Stability and Immunization Analysis of a Malware Spread Model Over Scale-Free Networks

Aresh Dadlani, *Student Member, IEEE*, Muthukrishnan Senthil Kumar, Kiseon Kim, *Senior Member, IEEE*, and Khosrow Sohraby, *Senior Member, IEEE*

***Abstract*—The spreading dynamics and control of infectious agents primarily depend on the connectivity properties of underlying networks. Here, we investigate the stability of a *susceptible–infected–susceptible* epidemic model incorporated with multiple infection stages and propagation vectors to mimic malware behavior over scale-free communication networks. In particular, we derive the basic reproductive ratio (R_0) and provide results for stability analysis at infection-free and infection-chronic equilibrium points. Based on R_0, the effectiveness of four prevailing immunization strategies as countermeasures is studied and compared. The outperformance of proportional and targeted immunization is justified via numerical results.**

***Index Terms*—Malware modeling, epidemiology, scale-free network, basic reproductive ratio, stability analysis, immunization.**

I. Introduction

RECENT findings in the field of network science indicate that the inset of an endemic state is not only correlated with the infection schemes, but also with the network topology under study. This has led to a surge in scrutinizing the spreading behavior of new epidemics over networks [1]. Disordered networks with extreme heterogeneity such as the Internet and the World Wide Web are shown to be *scale-free* (SF), i.e., they exhibit a power-law degree distribution. These networks, unlike homogeneous networks, are more vulnerable to the spread and persistence of malwares due to their diverging connectivity fluctuations [2]. In virtue of such weakness, tailored immunization strategies are in high demand to ensure the stability of large-scale networking systems. In fact, the ability to mathematically model, predict, and restrain the continuing threat of computer malware proliferation over such technological networks is now a challenging task.

A handful of existing works address the complex behavior of malwares using compartmental differential equations [3]–[6]. While built upon the basic SIS theme, these models do not consider the joint impact of additional refinements to evaluate the efficiency of immunization strategies. In reality, a multipartite malware may not only infect a network node in multiple phases (also known as *infection delay*), but also propagate via email attachments, file sharing, and phishing schemes. Though epidemic thresholds for the SIS model with time delay and propagation vectors are derived in [7], no study on the potency of immunization tactics for the model exists.

In this paper, we derive the basic reproductive ratio (R_0) [8] to analyze the equilibrium stability of the model in [7]. Being a key concept widely-used in epidemic theory, R_0 refers to the total number of secondary infections produced as the result of introducing an infected node in the infection-free population. Using this notion, we examine the model with respect to the effects of various immunization strategies using mean-field approximation (MFA) over SF networks.

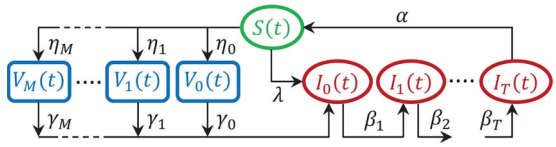

Fig. 1. The refined SIS model for malware spread.

II. Model Formulation and Stability Analysis

The malware propagation model adopted from [7] is shown in Fig. 1, where $S(t)$, $I_i(t)$, and $V_j(t)$ denote the densities of susceptible nodes, nodes in the i^{th} infection stage ($i = 0, 1, \ldots, T$), and the j^{th} propagation vector ($j = 0, 1, \ldots, M$) at time t, respectively. Since infection delay and propagation vector can be found simultaneously within various infectious cases, a susceptible node enters the initial infection stage (I_0) with rate λ or via some infective vector, say V_j, with rate γ_j. Once infected, the infection progresses from stage I_{m-1} to I_m with rate β_m ($m = 1, 2, \ldots, T$). Finally, the infected node returns to the susceptible class with rate α. Also, the transition rate of V_j from S to I_0 is given as η_j. Using the Barabási-Albert (BA) model [9] as the prototype example of SF networks, the malware model can be formulated as:

$$\begin{cases} \dot{I}_{k,0}(t) = -\beta_1 I_{k,0}(t) + \lambda k S_k(t)\Theta(t) + \sum_{j=0}^{M} \gamma_j S_k(t) V_j(t), \\ \dot{I}_{k,i}(t) = -\beta_{i+1} I_{k,i}(t) + \beta_i I_{k,i-1}(t); \quad i = 1, \ldots, T-1, \\ \dot{I}_{k,T}(t) = -\alpha I_{k,T}(t) + \beta_T I_{k,T-1}(t), \\ \dot{V}_j(t) = -V_j(t) + \eta_j \left(1 - V_j(t)\right) \Theta(t); \quad j = 0, \ldots, M, \end{cases} \tag{1}$$

where k refers to the degree such that $S_k(t) + \sum_{m=0}^{T} I_{k,m}(t) = 1$ and $\Theta(t)$ is the probability that any given link points to an infected node. With $\langle k \rangle$ and $P(k)$ defined as the average degree of

