Hindawi
Security and Communication Networks
Volume 2018, Article ID 4691203, 7 pages
https://doi.org/10.1155/2018/4691203

Research Article

Epidemic Model of Algorithm-Enhanced Dedicated Virus through Networks

Yi-Hong Du and Shi-Hua Liu

Department of Information Technology, Wenzhou Vocational & Technical College, Wenzhou 325035, China

Correspondence should be addressed to Yi-Hong Du; bigdata@wzvtc.edu.cn

Received 3 April 2018; Accepted 14 May 2018; Published 7 June 2018

Academic Editor: Angel M. Del Rey

Wi-Fi networks almost cover all active areas around us and, especially in some densely populated regions, Wi-Fi signals are strongly overlapped. The broad and overlapped coverage brings much convenience at the cost of great security risks. Conventionally, a worm virus can infect a router and then attack other routers within its signal coverage. Nowadays, artificial intelligence enables us to solve problems efficiently from available data via computer algorithm. In this paper, we endow the virus with some abilities and present a dedicated worm virus which can pick susceptible routers with kernel density estimation (KDE) algorithm as the attacking tasks automatically. This virus can also attack lower-encryption-level routers first and acquire fast-growing numbers of infected routers on the initial stage. We simulate an epidemic behavior in the collected spatial coordinate of routers in a typical area in Beijing City, where 56.0% routers are infected in 18 hours. This dramatical defeat benefits from the correct infection seed selection and a low-encryption-level priority. This work provides a framework for a computer-algorithm-enhanced virus exploration and gives some insights on offence and defence to both hackers and computer users.

1. Introduction

Wi-Fi technology, dating from the 90*s*, has proceeded in an explosive manner. Nowadays, almost all electronic devices are equipped with Wi-Fi modules and the number of routers has increased rapidly and complementarily. Widely deployed routers, with broad radiating areas, have their Wi-Fi signals overlapped in free space and, on the other hand, people usually set vulnerable passwords for convenience. Wi-Fi networks provide a target-rich epidemic spread platform for cybercriminals.

Traditional attacks are to plant viruses or worms having malicious or fraudulent motivation on personal computers [1–3]. Actually, Wi-Fi routers are perfect target platforms since routers are always on and connected to the Internet with a usually low security level or sometimes even no firewall software. Routers emit Wi-Fi signals all over the space within the range of tens of meters. The relatively close proximity is appreciable enough to perform an attack in such a densely populated environment [4]. In infection dynamics, an end user is infected as a seed or a broiler where the worm virus analyzes the Wi-Fi router and probes potential devices in the coverage. In this manner, worm viruses can spread through the router network [5].

Flaws in wireless protocol or misconfiguration of the access devices [6] are potentially utilized to control the Wi-Fi routers. As is popularly known, a public trap Wi-Fi router is set up to provide a free Internet and attract connection. Once one user falls into the trap, several types of attack, including man-in-the-middle attack and denial-of-service attack, can be conducted by virtue of the infected router [7]. In late 2017, the four-way handshake, said to be free from attacks, is vulnerable to a key reinstallation attack [8].

An increasing trend to investigate epidemic spread models in networks follows the explosive increase in the amount of Wi-Fi routers and mobile terminals. Mobile phone virus via multimedia messaging services is presented and it is predicted that viruses will break out when mobile phone market reaches a certain threshold [9]. A Susceptible-Infected-Recovered (SIR) model is constructed to simulate the spread of hypothetical Wi-Fi malwares in real-world router locations and the Wi-Fi networks are demonstrated to be potential

