

Received November 26, 2020, accepted December 8, 2020, date of publication December 14, 2020, date of current version December 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3044277

Stochastic Modeling of IoT Botnet Spread: A Short Survey on Mobile Malware Spread Modeling

ARASH MAHBOUBI¹, SEYIT CAMTEPE², (Senior Member, IEEE), AND KEYVAN ANSARI³

¹School of Computing and Mathematics, Charles Sturt University, Port Macquarie, NSW 2444, Australia

²CSIRO Data61, Marsfield, NSW 2122, Australia

³School of Science, Technology and Engineering, University of the Sunshine Coast, Maroochydore, QLD 4556, Australia

Corresponding author: Arash Mahboubi (amahboubi@csu.edu.au)

ABSTRACT The Internet of Things (IoT) devices are being widely deployed and have been targeted and victimized by malware attacks. The mathematical modelling for an accurate prediction of malicious spreads of botnets across IoT networks is of great importance. Suppose the spread of IoT botnets can be predicted using mathematical models, the security community can then take the necessary steps to deter an outbreak of botnet attacks and minimize the damage caused by malware. This paper surveys mobile malware epidemiological models to understand the mechanisms and dynamics of malware spread for IoT botnets. We describe the characteristics of IoT botnets based on the Susceptible-Infection-Recovery-Susceptible and Susceptible-Exposed-Infection-Recovery-Susceptible epidemic models. These models extend the traditional SIR (Susceptible-Infection-Recovery) model by adding extra states and parameters specific to the epidemic spread of IoT botnets. We use mathematical modelling to simulate complex spreading processes of IoT botnets and interpret the influence of an epidemic on distributed denial of service attacks. We use MATLAB and R to illustrate the use of a stochastic IoT botnet transmission model in the identification and mitigation of challenges towards minimizing the impact of devastating IoT botnet epidemics.

INDEX TERMS IoT malware, botnet, Mirai, propagation modeling, information-theoretic security, malware detection and mitigation

I. INTRODUCTION

Interests of cyber criminals are diversifying to a widespread adaptation of technologies for malicious activities. They are constantly adapting and maliciously evolving their methods of using modern technologies, including the Internet of Things (IoT) and its industrial variant that is referred to as IIoT. IoT devices have evolved with many business sectors due to the convergence of real-time analytics and integrated sensors and data collection systems, which collect and transmit data from on-site devices to the above layers for edge and fog computing. Since most standards governing IoT networks are de facto and applied to many industries today, IoT and IIoT devices are amongst systems targeted by malware developers. IoT sensors are generally connected to other edge devices across heterogeneous networks, including small/macro base stations. As edge computing becomes more widely distrusted, IoT and IIoT devices become more susceptible to malware infections that spread to the rest of network devices.

The associate editor coordinating the review of this manuscript and approving it for publication was Giovanni Pau¹.

The majority of IoT and IIoT devices, due to their limited processing and memory resources, lacks basic security and protection mechanisms [1]. The Mirai botnet malware [2] has proved that IoT devices, including IIoT and Social Internet of Things (SIoT) devices, can be compromised by a basic password attack [3]. Mirai is an IoT botnet attack that targeted French Web-Host and cloud service providers. Mirai manipulates millions of IoT systems to trigger the most complex decentralized Distributed Denial of Service (DDoS) attack known to date, with data traffic spikes of 1.1 terabytes per second [2]. A combination of factors, such as the use of free and powerful Internet-wide scanning tools and the widespread practice of weak default passwords on IoT devices, simplifies the operation of the botnet and causes numerous heterogeneous devices to be infected. Mirai has produced many variants that are still unknown to the security communities [4].

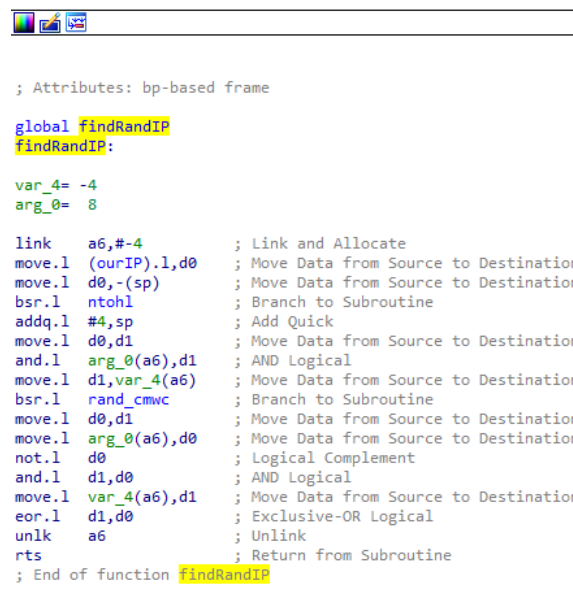
DDoS attacks on IoT networks have reached an unprecedented level [2], [5], and therefore, the demand for detecting IoT botnet attacks in a minimal time has become crucial to minimize the risks associated with such advanced attacks. Advanced hybrid peer-to-peer IoT botnets have been a threat

to all connected mobile devices [6], [7]. The combination of IoT devices and other wireless-enabled devices that use short or long-range wireless technologies such as Bluetooth (BT) or LoRa enables a mixture of heterogeneous devices to communicate. This can result in a four-time faster spread of botnet malware than a classic mobility model for homogeneous systems, such as botnet spreading presented in [8] and [9] that applied the random walking model for the launch of an attack. Immediate detection help to increase network security by triggering alerts and disconnecting infected devices such as internet-accessible video cameras from the network, which may prevent the botnet from spreading, thereby limiting network outbound attacks. But is it possible to immediately detect the spread of botnets and disconnect infected devices?

IoT-based botnets evolve in a series of operating phases for communicating and fetching malicious packages [2]: (1) reconnaissance, (2) report, (3) initial infection, (4) spread, (5) control of command and control (C&C) servers, and (6) execution of attacks. Most previous studies on botnet malware detection applied supervised machine learning to predict malicious traffic [10]. IoT botnets C&C servers, however, use a variety of communications protocols, including HTTP/S, Internet relay chat (IRC), and peer-to-peer protocols, which do not necessarily contain traits of malicious traffic. Also, modern IoT botnets use various resilience techniques such as cryptographic algorithms, obfuscation, mutation, fast-flux technique [11] and algorithms that generate a significant number of domain names, i.e., Domain Generation Algorithm [12]. These techniques can be fetched to actual IoT botnet malware as a payload by a C&C server. The C&C servers can then remotely manage the operations of the botnet malware. To minimize threats of IoT botnet malware, an effective detection method with high precision is required. Such a detection method must also be adaptable for low-cost computational devices and take no time to detect malicious operations. For early detection, though it is vital to investigate and understand how botnet malware is spread.

The aim of this study is to critically review the mathematical models borrowed from biology to illustrate the spread of malware. Biological models have played an important role in identifying strong and weak malware attributes for the elimination of outbreaks. In this study, we also adopt the Susceptible-Infection-Recovery-Susceptible (SIRS) and Susceptible-Exposed-Infection-Recovery-Susceptible (SEIRS) models for determining (alternative) future lines of investigation as a possible countermeasure of initial infection of IoT botnet malware and its spread. In general, we aim to investigate the following research challenges with the spread of IoT botnets: given the presence of an initial botnet in uncertain locations, how do we predict post-initial infection and minimize the severity of the DDoS attack?

Prior research including [13] has applied deterministic models for accurate prediction of malicious propagation over networks. However, in the case of Layer-3 IoT botnets,



```

; Attributes: bp-based frame

global findRandIP
findRandIP:

var_4= -4
arg_0= 8

link a6,#-4 ; Link and Allocate
move.l (ourIP).l,d0 ; Move Data from Source to Destination
move.l d0,-(sp) ; Move Data from Source to Destination
bsr.l ntohl ; Branch to Subroutine
addq.l #4,sp ; Add Quick
move.l d0,d1 ; Move Data from Source to Destination
and.l arg_0(a6),d1 ; AND Logical
move.l d1,var_4(a6) ; Move Data from Source to Destination
bsr.l rand_cmw ; Branch to Subroutine
move.l d0,d1 ; Move Data from Source to Destination
move.l arg_0(a6),d0 ; Move Data from Source to Destination
not.l d0 ; Logical Complement
and.l d1,d0 ; AND Logical
move.l var_4(a6),d1 ; Move Data from Source to Destination
eor.l d1,d0 ; Exclusive-OR Logical
unlk a6 ; Unlink
rts ; Return from Subroutine
; End of function findRandIP

```

FIGURE 1. Mirai looking for Random IP addresses globally.

the initial infection and propagation are not deterministic as IoT botnets aggressively scan for IP addresses to randomly find their victims. We consider the application of the stochastic differential equation (SDE) to find optimal control strategies for future outbreaks of IoT botnet malware. Specifically, we investigate the time required for detecting malicious traffic and disconnecting infected devices from networks (HoneyBot). The remainder of the paper is structured as follows. Section II describes the Mirai botnet and its new variants. Section III presents an analysis of mathematical models and their related literature. Section Section IV formalizes the IoT malware propagation. Section V simulates predictions of the IoT botnet outbreak. Finally, Section VI summarizes and provides concluding remarks and directions for future research.

