An Infectious Disease Outbreak Simulator Based on the Cellular Automata Paradigm

Sangeeta Venkatachalam and Armin R. Mikler

Department of Computer Science, University of North Texas
Denton, TX 76207, USA
{venkatac,mikler}@cs.unt.edu

Abstract. In this paper, we propose the use of Cellular Automata paradigm to simulate an infectious disease outbreak. The simulator facilitates the study of dynamics of epidemics of different infectious diseases, and has been applied to study the effects of spread vaccination and ring vaccination strategies. Fundamentally the simulator loosely simulates SIR (Susceptible Infected Removed) and SEIR (Susceptible Exposed Infected Removed). The Geo-spatial model with global interaction and our approach of global stochastic cellular automata are also discussed. The global stochastic cellular automata takes into account the demography, culture of a region. The simulator can be used to study the dynamics of disease epidemics over large geographic regions. We analyze the effects of distances and interaction on the spread of various diseases.

1 Introduction

Nowadays, the problem of emergent diseases and re-emergent diseases like influenza and SARS, have caused increased attention towards public health in general and epidemiology specifically. With the ever-increasing population and ability to travel longer distances in short time, the spread of communicable diseases in a society has been accelerated [16,17]. Growing diversity of the population, and globalization are leading towards increasing interaction among individuals. Constant exposure to public health threats is raising people’s concern and necessitates pro-active action towards preventing disease outbreaks. Further, greater emphasis on infections and epidemics is rooted in the imminent threat arising from bioterrorism. As a result, Public Health professionals have been focusing on identifying the factors in the social, physical and epidemiological environment which aid to faster spread of diseases.

As the significance of Public Health is being recognized, the role of epidemiologists has become more prominent. Epidemiology deals with the study of cause, spread, and control of diseases. The goal of epidemiologists is to implement mechanisms for surveillance, monitoring, prevention and control of different diseases. To accomplish the above mentioned, epidemiologists need to deal with large data

* This research is in part supported by National Science Foundation award: NSF-0350200
sets of disease outbreaks. These data sets are often spatially and/or temporally distributed. It is in fact ironic that, for epidemiologists to study the dynamics of different diseases, it is vital for an outbreak to occur. Epidemiologists have been studying and analyzing the data sets using primarily statistical tools. In the vast variety of infectious diseases, expertise is needed in terms of epidemiologists for every disease. Statistical tools, prove to be inadequate and fragmentary, when focusing on large spatial domains. These tools have been deemed limited, particularly in view of an emerging global computational infrastructure that facilitates high performance computing. Hence, it is imperative to develop new tools that take advantage of today’s computational power, and help epidemiologists to analyze and understand the spatial spread of diseases. The computational tools also enhance the quality of information, accelerate the generation of answers to specific questions and facilitate in prediction. Such tools will take on an important role in surveillance, monitoring, prevention and control of different diseases.

1.1 Cellular Automata

In the domain of computational tools, the Cellular Automata paradigm has been in use for several decades [14]. Nevertheless, in the field of modeling epidemics, this paradigm has rarely been utilized to its full potential [10,14,18]. A cellular automata as defined by Lyman Hurd is a discrete dynamical system, where space, time, and the states of the system are distinct [15]. CA has been exemplified as an array of similar processing units called cells. The cells arranged in a regular manner constitute a regular spatial lattice. Figure 1 shows a regular lattice of cells. The fundamental property of each cell is a state, where the states of cells change based on an update rule, either local or global. The update rule is applied synchronously throughout the lattice and the state transitions of the cells are based on few of the close by cells, known as the neighborhood. For a two-dimensional lattice the most common neighborhoods defined are von Neumann and Moore neighborhood as shown in figure 1-15. In the von Neumann neighborhood, the state of cell \( C_{i,j} \) depends on the states of the four neighborhood cells namely \( C_{i+1,j}, C_{i-1,j}, C_{i,j+1}, C_{i,j-1} \). In the Moore neighborhood, the state of cell \( C_{i,j} \) depends on the states of the eight neighborhood cells namely \( C_{i+1,j}, C_{i-1,j}, C_{i,j+1}, C_{i,j-1}, C_{i+1,j+1}, C_{i-1,j-1}, C_{i+1,j-1}, C_{i-1,j+1} \).