Manuscript received June 10, 2014; revised August 19, 2014; accepted September 21, 2014. Date of publication October 3, 2014; date of current version November 7, 2014. This work was supported in part by the Basic Research Project through a grant provided by the Gwangju Institute of Science and Technology (GIST) and in part by Grant K20901002277-12E0100-06010 funded by the Ministry of Science, ICT and Future Planning (MSIP) under Grant 2009-00422. The associate editor coordinating the review of this paper and approving it for publication was B. Rong.

A. Dadlani and K. Kim are with the School of Information and Communications, Department of Nanobio Materials and Electronics, Gwangju Institute of Science and Technology, Gwangju 500-712, Korea (e-mail: dadlani@gist.ac.kr; kskim@gist.ac.kr).

M. S. Kumar is with the Department of Applied Mathematics and Computational Sciences, PSG College of Technology, Coimbatore 641-004, India (e-mail: msk@amc.psgtech.ac.in).

K. Sohraby is with the Department of Computer Science and Electrical Engineering, School of Computing and Engineering, University of Missouri–Kansas City, Kansas, MO 64110-2499 USA (e-mail: sohrabyk@umkc.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LCOMM.2014.2361525

the network and the connectivity distribution [2], respectively, and $B = \left(\frac{1}{\alpha} + \frac{1}{\beta_1} + \ldots + \frac{1}{\beta_T}\right)$, the self-consistency equality in steady-state is derived to be [7]:

$$F(\Theta) = \frac{B\beta_1}{\langle k \rangle} \sum_k kP(k)I_{k,0}. \tag{2}$$

Given (2), it is easier to obtain the R_0 measure than the critical epidemic threshold for the malware model. Thus, R_0 is:

$$R_0 = \left.\frac{dF}{d\Theta}\right|_{\Theta=0} = \frac{B}{\langle k \rangle}\left(\lambda\langle k^2\rangle + \langle k\rangle \sum_{j=0}^{M} \eta_j\gamma_j\right), \tag{3}$$

where $\langle k^2\rangle = \sum_k k^2P(k)$. For stability analysis, we define Γ as the non-negative region in which the system has equilibrium points. Therefore, for some maximum degree K $(1 \le k \le K)$:

$$\Gamma = \Bigg\{ (I_{1,0}, \ldots, I_{1,T}, \ldots, I_{K,0}, \ldots, I_{K,T}, V_0, \ldots, V_M) \in R^{(T+1)K+(M+1)} | S_k + \sum_{i=0}^{T} I_{k,i} \le 1; V_j \le \frac{\eta_j}{1+\eta_j} \Bigg\}. \tag{4}$$

Consequently, if $R_0 < 1$, there exists an infection-free equilibrium (IFE) E^0 and if $R_0 > 1$, the system has an infection-chronic equilibrium (ICE) E^* inside region Γ.

Theorem 1: If $R_0 < 1$, the IFE E^0 is globally asymptotically stable in Γ. Otherwise, if $R_0 > 1$, E^0 is unstable and the system is uniformly persistent in Γ.