II. MIRAI BOTNET AND ITS NEW VARIANTS

Mirai is worm-like malware designed to contaminate IoT devices and has begun to infect IoT devices in major attacks since August 2016 [2]. Mirai made international headlines in September 2016 with massive DDoS attacks targeting Krebs-On-Security and OVH [14]. Mirai spreads by first entering into a quick scanning phase where it scans pseudo-random IPv4 addresses on different application protocols such as Telnet. Once Mirai detects a vulnerable (exposed) device, it initiates a brute-force attack through a Telnet connection using ten (10) username and password pairs selected randomly from a pre-configured set of 62 passwords. Mirai is equipped with free Linux utilities as an arsenal of weapons to find random IP addresses (scanner) to plan brute-force attacks by the botnet. Figure 1 illustrates Mirai's function of finding random IP addresses globally. The process of discovering random IP addresses of IoT devices indicates that Mirai behaves stochastically and not deterministically [15]. The goal of the Mirai botnet is to spread an initial

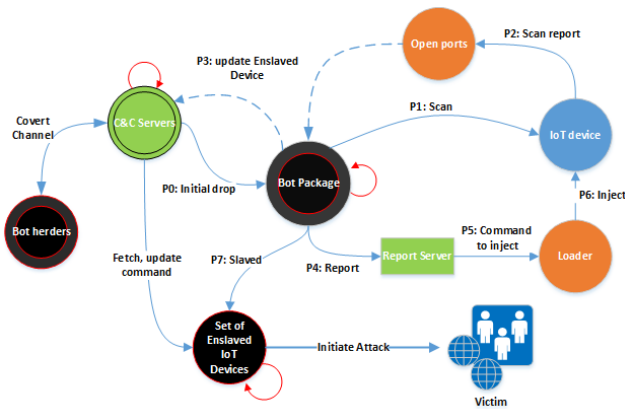


FIGURE 2. Mirai conceptual model in a finite-state machine. P0: Initial infection, P1: Quick scanning, P2: Reporting open ports back to C&C server, P3: Updating the list of enslaved devices, P4: Reporting vulnerable devices, P5: Loading malicious package, P6: Uploading and executing malware package and P7: Collecting slaves.

infection to misconfigured apps and devices and then to attack a target server as soon as it receives a corresponding command from the botnet that is controlled by the author or the bot-master.

After the first successful authentication, Mirai sends the victim's IP address and related password to a defined report server. A different loader software infects the identified exposed IoT devices asynchronously through logging in, evaluating the underlying machine environment, and finally by uploading and executing the malware package [3]. Through an effective infection, Mirai tries to hide its existence by deleting the uploaded malware file and obfuscating its malicious process name in a pseudo-random alphanumeric sequence. The major IoT devices that Mirai has compromised include but are not limited to routers, IP camera/DVR, storage and firewall over the CWMP, Telnet, HTTPS, FTP and SSH protocols. Mirai had infected more than 65,000 IoT devices by the end of its first day of being active [3]. In epidemiology, this means that the transmission coefficient of Mirai infection was 0.75231 devices per second. However, during its peak infection, Mirai has globally spread and compromised more than 600,000 IoT devices. Figure 2 illustrates a conceptual model of Mirai in a finite-state machine.

Once Mirai's source code was released to the research community, it quickly led to the development of some effective detection and defense mechanisms. However, within only two months of the release of the source code, the number of botnet instances increased and a wide range of Mirai variants emerged [2]. As of the time of writing of this article, Mirai continues to exploit new vulnerabilities in configurations for the same or cross-platform types of IoT devices. A new Mirai variant called Mukashi has used new vulnerabilities of CVE-2020-9054 and has infected Zyxel's network-connected storage (NAS) devices [16]. This remote code execution vulnerability of CVE-2020-9054 received a critical rating of 9.8 out of 10, which has since been patched. It is not surprising that the malicious actors were taking advantage of this vulnerability to wreak havoc on the IoT domain. It was

originally discovered after the selling of its exploit code as a 0-day attack.

III. A SHORT REVIEW ON MALWARE SPREAD AND MATHEMATICAL MODELS

The aim of mathematical modeling is to transfer problems that arise within a real world domain to mathematical languages so that the conceptual and numerical analyses can be carried out on the issues. In reality, mathematical modeling is the best solution when trying to gain knowledge from experiments that are either very expensive to perform or that would require an excessive amount of time. The main objective of using mathematical models of the spread of malware is to provide a critical analysis to identify weaknesses and strengths of the malware and, from there, to identify feasible future defense lines. The majority of proposed mathematical models are based on differential equations, which explain the behavior of malware spread [17]. Analytical epidemic models have been considered by many technology researchers to investigate and describe traditional malware propagation behaviors [13], [18], [19]. In that respect, traditional epidemic models applied to cyber security problems include SI (Susceptible-Infection), SIS (Susceptible-Infection-Susceptible), and SIR (Susceptible-Infection-Recovery).

In short, the SI model assumes that mobile devices are susceptible to malware infections after successful communications with an initially infected mobile device, and that the newly infected device can not be immunized. The SIS model assumes that a susceptible mobile device would change its condition after an initial contact with an infected device, which makes it becomes infected. The newly infected device could not develop an infection immunization and would become susceptible to new infections. As per the SIR model, a susceptible device becomes infected from previously infected devices and then the system develops immunity and recovers from the infection, i.e., this state is called R. In fact, IoT devices are less likely to gain immunity and are continuously vulnerable to new infections, and are threatened by either a zero-day botnet or new versions of a current botnet. Think of this as the flu season during which everyone is susceptible to the flu virus, irrespective of having a flu vaccine. The stochastic epidemic model differs according to the basic concepts of the time unit and the status of random variables. For illustration, the model in the stochastic differential equation (SDE) focuses on a diffusion method where both state and time variables are assumed to be constant. The rest of this section provides the details of several models employed to predict the spread of mobile malware.

A. BLUETOOTH WORM YAN ANALYTICAL MODEL

A theoretical model that characterizes the transmission dynamics of Bluetooth (BT) worms has been thoroughly studied by Yan and Eidenbenz [9], [20]. In the proposed system, the influence of mobility patterns on the spread of BT worms can be reviewed by adopting input parameters, such as the average node degree, the average node rate and

the distribution of the link duration. The authors claimed that their simulation results show that their model precisely predicts the complex distribution of BT worms. The formula is used to accurately estimate the distribution curve of BT worms in large cities like Los Angeles. For this model, the average number of infected mobile devices is denoted by $i(t)$ in the period (t) , the user population density is denoted by $p(t)$ at time t and the infection rate is denoted by $i'(t_k)$, where the $t_k (k \geq 0)$. The next mobile infections because of the BT worm spread is denoted by $t_k + 1$ and $i(t_k + 1)$. Hence, the slope of a curve of the spread of the BT worm is calculated as follows:

$$i(t_k + 1) = \frac{i(t_k)p(t_k)}{i'(t_k) + (p(t_k) - i'(t_k))e^{-\psi}} \quad (1)$$

$$a(t) = \Omega(t, 1, T_{ing}(t), < 0, 0, 0, 1, 0 >), \quad (2)$$

where $\psi = -\alpha(t_k) \cdot p(t_k) / (p(t_k) - i'(t_k))$, $p(t_k)$ indicates the average density of infected devices at time t_k . This model is limited to the large number of population and it is unlikely to present heterogeneous IoT devices.

