On Modeling Malware Propagation in Generalized Social Networks

Shin-Ming Cheng, *Member, IEEE,* Weng Chon Ao, Pin-Yu Chen, and Kwang-Cheng Chen, *Fellow, IEEE*

***Abstract*—A hybrid malware on smart phones can be propagated by both end-to-end messaging services via personal social communications and short-range wireless communication services via spatial social interactions. Inspired from epidemiology, we propose a novel differential equation-based model to analyze the mixed behaviors of delocalized infection and ripple-based propagation for the hybrid malware in generalized social networks consisting of personal and spatial social relations. Validated by simulations, our model serves as the very first analytical model successfully approximating the complicated propagation behaviors of the hybrid malware.**

***Index Terms*—Epidemiology, malware propagation, proximity malware, social network.**

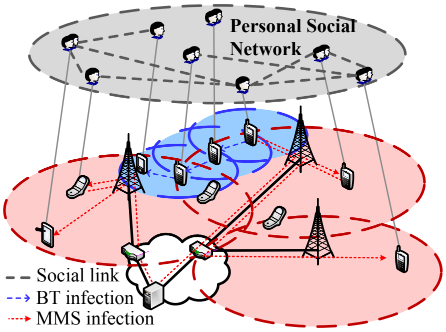

Fig. 1. Propagation of a hybrid malware in the generalized social network.

I. Introduction

THE popularity of mobile smart phones with richer wireless communication capabilities allows extensive social interactions in the following aspects. First, the communication between an individual and his friends in *personal social network* interconnected by call records and contacts is facilitated by portability of handset. Second, smart phones equipped with short-range wireless communication (SRWC) technology such as WiFi or Bluetooth (BT) realize peer-to-peer communication between individuals in geographic proximity, building a *spatial social network*. Such geographical interdependency for the individuals in cellular infrastructure couples personal social network with spatial social network as a generalized social network, which amplifies the opportunities for attacks from self-replicating malware.

The malware on handsets typically exploits messaging services [1] or uses SRWC services to propagate. The differential equation-based approach characterizing virus spreading in Internet [2], [3] is feasible to model the messaging malware dissemination due to homogeneity holds in person social network. On the other hand, the behavior of malware spreading by SRWC services was approximated by differential equation [4], [5] or investigated by agent-based simulation [6], [7].

In the generalized social network consisting of personal and spatial social relations, as shown in Fig. 1, a hybrid malware can exploit both messaging and SRWC services to spread. To the best of our knowledge, it is desirable to have a mathematical model analyzing the mixed behaviors of long-range infection pattern from spreading via messaging service and ripple-based infection pattern from propagating via SRWC. The existing investigations conducted by agent-based model [7] or by simulation [8] try to precisely capture attributes of all individuals in the network and the interactions among them. However, the complexity of modeling individual-level details significantly increases computational costs [9] and thus such agent-based simulations are unable to act as a quick reference to identify such malware in large networks.

This letter proposes a novel analytical model to efficiently analyze the speed and severity for spreading the hybrid malware such as Commwarrior that targets multimedia messaging service (MMS) and BT. Validation against conducted simulation experiments reveals that our model developed from the Susceptible-Infected (SI) model in epidemiology accurately approximates mixed spreading behaviors in large areas without the huge computational cost, which helps estimate the damages caused by the hybrid malware and aids in the development of detection and containment processes.

II. System Model

The proposed model is originated from the SI model in epidemic theory [10] to measure propagation of infections within a population under risk. The communication between a compromised and a noncompromised handset is modeled as a contact between an infected individual and a susceptible one, where a susceptible node acquires infection and never becomes susceptible again. This is due to the users' lack of concern about the threat of malwares and the limited capability of current antiviral software. The population in our model is the total number of nodes N in the network which are assumed to be stationary and uniformly distributed with node density ρ. We assume that all nodes are MMS- and BT-enabled to maintain the homogeneous mixing property. Denote subpopulation function $I(t) = I_{BT}(t) + I_{MMS}(t)$ as the total number of compromised handsets at time t, where $I_{BT}(t)$ and $I_{MMS}(t)$ are those that have been infected via BT and MMS at time t, respectively. Likewise, $S(t)$ denotes the set

Manuscript received May 16, 2010. The associate editor coordinating the review of this letter and approving it for publication was J. Holliday.

