Towards an Analytic Model of Epidemic Spreading in Heterogeneous Systems

Xiaohu Li T. Paul Parker Shouhuai Xu
Department of Computer Science, University of Texas at San Antonio
{xli,tparker,shxu}@cs.utsa.edu

ABSTRACT

Mathematical models have been utilized to help understand the epidemic spreading of malicious codes (e.g., computer virus and worms). However, existing such models are either adapted from the ones developed to capture the epidemic spreading of biologically infectious diseases in homogeneous systems, or suitable only for a very specific class of heterogeneous systems. In this paper we present an attempt at building an analytic model of epidemic spreading of malicious codes in arbitrary heterogeneous systems.

Categories and Subject Descriptors

K.6.5 [**MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS**]: Security and Protection

General Terms

Security

Keywords

Modeling, Epidemic Spreading, Heterogeneous Systems, SIR

1. INTRODUCTION

Motivated by the need for a deep understanding of the spreading of computer viruses, Kephart-White proposed perhaps the first epidemic models of infections in computer systems [6, 7]. Such a need has been reinforced by the recent incidents of large-scale self-spreading of malicious codes. However, most existing such models (e.g., [6, 7, 22, 4]) are obtained by modifying the well-known ones developed to capture the epidemic spreading of biologically infectious diseases in *homogeneous* systems [1]. In these systems, it is reasonable to make what is now known as the *homogeneous assumption*, namely that every individual (i.e., vertex in terms of graphs, or node in terms of networks) has equal contact to everyone else in the population, and the rate of infection is largely determined by the density of the infected individuals. Such homogeneous models could still be useful under certain circumstances. For example, worm spreading may still abide by the homogeneous assumption if a *truly random scanning* strategy is utilized; this may be justified by the fact that a classic homogeneous model does give a pretty accurate approximation of Code Red [15], except perhaps for the effect due to the factors investigated in [22].

It is clear that models for epidemic spreading in homogeneous systems do not apply to heterogeneous systems, which invalidate the aforementioned homogeneous assumption. Nevertheless, heterogeneous systems may be more important and difficult to protect since future sophisticated worms would not deploy the random scanning strategy because, otherwise, the presence of spreading could be easily detected using techniques such as those explored in [21, 18, 19]. Instead, they could spread along a predetermined network topology [15, 20].

There have been some models for epidemic spreading in *heterogeneous* systems. For a very specific class of heterogeneous systems known as the Barabasi-Albert power-law networks [2], models were investigated in [9, 11, 12, 10]. A model of spreading in a more general class of heterogeneous systems was proposed in [17], and a similar model was independently proposed in [16] for a different purpose in a different context. The discrete-time model of [17] was later confirmed by a continuous-time analysis in [5]. However, the systems accommodated in all these investigations are heterogeneous *only* in the sense that the nodes could have different degrees; i.e., they still assume (at least implicitly) that the nodes are equally subject to infection and have the same probability of being cured. This motivates us to explore a model for epidemic spreading in arbitrary heterogeneous systems, where different nodes exhibit different characteristics. The arbitrary heterogeneity may be induced by the underlying heterogeneous networks.

1.1 Our Contributions

We investigate a susceptible-infected-removed (SIR) model for epidemic spreading, in which an individual is always in the state of susceptible, infected, or removed. The model accommodates arbitrary heterogeneous systems, which may correspond to physical networks or to logical ones that reflect some kinds of real-life relationships between the nodes. Specifically,

1. Our model captures the fact that an individual could become infected because of its' own behaviors. This is realistic because, for instance, Alice's computer could become compromised when she downloads and runs

Qshine'07 August 14-17, 2007, Vancouver , Canada

a malicious code (which may be embedded into some innocent code such as an open source program). Furthermore, the model accommodates that the degrees that users suffer from this vulnerability vary (depending on, say, how cautious the users are).

2. Besides making no topological restrictions, our model further accommodates the following aspects of system heterogeneity: (1) Different nodes or users have different capabilities in detecting and curing compromises; for instance, some users may patch their systems more often than the others, and different users may use different softwares. (2) Different edges have different capabilities in spreading malicious codes.

As we will see, the computation-oriented utility of our model is validated via simulations, but the analysis-oriented utility of our model remains to be explored. Nevertheless, we believe that this paper presents a significant step towards the ultimate goal of a closed-form analytic model.

1.2 Related Work

As said before, many existing epidemic models for the spreading of computer viruses or worms [6, 7, 22, 4] are obtained by adapting the models in the biological epidemiology literature [1], and thus, in general, suffer from the *homogeneous assumption* mentioned above. [9, 11, 12, 10] presented analytic models for a specific class of heterogeneous systems, known as the Barabasi-Albert power-law topology [2]. Epidemic spreading in some non-Barabasi-Albert power-law topologies have been investigated in [16, 17, 5]. All these models do not accommodate arbitrarily heterogeneous systems, because they only allow the nodes to have different degrees (but no heterogeneity in any other sense).

Outline: In Section 2 we specify the system setting as well as notations. In Section 3 we explore the model. We compare the full-fledged model to simulations in Section 4, and conclude this paper in Section 5.