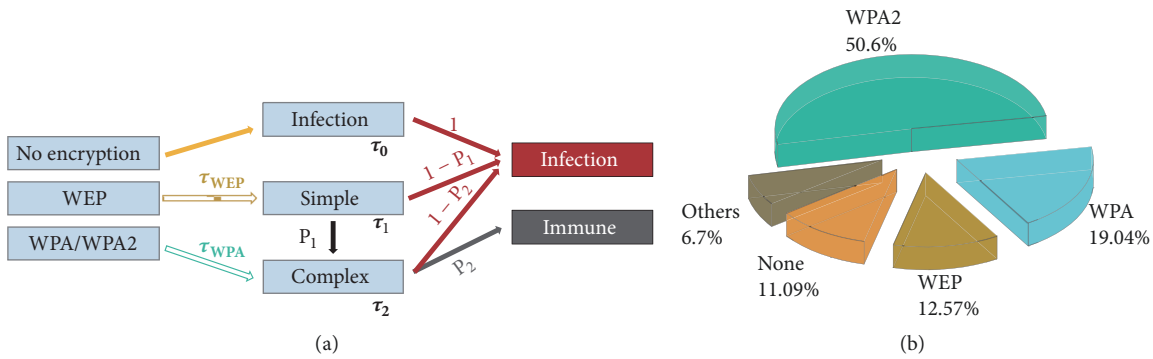

Figure 1: (a) **Flow diagram of the epidemic spread model.** No encryption routers can be affected in τ_0. Router encryption protocols are broken once the routers are attacked for τ_{WEP}(or τ_{WPA}) and then they step into the cracking password dynamics. A simple (complex) password can be cracked in τ_1(τ_2), whose failure rate is P_1 (P_2). (b) **Encryption type ratio in China** Data refers to the website *Wigle.net*.

and vulnerable platforms [10, 11]. In the network [12], the scale of access point connectivities in the victim population is a more important factor than others [4]. An enhanced model takes the vulnerabilities in Wi-Fi routers and protocol into consideration but no big progress in dynamics is made [6]. Inclusion of end terminals leads to a different epidemic spread model [13].

Recently, a diversity of interesting results on this epidemic spread model [14–21] has been demonstrated; however, plentiful fascinating works, including barely developed viruses endowed with abilities, need to be exploited. Thanks to the great advance of algorithm and computing capability, artificial intelligence develops rapidly. Viruses are destined to have the ability of identifying the environmental property from acquired data and make the optimal decision. Specifically, a virus chooses the appropriate router as the seed to begin its infection process according to the local router information. The victim candidate it prefers should be a router located in a crowded region (or a hub) with an outward-spreading potential. Kernel density estimation (KDE) algorithm estimates the probability density distribution directly from a set of spatial data without a prior distribution assumption [22]. This algorithm serves a simple and visualized approach to the selection of the infection seed.

2. Methods

2.1. Epidemic Spread Model Illustration. The epidemic model is established based on the following simplification. As shown in Figure 1(a), a router with no encryption can be infected directly in τ_0 and routers with encryption are usually divided into two types, WEP and WPA/WPA2. WEP-encrypted router can be broken when it is attacked for τ_{WEP} and then follows the password crack dynamics. The attacker attempts to crack the router with the simple password library in τ_1. There are two cases after that: the router is infected with the probability $(1 - P_1)$ or with P_1 the attacker has to change to crack it with the complex password library in τ_2 if it is not successful. Then again there will be two cases; that is, either the router is infected in $(1-P_2)$ or not infected in P_2 (immune to attackers). WPA/WPA2 encryption has long been thought of as immune to attackers until the work in 2017 appeared [8]. An analogy is made with our model that the WPA/WPA2 encryption will break down in τ_{WPA} and the password will be cracked in τ_2 with the probability of a successful infection $(1 - P_2)$.

Table 1: **Parameter assignment.** The first three lines show the encryption types and their corresponding typical time scales and the last two lines show typical time scales to crack the simple or complex password and the probability to fail cracking.

τ_0	$1min$	p_{nop}	11.9%
τ_{WEP}	$20min$	p_{WEP}	13.5%
τ_{WPA}	$45min$	p_{WPA}	74.6%
τ_1	$1min$	p_1	50%
τ_2	$5min$	p_2	40%

Typical time scales shown in Table 1 refer to the previous literature in the year 2009, also in consideration of the computing capacity leap in recent years [10]. The probability distribution of encryption types (p_{nop}, p_{WEP}, and p_{WPA} shown in Table 1) are normalized from the data displayed in Figure 1(b) without the category "others."

2.2. Data Acquisition and Processing. We collect 5001 pieces of raw data in a region in Beijing City (roughly latitude: 39.9141° N; longitude: 116.4050° E) from the website *wiglet.net*. We attempt to clean these data in two steps. First, some data with extremely unrealistic location information is deleted (e.g., Router 3901). Second, we delete the duplicate information whose data does not exist via identifying the Media Access Control (or MAC) address despite the same location information. After cleaning the acquired data, we label each router with an encryption type according to the ratio the website provides. Subsequently, we pick each router and collect router information within its radiating radius to construct this router's infection candidate set. At last, the encryption types are sorted from no encryption to low-encryption type and high-encryption type to determine the infection order.

