation.

References

[Barrat 2001] Barrat, Rod, "Atmospheric dispersion modeling:an introduction to practical applications", Earthscan, 2001, ISBN 1 85383 642 7

[Buckeridge et. al. 2002] Buckeridge, DL, Graham, J, O'Connor, MJ, Choy, MK, Tu, SW, Musen, MA., "Knowledge-Based Bioterrorism Surveillance", American Medical Informatics Association Symposium, San Antonio, TX, November 9-13, 2002.

[Carley 1991] Carley, K.M., "A theory of group stability", American Sociological Review, 56(3): 331-354, 1991

[Carley and Gasser 1999] Carley, K.M., & Gasser, L., "Computational organization theory", in G.Weiss (ed.), *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, pp. 299-330, Cambridge, MA: MIT Press, 1999.

[Carley and Hill 2001] Carley, K. M. and Hill, V., "Structural Change and Learning Within Organizations", in A. Lomi and E. R. Larsen (eds), *Dynamics of Organizations: Computational Modeling and Organization Theories*. Menlo Park, CA: MIT Press/AAAI, 2001

[Carley and Schreiber 2002] Carley K. M., & Schreiber, C., "Information Technology and Knowledge Distribution in C^3I Teams", 2002 Command and Control Research and Technology Symposium, Monterey, CA, 2002.

[Gimblett 2002] Gimblett, H. Randy, "Integrating Geographic Information Systems and Agent-based Modeling Techniques", Santa Fe Institute, Oxford University Press, 2002.

[Lawson 2001] Lawson, Andrew B., "Statistical Methods in Spatial Epidemiology", John Wiley & Sons Publisher, 2001.

[Meselson 2001] Meselson, Matthew, "Note Regarding Source Strength", ASA Newsletter, article 01-6a (www.asanltr.com).

[Miller 1982] Miller RA, Pople HE, Myers JD, "Interist-I, An Experimental Computer-based Diagnostic Consultant for General Internal Medicine", N Engl J Med 1982, 307:468-76.

[Turner 1994] Turner, D. Bruce, "Workbook of Atmospheric Dispersion Estimates: An Introduction to Dispersion Modelling", Lewis Publishers, 1994, ISBN 1-56670-023-X.

APPENDIX

ER VISIT CALCULATION SHEET

This is based on CDC Advance Data, From Vital and Health Statistics, no. 326, 2002 (file ad326.pdf)

Table 1 of the report shows MSA (metropolitan areas) have 37.6 visits per 100 persons per year

Lower Bound Calculation

The lower bound is based on the major disease groups in our 62-disease QMR. Major is defined as having more that a few diseases.
Table 7 of the report gives us
Infectious and parasitic diseases 3.0%
Diesease of the respiratory system 11.8%
totaling 14.8%.
So we have
14.8% X 37.6/100 X number of agents
or
0.055648 X number of agents
or
0.056 visits per agent per year

Higher Bound Calculation

The higher bound is based on all disease groups in our 62-disease QMR.

Table 7 of the report gives us
Infectious and parasitic diseases 3.0%
Mental disorders 2.8%
Diseases of the nervous system and sense organs 5.4%
Diseases of the circulatory system 4.1%
Diseases of the respiratory system 11.8%
Diseases of the musculoskeletal system and connective tissue 5.4%
Symptoms, signs, and ill-defined conditions 16.4%
Injury and poisoning 28.8%

Total 77.7%.

So we have
77.7% X 37.6/100 X number of agents
or
0.232152 X number of agents
or

0.232 visits per agent per year

DOCTOR OFFICE VISIT CALCULATION SHEET

This is based on CDC Advance Data, From Vital and Health Statistics, no. 328, 2002 (file ad328.pdf)

Table 1 of the report shows MSA (metropolitan areas) have 294.6 visits per 100 persons per year.

Lower Bound Calculation

The lower bound is based on the major disease groups in our 62-disease QMR. Major is defined as having more that a few diseases.
Table 11 of the report gives us
Infectious and parasitic diseases 3.1%
Diseases of respiratory systems 11.0%
totaling 14.1%.
Thus we have
14.1% X 294.6/100 X the number of agents
or
0.415386 X the number of agents
or
0.415 visits per agent per year

Higher Bound Calculation

The higher bound is based on all disease groups in our 62-disease QMR.

Table 11 of the report gives us
Infectious and parasitic diseases 3.1%
Mental disorders 5.3%
Diseases of the nervous system and sense organs 8.4%
Diseases of the circulatory system 8.0%
Diseases of the respiratory system 11.0%
Diseases of the musculoskeletal system and connective tissue 7.2%
Symptoms, signs, and ill-defined conditions 6.2%
Injury and poisoning 5.5%

Total 54.7%

Thus we have
54.7% X 294.6/100 X the number of agents
or
1.611462 X the number of agents
or

1.611 visits per agent per year