2. SYSTEM SETTING AND NOTATIONS

Let $G = (V, E)$ be a finite graph (or interchangeably, network), where $V = \{1, 2, \ldots, n\}$ is the fixed set of vertices, and E is the set of edges. A graph G may be directed or undirected. The justification for assuming a finite graph of a fixed size is that malicious code typically spreads in networked systems of a short period of time, compared with the life-time of the systems. We stress that the above conventions allow us not to specify whether a graph G is directed or not; that is, all the results are equally applicable to both undirected and directed graphs.

We are interested in a discrete time, $t = 0, 1, 2, \ldots$, susceptible-infected-removed (SIR) model, in which a node (or interchangeably, vertex or individual) $v \in V$ is always in one of the three states: susceptible, infected, or removed. In particular, a removed node will not be infected any further, meaning that removed is an absorbing state; this is typical to epidemic spreading in computer systems. Unlike existing epidemic models, our model captures that a node v may become infected because of its owner's (or user's) behavior (e.g., downloading and running a malicious code) at each discrete time step; this is specified by a parameter α_v. Each node v is also associated with a parameter β_v, which captures the probability that an infected node v becomes removed at a discrete time step. Similarly, each edge $(u, v) \in E$ is also associated with a parameter γ_{vu}, which captures the capability of an infection occurring from an infected node u to a susceptible node v. Naturally, in the case of undirected graphs, $\gamma_{vu} = \gamma_{uv}$ for all $(u, v) \in E$; whereas, it is not necessarily true in the case of directed graphs.

Let S_t be the random variable indicating the total number of susceptible nodes at time t, I_t be the random variable indicating the total number of infected nodes at time t, and R_t be the random variable indicating the total number of removed nodes at time t. Furthermore, let $s_v(t)$ be the probability that node v becomes susceptible at time t, $i_v(t)$ be the probability that node v becomes infected at time t, and $r_v(t)$ be the probability that node v becomes removed at time t. It is reasonable to assume that, $\forall v \in V$, $s_v(0) = 1$, $i_v(0) = 0$, and $r_v(0) = 0$. This means $S_0 = n$, $I_0 = 0$, and $R_0 = 0$. We note that for all t, $s_v(t) + i_v(t) + r_v(t) = 1$ as well as $S_t + I_t + R_t = n$, and that, by definition, $\mathsf{E}[S_t] = \sum_{v \in V} s_v(t)$, $\mathsf{E}[I_t] = \sum_{v \in V} i_v(t)$, and $\mathsf{E}[R_t] = \sum_{v \in V} r_v(t)$.

We summarize the notations below.

$s_v(t)$	the probability a node v is susceptible at time t
$i_v(t)$	the probability a node v is infected at time t
$r_v(t)$	the probability a node v is removed at time t
S_t	the total number of susceptible nodes at time t
I_t	the total number of infected nodes at time t
R_t	the total number of removed nodes at time t
α_v	the probability that node v becomes infected because of its own behavior
β_v	the probability that an infected node v becomes removed
γ_{vu}	the probability that a susceptible node v is infected by an infected neighbor u (i.e., the probability an infection occurs over edge $(u, v) \in E$)
$\eta_v(t)$	the probability that a susceptible node v becomes infected at time $t+1$ because of its neighbors that are in the state of infected at time t

3. MODEL

3.1 Warmup

In this section we consider a SIR model (called the *simple model* thereafter) for simplified heterogeneous systems with $E = \emptyset$, which means that the vertices become infected only because of their users' behaviors. The derived results will be used to help bound certain metrics in the full-fledged case investigated in the next section. Figure 1 shows the state transitions of the nodes. Basically, it says:

1. At each discrete time step, vertex v changes its state from susceptible to infected with probability α_v.

2. At each discrete time step, vertex v changes its state from infected to removed with probability β_v.

3. The vertices in removed state will remain in this state (i.e., removed is an absorbing state).

Within this model, we observe the following, for all $v \in V$:

$$\begin{cases} s_v(t) &= (1 - \alpha_v)\, s_v(t-1), \\ i_v(t) &= \alpha_v\, s_v(t-1) + (1 - \beta_v)\, i_v(t-1), \\ r_v(t) &= \beta_v\, i_v(t-1) + r_v(t-1). \end{cases} \quad (1)$$

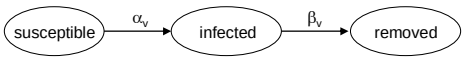

Figure 1: State transition of vertex v in the simplified case of $E = \emptyset$

We are interested in the analytic formula of $\mathsf{E}[S_t]$ (i.e., the expected number of susceptible nodes), $\mathsf{E}[I_t]$ (i.e., the expected number of infected nodes), and $\mathsf{E}[R_t]$ (i.e., the expected number of removed nodes) at time t.