Proof: Following the approach in [5], the Lyapunov function for the system is expressed as follows:

$$L(t) = \sum_k \left(a_k \sum_{i=0}^{T} I_{k,i}(t)\right) + \sum_{j=0}^{M} V_j(t), \tag{5}$$

where $a_k = kP(k)/\langle k\rangle$. Differentiating (5) w.r.t. t and substituting for $\dot{I}_{k,i}(t)$ and $\dot{V}_j(t)$ from (1) eventually yields:

$$L'(t) = -\alpha \sum_k a_k I_{k,T}(t) - \frac{\Theta(t)}{B}(1 - R_0) - \sum_{j=0}^{M} \left(V_j(t) - \eta_j\left(1 - V_j(t)\right)\Theta(t)\right). \tag{6}$$

Hence, $L'(t) < 0$ only if $R_0 < 1$ and $V_j \le \frac{\eta_j}{1+\eta_j}$. Furthermore, if $R_0 > 1$ and $V_j(t) > \frac{\eta_j}{1+\eta_j}$, it follows that $L'(t) > 0$ in a small enough neighborhood of E^0 in Γ. Therefore, for $R_0 > 1$, all solutions in Γ sufficiently close to E^0 move away from E^0 thus, making it unstable. This completes the proof. ■

Theorem 2: If $R_0 > 1$, the system has a unique ICE E^*.

Proof: To get E^*, we have $\dot{I}_{k,i}(t) = 0$ and $\dot{V}_j(t) = 0$ as t tends to infinity. Undertaking the same approach as in [6], the equilibrium $E^*(I^\infty_{k,0}, \ldots, I^\infty_{k,T}, V^\infty_0, \ldots, V^\infty_M)$ should satisfy:

$$\begin{cases} -\beta_1 I^\infty_{k,0} + \lambda k S^\infty_k \Theta^\infty + \sum_{j=0}^{M} \gamma_j S^\infty_k V^\infty_j = 0, \\ -\beta_{i+1} I^\infty_{k,i} + \beta_i I^\infty_{k,i-1} = 0; \quad i = 1, \ldots, T-1, \\ -\alpha I^\infty_{k,T} + \beta_T I^\infty_{k,T-1} = 0, \\ -V^\infty_j + \eta_j\left(1 - V^\infty_j\right)\Theta^\infty = 0; \quad j = 0, \ldots, M, \end{cases} \tag{7}$$

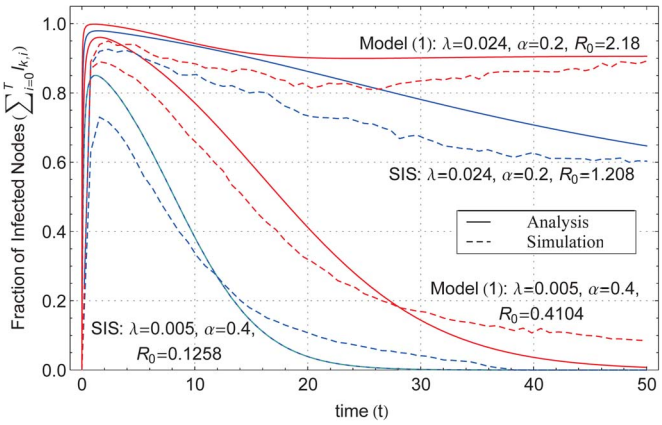

Fig. 2. Numerical results and simulation validation for model (1) and the basic SIS model.

where $\Theta^\infty = (\sum_n nP(n)I^\infty_n)/\langle k\rangle$ and $I^\infty_n = \sum_{i=0}^{T} I^\infty_{n,i}$. From the second and fourth equations of (7), we arrive at $V^\infty_j = \eta_j\Theta^\infty/(1+\eta_j\Theta^\infty)$ and $\sum_{i=0}^{T} I^\infty_{k,i} = B\beta_1 I^\infty_{k,0}$, which when substituted in the first equation of (7) results in:

$$I^\infty_{k,0} = \frac{\lambda k\Theta^\infty + \sum_{j=0}^{M} \frac{\gamma_j\eta_j\Theta^\infty}{1+\eta_j\Theta^\infty}}{\beta_1\left(1 + \lambda kB\Theta^\infty + B\sum_{j=0}^{M} \frac{\gamma_j\eta_j\Theta^\infty}{1+\eta_j\Theta^\infty}\right)}. \tag{8}$$

On substituting (8) in the definition of Θ^∞ and denoting it as $f(\Theta)$ for simplicity, we get the following expression:

$$f(\Theta) = B\sum_k P(k)\frac{\lambda k\Theta + \sum_{j=0}^{M} \frac{\gamma_j\eta_j\Theta}{1+\eta_j\Theta}}{\left(1 + \lambda kB\Theta + B\sum_{j=0}^{M} \frac{\gamma_j\eta_j\Theta}{1+\eta_j\Theta}\right)}. \tag{9}$$