B. RHODES AND NEKOVEE WIRELESS WORM MODEL

The impact of demographic factors and node behavior on dynamic outbreaks of BT worms, which use the *SIP* mathematical model, has been considered in [21]. Considering the *SIP* model with a population of N , where ρ is the number of active individuals residing in different areas and v is the mean of communications speeds, it is said that a BT worm can spread with a probability of p to other susceptible nodes within the BT range (R) if there exists a single infected node in each area. Therefore, when a BT worm is spread to a network, each node can be in any of the following states: (S :) susceptible, (I :) infected, or (P :) recovered. Therefore, in a fixed size population, the spread of BT worms is defined as the following equations:

$$SIP = \begin{cases} \frac{ds}{dt} = -2 R_{pvp} \frac{SI}{N} \\ \frac{dI}{dt} = 2 R_{pvp} \frac{SI}{N} - \delta I \\ \frac{dP}{dt} = \delta I \end{cases} \quad (3)$$

C. MARTIN SIS MODEL

Martin *et al.* [22] presented a *SIS* epidemic model that measured the propagation of a mobile worm. In this model, let I be the number of mobile devices infected by other infected devices in its vicinity. As a result of the infection, a newly infected device is moved from the susceptibility state to the infection state. Let S be a set of unprotected mobile devices within a total population of N devices where N equals to $S + I$. The transition parameters between the states are α and β , where α reflects the rate at which infected devices are recovered and returned to the state of susceptibility over a period of t , and the β parameter represents the transmission of infections among susceptible and infected mobile devices based on discrete contacts. Therefore, a model of *SIS* is given

in Equation 4.

$$SIS = \begin{cases} \frac{dS(t)}{dt} = \frac{\beta S(t)I(t)}{N} \\ \frac{dI(t)}{d(t)} = \frac{\alpha S(t)I(t)}{N} \end{cases} \quad (4)$$

D. THE PROBABILISTIC QUEUING MODEL (MICKENS MODEL)

A deterministic queuing system to illustrate the spread of smart-device worms through a short-range wireless network is presented in [23]. The authors argued that classical epidemiological models fail to show the unique characteristics of mobile networks. Node mobility presents non-homogeneous communications distributions that cannot be depicted using simple mobile environments and networks, because previous models ignore node velocity and the non-homogeneous connectivity. Therefore, they introduced a new infectious mobile worm spread framework using a probabilistic queue prediction model (for abstract data type) to of infected nodes. If $P(k)$ shows the distribution of cellular connectivity, the queuing system is computed by placing $N_k = P(k)N$ networks within each Q_k queuing system. Therefore, the general homogeneous characteristics of each probabilistic queue infection are modeled as follows.

$$\frac{dI_k}{dt} = \beta K_i I_k (1 - I_k) - \delta I_k \quad (5)$$

The individual node infection in the population is derived by $\sum_{k=0}^{N-1} [I_k N_k]$. The researchers demonstrated the influence of node speeds on the constant-state level of network infections and described a conceptual stochastic equivalent to the deterministic model.

E. TWO-LAYER PREDICTION MODEL OF SMARTPHONE MALWARE PROPAGATION BY HUMAN BEHAVIORS

A two-layer geographical model to illustrate the spread of BT/SMS (Short Message Service) based mobile worms within a limited geographical area, consisting of cellular towers and a practical communications channels consisting of mobile devices is studied in [24], [25]. The inner layer of the model is defined as a mobile network allowing the recognition of geographical distributions of nodes. The BT worms can propagate through this layer using location data of mobile nodes. The outer layer of the model is a logical network that relays each phone's contact list. In this model, $G[N][N]$ represents a two-dimensional network with respect to the geographical network, therefore a network node T_i is represented using a tuple list $\langle r, p(x, y), n_{tp}, T_{link} \rangle$ in which r is defined as the coverage area of the network node, $p(x, y)$ is an infection vector drives form T_i , n_{tp} is defined as the total number of mobile devices connecting to T_i , and T_{link} is a set of connections T_i establishes within its vicinity.

In the logical network, an individual v_i device is defined as $\langle T_{id}, l(x, y), on(1) - off/0, t_{on}, p_{click}, P_{link} \rangle$ in which T_{id} is the *ID* of the cellular tower that extends a

cellular/wireless communications service of the provider for v_i , $l(x, y)$ is the location of v_i , on or off is a binary function used to validate whether or not v_i is available, t_{on} logs the time when v_i is available, p_{click} is the likelihood of a user clicking and opening an email, and P_{link} is the contact list of v_i . In this model, the impacts of human behavior and their interactions are measured for propagation of SMS and BT-based mobile worms using the *SIR* epidemic model. However, this model is not truly random and cannot handle the large number of nodes and spreading IoT botnet malware.

F. SI ANALYTICAL MODEL FOR HYBRID MOBILE MALWARE

An SI mathematical model to examine the propagation velocity and severity of hybrid mobile worms targeting wireless communications channels, such as Commwarrior worm, was introduced in [26]. In this research, authors considered Multimedia Messaging Service (MMS) communications for spreading a mobile worm from Node A to Node B as an initial infection at the time t . The BT communications may then be used to propagate the worm to Node C who is in the vicinity of Node B. The differential equation in 6 represents an MMS/BT epidemic.

$$\frac{dI_{MMS}(t)}{dt} = \beta_{MMS} \frac{S(t)(\eta_{MMS} - 1)}{N} I(t) \tag{6}$$

The authors assumed that an individual infection period may start at time r by being in the locality of an origin infection by MMS and it continues to infect other Node toward S time units. However, The incremental spatial node infection at time t of all infection periods presents in the following equation:

$$\frac{dI_{BT}(t)}{dt} = \int_t^0 I'_{MMS}(\tau) G'(\tau, t - \tau) d\tau \tag{7}$$

The SI analytical model is based on Node vicinity and it assumed that malware spread through MMS and BT, this infection method is not compatible for the large population specially in case of IoT botnet.

G. RAMACHANDRAN AND SIKDAR'S SEIR MODEL

Ramachandran and Sikdar [27] studied the SEIR analytical model to investigate the influence of different spread methods, like downloading a virus package from either the Internet or peer-to-peer (P2P) network channels, and transferring via BT, Wi-Fi, infrared, MMS, or SMS. The proposed model consists of four equations which have been used to characterize each location where the individual with smartphone may visit. The location are denoted by P (patches), let m_{pq} denotes the transfer frequency from the state patch q to patch p . Where the $S_p, E_p, I_p,$ and R_p denote the rate of location changing from susceptible to exposed, exposed to infected or to the recovered state of the populations from the patch state $p(1 \leq p \leq P)$. This SEIR model is described in Equation 10. The new malware are mutating and they can be susceptible to new

variant of the same malware family.

$$\left\{ \begin{aligned} \frac{dS_p}{dt} &= d_p(N_p - S_p) - P_{on}^p \gamma_p(t) S_p - P_{on}^p \beta_p S_p \frac{I_p}{N_p} \\ &\quad - \sum_{i=1}^P \alpha(1-p) S_p \frac{I_t}{N_t} + \sum_{q=1}^P m_{pq} S_q - \sum_{q=1}^P m_{pq} S_p \\ \frac{dE_p}{dt} &= \sum_{i=1}^P \alpha(1-p) S_p \frac{I_t}{N_t} + \sum_{q=1}^P m_{pq} S_q - m_{pq} E_q \\ \frac{dI_p}{dt} &= P_{on}^p \gamma_p(t) S_p + P_{on}^p \beta_p S_p \frac{I_p}{N_p} - (d_p + \delta_p) I_p \\ &\quad + \varepsilon_p E_p + \sum_{q=1}^P m_{pq} I_q - \sum_{q=1}^P m_{pq} I_p \\ \frac{dR_p}{dt} &= \delta_p I_p - d_p R_p + \sum_{q=1}^P m_{pq} R_q - \sum_{q=1}^P m_{pq} R_p \end{aligned} \right. \tag{8}$$

where $N_p = S_p + E_p + I_p + R_p; S_p, E_p, I_p, R_p \geq 0$ at $t = 0$

H. XIA'S SEIRD MODEL

Xia et al. [28] studied a mathematical model for MMS and BT hybrid malware spread, i.e., Commwarrior. They argued that mobile devices can fall into one of five states of Susceptible (S), Exposed (E), Infected (I), Recovered (R), and Dormancy (D). The authors defined 12 transitions between states: $S \implies I, I \implies D, D \implies I$ and $E \implies I(\beta)$ in the case of BT spread; $S \implies E, E \implies S, E \implies R, R \implies E$ and $E \implies I_\mu$ in the case of MMS/SMS propagation, and $S \implies R, I \implies S$ and $I \implies R$ for both SMS/MMS and BT spread modes. The following equation describes the SEIRD model.