LEMMA 1. *In this model, for all $t = 1, 2, \ldots$:*

$$\mathsf{E}[S_t] = \sum_{v=1}^{n} (1-\alpha_v)^t,$$

$$\mathsf{E}[I_t] = \sum_{v=1}^{n} \alpha_v \sum_{j=0}^{t-1} (1-\beta_v)^j (1-\alpha_v)^{t-j-1},$$

$$\mathsf{E}[R_t] = n - \sum_{v=1}^{n} (1-\alpha_v)^t - \sum_{v=1}^{n} \alpha_v \sum_{j=0}^{t-1} (1-\beta_v)^j (1-\alpha_v)^{t-j-1}.$$

PROOF. For $v = 1, 2, \cdots, n$, denote by

$$\Lambda_v = \begin{pmatrix} 1-\alpha_v & 0 & 0 \\ \alpha_v & 1-\beta_v & 0 \\ 0 & \beta_v & 1 \end{pmatrix}.$$

Then, (1) can be rephrased as

$$\begin{pmatrix} s_v(t) \\ i_v(t) \\ r_v(t) \end{pmatrix} = \Lambda_v^t \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Since

$$\Lambda_v^t = \begin{pmatrix} (1-\alpha_v)^t & 0 & 0 \\ \frac{\alpha_v(1-\beta_v)^t}{\alpha_v-\beta_v} - \frac{\alpha_v(1-\alpha_v)^t}{\alpha_v-\beta_v} & (1-\beta_v)^t & 0 \\ 1 - \frac{\alpha_v(1-\beta_v)^t}{\alpha_v-\beta_v} + \frac{\beta_v(1-\alpha_v)^t}{\alpha_v-\beta_v} & 1-(1-\beta_v)^t & 1 \end{pmatrix},$$

it holds that, for all $v = 1, 2, \cdots, n$,

$$s_v(t) = (1-\alpha_v)^t, \tag{2}$$

$$\begin{aligned} i_v(t) &= \frac{\alpha_v(1-\beta_v)^t}{\alpha_v-\beta_v} - \frac{\alpha_v(1-\alpha_v)^t}{\alpha_v-\beta_v} \\ &= \frac{\alpha_v}{\alpha_v-\beta_v}\left[(1-\beta_v)^t - (1-\alpha_v)^t\right], \end{aligned} \tag{3}$$

$$r_v(t) = 1 - \frac{\alpha_v(1-\beta_v)^t}{\alpha_v-\beta_v} + \frac{\beta_v(1-\alpha_v)^t}{\alpha_v-\beta_v}. \tag{4}$$

As a result, we have

$$\mathsf{E}[S_t] = \sum_{v=1}^{n} (1-\alpha_v)^t,$$

$$\begin{aligned} \mathsf{E}[I_t] &= \sum_{v=1}^{n} \frac{\alpha_v}{\alpha_v-\beta_v}\left[(1-\beta_v)^t - (1-\alpha_v)^t\right] \\ &= \sum_{v=1}^{n} \alpha_v \sum_{j=0}^{t-1} (1-\beta_v)^j (1-\alpha_v)^{t-j-1}, \end{aligned}$$

$$\begin{aligned} &\mathsf{E}[R_t] \\ &= \sum_{v=1}^{n} \left[1 - \frac{\alpha_v(1-\beta_v)^t}{\alpha_v-\beta_v} + \frac{\beta_v(1-\alpha_v)^t}{\alpha_v-\beta_v}\right] \\ &= n - \sum_{v=1}^{n} (1-\alpha_v)^t - \sum_{v=1}^{n} \alpha_v \sum_{j=0}^{t-1} (1-\beta_v)^j (1-\alpha_v)^{t-j-1}. \end{aligned}$$

□

Since $0 \le 1-\alpha_v \le 1$ and $0 \le 1-\beta_v \le 1$ for any $v \in V$, (2), (3) and (4) imply that $s_v(t) \to 0$, $i_v(t) \to 0$ and $r_v(t) \to 1$ as $t \to \infty$. This means that both susceptible and infected states are transient, only removed is absorbing; i.e., the evolution procedure will tend to the absorbing state removed eventually. Furthermore, it is straightforward to obtain the following lemma:

LEMMA 2.

$$\Pr[S_t = 0] = \prod_{v=1}^{n} \left[1 - (1-\alpha_v)^t\right],$$

$$\begin{aligned} &\Pr[R_t = n] \\ &= \prod_{v=1}^{n} \left[1 - \frac{\alpha_v(1-\beta_v)^t}{\alpha_v-\beta_v} + \frac{\beta_v(1-\alpha_v)^t}{\alpha_v-\beta_v}\right] \\ &= \prod_{v=1}^{n} \left[1 - (1-\alpha_v)^t - \alpha_v \sum_{j=0}^{t-1} (1-\beta_v)^j (1-\alpha_v)^{t-j-1}\right]. \end{aligned}$$

3.2 Full-fledged Case

In this model, we further consider edge-infections, meaning that vertices could become infected also because of their infected neighbors. Thus, each edge $(u, v) \in E$ is associated with an *infection probability*, γ_{vu}. Figure 2 specifies the state transition. It basically says:

1. At each discrete time step, vertex v changes state from susceptible to infected because of (1) its own behavior (with probability α_v) or (2) of its neighbors' infection (with probability η_v), where η_v depends on the edge-infection probability γ_{vu}.
2. At each discrete time step, vertex v changes state from infected to removed with probability β_v.
3. Vertices in removed state will remain in this state (i.e., removed is an absorbing state).