This research is sponsored by the National Science Council under the contract of NSC 98-2221-E-002-065-MY3 and NSC 99-2219-E-002-024.

The authors are with the Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan (e-mail: smcheng@cc.ee.ntu.edu.tw, {r97942044, r98942052}@ntu.edu.tw, chenkc@cc.ee.ntu.edu.tw).

Digital Object Identifier 10.1109/LCOMM.2010.01.100830

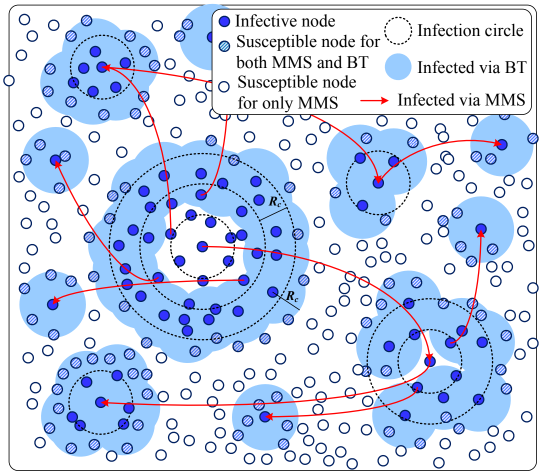

Fig. 2. The spreading phenomenon of a hybrid malware.

of susceptible nodes at time t. Obviously, we have

$$I(t) + S(t) = I_{BT}(t) + I_{MMS}(t) + S(t) = N, \tag{1}$$

and

$$\frac{dI(t)}{dt} = \frac{dI_{BT}(t)}{dt} + \frac{dI_{MMS}(t)}{dt}. \tag{2}$$

Without loss of generality, we assume that only one handset is infected at the initial stage, that is, $I(0) = I_{MMS}(0) = 1$ and $I_{BT}(0) = 0$. The rates of malware infection β_{BT} and β_{MMS} respectively represent the probabilistic rates at which an infective node communicates with and compromises a susceptible node through BT and MMS. The average degrees of a node connecting via BT and MMS are denoted as η_{BT} and η_{MMS}, respectively.

III. A Differential Equation-Based Analytical Model

In this section, we present the analysis of the mixed propagation behaviors by deriving $I_{MMS}(t)$ and $I_{BT}(t)$ describing the time dynamics of infected subpopulations.

A. Spreading Dynamics via MMS

When an infected node tries to spread malware via MMS, it behaves like traditional email virus seen on the Internet. Typically, it sends MMS messages to phone numbers found in the address book since message from an acquaintance has higher possibility to be opened and further activated [1]. As shown in Fig. 2, malware propagating via MMS follows a delocalized pattern since that contacts in address book are often far away.

By extracting the contacts in the address book of a handset, a personal social network is constructed describing the social relationships between handsets, which is exploited by MMS malware for spreading. The average degree η_{MMS} of the network means the average number of contacts in the address book and infection rate β_{MMS} indicates the probability that a susceptible node becomes infected after receiving the malware. Please note that the probability that victim confirms and installs the malware may affect β_{MMS}. Under the assumption of homogeneous mixing in the personal social network, the basic differential equation that describes the dynamics of infected subpopulation by MMS with time is

$$\frac{dI_{MMS}(t)}{dt} = \beta_{MMS}\frac{S(t)(\eta_{MMS} - 1)}{N}I(t), \tag{3}$$

where $\eta_{MMS} - 1$ accounts for the fact that one infected node implies at least one of its neighbors being infected.

B. Spreading Dynamics via BT

When an infected node intends to spread malware via BT, it first scans to search the nearby nodes within its transmission range R_c and connects to the neighbor so as to determine the susceptible neighbors for propagating. In this case, the average number of neighbors η_{BT} equals $\rho\pi R_c^2$. The probability that a susceptible node becomes infected after receiving the malware β_{BT} depends on probability that it confirms and opens the malware. Comparing with MMS with delocalized pattern, the spreading effect via BT facilitated by mobility is much small and thus human mobility is ignored in our model.