As mentioned before, the CA’s evolution is based on a global update rule applied uniformly to all the cells. The signature of this rule can be thought of as a state transition from time \( t-1 \) to \( t \). As shown in the figure 2 the state of the center cell changes to a state, which is in majority among the cells in the neighborhood. The update rule determines the deterministic or stochastic behavior of a CA. Stochastic behavior is seen by probabilistic update rules in non-deterministic state transitions.

Our efforts to design and implement a Cellular Automata based simulator has been necessitated by the need to study the dynamic of spread of a vast number of infectious diseases. Towards this goal, this paper proposes the use of CA paradigm to simulate an infectious disease outbreak. Specifically, this paper focuses on the design and evaluation of EPI-SIM, a global disease outbreak sim-
ulator. The following section summarizes some of the research effort in modeling disease epidemic and highlights principle approaches. The design of EPI-SIM is discussed in Section 3. Section 4 presents the experimental analysis and results of the simulator. Section 5 discusses the Geo-Spatial model and the approach towards the global model to account for different demographics. Section 6 concludes the paper with a summary and direction for future work in the area of modeling infectious diseases outbreaks.

2 Related Work

Most of the work in modeling infectious disease epidemics is mathematically inspired and based on differential equations and SIR/SEIR model [3]. Differential equation, SIR modeling rely on the assumption of constant population and neglect the spatial effects [5,6]. They often fail to consider individual contact/interaction process and assume populations are homogeneously mixed and do not include variable susceptibility. Considerable research has been conducted in SIR(Susceptible, Infectious, Recovered) modeling of infectious diseases using a set of differential equations. Both partial and ordinary differential equation models are so deterministic in nature that they neglect the stochastic or probabilistic behavior [8]. Nevertheless, these approaches/models have been shown to be effective in regions of small population [8]. Other approaches for modeling disease epidemics have been using mean field type approximations [12]. Even though the MFT models are similar to the differential equations, they add a probabilistic nature by adding different probabilities for the mixing among individuals. Although, according to Boccara [5] mean field approximations tend to
neglect spatial dependencies and correlations and assume that the probability of the state of cell being susceptible or infective is proportional to the density of the corresponding population. This approach relies on the quantitative measures to predict local interaction. Boccara and Cheong [5] study the SIS model of spread of infectious disease in a population of moving individuals, thereby introducing non-uniform population density. In every update the cells take up a state of being either susceptible or infectious and randomly choose a cell location to move to.

Ahmed et al [2] model variations in population density by allowing cyclic host movement. Other approaches in modeling variable susceptibility of the population, have been done by inducing immunity in the population. Ahmed et al [1] introduce incubation and latency time, and suggest that the parameters have an accelerating impact on the spread of a disease epidemic. Nevertheless, the underlying assumption is spontaneous infection of individuals. Boccara and Cheong [6] concentrate on SIR epidemic models and take into consideration the fluctuation in the population by births and deaths, exhibiting a cyclic behavior with primary emphasis on moving individuals. Di Stefano et al [8] have developed a lattice gas cellular automata model to analyze the spread of epidemics of infectious diseases. The model is based on individuals, where individuals can change their state independent of others and can move from one cell to other. However, this approach does not consider the infection time-line of latency, incubation period, and recovery which have been shown to be important to model a disease epidemic.

3 EPI-SIM Disease Outbreak Simulator

In our model the basic unit of cellular automata is a cell, which may represent an individual or a small sub-population. For each cell we use the Moore (8) neighborhood definition. Each cell can be characterized with its own probability for risk of exposure, probability of contracting the disease and state. Unlike the SIR model, every cell comes in contact with the cells in its defined neighborhood. The time-line for infection that we consider is shown in figure 3. However, the moore neighborhood is restricted in modeling population demographics and travel patterns. The limitation is eliminated in the next version of the simulator with a global neighborhood which will be proposed in future publications. The following sections discuss the definitions, features and rules of the model and simulator.
3.1 Definitions

In order to understand the functioning of the simulator, we define definite number of states a can exists in, and define the infectious time-line. The following section describes the different states and definitions considered in the model.

States of a Cell
State ‘S’ for Susceptible is defined as the state where, the cell is capable of contracting a disease from its neighbors. In the infectious state, ‘I’ the cell is capable of passing on the infection to its neighbors. In the recovery state, ‘R’ the cell is neither capable of passing on the infection, nor is capable of contracting the infection.