$$\left\{ \begin{aligned} \frac{dS(t)}{dt} &= P_{ES} E(t) + P_{IS} I(t) - \beta k S(t) I(t) - \lambda(t) S(t) \\ &\quad - P_{SR} S(t) \\ \frac{dE(t)}{dt} &= \lambda(t) S(t) - \beta k E(t) I(t) - (\mu + P_{ES} + P_{ER}) E(t) \\ \frac{dI(t)}{dt} &= \beta K(S(t) + E(t)) I(t) + \mu E(t) + \theta D(t) \\ &\quad - (\gamma + P_{IS} + \varepsilon) I(t) \\ \frac{dR(t)}{dt} &= \gamma I(t) + P_{ER} E(t) + P_{SR} S(t) \\ \frac{dD(t)}{dt} &= \varepsilon I(t) - \theta D(t) \\ N &= S(t) + E(t) + I(t) + R(t) + D(t) \\ \kappa &= \sigma(\sqrt{4r_2 - (\Delta t)^2 v^2} + \pi r^2 - \frac{\pi}{4} (\Delta t)^2 v^2) - 1 \\ \lambda(t) &= \omega \eta \frac{I(t)}{N} \frac{S(t)}{S(t) + I(t)} \end{aligned} \right. \tag{9}$$

where β is the rate of malware infection; κ is the average of the number of mobile nodes; η is the likelihood that an infected node propagates a wireless worm to its contact list; γ is the likelihood that an infected mobile phone receives security support to eliminate the infection; μ is the likelihood that an exposed mobile phone will transition to the

infected state; ε is the likelihood that an infected phone moves to the dormancy state when the battery is exhausted by the BT module for example, and θ is the possibility that a dormant mobile node becomes infected after being recharged.

I. FAN'S SEIR MODEL

A SEIR mathematical model, i.e., Susceptible (S), Exposed (E), Infected (I) and Recovered (R), for hybrid SMS/MMS/BT malware propagation was studied in [29]. Their framework primarily focuses on a preventive immunization mechanism and the mutation of mobile malware. Also, the authors discussed the impact of transmission parameters, such as preventive immunity for users of smart devices, virus replication, immune composition of MMS/SMS communications network, and average rate of spreading virus across BT network.

Smart devices can be in four states, which enables eight categories of state transitions: $S \implies E$, $E \implies I_{\beta_2}$, and $E \implies R$ are the transitions for the SMS/MMS spread mode, $S \implies I$ and $E \implies I_{(\beta_1)}$ for the BT spread mode, and $S \implies R$, R implies S , and $I \implies R$ for hybrid malware that use both BT and SMS/MMS communications modes. Accordingly, the SEIR model is described in Equation 10.

$$\left\{ \begin{array}{l} \frac{dS(t)}{dt} = -\beta_1 \kappa S(t) I(t) - P_{SE} S(t) - \mu_1 S(t) + P_{RS} R(t) \\ \frac{dE(t)}{dt} = P_{SE} S(t) - \beta_1 \kappa E(t) I(t) - (\mu_2 + \beta_2) E(t) \\ \frac{dI(t)}{dt} = \beta_1 \kappa (S(t) + E(t)) I(t) + \beta_2 E(t) - \delta I(t) \\ \frac{dR(t)}{dt} = \mu_1 S(t) + \mu_2 E(t) + \delta I(t) - P_{RS} R(t) \\ N = \rho \pi \gamma^2 (1 - \alpha) + \rho [3 \Delta_{tv} \sqrt{r^2 - \frac{1}{4} (\Delta t)^2 v^2} \\ + 2 r^2 \arccos(\frac{\frac{1}{2} \Delta_{tv}}{r})] \alpha - 1 \\ P_{SE} = \gamma(t) = \omega \eta \frac{I(t)}{N} \frac{S(t)}{S(t) + I(t)} \\ P_{RS} = f(t - t_0) \varepsilon \end{array} \right. \quad (10)$$

where k is the average of the number of nodes; β_1 is the frequency at which vulnerable or unprotected smartphones are compromised by BT malware; η is the likelihood that a contaminated mobile device can transmit malware to its contact list; β_2 is the frequency at which exposed smartphones become infected by SMS / MMS; δ is the likelihood that infected devices will be protected from viruses using security software applications and patches; μ_1 is the probability at which susceptible smart devices would achieve pre-immunity utilizing security techniques such as upgrading their anti-virus signature database or patching server; μ_2 is the possibility that infected devices can achieve pre-immunity by security software, such as upgrading their virus registry and patching.

J. TWO-DIMENSIONAL (2D) CELLULAR AUTOMATA MODEL

A two-dimensional cellular automata describing the dynamics of worm spreading from a single node to an entire network was analyzed in [30], [31]. The model incorporates an infection parameter that calculates the degree of the spread amongst infected nodes, and an immunity parameter that provides an approximation of the protection for susceptible nodes. Let N_u represents the number of vicinity for the u node. Let $\Phi_{C_{ij}}$ and C_{kl} denote the coefficients of interactions between cellular C_{ij} and its vicinity, that described the probability of infections between cells vicinity. Let δ be an infection factor where calculate by the ratio of communication factors between cellular C_{ij} and its vicinity to the resistance factor (i.e., communication detected as malicious), Whereas $\Phi_{C_{ij}}$, C_{kl} and δ are listed as continues to follow:

$$\Phi_{C_{ij}}, C_{kl} = \sum_{v=N_u}^{v=1} \frac{IF_{vu}}{\sqrt{(i-k)^2 + (i-l)^2}} \quad (11)$$

$$\delta = \frac{\Phi_{C_{ij}}, C_{kl}}{RF} \quad (12)$$

where IF_{vu} is the infection parameter that represents the rate of infections from the node v to the node u ($0 \leq IF \leq 1$); and RF is the defense factor that represents the level of immunity of a node to an infection from another node ($0 \leq RF \leq 1$).

K. WANG'S SI MODEL

In addition to the mobile malware propagation hybrid model, i.e., Bluetooth and MMS communications, Wang et al. [32] described the SI epidemic model for mobile malware propagation and its initial infection. Let I represents infected mobile users and S susceptible mobile users where (t) is the total number of mobile users infected over the entire time. The SI model is defined as follows:

$$\frac{dI}{dt} = \frac{\beta SI}{N} \quad (13)$$

where $\beta = \langle k \rangle$ is an effective rate (i.e., successfully contact and infect mobile devices) of infecting mobile devices of N population; $m = 1$ is the number of individuals situated in the vicinity of the mobile tower; $\langle k \rangle = RA = NA/Atower$ is the total number of contact mobile devices. $A = pr^2$ denotes the BT mobile contact in the area and the population density in a transmitter's coverage area is denoted by $r = N/Atower$.

When an infected mobile user travels to a new mobile tower zone, this appears as a source of BT infection at the new location. The researchers suggested more functional spread models should be considered to evaluate MMS and mobile viruses by analyzing and predicting the communications connectivity channel in the real-world example.

L. PENG'S DTMC STOCHASTIC MODEL

A discrete stochastic process model for analyzing the propagation of SMS/MMS based mobile malware was

presented in [33], which employed the semi-Markov processes model and the social relationship graphs. The researchers identified a state transition framework to understand the nature and complexity of the spread of the mobile malware, where the semi-Markov method and an active empiric derivation demonstrates how to integrate human interactions and the semi-Markov process, i.e., stochastic process, by using a conceptual probability limitation analysis. The researchers also implemented an updated program to receive messages from existing mobile networks. They obtained a large range of real-world data-set of mobile communications of the main mobile service providers in Mainland China. The data-set was created from 20,000,000 SMS/MMS messages obtained from 400,000 mobile users in three weeks.