The behavior of such spontaneous spreading can be regarded as a ripple centered at the source infected node which grows with time [5]. It is approximated into our model by having only the infected nodes that lie on the periphery of an infection circle can communicate with the susceptible nodes located at a distance of at most R_c outside the infection circle, and thus have the potential to infect them. In other words, as illustrated in Fig. 2, the spatial spreading of the epidemics through BT is only contributed from the wavefronts of infection circles, while the infected nodes located in the interior of the infection circles are not engaged in further spatial infections. This phenomenon should be carefully modeled otherwise the overestimation problem [3] leads significant deviation.

Without loss of generality, we assume that a single infection circle is generated at time r by a point source infected through MMS and kept stretching for s time units. Then its incremental spatial infection at time $r + s$ is

$$G'(r, s) \triangleq \frac{dG(r, s)}{ds} = \beta_{BT}\frac{S(r+s)\cdot\frac{1}{2}\eta_{BT}}{N}c\sqrt{G(r, s)}, \tag{4}$$

where $\frac{1}{2}\eta_{BT}$ accounts for the fact that for an infected node on a periphery, roughly half of neighbors outside the infection circle are susceptible. Under the assumption of uniform distribution for nodes, $c = 2R_c\sqrt{\rho\pi}$ is the proportionality constant [5]. The incremental spatial infection at time t of all infection circles is given by

$$\frac{dI_{BT}(t)}{dt} = \int_0^t I'_{MMS}(\tau)G'(\tau, t - \tau)d\tau. \tag{5}$$

It means that there are $I'_{MMS}(\tau)d\tau$ point sources originated through MMS-infection at time τ and each contributes $G'(\tau, t - \tau)$ incremental spatial infection at time t.

IV. Numerical Results

A. Analysis Discussion

Fig. 3 illustrates the analytical plots depicting the propagation dynamics of a hybrid malware spreading via only BT,

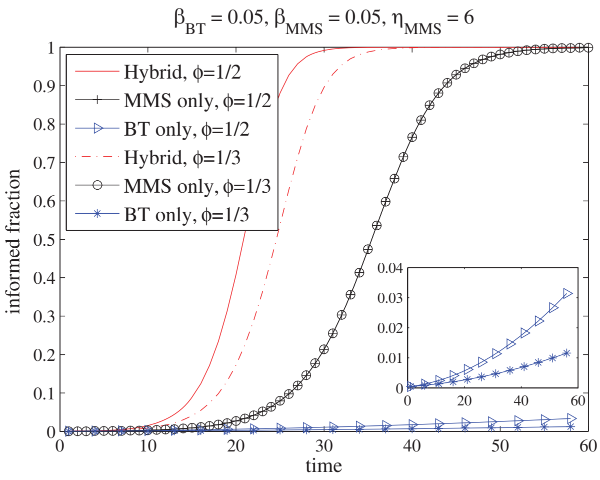

Fig. 3. Analytical results of propagation dynamics of spreading via BT, MMS, and both under $\phi = 1/2$ and $1/3$.

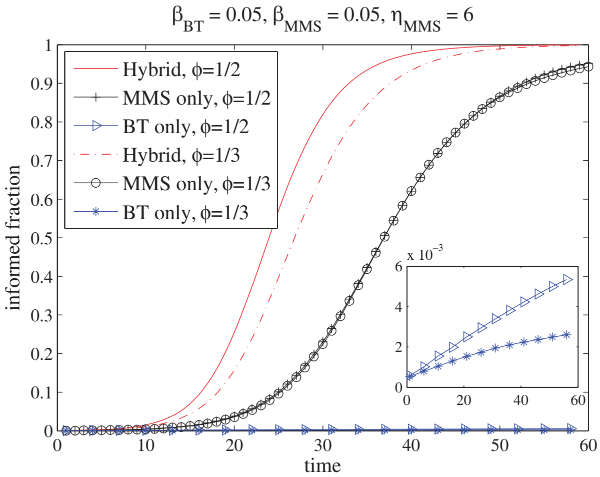

Fig. 4. Simulation results of propagation dynamics of spreading via BT, MMS and both under $\phi = 1/2$ and $1/3$.

only MMS, and both among 2000 nodes under $\rho = 0.8$. We consider the impact of infection degree ratio (defined as $\phi = \eta_{BT}/\eta_{MMS}$) on the the propagation process in terms of speed and reachability. The parameter setups are $\beta_{BT} = \beta_{MMS} = 0.05$ and $\eta_{MMS} = 6$.