Parameters for the Simulator
Infectivity $\psi$, at any given time is defined as the probability of an susceptible individual to become infectious, if it has an infectious cell as a neighbor. Latency $\lambda$, is defined as the time period between, the cell becoming infected and it becoming infectious. Infectious period $\theta$, is the period of time, when the infected cell is capable of spreading the disease to other cells. Recovery period $\rho$ is defined as the time period, the cell takes to recover, wherein it is neither capable of passing on the infection, nor is capable of catching the infection.

3.2 Rules for Spread of Disease

The following rules are applied to the CA for simulating the spread of the disease. The rules describe the state transitions of individual cells.

1. A cell’s state changes from susceptible S to Latent L when it comes in contact with an infected cell in its defined neighborhood. The cell acquires the disease from the infected neighbor based on the probability of given by the parameter of infectivity $\psi$. The cell remains in the latent state for the number of time steps (updates) as defined by the parameter latency $\lambda$.

2. The state of the cell changes from latent L to infectious I after being in state L for the given $\lambda$. In this model we assume for simplicity, that every cell exposed to the pathogen, will become infectious. In the state I, the cells are capable of passing on the infection to neighborhood cells. For example if for a disease $\mathcal{D}$, $\lambda = 2$ units, then after two time steps the cell will enter the infectious state I.

3. After a time period, defined by the infectious period $\theta$, the state of the cell changes from infectious I to recovered or removed R. Once the cells enter the state R, the cell is no more capable of passing on the infection.

4. From the state R, the cell’s state changes back to either susceptible S or it remain in state R, signifying complete immunity. The ‘healing mode’ turned on determines the transition from state R to state S and vice versa.
3.3 Features
While modeling a disease epidemic, few parameters that are considered important are neighborhood radius, contact between individuals, infection probability (variable susceptibility), immunity, latency, infectious period and recovery period. The simulator is highly parameterized to let the user change and modify the above parameters. The neighborhood of every cell can be changed from a 8 neighborhood to 4 neighborhood depending on the region being simulated and the contacts among the individuals of the region. As mentioned, the infection probability represented as infectivity $\psi$ is a significant parameter for the spread of a disease. In the case of our model, $\psi$ is based on the virulence of the disease and contact rate among individuals. For some diseases individuals attain lifetime immunity, after being infected, while for disease like common cold, individuals attain temporary immunity. Thus, to take this fact into consideration, the simulator has a feature of healing mode. With the healing mode enabled the simulation is executed in a mode that forces cells to turn into susceptible after the recovery state and with healing turned off, the cell attains complete lifetime immunity.

As mentioned above, the infection time-line is also an important factor in modeling a disease epidemic. Thus the time periods of latency $\lambda$, infectious $\theta$, and recovery $\rho$ are all expressed as time units, for example, latency of two days, can be represented as $\lambda=2$ units. The simulator allows the user step through the simulation at each time step, or execute it continuously. We will see in the next section, how changing these parameters, can change the dynamics of spread of diseases.

4 Experiments and Results
An epidemic is a severe outbreak of an infectious disease which spreads rapidly to many people. For example, the occurrence of Influenza in a region is considered as an epidemic. When a disease spreads to larger geographic regions or throughout the world it is known as pandemic.

Moving along the same direction an endemic is defined as a disease that is always present in certain group of the population. Using our model we show both an epidemic and endemic. An epidemic is characterized by an exponential growth of the infected individuals in a population. In the case of an endemic the number of infected individuals fluctuates around a mean, there is no exponential growth.

Experiments were conducted on a 140 by 140 grid cellular automata with different values of $\psi$, $\lambda$ and $\theta$. The results in this section represent the mean over multiple random experiments and different random graphs of the same type. The analysis of results in this section have been conducted with reference to the above definitions.

4.1 Analysis of Variation in Infectivity $\psi$
As mentioned earlier, $\psi$ is an important factor in the analysis of spread of a disease. Figure 4(a) and 4(b) show the results of executing the simulation of a
The growth of number of infected individuals per time step is represented for $\psi = 7$.

The growth of number of infected individuals per time step is represented for $\psi = 15$.