M. LIU'S WSIS MODEL

The research presented in [34] demonstrated that the propagation of mobile malware is heterogeneous rather than homogeneous, and therefore should not be seen as an individual susceptible state of infections as per the past models. The authors claimed that seeing mobile malware spreading as a simplistic homogeneous pattern could not explain the effect of different levels of anti-virus protection on a computer system. Based on this fact, the authors introduced a theoretical *WSIS* model to illustrate malware spreads throughout a mobile population. In the *WSIS* model, additional states including strongly-protected and weakly-protected are defined. It is assumed that mobile devices with up-to-date security software packages are in a strongly-protected state while devices with no security protection are assumed to be in a weakly-protected state.

$$WSIS = \begin{cases} \frac{dW(t)}{dt} = -\varepsilon W(t) + \alpha S(t) - \beta_w W(t)i(t) \\ \frac{dS(t)}{dt} = \varepsilon W(t) - \alpha S(t) - \beta_s S(t)i(t) + \gamma i(t) \\ \frac{di(t)}{dt} = \beta_w W(t)i(t) + \beta_s S(t)i(t) - \gamma i(t) \end{cases} \quad (14)$$

N. LIU'S WSI MODEL

A theoretical and numerical simulation of networks with high degree of heterogeneity leading to the propagation of mobile malware and diverse networks with lower power exponents was studied in [34], [35]. The authors introduced a *WSI* epidemic model 15 for mobile malware propagation, in which *W* indicates poorly protected susceptible nodes, *S* indicates highly protected susceptible nodes, and *I* indicates infected nodes.

In this model, the details of smartphone malware infections have been neglected and the total size of the networks are considered to be fixed. The authors assumed for propagating a malware in complex networks where the nodes in the network are considered to be asymptotically, malware can propagates from one node to another node based on power law distribution. Figure 3 depicts the conceptual transition diagram of the

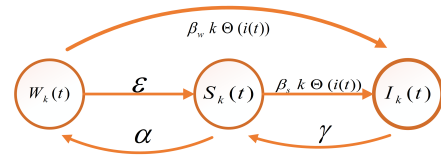


FIGURE 3. WSIS transition model.

WSI model in a complex-network.

$$\begin{cases} \frac{dW_k(t)}{dt} = \varepsilon W_k(t) + \alpha S_k(t) - \beta_w k W_k(t) \Theta(i(t)) \\ \frac{dS_k(t)}{dt} = \varepsilon W_k(t) - \alpha S_k(t) - \beta_s k S_k(t) \Theta(i(t)) - \gamma I_k(t) \\ \frac{dI_k(t)}{dt} = \beta_w k W_k(t) \Theta(i(t)) + \beta_s k S_k(t) \Theta(i(t)) - \gamma I_k(t) \end{cases} \quad (15)$$

whereby the preliminary states are $W_k(0), S_k(0), I_k(0) \geq 0, k = 1, 2, \dots, \Delta$.

The SEIRS model has been studied for mobile malware spread in [36]. The author presented a time-space framework to model the spread of mobile malware based on local Boolean rules that fix the evolution of nodes placed in a square grid. Relationships between entities are defined by different types of relations, such as the Moore or Von Neumann models. The author claimed that their method was effective in achieving the same behavior patterns as the ordinary differential equation (ODE) simulation framework.

Traditionally, the Android App market is one source of mobile malware spread. Meng et al. [39] studied malware spread between Android markets. This study borrowed the *SI* biological model to illustrate malware spread between android markets such as GooglePlay, QQ, ANZHI, GetJar, Xiaomi, Mumayi, and APPCHINA with a focus on Chinese app markets. They have used the Gillespie algorithm for predicting infection orders. The Gillespie algorithm indicates the next infection occurrence, and the period until another event occurs dynamically, based on a sequence of inclination values $e_1 \dots e_n$. In each step, the algorithm generates two-interval random numbers $r_1, r_2 \in [0, 1]$. r_1 is used to evaluate the next α infected market by testing markets depending on their inclination, where $\sum_{x=1}^{\alpha-1} e_x < r_1 \sum_{x=1}^{\alpha} e_x < \sum_{x=1}^{\alpha} e_x$. r_2 is used to predict a time (t) for the next infection, with ratio of $\sum e_x$ which e_x is the likelihood of the event e .

IV. FORMALIZING THE PROPAGATION OF IOT BOTNET MALWARE

A summary of the previous research papers considered in the previous section is given in Table 1 with their important infection parameters being highlighting. Previous studies of malware epidemic have rarely dealt with the spread of botnet amongst IoT devices using stochastic models. One of the commonly used models in previous studies is the random walk model. Generally, random walk models can be classified into restricted and unrestricted models. In a restricted random walk model at least there is one boundary so that either

TABLE 1. The infection, immunity, and population parameters used to simulate epidemiological models in different research.

| Research | Epedemic Model | Contact rate | Immunity rate | Population | Comments |
|------------------------|---|--|--|------------------------------------|---|
| [23] (2005) | Probabilistic queuing | 0.5 & 0.7 | 0.25 | 60 | |
| [27] (2007) | SEIR | 0.1 | 0.05 | 1000 | |
| [20] [9] (2007 , 2009) | Random walk model | N/A | N/A | 800 | the random variables $(X_t) = (X_1, X_2, \dots)$, the infection defined by X_t for all times $t \in [0, +\infty)$. |
| [21] (2008) | SIR | 0.2 | 0.1 | 3000 | |
| [32] (2009) | SI | Infection rate is $\beta = \mu < k >$ with $\mu = 1$ and the average number of contacts is $<k>$ | N/A | NA | immunity rate will not considered in this model. $\frac{N}{A_{tower}}$ is the total population within the coverage area of the tower. |
| [29] (2010) | SEIR | $\beta_1 = 0.5$ & $\beta_2 = 0.007$ | 0.001 | 10^6 | $S \Rightarrow I, E \Rightarrow I(\beta_1)$ are BT spread mode; $S \Rightarrow E, E \Rightarrow I(\beta_2), E \Rightarrow R$ are SMS/MMS spread mode; $S \Rightarrow R, R \Rightarrow S, I \Rightarrow R$ are the combination of the both BT and SMS/MMS communications |
| [24] (2011) | SIR | Inverse of time* $<$ average number of phones that can contact with each others $>$, | depends on user's own security awareness | 8000 | |
| [26] (2011) | SIS | 0.05 | NA | interval [0, 0.1, 0.2, ... 0.9, 1] | |
| [30] (2011) | SEIDR | 0.2 & 0.4 | 0.1 & 0.2 | 2000 | Probability with which a node in state D becomes a node in state R |
| [25] (2013) | SIR | Inverse of time* $<$ average number of phones that can contact with each others $>$ | depends on individual security awareness | 8000 - 10000 | |
| [33] (2014) | SEIR | 0.54 | 0.06 | 4500 | P_{SE} represents the probability of a node transform from state S to state E, and P_{EI} represents the probability of a node transform from state E to state I and p_{IR} probability of a node transform from state I to state R |
| [34] (2016) | WSIS, WSI, WIW | open interval [0.2,1) | NA | interval (0,1) | three models, W = Weakly-protected, S = Strongly-protected susceptible, I = Infected node |
| [13] (2017) | SIRS | 0.76 | 0.05 | 4000 | |
| [36] (2018) | SEIRS | 0.9 | 0.6 | 100, 2500, 10000 | |
| [37] (2020) | SNIRD | 0.01 | 0.1 | 1500 | Spreading malware in heterogeneous WSNs network |
| [38] (2020) | vulnerable-compromised-quarantined-patched-scrapped (VCQPS) model | N/A | N/A | 100 | |
| This paper | SEIRS | 0.419 | 0.7 | 10^5 | Stochastic model where the N is derived from $S + E + I + R$ and β is a worst case scenario of Mirai infected IoT devices, i.e., 41.93% in 20 hours. |

the state space of the chain is finite, $\{0, 1, 2, \dots, N\}$ with a boundary of 0 and N , or semi-finite, $\{0, 1, 2, \dots\}$ with a boundary of 0. In an unrestricted random walk, the model has no boundaries, but random walking models use all the nodes fairly ($1/|N|$). Thus, in the case of the IoT bot malware, the spread is random, and that's because of the botnet scanning mechanism. In the literature, most studies used deterministic models to present malware propagation, while in this type of models randomness is not factored in for the development of the future states of the system under study. Stochastic models are on the other hand considered promising to randomly determine the IoT botnet infection system. The blow subsection considers a stochastic model based on which we will present the results of a series of simulation studies in the subsequent section.

A. A STOCHASTIC SIRS MODEL

To consider a stochastic SIRS model, we assume the growth of IoT deployments continues and will never end and hereafter new IoT malware, developed or mutated, would continue to infect new deployments or current operational devices within the IoT population. Amongst IoT devices, low-power devices may be particularly compromised by IoT malware since bot-based malware would heavily and quickly drain their batteries. Considering biological models, such devices can be considered dead.