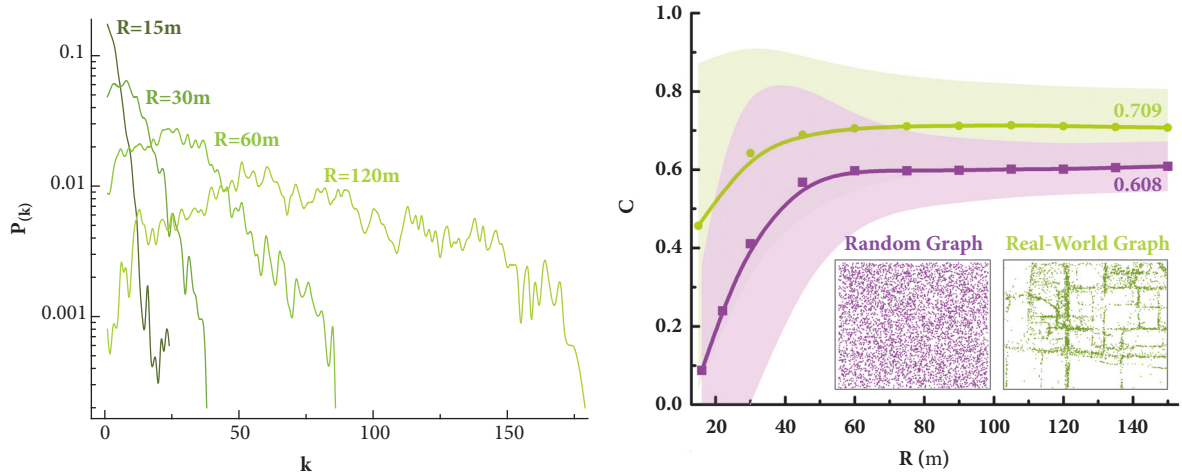

FIGURE 2: **Degree distribution and clustering coefficient for different radiating radius.** (a) The degree distribution shows an exponential decay when radius is small and a flat distribution when radius gets larger. (b) This plot compares clustering coefficients of a random generated graph and the real-world region in Beijing City (maps shown in the inset) and the shadow indicates the error bars.

3. Results

In this paper, we assume a dedicated worm virus which can pick a more susceptible router region from the router network with a KDE algorithm and carry out the whole attacking procedure automatically, ranging from searching victims to installing the malware. The infection model refers to the previous literature where malicious worm is spreading directly from one wireless router to another via free space wireless propagation [10]. This virus can also query the encryption information, such as No Encryption, Wired Equivalent Privacy (WEP), Wi-Fi Protected Access (WPA/WPA2) encryption protocol, and attack routers with relatively low-encryption-level routers first at the same condition. The virus is triggered to stop attacking to enhance the task efficiency if the attacking duration reaches the predetermined threshold.

3.1. Real-World Network Characterization. We sample the real-world geographic location data for wireless routers from the wireless network mapping site (*wigle.net*). The detailed data acquisition and processing are shown in the Methods section. For notational convenience, routers are labelled in an identifier number.

In the Wi-Fi network, the radiating coverage of a router, ranging from tens of meters to more than a hundred meters [23], depends strongly on both the internal factors (such as the radiating power and the antenna orientation) and the external factors (such as local barriers and signal interference). For simplicity, we keep the radiating radius R as constant and consider four different values of the maximum radiating radius which are $15m, 30m, 60m$, and $120m$ to analyze the degree distribution [24] in real-world router network. Figure 2(a) describes the probability distribution that there are k other routers locating within the range R. The Wi-Fi network, whose degree distribution follows an exponential decay, is a scale-free network when R is set to $15m$ [25]. Note that, with an increased R, the distribution becomes more and more flat, which indicate the fact that a larger radiating radius will overcome geographic obstacles.