Fig. 3 also shows that the propagating via only BT is relatively slow compared with that via only MMS due to spatial spreading characteristics. We also observe the same phenomenon on the hybrid malware with much faster propagation speed, where the rapid invasion via MMS dominates the propagation dynamics. When ϕ increases from $1/3$ to $1/2$ (i.e., η_{BT} increases from 2 to 3), our model indicates a significant increase in the propagation speed in early stages of spreading process. This is in accordance with the fact that a larger η_{BT} results in a larger infected subpopulation who could exploit both BT and MMS to spread, increasing propagation severity.

Note that agent-based emulation model [7] and simulation [8] try to characterize behaviors of the N nodes and all interactions among them, which requires huge computation overhead. In contrast, our model aggregates the N nodes into two states and only tracks the behavior of these two states and the interactions between them, such that our model can be more computationally effective.

B. *Simulation Study*

To validate the analytical model, we develop experiments to simulate malware spreading via social and spatial interactions among 2000 individuals uniformly deployed in a 50×50 plane. The constructions of social networks and setup of parameters (e.g., $\eta_{MMS} = 6$) follow the data sheet in [8]. Fig. 4 illustrates the time for a hybrid malware to infect a given fraction of the network, spreading via only BT, only MMS and both, against different values for ϕ. Each reported result is averaged over 300 simulation runs.

We observe that the curves of propagation dynamics closely match our analytical model, where limited discrepancy exists mainly due to that the hybrid malware may propagate to individuals who have already been infected and uncertain boundary conditions could not be considered in the analysis.

V. Conclusion

Comparing with the existing agent-based model or simulation with computation burden, our analytical model basing on differential equations works more efficiently and could act as a quick reference to gather approximate knowledge of propagation speed and severity of hybrid malwares with various settings of infection rates and average node degrees in generalized social networks. The security assessment could adopt such results to develop detection and containment strategies and processes so as to avoid vital outbreak.

References

[1] Z. Zhu, G. Cao, S. Zhu, S. Ranjan, and A. Nucci, “A social network based patching scheme for worm containment in cellular networks,” in *Proc. 28th IEEE Int’l Conf. Comput. Commun. (INFOCOM ’09)*, Apr. 2009, pp. 1476–1484.

[2] R. Pastor-Satorras and A. Vespignani, “Epidemic spreading in scale-free networks,” *Phys. Rev. Lett.*, vol. 86, no. 14, pp. 3200–3203, Apr. 2001.

[3] C. C. Zou, D. Towsley, and W. Gong, “Modeling and simulation study of the propagation and defense of Internet e-mail worms,” *IEEE Trans. Dependable Secure Comput.*, vol. 4, no. 2, pp. 105–118, Apr.-Jun. 2007.

[4] J. W. Mickens and B. D. Noble, “Modeling epidemic spreading in mobile environments,” in *Proc. 4th ACM Workshop Wireless Security (WiSe ’05)*, Sep. 2005, pp. 77–86.

[5] P. De, Y. Liu, and S. K. Das, “An epidemic theoretic framework for vulnerability analysis of broadcast protocols in wireless sensor networks,” *IEEE Trans. Mobile Comput.*, vol. 8, no. 3, pp. 413–425, Mar. 2009.

[6] G. Yan, L. Cuellar, and S. Eidenbenz, “Bluetooth worm propagation: mobility pattern matters!” in *Proc. 2nd ACM Symp. Inf., Comput. and Commun. Security (ASIACCS ’07)*, Mar. 2007, pp. 32–44.

[7] A. Bose and K. G. Shin, “On capturing malware dynamics in mobile power-law networks,” in *Proc. 4th Int’l Conf. Security and Privacy in Commun. Netw. (SecureComm ’08)*, no. 12, Sep. 2008.

[8] P. Wang, M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, “Understanding the spreading patterns of mobile phone viruses,” *Science*, vol. 324, no. 5930, pp. 1071–1075, May 2009.

[9] H. Rahmandad and J. Sterman, “Heterogeneity and network structure in the dynamics of diffusion: comparing agent-based and differential equation models,” *Manag. Science*, vol. 54, no. 5, pp. 998–1014, May 2008.

[10] D. J. Daley and J. Gani, *Epidemic Modelling: An Introduction*. Cambridge University Press, 2001.