Obviously, the trivial solution of (9) is $f(0) = 0$. For a non-trivial solution $(0 < \Theta \le 1)$, conditions $\left.\frac{df(\Theta)}{d\Theta}\right|_{\Theta=0} > 1$ and $f(1) \le 1$ must be satisfied. So, we have:

$$\frac{B}{\langle k\rangle}\left(\lambda\langle k^2\rangle + \langle k\rangle\sum_{j=0}^{M}\eta_j\gamma_j\right) > 1. \tag{10}$$

Putting (10) in (8) yields $0 < I^\infty_{k,i}, V^\infty_j \le 1$ for $i = 0, 1, \ldots, T$ and $j = 0, 1, \ldots, M$. Hence, E^* is well-defined and there exists one and only one ICE for (1) when $R_0 > 1$. ■

Fig. 2 depicts the numerical results for propagation dynamics of the malware model. For a BA network of 1000 nodes and parameter set-up of $(M, T) = (1, 1)$, $\beta_1 = 0.3$, $\eta_1 = 0.2$, $\gamma_1 = 0.1$, and $\sum_{i=0}^{T} I_{k,i}(0) = 10$, we observe that the fraction of infected nodes reaches zero when $R_0 = 0.4104$, while an epidemic outbreaks at $R_0 = 2.18$. Thus, malware infection dies out when $R_0 < 1$ and persists in the network otherwise, which justifies the above-stated theorems. Validated by simulation, we compare the analytical plots with that of the basic SIS model to demonstrate the impact of additional refinements on the increase in the spreading process. Such increase in the infected density justifies the reduction of the epidemic threshold for the malware model as studied in [7]. Each simulation result in Fig. 2 is averaged over 200 runs.

III. Impact of Immunization Strategies

The practice of investigating spreading patterns of malevolent softwares in SF networks is complemented by the formulation of mitigation strategies tailored to topological specificities of such networks. Due to the vanishingly small epidemic threshold, SF networks show higher susceptibility to infection. As a result, the ineffectiveness of the intuitive uniform immunization in eradicating malwares over these networks has led researchers to design new tactics that take the rank of constituent nodes into account [2], [5], [6]. Henceforth, we examine the impact of the proportional, targeted, acquaintance, and active immunization strategies for the above malware model. These strategies, among many, have been highly appreciated for their performance over SF networks.

A. Proportional Immunization

As stated in the name, the immunizing probability of each node is proportional to its degree. If we denote the fraction of immune nodes by g_k $(0<g_k<1)$ and assume $\bar{\lambda}=\lambda k(1-g_k)$, the first equation of system (1) changes to:

$$\dot{I}_{k,0}(t)=-\beta_1 I_{k,0}(t)+\bar{\lambda}S_k(t)\Theta(t)+\sum_{j=0}^{M}\frac{\gamma_j\bar{\lambda}S_k(t)V_j(t)}{\lambda k}. \quad (11)$$

Solving the above system in steady-state gives us:

$$\begin{cases}\dot{I}_{k,0}=\dfrac{\bar{\lambda}\Theta+\sum_{j=0}^{M}\frac{\gamma_j\bar{\lambda}V_j}{\lambda k}}{\beta_1\left(1+\bar{\lambda}B\Theta+B\sum_{j=0}^{M}\gamma_j\frac{\bar{\lambda}}{\lambda k}V_j\right)},\\ \dot{V}_j=\dfrac{\eta_j\Theta}{1+\eta_j\Theta}.\end{cases} \quad (12)$$

Substituting (12) in (2) gives the self-consistency equation, which when differentiated w.r.t. Θ at $\Theta=0$ gives the basic reproductive ratio under proportional immunization, R_0^P, as:

$$R_0^P=\frac{B}{\langle k\rangle}\left(\lambda\left\langle(1-g_k)k^2\right\rangle+\langle(1-g_k)k\rangle\sum_{j=0}^{M}\eta_j\gamma_j\right). \quad (13)$$

If g_k is some constant g, then we have $R_0^P=(1-g)R_0$.