Let $\{M (time) \mid time \in [1, \infty)\}$ be a set of discrete random numbers that belong to the natural numbers set $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ and has the Markov property where M denotes bot-based malware. It is assumed that the time between infection events is exponentially distributed. Let $p_{xi}(\delta time)$ denote

the probability of M being transited from Node x to Node i during the time period of $\delta time$: $p_{xi}(\delta time) = \mathbb{P} = (M(time + \delta time) = j \mid M(time))$. The following equation models the transmission probability of a new infection:

$$\sum_{i=0}^{\infty} p_{xi}(M(time))s^i = \left(\sum_{i=0}^{\infty} p_1 i(M(time))s^i \right)^x, \quad s \in [0, 1]. \tag{16}$$

In general, in the case of Mirai outbreak, the number of infected nodes represents the state of the branching process and the concept of birth implies latent infections that have occurred. Also, in Equation 16, the death is not considered simply because of the nature of IoT botnet. The assumption is that although the botnet malware (M) infects new IoT devices, the infection remains inactive and waits for a major attack. Therefore, Equation 16 implies that the propagation of a botnet is additive and that botnets collect zombies in planning for major attacks. Detecting botnets is hard due to malware mutation and evolution and also zero-day vulnerabilities.

Introducing infectious entities into susceptible entity pools can lead to an epidemic, contributing to an increase in the number of infected entities. Whether an epidemic occurs depends on the rate of infections, regeneration, mortality, and the extent of vulnerable population. In a basic SIR epidemic model, the population is split into susceptible (S), infected (I), and recovered (R) entities. Let the variables β and γ imply the rate of transmission and recovery, respectively. In the case of a serious condition, an increase in the incidence of infection results in a mortality rate of α . In fact, if a recovered entity

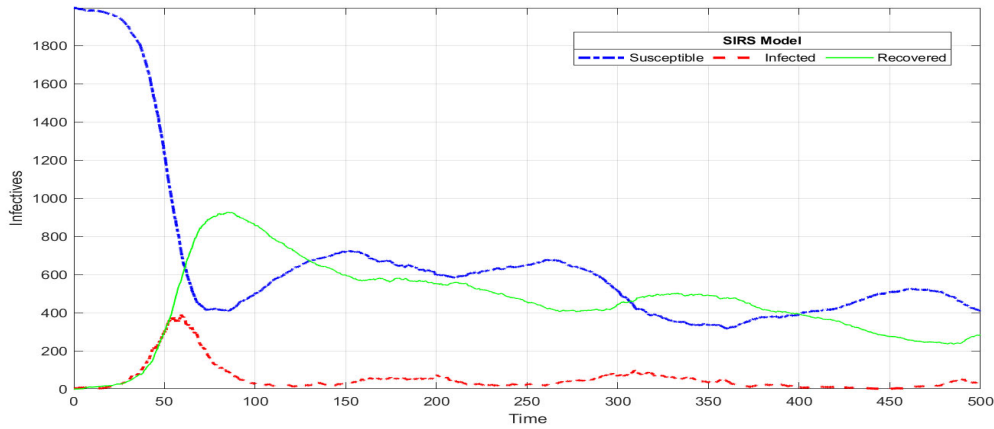


FIGURE 4. IoT devices losing immunity and becoming susceptible again to new variant of the botnet. Considering the SIRS model with the parameters $\gamma = 0.3$, $\alpha = 0.02$, $\delta = 0.05$, $\beta = 0.75$, with an initial infection of 2 individuals in a population of 2000 devices.

has developed only temporary immunity to re-infection, with a declining immunity rate of δ , the recovered entity returns to being susceptible. As the botnet updates frequently to avoid detection and remain obfuscated, the SIRS epidemic model shows botnets propagate stochastically. In the case of the SIRS epidemic model, the ratio R_0 known as the basic reproduction number is an important parameter in the epidemic theory. Let all the three random parameters for all the states (S, I, R) in the continuous-time Markov epidemic model be denoted as $(S_1, S_2, S_3) = S$ with $N = \sum_{i=1}^3 S_i$ directly correlated to the random variable for the overall size of the population. Therefore, the SIRS model 17 is described in the following equation to illustrate the botnet propagation:

$$\begin{cases} \frac{dS(t)}{dt} = -\beta \frac{S(t)}{N(t)} S(t) I(t) + \delta R(t) \\ \frac{dI(t)}{dt} = \beta \frac{S(t)}{N(t)} S(t) I(t) - \gamma I(t) - \alpha I(t) \\ \frac{dR(t)}{dt} = \gamma I(t) - \delta R(t), \end{cases} \quad (17)$$

where the S and $I > 0$, $R = 0$, and $S + I + R = N$. Figure 4 illustrates a botnet outbreak using the SIRS stochastic model with a population of 2000 devices.

V. SIMULATION AND PREDICTION OF IOT BOTNET OUTBREAK WITH THE SEIRS STOCHASTIC MODEL

IoT botnets can be presented using the SEIRS epidemic model with additional compartments. In the case of IoT botnet, there is a significant incubation period between scanning of IoT devices, i.e., the scanning phase, and the C&C server report that was not infected with the initial malware. This is called the exposed state or latency period. The SEIRS model consists of four distinct states: susceptible (S), exposed (E), infected (I), and recovered (R) with six parameters: β , σ , γ , Ω , μ and α where β is the contact rate and controls how fast botnet spreads from being susceptible to exposed, σ gives the transition from being exposed to infected, γ gives the transition from infection to recovery, and Ω is the rate of recovered IoT devices becoming susceptible again due to new botnet variants or new malware. The SEIRS compartments

TABLE 2. Summary of parameters and compartments used in the SEIRS system.

| Parameters | Definition |
|---------------------------|---|
| β | Contact rate |
| σ | Incubation period |
| γ | Recovery period |
| Ω | Immunity duration |
| μ | Mean of expectancy of device availability |
| α | Infection induced death rate |
| p | Fraction of population received security patches |
| $1/\mu + \gamma + \alpha$ | Mean of infection period |
| S | Susceptibility State |
| E | Exposure State (Device infected but stays stealth for receiving commands from C&C server) |
| I | Infection State |
| R | Recovery State |

and parameters are summarized in Table 2.

$$\begin{cases} \frac{dS}{dt} = \underbrace{(1-p)\mu N}_a - \underbrace{\beta SI/N}_b + \underbrace{\Omega R}_c - \underbrace{\mu S}_d \\ \frac{dE}{dt} = \underbrace{\beta S(\text{router} + \text{camera} + \text{others})I}_a - \underbrace{\sigma E}_b - \underbrace{\mu E}_c \\ \frac{dI}{dt} = \underbrace{\sigma E}_a - \underbrace{\gamma I}_b - \underbrace{(\mu + \alpha)I}_c \\ \frac{dR}{dt} = \underbrace{p\mu N}_a + \underbrace{\gamma I}_b - \underbrace{\Omega R}_c - \underbrace{\mu R}_d \end{cases} \quad (18)$$

We classify these parameters into three categories, botnet-centric, attack-centric, and immunity-centric. β belongs to the botnet-centric category, σ and α parameters are in the attack-centric subset, and γ , Ω , μ and P belong to the immunity-centric subset. The severity of the contact rate and incubation rate depends on how botnet is designed including code complexity, scanning, and vulnerabilities preloaded in the botnet or possibly receive payloads from the C&C servers. In the case of the Mirai outbreak, the contact rate included 65,000 IoT devices infected within the first 20 hours of its activation, with Brazil recorded 41.93%, Iran 10.17% and China 5.14% of infected IoT devices [3], [40]. Here in this