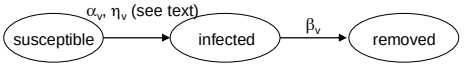

Figure 2: State transition of vertex v in the full-fledged case

Let $\eta_v(t)$ be the probability that at time $t+1$ a node v becomes infected because of its neighbors that are in infected state at time t. We have

$$\eta_v(t) = 1 - \prod_{(u,v)\in E} [1 - \gamma_{vu}\, i_u(t)]. \tag{5}$$

Thus, the model implies:

$$\begin{aligned} s_v(t) &= [1-\alpha_v-(1-\alpha_v)\,\eta_v(t-1)]\,s_v(t-1), \quad (6)\\ i_v(t) &= [\alpha_v+(1-\alpha_v)\,\eta_v(t-1)]\,s_v(t-1)+ \\ &\quad (1-\beta_v)\,i_v(t-1), \quad (7)\\ r_v(t) &= \beta_v\,i_v(t-1)+r_v(t-1). \quad (8)\end{aligned}$$

We utilize the following mathematical tools: stochastically increasing/decreasing [3], stochastically larger/smaller [14], and convergence in probability (denoted by "$\xrightarrow{P}$") [13]; we refer to the respective texts for details. We also need the following lemma.

LEMMA 3. *For $0<\delta<1$ and $d>0$,*

$$\lim_{n\to\infty} n\cdot(1-\delta)^{d\ln n}=0 \quad \textit{if and only if} \quad d>-\frac{1}{\ln(1-\delta)}.$$

PROOF. Set $w=\frac{1}{1-\delta}$, then

$$\begin{aligned} n\cdot(1-\delta)^{d\ln n} &= \frac{n}{w^{\ln(n^d)}}=\frac{n}{w^{\frac{\log_w(n^d)}{\log_w e}}}\\ &= \frac{n}{(w^{\log_w(n^d)})^{\frac{1}{\log_d e}}}=\frac{n}{n^{\frac{d}{\log_w e}}}.\end{aligned}$$

Since $\log_w e>0$, $\frac{d}{\log_w e}>1$ if and only if $d>\log_w e=-\frac{1}{\ln(1-\delta)}$, the desired result follows immediately. □

□

Since $-\ln(1-\delta)>\delta+\frac{\delta^2}{2}$, or $\frac{1}{-\ln(1-\delta)}<\frac{1}{\delta+\frac{\delta^2}{2}}$, we obtain the following corollary.

COROLLARY 1. *For $0<\delta<1$, as $n\to\infty$,*

$$n(1-\delta)^{\frac{\ln n}{\delta-\delta^2/2}}\to 0. \quad (9)$$

To help derive the desired results, denote by

$$S(t)=\begin{pmatrix} s_1(t)\\ \vdots\\ s_n(t)\end{pmatrix}, I(t)=\begin{pmatrix} i_1(t)\\ \vdots\\ i_n(t)\end{pmatrix}, R(t)=\begin{pmatrix} r_1(t)\\ \vdots\\ r_n(t)\end{pmatrix},$$

$$\mathsf{D}_\alpha=\begin{pmatrix} \alpha_1 & \cdots & 0\\ 0 & \vdots & 0\\ 0 & \cdots & \alpha_n\end{pmatrix}, \mathsf{D}_\beta=\begin{pmatrix} \beta_1 & \cdots & 0\\ 0 & \vdots & 0\\ 0 & \cdots & \beta_n\end{pmatrix},$$

$$\mathsf{I}=\begin{pmatrix} 1 & \cdots & 0\\ 0 & \vdots & 0\\ 0 & \cdots & 1\end{pmatrix}.$$

Set

$$\mathsf{D}_\eta(t)=\begin{pmatrix} (1-\alpha_1)\,\eta_1(t) & 0 & \cdots & 0\\ 0 & (1-\alpha_2)\,\eta_2(t) & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & (1-\alpha_n)\,\eta_n(t)\end{pmatrix},$$

Eqs. (6), (7) and (8) may be rephrased as

$$\begin{aligned} S(t) &= S(t-1)-\mathsf{D}_\alpha\times S(t-1)-\\ &\quad \mathsf{D}_\eta(t-1)\times S(t-1), \quad (10)\\ I(t) &= \mathsf{D}_\alpha\times S(t-1)+\mathsf{D}_\eta(t-1)\times S(t-1)+\\ &\quad (\mathsf{I}-\mathsf{D}_\beta)\times I(t-1), \quad (11)\\ R(t) &= \mathsf{D}_\beta\times I(t-1)+R(t-1). \quad (12)\end{aligned}$$

Denote by $\mathbf{1}=(\ 1\ \ 1\ \ \cdots\ \ 1\ )$. Then we obtain $\mathsf{E}[S_t]=\mathbf{1}\times S(t)$, $\mathsf{E}[I_t]=\mathbf{1}\times I(t)$, and $\mathsf{E}[R_t]=\mathbf{1}\times R(t)$.