B. Targeted Immunization

A better approach to overcome the vulnerability of SF networks to selective attacks is to immune the most highly connected nodes. Hence, unlike proportional immunization, the immunization rate in targeted immunization is defined as:

$$g_k=\begin{cases}1, & \text{if } k\geq\kappa_2,\\ p_k, & \text{if } \kappa_1\leq k<\kappa_2,\\ 0, & \text{if } k<\kappa_1,\end{cases} \quad (14)$$

where κ_1 and κ_2 are lower and upper thresholds, respectively, such that all nodes with degree k are immunized if $k\geq\kappa_2$, while a portion p_k $(0<p_k<1)$ of k degree nodes are immunized if $\kappa_1\leq k<\kappa_2$. Repeating the same procedure as in the preceding sub-section, the basic reproductive ratio under targeted immunization, R_0^T, is derived as below:

$$R_0^T=(1-\langle g_k\rangle)R_0-\frac{B}{\langle k\rangle}\left(\lambda\sigma_{k^2,g_k}+\sigma_{k,g_k}\sum_{j=0}^{M}\eta_j\gamma_j\right), \quad (15)$$

where $\sigma_{k^r,g_k}=\langle(k^r-\langle k^r\rangle)(g_k-\langle g_k\rangle)\rangle$ for $r\in\{1,2\}$ and $\langle g_k\rangle=\sum_k g_kP(k)$. Thus, we get $R_0^T<(1-\langle g_k\rangle)R_0^P/(1-g)$.

C. Acquaintance Immunization

Applying targeted immunization requires global information about the network structure (i.e., σ_1 and σ_2), which renders practical applicability in real-world. Acquaintance immunization is an alternative approach in which a random acquaintance of a randomly chosen fraction q of nodes is selected for immunization. Thus, the immunization rate is:

$$g_k=q\frac{kP(k)}{\langle k\rangle};\qquad 0<q<1. \quad (16)$$

Defining ω_r as $\langle k^rP(k)\rangle$, the basic reproductive ratio for this immunization strategy, R_0^{ACQ}, is derived to be:

$$R_0^{ACQ}=R_0-\frac{qB}{(\langle k\rangle)^2}\left(\lambda\omega_3+\omega_2\sum_{j=0}^{M}\eta_j\gamma_j\right). \quad (17)$$

D. Active Immunization

In this immunization strategy, we choose an infected node and immunize its neighbors with degree $k\geq\kappa_2$. With g_k defined as in (14), the first three equations of (1) become:

$$\begin{cases}\dot{I}_{k,0}(t)=-\delta_1 I_{k,0}(t)+\lambda kS_k(t)\Theta(t)+\sum_{j=0}^{M}\gamma_jS_k(t)V_j(t),\\ \dot{I}_{k,i}(t)=-\delta_{i+1}I_{k,i}(t)+\delta_iI_{k,i-1}(t),\ i=1,\ldots,T-1,\\ \dot{I}_{k,T}(t)=-\mu I_{k,T}(t)+\delta_TI_{k,T-1}(t),\ j=0,\ldots,M,\end{cases} \quad (18)$$

where $\delta_i=(\beta_i+\bar{g}_k)$, $\mu=(\alpha+\bar{g}_k)$, $\bar{g}_k=\langle kg_k\rangle/\langle k\rangle$, and $\langle kg_k\rangle=\langle k\rangle\langle g_k\rangle+\sigma_{k,g_k}$. It can be easily show that the basic reproductive ratio for active immunization, R_0^{ACT}, is:

$$R_0^{ACT}=\frac{B_1}{\langle k\rangle}\left(\lambda\langle k^2\rangle+\langle k\rangle\sum_{j=0}^{M}\eta_j\gamma_j\right)=\frac{B_1}{B}R_0, \quad (19)$$

where $B_1=\left(\frac{1}{\alpha+\bar{g}_k}+\frac{1}{\beta_1+\bar{g}_k}+\ldots+\frac{1}{\beta_T+\bar{g}_k}\right)$. Hence, R_0^{ACT} is comparable to R_0 in effectiveness for small values of $\bar{g}_k$.

IV. Numerical Results and Discussions

We validate our analysis via numerical simulations for a BA network comprising of $N=1000$ nodes with $\langle k\rangle\simeq 2m$, $\langle k^2\rangle\simeq 2m^2\log(\sqrt{N})$, and $P(k)=2m^2k^{-3}$ as given in [2]. In our settings, we consider $m=2$ and $(\kappa_1,\kappa_2)=(600,800)$.