Local interconnectedness can be characterized by the clustering coefficient

$$C = \frac{1}{n}\sum C_i \tag{1}$$

where C_i represents the fraction of the neighbors of Router i that are also interconnected. It is mathematically expressed as $(2 \times l_i)/k_i(k_i - 1)$, with l_i indicating the number of links bridging neighbors of Router i and $k_i(k_i - 1)/2$ representing the number of all possible connections between these neighbors. In Figure 2(b), we compare clustering coefficients of a random generated graph and the real-world region in Beijing City and the shadows indicate the corresponding error bars. The results show that the real-world network has a stronger clustering property than a random one. It makes sense that a network with larger clustering coefficient is vulnerable to the epidemic virus.

3.2. Algorithm-Enhanced Hunt for Infectious Source. With the knowledge of the network property, we could inject the worm virus into the network purposefully. Our dedicated virus acquires spatial distribution information of routers from some interface first and analyze the appropriate injection candidate [26]. What viruses prefer should be a router located in a crowded region (or a hub) with an outward-spreading potential. A simple and visualized approach to find these routers is the KDE algorithm, which can estimate the probability density distribution directly from a set of given location data without a prior distribution assumption from a macroscopic perspective [22].

In Figure 3(a), we map all position information directly on the map with brighter color indicating more routers and

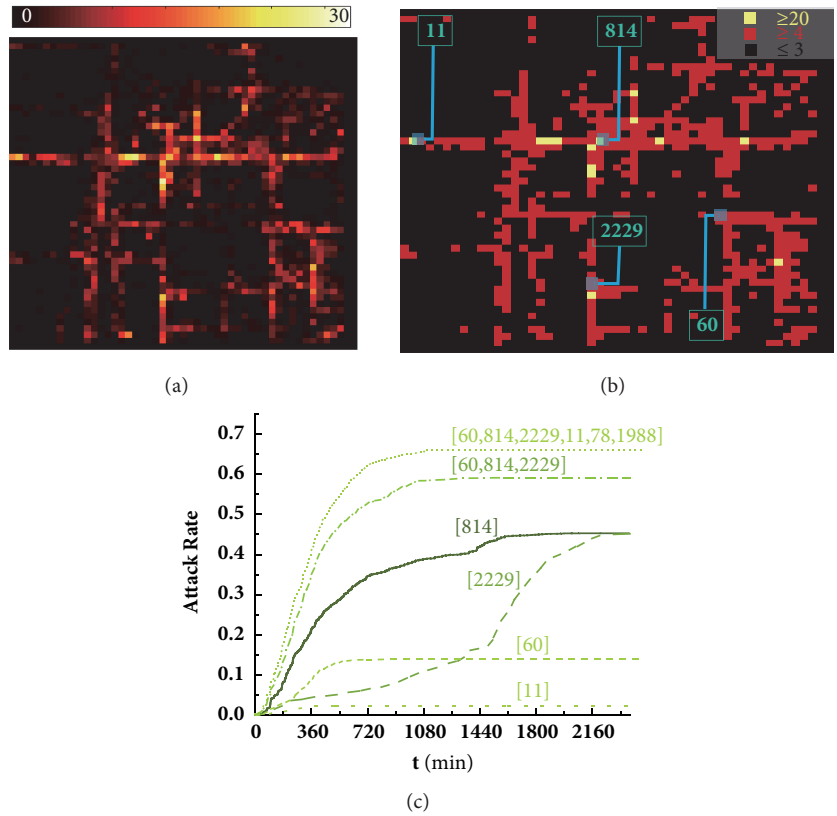

FIGURE 3: **Algorithm assists the viruses to infect the network efficiently and the corresponding performance test.** (a, b) The dedicated worm virus acquires the router network distribution and carries out the KDE algorithm to pick the appropriate infection seed. (b) The matrix element in light yellow (red or black) indicates that there are over 20 (over 4 or no more than 3) routers in the area. (c) Different seed choices lead to different trend and the final attack rate.