**Fig. 4.** Variation in spread for different $\psi$’s

disease $D$ with $\psi$ of 7 and 15 respectively. Figure 4(a) depicts that the number of infected reached around 500 in 300 time steps, whereas in figure 4(b) with $\psi$ of 15 the number of infected reached around 10000 in 300 updates. Thus, this depicts that the growth is rapid with $\psi$ of 15 as compared to $\psi$ of 7. Figure 5(a) shows the comparison of the different values $\psi$’s, which are 7,10,12,15. The curves represent the growth rate. The curve is much steeper for $\psi$ of 15 as compared to others. The experiment was conducted with the $\lambda = 2$ units, $\theta = 3$ units and $\rho = 2$ units. Healing option was turned off. This shows the sensitivity of the parameter $\psi$.

### 4.2 Effects of Vaccination

Vaccination has contributed significantly towards the eradication and reduction of effect of many infectious diseases [7]. The following experiments were con-
(a) Comparison of $\psi$'s of 7,10,12,15 is shown. The curve with $\psi$ represents faster growth of infected individuals than $\psi$ of 7,10,12

(b) Comparison of Random Vaccination (5% of population was vaccinated) and no Vaccination

Fig. 5. Comparison of $\psi$'s and Comparison of Random Vaccination and no Vaccination

ducted on the simulator by vaccinating about 5% of the population at random and infecting few cells. Figure 5(b) shows the growth of infected individuals in a vaccinated and non vaccinated population. Figure 5(b) depicts that the growth of infected individuals in a population with only 5% of the population vaccinated, is considerably less as compared to the growth in a non-vaccinated population.

We study the effects of spatial distribution of population, by vaccinating a part of the population using the random vaccination and ring vaccination. Every time a new vaccine is discovered, the question arises as to how should the vaccine be distributed to minimize the spread of a disease and maximize the effect of vaccination. Thus, in this experiment we compare the random vaccination, which
is also known as uniform strategy [9], and ring vaccination. The doses of vaccine available at our disposal is often limited, thus for the purpose of experiment we consider N doses of vaccine to be available to vaccinate the population, where N is about 5% of the population. In random vaccination, the N vaccines, are randomly distributed to individuals in a population, independent of the other. In the ring vaccination, individuals are vaccinated in a ring surrounding an area. The thickness and circumference of the ring depends on N. As Figure 6 shows, using random vaccination many more individuals are infected as compared to the ring vaccination. This experiment validates the result shown by Fukš and Lawniczak in [9].

4.3 Conclusion from Experiments
The previous model described poses a limitation of neighborhood. The model considers a neighborhood of 8 cells, because of which after a time period the number of susceptibles reduce and saturate the neighborhood. In such a situation the variance of infectivity parameter plays no role and has the same effect on the spread of the disease. Also, the need to simulate a disease, where an infective can spread the disease to twelve other individuals in one time step, will not be possible to simulate. Another important issue to note is the movement of people, migration, or travel is not considered. Some models, we saw in the previous section deal with movement of individuals from one cell to another in the defined neighborhood, where again the neighborhood is restricted. The saturation of neighborhood occurs due to overlapping of neighborhood, when more than one cell is infected in a neighborhood. Cells in a neighborhood may get infected more than once in one time unit.

5 Geo-spatial Model
The Geo-spatial model is designed for simulating global outbreak of a disease in an environment with global interaction. Even for this model the basic unit of CA
is a cell, which represents an individual. The neighborhood as defined for this model is global, where in a region of n cells every cell has n-1 neighbors.

For the functioning of this model, the definitions for the states of cell, parameters for simulation are same as the ones for SIR model discussed earlier. This model has an additional parameter of contact rate and the definition is as follows.

**Contact rate** parameter defines the number of contacts made by an individual per time unit. Instead of having the same contact rate parameter for every cell in the lattice, for simulation purposes this parameter has a Poisson distribution over the cells. The simulation of spread of disease is discussed further.

1. In a time step a cell chooses k cells at random from the pool of n-1 neighbors, where k is the contact rate defined for that cell. Thus the cell has now established contacts with k cells.
2. Once a contact has been established between cell ‘a’ and cell ‘b’, depending on the virulence of the disease defined by the infectivity parameter, cell ‘a’ can pass the infection to cell ‘b’ if cell ‘b’ is in a susceptible state S. If cell ‘a’ is not infected currently and cell ‘b’ is infected then cell ‘a’ can acquire the infection from ‘b’. Thus the infection can pass on in both directions.