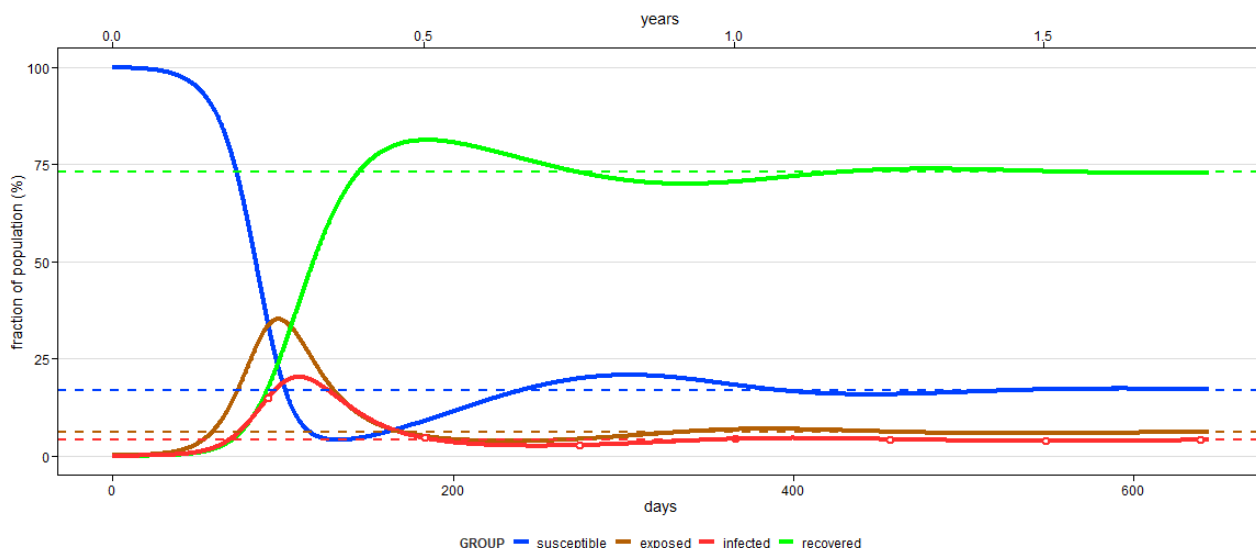


FIGURE 5. A simulation of the SEIRS system with $R_0 = 5.9$, $\beta = 0.419$, $\gamma = 0.071$, $\sigma = 0.048$, $\Omega = 0.7$, $\mu = 0.01$ and $\alpha = 0.00$ including no security patches on the IoT devices. This graph shows the botnet’s incubation period while collecting zombies. We assumed that the botnet uses zero-day vulnerabilities to collect zombies for its attack, and there is no security patches available at the time of infection.

TABLE 3. Components of the SEIRS differential equation.

| Intended notions | | |
|------------------|---|--|
| S | a | Number of IoT devices without security patches |
| | b | Number of individual IoT Infections |
| | c | Number of IoT devices losing immunity due to existing or new vulnerabilities |
| | d | Number of dead IoT devices due to power failure |
| E | a | Number of infected IoT devices |
| | b | Number of devices infected during incubation period, devices can stay stealth in periods of time |
| | c | Number of dead IoT devices due to power failure or update security patches |
| I | a | Number of exposed devices (receiving new malicious packages from C&C servers) |
| | b | Number of recovered devices after death or having security patches applied) |
| | c | Number of dead IoT devices due to power failure or security patch applied |
| R | a | Number of IoT devices with security patches applied |
| | b | Number of IoT devices recovered from infection |
| | c | Number of recovered IoT devices losing immunity again due to new vulnerabilities |
| | d | Number of dead IoT devices due to power failure or disconnected from network or had security patches applied |

study, we use the worst case scenario of Mirai infected IoT devices, i.e., 41.93%, for the β infection rate parameter for our SEIRS model. The incubation state can be potentially used to hold a large pool of zombies for larger attacks. The SEIRS governing differential equation is described in 18 and each of its functions are explained in Table 3. We assume the infection rate β is fixed where the attack- and immunity-centric parameters may vary.

$$R_0 = \frac{\sigma}{\omega + \mu} + \frac{\beta}{\gamma + \mu + \alpha} \quad (19)$$

The SEIRS models generally consider homogeneous IoT devices, in which the infection is likely to equally initiate communications with individual IoT devices and therefore with each susceptible device at a β rate. This represents the

spread of the botnet malware in IoT networks where the infectious servers try to randomly infect and evenly generate addresses for next infections. The β and σ depend on the rate at which the botnet server scans (see Figure 1) for susceptible nodes, their distribution ranges and densities, and up and down links as well as communications rates [41]. Therefore, to reach a saddle point of IoT botnet infection and reduce the speed of infections, it is crucial to control β and σ . We assume all heterogeneous IoT devices are losing immunity and they will be susceptible for new infection with zero-day payload, considering infinite cycle.

Our SEIRS model is based on a susceptible state that includes homogeneous and heterogeneous IoT devices. However, susceptible IoT devices may transit to the exposed state if one of their individual connections gets infected by bot-malware. In the case of an initial botnet infection, the botnet author(s) has control over master servers where they are responsible for exposing vulnerable devices. In this state, the infected devices (Zombies) are under the control of the masters and are inactive for a period of time until they are sold in which case they would receive additional payloads from C&C servers for the next stage of an DDoS attack. In the SEIRS system, collecting zombies can be represented by R_0 , which plays a critical role in collecting zombies. Once an attack begins, the first response from the security point of view would be to block the network domain, change network addresses including URLs, or simply kill IoT devices who may also die from the loss of power. In the SEIRS system, the α parameter identifies the death rate for IoT devices. After an immunity period represented by Ω , IoT devices may be losing their immunity p due to new vulnerabilities or mutations. We simulated the SEIRS model using the R language [42] to show IoT botnet can initiate infections following its release into the Internet.

Figures 5 and 6 illustrate that the botnet’s initial infections evolved over time and developed new transitions from the

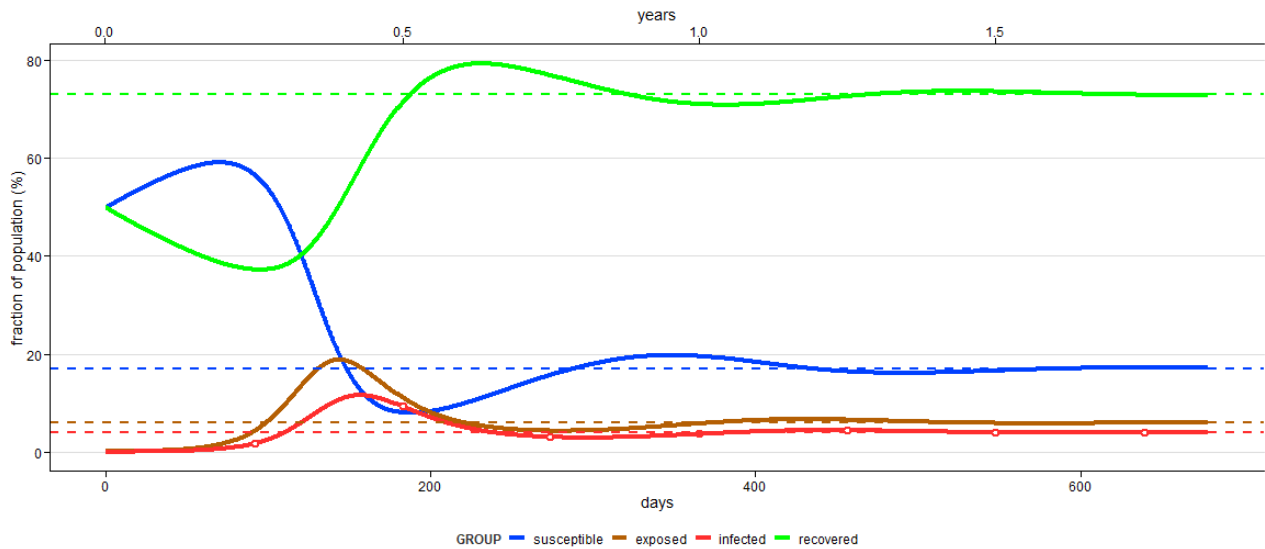


FIGURE 6. A simulation of the SEIRS system with $R_0 = 5.9$, $\beta = 0.419$, $\gamma = 0.071$, $\sigma = 0.048$, $\Omega = 0.7$, $\mu = 0.01$ and $\alpha = 0.00$ including the rate of 0.5 for the security patches on IoT devices. This graph shows the botnet's incubation period while collecting zombies for a major attack. We assumed half of the IoT devices updated with the security patches that takes the edge off collecting zombies.

susceptibility to incubation state. The simulation predictions illustrate that the botnet collected IoT devices for a major attack. The incubation period depends on attack aims of the botnet or the sell of zombies on the darknet. The incubation is a continues process that is shown as $E(\infty)$. Figure 5 shows the infection rate, i.e., $I(\infty)$, never flattens and devices are continuously infected or victim servers are attacked. Hence, as the infection continues to contaminate susceptible IoT devices, an equilibrium point $S(\infty)$ ¹ is never reached although security patches $R(\infty)$ continue to be developed. This is due to the continuous development of new botnet variants that take advantage of unknown vulnerabilities.