darker color indicating fewer ones and in Figure 3(b) we generate a two-dimensional matrix sized 51×51. Each matrix element means the number of routers in the corresponding area of about $50 \times 50m^2$ and three levels of router density are characterized in different colors. A virus learns the network from the reduced matrix. It finds a connected giant component and checks its neighbor environment. The right candidate is the one lying in the largest giant components and having the outward-spreading potential and it itself has enough neighbors to attack. In Figure 3(b), the identifier number and the location of some routers are labelled in blue and green, respectively. We choose six different sets of infection seeds and the relation between the attack rate and evolution time is shown in Figure 3(c). If the virus is seed only in single router among the four routers (11, 60, 814, and 2229), Router 814 shows the highest attack rate within the shortest time. Note that Routers 11 and 60 are in isolated regions and Routers 814 and 2229 are in the same largest clusters. This gives also the reason for different attack rates. The trend of Router 2229 on the initial stage is flat, for small amount of seed neighbors do not accumulate enough broilers. From this plot, we conclude that an appropriate selection of initial seed can dramatically influence the attack rate and efficiency. Specifically, a randomly selected infective seed probably drops in an isolated region and has a restricted spreading region. There is also possibilitiy that the seed has a small amount of neighbors and spreads slowly in the initial stage, though in a large cluster.

3.3. Visualization of the Epidemic Dynamics. To visualize the time evolution of the epidemic behaviors, we set the radiating radius R to $45m$ and the virus picks the router set 60,814 and 2229 as the initial infection seed. Figures 4(a)–4(f) present the epidemic behaviors where infected routers are labelled yellow and the infection time and attack rate are displayed on the top left. The overall attack rate is shown in Figure 4(g). We see that 36.2% routers are infected in just 6 hours and 56.0% are infected in 18 hours. We also present the concept of threads, which means the number of viruses that are attacking routers concurrently in a given time. The gray curve in Figure 4(g) shows the number of threads changing with infection time. A sharp exponential increase from 0 to more than 300 in the first half hour is attributed to the dedicated virus attacking routers with no encryption within its range as the first priority and a good choice of initial broiler seed is set. This rapid expansion in the beginning lays a foundation of high infection efficiency.

4. Discussion

This paper provides insights on the offence and defence of a dedicated virus. As a hacker, to develop an efficient

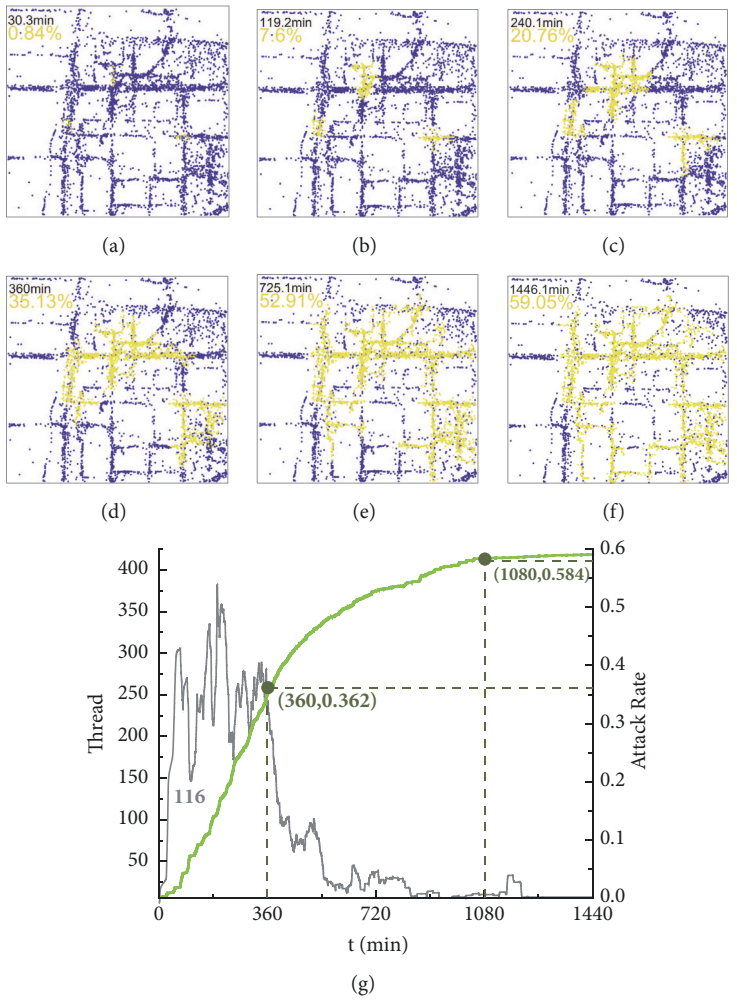

Figure 4: (a–f) **Time evolution of the infection process.** Several time moments are extracted from the whole simulated infection process in a region in Beijing City with the radiating radius set to 45 meters. The infection time and the attacking rate are on the top left of each picture. (g) **Concurrent multiple threads and attack rate changing with time.** Multiple virus attacks are executed simultaneously in the router network and synchronous attack rate is recorded in the same plot.