THEOREM 1. (characterizing S_t) *Random variable S_t can be characterized as follows:*

1. *$\mathsf{E}[S_t]$ can be bounded as follows:*

$$\sum_{v=1}^n(1-\alpha_v)^t\ge\mathsf{E}[S_t]\ge\sum_{v=1}^n\left[(1-\alpha_v)^t\prod_{(u,v)\in E}(1-\gamma_{vu})^t\right].$$

2. *S_t is stochastically decreasing with respect to time t, and indeed $S_t\xrightarrow{P}0$ as $t\to\infty$. Furthermore, under reasonable approximation, there are no* susceptible *nodes since time step*

$$\left\lceil\frac{\ln n}{\min\limits_{v\in V}\alpha_v-\frac{1}{2}\left(\min\limits_{v\in V}\alpha_v\right)^2}\right\rceil.$$

3. *Under reasonable approximation, $\mathsf{E}[S_t]<1$, meaning that the expected number of* susceptible *nodes becomes less than one, since time step:*

$$\left\lfloor\frac{\ln n}{-\ln\left(1-\min\limits_{v\in V}\alpha_v\right)}\right\rfloor+1$$

PROOF. Note that the total number of susceptible nodes in the present model at time t is stochastically smaller than the total number of susceptible nodes in the above *simple model* at time t. Combined with Lemma 1, we get the following upper bound:

$$\mathsf{E}[S_t]\le\sum_{v=1}^n(1-\alpha_v)^t. \quad (13)$$

On the other hand, by Eq. (6),

$$\begin{aligned} s_v(t) &= (1-\alpha_v)[1-\eta_v(t-1)]\,s_v(t-1)\\ &= (1-\alpha_v)\left[\prod_{(u,v)\in E}(1-\gamma_{vu}\,i_u(t))\right]s_v(t-1)\\ &\ge (1-\alpha_v)\left[\prod_{(u,v)\in E}(1-\gamma_{vu})\right]s_v(t-1)\\ &\ge (1-\alpha_v)^t\prod_{(u,v)\in E}(1-\gamma_{vu})^t.\end{aligned}$$

So we obtain the lower bound

$$\mathsf{E}[S_t]=\sum_{v\in V}s_v(t)\ge\sum_{v=1}^n\left[(1-\alpha_v)^t\prod_{(u,v)\in E}(1-\gamma_{vu})^t\right]. \quad (14)$$

Now we prove the second part of the theorem. According to Eq. (10), all the elements of the vector

$$S(t)-S(t-1)=-[\mathsf{D}_\alpha+\mathsf{D}_\eta(t-1)]\times S(t-1)$$

are negative, so S_t is stochastically decreasing [3] with respect to time t.

Since a node may be infected by one of its neighbors, the probability for a susceptible node in a network without edge-infection (i.e., the above *simple model*) to become infected is smaller than that for a susceptible node in a network with edge-infections (i.e., the present model). As a result, the total number of susceptible nodes in the present model at time t is stochastically smaller than the total number of susceptible nodes in the *simple model* at time t. Combining with Lemma 1, we obtain

$$\Pr[S_t = 0] \geq \prod_{v=1}^{n} \left[1 - (1-\alpha_v)^t\right].$$

Since for any $\varepsilon > 0$,

$$\begin{aligned}\Pr[S_t > \varepsilon] &= 1 - \Pr[S_t \leq \varepsilon] \leq 1 - \Pr[S_t = 0] \\ &\leq 1 - \prod_{v=1}^{n} \left[1 - (1-\alpha_v)^t\right] \\ &\to 0, \quad \text{as } t \to \infty.\end{aligned}$$

Thus,

$$S_t \xrightarrow{P} 0, \quad \text{as } t \to \infty.$$

Now we give an estimation when S_t will become almost 0. Since S_t is stochastically decreasing and $S_t \to 0$ as $t \to \infty$,

$$\begin{aligned}\Pr[S_t = 0] &= \prod_{v\in V}(1 - s_v(t)) \approx 1 - \sum_{v=1}^{n} s_v(t) \\ &\geq 1 - \sum_{v\in V, \alpha_v \neq 0} (1-\alpha_v)^t \\ &\geq 1 - \sum_{v\in V, \alpha_v \neq 0} \left(1 - \left(\min_{v\in V, \alpha_v \neq 0} \alpha_v\right)\right)^t \\ &\geq 1 - n\left(1 - \left(\min_{v\in V, \alpha_v \neq 0} \alpha_v\right)\right)^t\end{aligned}$$

By Corollary 1, there are no susceptible nodes since time step

$$\left\lceil \frac{\ln n}{\left(\min_{v\in V, \alpha_v \neq 0} \alpha_v\right) - \frac{1}{2}\left(\min_{v\in V, \alpha_v \neq 0} \alpha_v\right)^2} \right\rceil.$$

Now we show the third part of the theorem. Since

$$\begin{aligned}\mathsf{E}[S_t] &\leq \sum_{v=1}^{n}(1-\alpha_v)^t \leq \sum_{v=1}^{n}\left(1 - \min_{v\in V}\alpha_v\right)^t \\ &= n\left(1 - \min_{v\in V}\alpha_v\right)^t.\end{aligned}$$

By setting

$$n\left(1 - \min_{v\in V}\alpha_v\right)^t < 1,$$

we obtain

$$t > \frac{\ln n}{-\ln\left(1 - \min_{v\in V}\alpha_v\right)},$$

which immediately leads to the desired result. □

□

THEOREM 2. (characterizing R_t) *Random variable R_t can be characterized as follows:*