Fig. 3 illustrates the effectiveness of the four immunization strategies in terms of their basic reproductive ratios (R_0^*) for different values of infection rate (λ) and refinements (M and T). We first investigate the relationship between R_0^* in terms of λ for $(M,T)=(3,2)$, $\alpha=0.7$, $q=0.9$, and $p_k=k/N$ in Fig. 3(a). For a fair comparison, we fix the fraction of immunized nodes to 0.4 and compare the results with that of the basic SIS model where $(M,T)=(0,0)$. We observe that for any λ value, all four immunization schemes reduce the number of secondary infections introduced in the network as compared to the non-immunized case (R_0). Among the four, however, the striking effectiveness of the proportional and targeted schemes is evident. As λ increases, the rate at which

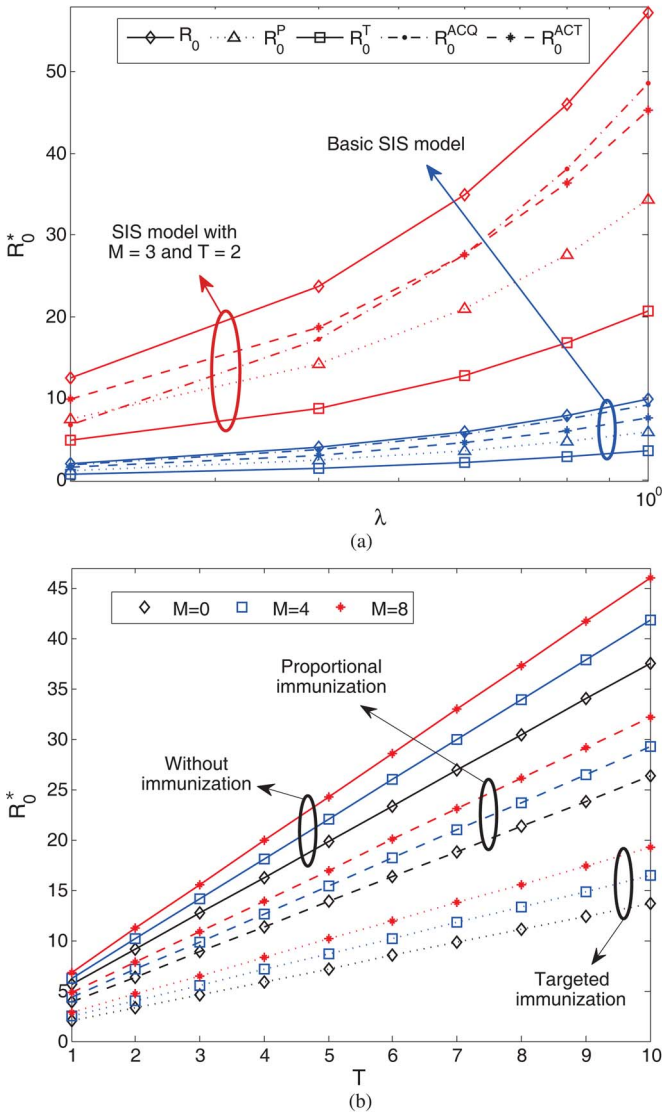

Fig. 3. Reduced basic reproductive ratio (R_0) of the malware model in presence of immunization. (a) Basic reproductive ratios (R_0^*) versus spreading rate (λ). (b) Basic reproductive ratios (R_0^*) versus infection delay (T).

nodes are newly infected drastically drops under proportional and targeted immunizations as compared to the other two immunization tactics. This is in accordance with the fact that immunizing nodes with higher connectivity reduces the risk of infection spreading through them. Choosing proportional and targeted schemes as the best-case baselines, we plot their corresponding reproductive ratios against R_0 for varying values of M and T in Fig. 3(b). In here, we show how generalizing the malware model to multiple infection stages and propagation mediums would result in an epidemic outbreak. Even though imposing proportional and targeted immunization strategies does not bring the system to an infection-free state, their effectiveness in reducing the number of newly infected nodes is visibly distinguishable. For instance, at $(M, T) = (4, 5)$, R_0 is reduced by nearly 31% and 63% under proportional and targeted immunization schemes, respectively.