algorithm and to choose an appropriate initial seed can make the infection efficient. As a user, to have a router with higher encryption level can dramatically reduce the risk. We simulate this process in Figure 5 where P_1 and P_2 increase with a strengthened password. The upper dashed line shows the current situation ($P_1 = 0.5$ and $P_2 = 0.4$). When P_1 and P_2 are gradually increased until 0.85 and 0.8, the attack rate will be reduced to the lower dashed line. From this set of curves, we see a distinct suppression in every increase of the password strength.

5. Conclusion

In conclusion, we collect raw information of router location in a region in Beijing City and analyze the property of Wi-Fi networks, such as the degree distribution and the clustering coefficient, indicating that the real-world Wi-Fi network is vulnerable to the infection. In this paper, the biggest selling point is that we endow the virus with some abilities and present the dedicated worm virus which can pick a more susceptible router region with the KDE algorithm and perform the attacking tasks automatically. This virus can also search the encryption types of routers within its range and attack lower-encryption-level routers first. In this way, the virus gains a rapid expansion in the beginning and 56.0% of routers are infected in only 18 hours. We also present the concept thread to interpret the reason of high infection efficiency of our dedicated virus than a normal one.

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Disclosure

Electronic address is bigdata@wzvtc.edu.cn.

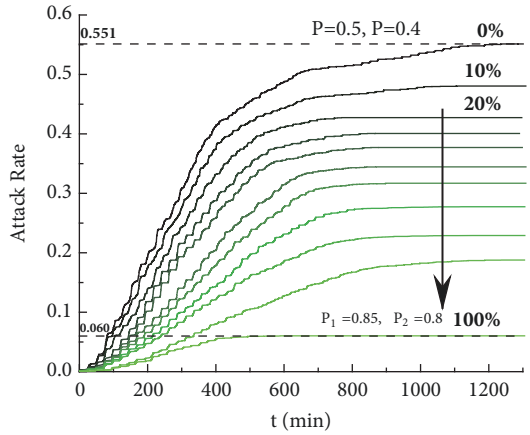

FIGURE 5: **Increasing the password strength suppresses viruses spreading.** In current situation, we set $P_1 = 0.5$ and $P_2 = 0.4$ (upper dashed line). We simulate the evolutionary attack rate with 10 different probabilities $P_1(P_2)$ from 0.5 (0.4) to 0.85 (0.8) in a 10% increase.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Yi-Hong Du conceived the work. Yi-Hong Du collected the raw data and designed the flow diagram. Shi-Hua Liu analyzed the data and performed simulation. Shi-Hua Liu and Yi-Hong Du both wrote the paper.

Acknowledgments

This research leading to the results reported here was supported by the Scientific Research Project of Zhejiang Provincial Education Department (*No.Y*201329845).

References

[1] B. Gu, X. Hong, and P. Wang, "Modeling worm propagation through hidden wireless connections," in *Proceedings of the GLOBECOM - IEEE Global Telecommunications Conference*, 2009.

[2] M. Ovelgonne, T. Dumitras, B. A. Prakash, V. Subrahmanian, and B. Wang, "Understanding the Relationship between Human Behavior and Susceptibility to Cyber-Attacks: A Data-Driven Approach," *ACM Transactions on Interactive Intelligent Systems & Technology*, vol. 8, no. 51, 2016.

[3] W. Liu and S. Zhong, "Web malware spread modelling and optimal control strategies," *Scientific Reports*, vol. 7, pp. 1–19, 2017.

[4] J. Milliken, V. Selis, and A. Marshall, "Detection and analysis of the Chameleon WiFi access point virus," *EURASIP Journal on Information Security*, pp. 1–14, 2013.

[5] P. Akritidis, W. Y. Chin, V. T. Lam, S. Sidiroglou, and K. G. Anagnostakis, "Proximity Breeds Danger: Emerging Threats in Metro-area Wireless Networks," in *Proceedings of the 16th USENIX Security Symposium*, pp. 323–338, 2007.