1. *$\mathsf{E}[R_t]$ can be bounded as follows:*

$$n - \sum_{v=1}^{n}\left[(1-\alpha_v)^t \prod_{(u,v)\in E}(1-\gamma_{vu})^t\right] \geq \mathsf{E}[R_t] \geq n - \sum_{v=1}^{n}\left[(1-\alpha_v)^t + \alpha_v \sum_{j=0}^{t-1}(1-\beta_v)^j(1-\alpha_v)^{t-j-1}\right].$$

2. *R_t is stochastically increasing with respect to time t, and indeed $R_t \xrightarrow{P} n$ as $t \to \infty$, Further, under reasonable approximation, $\Pr[R_t = n] = 1$, meaning that all the nodes become* removed*, since time step*

$$\left\lceil \frac{\ln n + \ln \max_{v\in V, \alpha_v \neq \beta_v} \frac{\max\{\alpha_v, \beta_v\}}{\max\{\alpha_v, \beta_v\} - \min\{\alpha_v, \beta_v\}}}{\Delta\left(1 - \frac{1}{2}\Delta\right)} \right\rceil$$

where $\Delta = \min\left\{\min_{v\in V, \alpha_v > \beta_v}\beta_v, \min_{v\in V, \alpha_v \leq \beta_v}\alpha_v^2\right\}$.

PROOF. Note that the total number of removed nodes in the present model at time t is stochastically larger than the total number of removed nodes in the *simple model* at time t. Combined with Lemma 1, we have the following lower bound,

$$\mathsf{E}[R_t] \geq \sum_{v=1}^{n}\left[1 - (1-\alpha_v)^t - \alpha_v \sum_{j=0}^{t-1}(1-\beta_v)^j(1-\alpha_v)^{t-j-1}\right]. \tag{15}$$

On the other hand, by Eq. (14) we obtain

$$\mathsf{E}[R_t] \leq n - \sum_{v=1}^{n}\left[(1-\alpha_v)^t \prod_{(u,v)\in E}(1-\gamma_{vu})^t\right]. \tag{16}$$

Now we prove the second part of the theorem. According to Eq. (12), all the elements of the vector

$$R(t) - R(t-1) = \mathsf{D}_\beta \times C(t-1)$$

are positive, so R_t is stochastically increasing with respect to time t.

Since a node may be infected by one of its neighbors, the probability for a susceptible node in a network without edge-infection (i.e., the above *simple model*) to become infected is smaller than that for a susceptible node in a network with edge-infections (i.e., the present model). As a result, the total number of removed nodes in the above *simple model* at time t is stochastically smaller than the total number of removed nodes in the *simple model* at time t. Combined with Lemma 1, we obtain

$$\begin{aligned}&\Pr[R_t = n] \\ &\geq \prod_{v=1}^{n}\left[1 - (1-\alpha_v)^t - \alpha_v \sum_{j=0}^{t-1}(1-\beta_v)^j(1-\alpha_v)^{t-j-1}\right].\end{aligned}$$

Since for any $\varepsilon > 0$,

$$\begin{aligned}
&\Pr[R_t > n-\varepsilon] \;\geq\; \Pr[R_t = n] \\
\geq\;& \prod_{v=1}^{n}\left[1-(1-\alpha_v)^t - \alpha_v \sum_{j=0}^{t-1}(1-\beta_v)^j(1-\alpha_v)^{t-j-1}\right] \\
\to\;& 1, \quad \text{as } t\to\infty.
\end{aligned}$$

Thus,

$$R_t \xrightarrow{P} n, \quad \text{as } t\to\infty.$$

Now we give an estimation when R_t will become almost n. Denote by $n_1 = |\{v\in V|\alpha_v > \beta_v\}|$, $n_2 = |\{v\in V|\alpha_v < \beta_v\}|$, and $n_3 = |\{v\in V|\alpha_v = \beta_v\}|$. It is clear that $n_1+n_2+n_3 = n$.