In Fig. 4, we show the outperformance of proportional and targeted immunizations in combating the malware spread. For $N = 1000$, $\lambda = 0.024$, $\alpha = 0.2$, $(M, T) = (1, 1)$, $\beta_1 = \gamma_1 = \eta_1 = 0.1$, $V_1(0) = 0.2$, and $\sum_{i=2}^{3} I_{k,i}(0) = 10$, we see that the

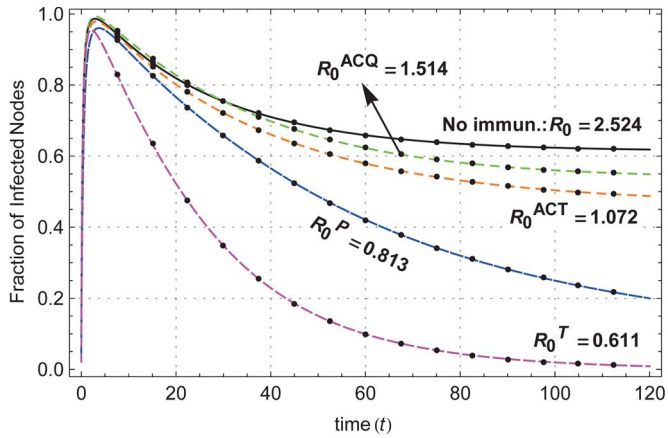

Fig. 4. Fraction of infected nodes under immunization.

infection-chronic system $(R_0 = 2.524)$ changes to infection-free states of $R_0^P = 0.813$ and $R_0^T = 0.611$ when targeted and proportional immunizations are applied, respectively. On the contrary, acquaintance and active immunization schemes fail to suppress the epidemic outbreak as R_0^{ACQ} and R_0^{ACT} maintain values greater than 1. Note that the rate at which the malware infection dies out for targeted immunization is higher than that for proportional immunization. This is because targeted immunization focuses on immunizing nodes that are more likely to spread the infection, i.e. nodes that are highly connected. Once these nodes are immunized, the capability of SF networks to carry the malware decreases substantially.

V. Concluding Remarks

We analyzed the stability of a network malware spread model with infection delay and vectors. Accounting for the heterogeneous nature of scale-free networks, we investigated and compared the impact of various schemes for immunizing the spreading model in terms of basic reproductive ratio. We showed the effectiveness and convenience of targeted and proportional immunization strategies in controlling the spread in comparison to the acquaintance and active counterparts. Assessing the model under different immunization cost functions can be considered as a potential future work.

References

[1] S. Peng, S. Yu, and A. Yang, "Smartphone malware and its propagation modeling: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 925–941, 2014.
[2] R. Pastor-Satorras and A. Vespignani, "Immunization of complex networks," *Phys. Rev. E, Stat. Nonlin. Soft Matter Phys.*, vol. 65, no. 3, pp. 036104-1–036104-8, Feb. 2002.
[3] M. Ajelli, R. L. Cigno, and A. Montresor, "Modeling botnets and epidemic malware," in *Proc. IEEE ICC*, May 2010, pp. 1–5.
[4] S.-M. Cheng, W. C. Ao, P.-Y. Chen, and K.-C. Chen, "On modeling malware propagation in generalized social networks," *IEEE Commun. Lett.*, vol. 15, no. 1, pp. 25–27, Jan. 2011.
[5] J.-P. Zhang and Z. Jin, "The analysis of an epidemic model on networks," *Appl. Math. Comput.*, vol. 217, no. 17, pp. 7053–7064, May 2011.
[6] T. Li, Y. Wang, and Z.-H. Guan, "Spreading dynamics of a SIQRS epidemic model on scale-free networks," *Commun. Nonlin. Sci. Numer. Simul.*, vol. 19, no. 3, pp. 686–692, Mar. 2014.
[7] A. Dadlani, M. S. Kumar, S. Murugan, and K. Kim, "System dynamics of a refined epidemic model for infection propagation over complex networks," *IEEE Syst. J.*, to be published.
[8] J. M. Heffernan, R. J. Smith, and L. M. Wahl, "Perspectives on the basic reproductive ratio," *J. R. Soc. Interface*, vol. 2, no. 4, pp. 281–293, Sep. 2005.
[9] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.