[6] A. Sanatinia, S. Narain, and G. Noubir, "Wireless spreading of WiFi APs infections using WPS flaws: An epidemiological and experimental study," in *Proceedings of the IEEE Conference on Communications and Network Security, CNS 2013*, pp. 430–437, October 2013.

[7] M. Li, Y. Meng, J. Liu et al., "When CSI meets public WiFi: Inferring your mobile phone password via WiFi signals," in *Proceedings of the CCS '16 (ACM Conference on Computer and Communications Security)*, pp. 1068–1079, October 2016.

[8] M. Vanhoef and F. Piessens, "Key Reinstallation Attacks: Forcing Nonce Reuse in WPA2," in *Proceedings of the CCS '17 (ACM Conference on Computer and Communications Security)*, pp. 1313–1328, Dallas, Texas, USA, October 2017.

[9] P. Wang, M. C. González, C. A. Hidalgo, and A. L. Barabási, "Understanding the spreading patterns of mobile phone viruses," *Science*, vol. 324, no. 5930, pp. 1071–1076, 2009.

[10] H. Hu, S. Myers, V. Colizza, and A. Vespignani, "WiFi networks and malware epidemiology," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 106, no. 5, pp. 1318–1323, 2009.

[11] T. Tala and S. Babak, "A stochastic model for the size of worm origin," *Security and Communication Networks*, vol. 9, pp. 1103–1118, 2015.

[12] W. Wang, Q. Liu, L. Zhong, M. Tang, H. Gao, and H. E. Stanley, "Predicting the epidemic threshold of the susceptible-infected-recovered model," *Scientific Reports*, vol. 6, pp. 1–12, 2016.

[13] H. Kavak, D. Vernon-Bido, J. J. Padilla, S. Y. Diallo, and R. J. Gore, "The spread of wi-fi router malware revisited," *Simulation Series*, vol. 49, pp. 80–89, 2017.

[14] C. Jin and X. Y. Wang, "Analysis and control stratagems of flash disk virus dynamic propagation model," *Security and Communication Networks*, vol. 5, pp. 226–235, 2011.

[15] Bose A. and Shin K. G., "Agent-based modeling of malware dynamics in heterogeneous environments," *Security and Communication Networks*, vol. 6, pp. 1576–1589, 2013.

[16] S. Gil, A. Kott, and A. L. Barabási, "A genetic epidemiology approach to cyber-security," *Scientific Reports*, vol. 4, pp. 1–7, 2014.

[17] P.-Y. Chen, S.-M. Cheng, and K.-G. Chen, "Optimal control of epidemic information dissemination over networks," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2316–2328, 2014.

[18] A. M. del Rey, "Mathematical modeling of the propagation of malware: a review," *Security and Communication Networks*, vol. 8, no. 15, pp. 2561–2579, 2015.

[19] C. Zhang, S. Zhou, J. C. Miller, I. J. Cox, and B. M. Chain, "Optimizing hybrid spreading in metapopulations," *Scientific Reports*, vol. 5, pp. 1–7, 2015.

[20] X. Zhang, B. Ge, Q. Wang, J. Jiang, H. You, and Y. Chen, "Epidemic spreading characteristics and immunity measures based on complex network with contact strength and community structure," *Mathematical Problems in Engineering*, vol. 2015, Article ID 316092, 12 pages, 2015.

[21] Z. Wang, H. Yao, H. Han, J. Du, and C. Ding, "Periodic epidemic spreading over complex systems: modeling and analysis," *Mathematical Problems in Engineering*, vol. 2016, Article ID 8423135, 7 pages, 2016.

[22] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *The Annals of Mathematical Statistics*, vol. 27, pp. 832–837, 1956.

[23] M. Gast, *802.11 Wireless Networks: The Definitive Guide*, O'Reilly Media, 2005, http://books.google.com/books?id=9rHnRzzMHLIC&pgis=1.

[24] J. Balthrop, S. Forrest, M. E. Newman, and M. M. Williamson, "Technological networks and the spread of computer viruses," *Science*, vol. 304, pp. 527–529, 2004.

[25] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.

[26] M. Kitsak, L. K. Gallos, S. Havlin et al., "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.