### 5.1 Experiments on Geo-spatial Model

The Geo-spatial model is different from the SIR type model in terms of the neighborhood. The neighborhood saturation problem posed by SIR type model is overcome by this model. However, this model is restricted in modeling population demographics and travel patterns. The choice of cells for contact, is random and is not based on distance from the cell or any other parameter.

To study the effects of **position of index case** on spread of a disease, the simulation was run with different initial positions. After a certain time unit it is seen that locations of new infected cases are not very different for the two
To analyse the contact rate the experiment was done with three different contact rates for cells. The result shows that as average contact rate increases, the number of infected individuals also grows. For this model the contact rate is directly proportional to number of infected individuals. It is important to note that the contacts made by cells are random. Figure 8(a) shows the comparison.

As seen before in the other model, as infectivity parameter \( \psi \) increases the number of infected individuals increases. The average contact rate was fixed for this experiment. Figure 8(b) shows the comparison.
5.2 Accounting for Different Demographics

The models described above may be used for simulating diseases over small regions with local interaction and global interaction respectively. As mentioned before, these models do not take into account the demographics of the region and may not be accurate for simulating disease spread over large geographic regions because of the neighborhood constriction posed by them. Thus the global stochastic cellular automata with demographics will facilitate to understand the effects of different demographics, the population density, socio-economics of a region and culture. It can also be used effectively for investigating different vaccination strategies and understanding the effects of travel.

5.3 Global Outbreak Simulator

In the following section we discuss the design of a global outbreak simulator with a global interaction and demography. Even for this model the basic unit of CA is a cell, which represents an individual or a small sub-population. The neighborhood as defined for this model is global, where in a region of n cells every cell has n-1 neighbors.

The neighborhood for a global SCA is defined using a fuzzy set neighborhood. The definition of Fuzzy set neighborhood is as follows.

The set $F \subset S$ where $S$ is a set of all the cells

$F: \{(s, p) | s \in S, 0 \leq p \leq 1\}$

$\langle s, 1 \rangle$: Total/Complete membership

$\langle s, 0 \rangle$: No membership

The variable $p$ maintains the state of infection, 1 if infected else 0.

5.4 Characteristics of a Cell

State of infection $\delta$ is defined as any number between 0 and 1, indicating the level of infection present in the cell. 0 indicates not infected, 1 indicates fully infected.

Interaction Coefficient $i$ for a particular cell is defined as the interaction between that cell and every other cell in the lattice space. It is calculated as the reciprocal of the euclidean distance between the cells. Euclidean distance as derived from the GIS gravity model.

$$i_{C_{i,j},C_{k,l}} = \frac{1}{\sqrt{(i-k)^2 + (j-l)^2}}$$

Global interaction coefficient $\Gamma$ of cell $C_{i,j}$ is the summation of all the individual (n-1) interaction coefficients of the cell. Every cell has one global interaction coefficient and n-1 interaction coefficients.

The infection factor $I$ is calculated as a fraction of the interaction coefficient to the global interaction coefficient $\Gamma$, for every cell to cell interaction. It is also based on the virulence of the disease and the state of infection of the infecting agent.

$$I_{C_{i,j}} = \sum_{\forall C_{k,l} \neq C_{i,j}} \frac{i_{C_{i,j},C_{k,l}}}{\Gamma} \times \delta_{C_{k,l}} \times \psi$$
5.5 Simulation Based on Population

The global interaction coefficient and the interaction coefficients are calculated based on the distance. As the distance in between the cells reduce, the interaction coefficients increase which indicates more chances of interaction between them.

\[ \Gamma_{C_{i,j}} = \sum_{\forall C_{k,l} \neq C_{i,j}} \frac{1}{\sqrt{(i-k)^2 + (j-l)^2}} \]

5.6 Simulation Based on Population and Distance

The global interaction coefficient and the interaction coefficients are calculated based on the distance and population. The distance between the cells and the populations of the cells are considered. For better understanding, the cells are considered to be small regions having certain populations. The product of the populations of the two cells, acts as a factor for the interaction coefficients. The population factor is directly proportional to the interaction coefficient and the distance between them is inversely proportional to the interaction coefficient. Thus two cells with high populations are assumed to interact more than two cells with low populations, when the distance between them is same.