$$\begin{aligned}
&\Pr[R_t = n] = \prod_{v=1}^{n} r_v(t) \\
=\;& \prod_{v\in V,\alpha_v\neq\beta_v}\left(1-\frac{\alpha_v(1-\beta_v)^t}{\alpha_v-\beta_v}+\frac{\beta_v(1-\alpha_v)^t}{\alpha_v-\beta_v}\right)\cdot \\
& \prod_{v\in V,\alpha_v=\beta_v}\left(1-(1-\alpha_v)^{t-1}(1+(t-1)\alpha_v)\right) \\
\geq\;& \prod_{v\in V,\alpha_v>\beta_v}\left(1-\frac{\alpha_v(1-\beta_v)^t}{\alpha_v-\beta_v}\right)\cdot \\
& \prod_{v\in V,\alpha_v<\beta_v}\left(1-\frac{\beta_v(1-\alpha_v)^t}{\beta_v-\alpha_v}\right)\cdot \\
& \prod_{v\in V,\alpha_v=\beta_v}\left(1-t(1-\alpha_v)^{t-1}\right) \\
\geq\;& \left[1-\max_{v\in V,\alpha_v>\beta_v}\frac{\alpha_v}{\alpha_v-\beta_v}\cdot\left(1-\min_{v\in V,\alpha_v>\beta_v}\beta_v\right)^t\right]^{n_1}\cdot \\
& \left[1-\max_{v\in V,\alpha_v<\beta_v}\frac{\beta_v}{\beta_v-\alpha_v}\cdot\left(1-\min_{v\in V,\alpha_v<\beta_v}\alpha_v\right)^t\right]^{n_2}\cdot \\
& \prod_{v\in V,\alpha_v=\beta_v}\left(1-(1-\alpha_v^2)^t\right) \\
& \left(\text{since } \frac{t(1-\alpha_v)^{t-1}}{(1-\alpha_v^2)^t}\to 0, \quad \text{as } t\to\infty\right) \\
\approx\;& \left[1-n_1\max_{v\in V,\alpha_v>\beta_v}\frac{\alpha_v}{\alpha_v-\beta_v}\cdot\left(1-\min_{\alpha_v>\beta_v}\beta_v\right)^t\right]\cdot \\
& \left[1-n_2\max_{v\in V,\alpha_v<\beta_v}\frac{\beta_v}{\beta_v-\alpha_v}\cdot\left(1-\min_{\alpha_v<\beta_v}\alpha_v\right)^t\right]\cdot \\
& \left[1-n_3\left(1-\min_{\alpha_v=\beta_v}\alpha_v^2\right)^t\right] \\
\approx\;& 1-n\cdot\max\left\{\max_{v\in V,\alpha_v>\beta_v}\frac{\alpha_v}{\alpha_v-\beta_v},\max_{v\in V,\alpha_v<\beta_v}\frac{\beta_v}{\beta_v-\alpha_v}\right\}\cdot \\
& \left(1-\min\left\{\min_{v\in V,\alpha_v>\beta_v}\beta_v,\min_{v\in V,\alpha_v<\beta_v}\alpha_v,\min_{v\in V,\alpha_v=\beta_v}\alpha_v^2\right\}\right)^t \\
=\;& 1-n\cdot\max_{v\in V,\alpha_v\neq\beta_v}\frac{\max\{\alpha_v,\beta_v\}}{\max\{\alpha_v,\beta_v\}-\min\{\alpha_v,\beta_v\}}\cdot \\
& \left(1-\min\left\{\min_{v\in V,\alpha_v>\beta_v}\beta_v,\min_{v\in V,\alpha_v\leq\beta_v}\alpha_v^2\right\}\right)^t.
\end{aligned}$$

Set $\Delta = \min\left\{\min_{v\in V,\alpha_v>\beta_v}\beta_v, \min_{v\in V,\alpha_v\leq\beta_v}\alpha_v^2\right\}$. Then, by Corollary 1 we immediately obtain the desired result. □

Regarding random variable I_t, we have the following corollary.

COROLLARY 2. *Under reasonable approximation, there are no* infected *node after time step:*

$$\left\lceil \frac{\ln n + \ln \max_{v\in V,\alpha_v\neq\beta_v}\frac{\max\{\alpha_v,\beta_v\}}{\max\{\alpha_v,\beta_v\}-\min\{\alpha_v,\beta_v\}}}{\Delta\left(1-\frac{1}{2}\Delta\right)}\right\rceil.$$

where $\Delta = \min\left\{\min_{v\in V,\alpha_v>\beta_v}\beta_v, \min_{v\in V,\alpha_v\leq\beta_v}\alpha_v^2\right\}$.

4. EXPERIMENTS

We are interested in knowing how far away $\mathsf{E}[S_t]$ and $\mathsf{E}[R_t]$ are from simulation, and from their lower and upper bounds given by Theorem 1 and Theorem 2, respectively. The simulation results are averaged over 20 runs. For this purpose, we generate a synthetic graph using the BRITE network topology generator [8]. The generated graph has 600,000 nodes with an average node degree 4. Let $x \in [a,b]$ denote that x is uniformly selected from interval $[a,b]$. To be concise, we only consider two sets of representative parameters for each data: (1) $\alpha_v \in [0.05, 0.1]$, $\gamma_{vu} \in [0.2, 0.3]$, and $\beta_v \in [0.2, 0.3]$; (2) $\alpha_v \in [0.05, 0.1]$, $\gamma_{vu} \in [0.2, 0.3]$, and $\beta_v \in [0.5, 0.6]$.

The comparisons are plotted in Figure 3 — Figure 4. They indicate the following. First, in every case the simulation results are very close to the model predictions. This validates the computation-oriented utility of our model. Second, the bounds are not tight. However, $\mathsf{E}[S_t]$ is somehow closer to its lower bound. The reason is unclear. This leaves some interesting future research problems (see below).

5. CONCLUSION

We presented a model of epidemic spreading in arbitrary heterogeneous systems. While the computation-oriented utility of our model is validated via simulations, the analysis-oriented utility of our model remains to be explored. We hope that this work will inspire more research in modeling epidemic spreading in heterogeneous systems, which might lead to a deeper understanding of what useful and effective quarantine techniques would be. We are investigating methods whereby we can exploit both the lower and upper bounds to enhance the analysis-oriented utility of our model.

Acknowledgments

This work was supported in part by ARO and UTSA.

6. REFERENCES

[1] N. Bailey. *The Mathematical Theory of Infectious Diseases and Its Applications*. 2nd Edition. Griffin, London, 1975.

[2] A. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286.

[3] R. Barlow and F. Proschan. *Statistical Theory of Reliability and Life Testing*. To Begin With, Silver Spring, MD, 1981.

[4] Z. Chen, L. Gao, and K. Kwiat. Modeling the spread of active worms. In *INFOCOM'03*.

[5] A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology on the spread of epidemics. In *Infocom'05*.

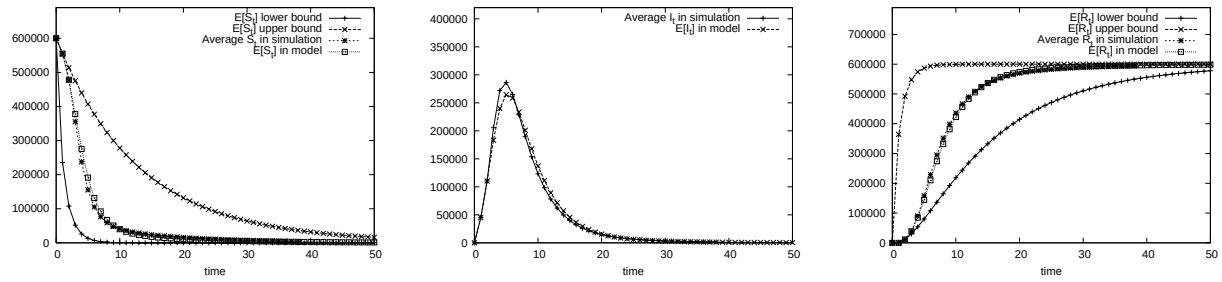

Figure 3: $\alpha_v \in [0.05, 0.10]$**,** $\gamma_{vu} \in [0.2, 0.3]$ **and** $\beta_v \in [0.2, 0.3]$

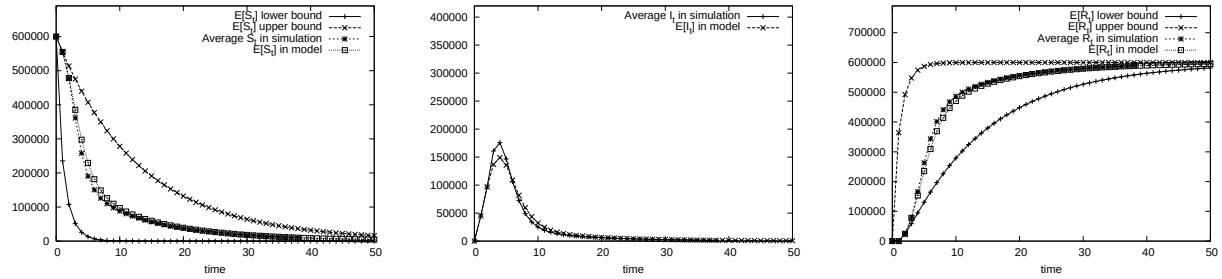

Figure 4: $\alpha_v \in [0.05, 0.10]$**,** $\gamma_{vu} \in [0.2, 0.3]$ **and** $\beta_v \in [0.5, 0.6]$

[6] J. Kephart and S. White. Directed-graph epidemiological models of computer viruses. In *IEEE Symposium on Security and Privacy'91*.

[7] J. Kephart and S. White. Measuring and modeling computer virus prevalence. In *IEEE Symposium on Security and Privacy'93*.

[8] A. Medina, A. Lakhina, I. Matta, and J. Byers. Brite: An approach to universal topology generation. In *MASCOTS'01*.

[9] Y. Moreno, R. Pastor-Satorras, and A. Vespignani. Epidemic outbreaks in complex heterogeneous networks. *European Physical Journal B*, 26:521–529, 2002.

[10] R. Pastor-Satorras and A. Vespignani. Epidemics and immunization in scale-free networks. *Handbook of Graphs and Networks: From the Genome to the Internet*.

[11] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63:066117, 2001.

[12] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65:035108, 2002.

[13] S. Ross. *Stochastic Processes*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc, 1996.

[14] M. Shaked and J. Shanthikumar. *Stochastic Orders and Their Applications*. Academic Press, San Diego (CA), 1994.

[15] S. Staniford, V. Paxson, and N. Weaver. How to own the internet in your spare time. In *USENIX Security Symposium'02*.

[16] J. Wang, L. Lu, and A. Chien. Tolerating denial-of-service attacks using overlay networks – impact of topology. In *ACM workshop on survivable and self-regenerative systems*, 2003.

[17] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *SRDS'03*.

[18] N. Weaver, S. Staniford, and V. Paxson. Very fast containment of scanning worms. In *Usenix Security Symposium'04*.

[19] D. Whyte, E. Kranakis, and P. van Oorschot. Dns-based detection of scanning worms in an enterprise network. In *NDSS'05*.

[20] L. Zhou, L. Zhang, F. McSherry, N. Immorlica, M. Costa, and S Chien. An effective architecture and algorithm for detecting worms with various scan. In *IPTPS'05*.

[21] C. Zou, L. Gao, W. Gong, and D. Towsley. Monitoring and early warning for internet worms. In *CCS'03*.

[22] C. Zou, W. Gong, and D. Towsley. Code red worm propagation modeling and analysis. In *CCS'02*.