\[ \Gamma_{C_{i,j}} = \sum_{\forall C_{k,l} \neq C_{i,j}} \frac{1}{\sqrt{(i-k)^2 + (j-l)^2}} \times P_{C_{i,j}} \times P_{C_{i,j}} \]

6 Conclusion and Future Work

This paper describes a disease outbreak simulator using the cellular automata paradigm. The results show the variation in the spread of the disease for different parameters of infectivity \( \psi \). The simulator has also facilitated the study of different vaccination strategies. Geo-spatial model helps us in simulating disease spread in an environment with global interaction including travel and migration. In the same direction the global model can be used to simulate disease spread over large geographic regions. It deals with global interaction and the demographics of the region. While still working on the development of computational tools to facilitate surveillance, monitoring, prevention and control of dynamics of different diseases, the current simulators prove as valuable tools to study the dynamics of different diseases. Global stochastic versions of the CA are currently being developed.

References

1. Ahmed E., Agiza H.N. On Modeling epidemics. Including latency, incubation and variable susceptibility *Physica A* 253 (1998), pp. 347-352
2. Ahmed E., Elgazzar A.S. On some applications of cellular automata *Physica A* 296 (2002), pp.529-538
3. Bagni, R., Berchi, R., and Cariello, P. A comparison of simulation models applied to epidemics. *Journal of Artificial Societies and Social Simulation* 5 (2002),3
4. Barfoot, T. D. and D’Eleuterio, G. M. T. Multiagent Coordination by Stochastic Cellular Automata. *Presented at the International Joint Conference on Artificial Intelligence* (2001)

5. Boccahar, N. Cheong K. Critical behavior of a probabilistic automata network SIS model for the spread of an infectious disease in a population of moving individuals. *Journal of Physics A: Mathematical and General* 26:5 (1993), pp. 3707-3717

6. Boccahar, N. Cheong K., Oram M. A probabilistic automata network epidemic model with births and deaths exhibiting cyclic behavior. *Journal of Physics A: Mathematical and General* 27 (1994), pp. 1585-1597

7. Del Giudice G., Vaccine. 21 Suppl 2:S83-8. (2003)

8. Stefano, B. D., Fukš, H., and Lawniczak, A. T. Object-oriented implementation of CA/LGCA modelling applied to the spread of epidemics. *In 2000 Canadian Conference on Electrical and Computer Engineering* 1, IEEE, pp. 26-31.

9. Henryk Fukš and Anna T. Lawniczak. Individual-based lattice model for spatial spread of epidemics. *Discrete Dynamics in Nature and Society* 6 (2001), pp. 191-200

10. Hokky Situngkir, Epidemiology through Cellular Automata Case of Study: Avian Influenza Indonesia. *Working Paper WPF2004*, Bandung Fe Institute.

11. James C. Thomas and David J. Weber. Epidemiologic Methods for the Study of Infectious Diseases. *Oxford Press* (2001)

12. Kleczkowski, A., and Grenfell, B. T. Mean-field-type equations for spread of epidemics: The ‘small world’ model. *Physica A* 274, 1-2 (1999), pp. 355-360.

13. Ricardo Mansilla, Jose L. Gutierrez. Deterministic site exchange cellular automata models for the spread of diseases in human settlements. *Bulletin of Mathematical Biology*

14. Shih Ching Fu and George Milne. Epidemic Modelling Using Cellular Automata. *To appear in Australian Conference on Artificial Life 2003*

15. Wolfram, S. Statistical Mechanics of Cellular Automata. *Reviews of Modern Physics* 55 pp. 601-644.

16. Yaganehdoost A, Graviss EA, Ross MW, et al. Complex transmission dynamics of clonally related virulent Mycobacterium tuberculosis associated with barhopping by predominantly human immunodeficiency virus-positive gay men. *Journal of Infect Diseases*. 180(4) (1999) pp.1245-51.

17. Youngblut, C. Educational uses of virtual reality technology. *Technical Report IDA Document D-2128 (1998)*, Institute for Defense Analyses, Alexandria, VA