VI. CONCLUSION

Many cyber criminals are financially motivated to develop new forms of malware. Botnet-based malware has seen a tremendous growth as a result of the widespread adoption of IoT devices. The Mirai botnet malware, mainly consisting of embedded systems and IoT devices, has invaded network services and overwhelmed several high-profile organizations with massive DDoS attacks. We therefore provided a short survey of mobile malware spread models in this study. SIRS and SEIRS mathematical models were used to illustrate how homogeneous and heterogeneous devices can infect and spread the infection in a large or small scale over comprehensive networks. We formulated the IoT botnet in a stochastic model that has the Markov property, and from there, determined that an initial botnet infection always exists. We have also developed a stochastic SIRS model in Matlab to prove that IoT devices are always susceptible to infection $I(\infty)$ even after acquiring immunity. Based on the malware analysis of the Mirai botnet, we formulated the botnet in a

¹Malwarefree equilibrium point refers to an equilibrium with all infection states being zero.

SEIRS epidemic model and simulated the model in the R language. The SEIRS stochastic model is a suitable model to illustrate the botnet-based malware, since there is the incubation state for collecting zombies in preparation sfor major attacks. In general, bot-based malware has a substantial incubation period, i.e., for collecting zombies and waiting for the master server to send attack commands, during which IoT devices may have been compromised but have not yet been triggered to attack. Finally, based on our observations from the SEIRS results, it can be concluded that all S , E , I , and R states are infinite in the SEIRS system and the curves would never reach a saddle point at any point in time, even after we assumed that 50% of IoT devices were updated with a security patch. Controlling R_0 is crucial to manage the infection and probably using network telescope can be a major defense mechanism against the incubation state in a DDoS attack while botnets send random probes to scan for security holes and harvesting zombies.

REFERENCES

- [1] J. Su, V. Danilo Vasconcellos, S. Prasad, S. Daniele, Y. Feng, and K. Sakurai, "Lightweight classification of IoT malware based on image recognition," in *Proc. IEEE 42nd Annu. Comput. Softw. Appl. Conf. (COMPSAC)*, vol. 2, Jul. 2018, pp. 664–669.
- [2] C. Koliadis, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and other botnets," *Computer*, vol. 50, no. 7, pp. 80–84, 2017.
- [3] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, N. Sullivan, K. Thomas, and Y. Zhou, "Understanding the Mirai Botnet," in *Proc. 26th USENIX Secur. Symp. USENIX Secur.*, Vancouver, BC, Canada: USENIX Association, Aug. 2017, pp. 1093–1110. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/antonakakis>
- [4] M. De Donno, N. Dragoni, A. Giaretta, and A. Spognardi, "DDoS-capable IoT malwares: Comparative analysis and mirai investigation," *Secur. Commun. Netw.*, vol. 2018, pp. 1–30, Jun. 2018.
- [5] E. Bertino and N. Islam, "Botnets and Internet of Things security," *Computer*, vol. 50, no. 2, pp. 76–79, Feb. 2017.

- [6] P. Wang, S. Sparks, and C. C. Zou, "An advanced hybrid peer-to-peer botnet," *IEEE Trans. Dependable Secure Comput.*, vol. 7, no. 2, pp. 113–127, Apr. 2010.
- [7] K. Singh, S. Sangal, N. Jain, P. Traynor, and W. Lee, "Evaluating Bluetooth as a medium for botnet command and control," in *Detection of Intrusions and Malware, and Vulnerability Assessment*, C. Kreibich and M. Jahnke, Eds. Berlin, Germany: Springer, 2010, pp. 61–80.
- [8] G. Yan, H. D. Flores, L. Cuellar, N. Hengartner, S. Eidenbenz, and V. Vu, "Bluetooth worm propagation: Mobility pattern matters!" in *Proc. 2nd ACM Symp. Inf. Comput. Commun. Secur. (ASIACCS)*, New York, NY, USA: Association Computing Machinery, 2007, pp. 32–44, doi: 10.1145/1229285.1229294.
- [9] G. Yan and S. Eidenbenz, "Modeling propagation dynamics of Bluetooth worms (extended version)," *IEEE Trans. Mobile Comput.*, vol. 8, no. 3, pp. 353–368, Mar. 2009.
- [10] M. Stevanovic and J. M. Pedersen, "An efficient flow-based botnet detection using supervised machine learning," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2014, pp. 797–801.
- [11] S. Abraham and I. Chengalur-Smith, "An overview of social engineering malware: Trends, tactics, and implications," *Technol. Soc.*, vol. 32, no. 3, pp. 183–196, Aug. 2010.
- [12] J. Woodbridge, H. S. Anderson, A. Ahuja, and D. Grant, "Predicting domain generation algorithms with long short-term memory networks," 2016, *arXiv:1611.00791*. [Online]. Available: <https://arxiv.org/abs/1611.00791>
- [13] A. Mahboubi, S. Camtepe, and H. Morarji, "A study on formal methods to generalize heterogeneous mobile malware propagation and their impacts," *IEEE Access*, vol. 5, pp. 27740–27756, 2017.
- [14] E. D. Jovanovic and P. V. Vuletic, "Analysis and characterization of IoT malware command and control communication," in *Proc. 27th Telecommun. Forum (TELFOR)*, Nov. 2019, pp. 1–4.
- [15] M. S. Pour, S. Torabi, E. Bou-Harb, C. Assi, and M. Debbabi, "Stochastic modeling, analysis and investigation of IoT-generated Internet scanning activities," *IEEE Netw. Lett.*, vol. 2, no. 3, pp. 159–163, Sep. 2020.
- [16] D. Kumar, K. Shen, B. Case, D. Garg, G. Alperovich, D. Kuznetsov, R. Gupta, and Z. Durumeric, "All things considered: An analysis of IoT devices on home networks," in *Proc. 28th USENIX Secur. Symp.* Santa Clara, CA, USA: USENIX Association, Aug. 2019, pp. 1169–1185. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity19/presentation/kumar-deepak>
- [17] A. M. D. Rey, "Mathematical modeling of the propagation of malware: A review," *Secur. Commun. Netw.*, vol. 8, no. 15, pp. 2561–2579, Jan. 2015.
- [18] W. Liu and S. Zhong, "Web malware spread modelling and optimal control strategies," *Sci. Rep.*, vol. 7, no. 1, p. 42308, Feb. 2017.
- [19] A. Mahboubi, S. Camtepe, and H. Morarji, "Reducing USB attack surface: A lightweight authentication and delegation protocol," in *Proc. Int. Conf. Smart Comput. Electron. Enterprise (ICSCSEE)*, Jul. 2018, pp. 1–7.
- [20] G. Yan and S. Eidenbenz, "Modeling propagation dynamics of Bluetooth worms," in *Proc. 27th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, 2007, p. 42.
- [21] C. J. Rhodes and M. Nekovee, "The opportunistic transmission of wireless worms between mobile devices," *Phys. A, Stat. Mech. Appl.*, vol. 387, no. 27, pp. 6837–6844, Dec. 2008. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378437108007772>
- [22] J. C. Martin, L. L. Burge, J. I. Gill, A. N. Washington, and M. Alfred, "Modelling the spread of mobile malware," *Int. J. Comput. Aided Eng. Technol.*, vol. 2, no. 1, pp. 3–14, 2010.
- [23] J. W. Mickens and B. D. Noble, "Modeling epidemic spreading in mobile environments," in *Proc. 4th ACM Workshop Wireless Secur.* New York, NY, USA: ACM, 2005, pp. 77–86.
- [24] C. Gao and J. Liu, "Modeling and predicting the dynamics of mobile virus spread affected by human behavior," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw.*, Jun. 2011, pp. 1–9.
- [25] C. Gao and J. Liu, "Modeling and restraining mobile virus propagation," *IEEE Trans. Mobile Comput.*, vol. 12, no. 3, pp. 529–541, Mar. 2013.
- [26] 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.
- [27] K. Ramachandran and B. Sikdar, "Modeling malware propagation in networks of smart cell phones with spatial dynamics," in *Proc. IEEE INFOCOM-26th IEEE Int. Conf. Comput. Commun.*, May 2007, pp. 2516–2520.
- [28] W. Xia, Z.-H. Li, Z.-Q. Chen, and Z.-Z. Yuan, "Commwarrior worm propagation model for smart phone networks," *J. China Universities Posts Telecommun.*, vol. 15, no. 2, pp. 60–66, Jun. 2008. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1005888508600850>
- [29] Y. Fan, K. Zheng, and Y. Yang, "Epidemic model of mobile phone virus for hybrid spread mode with preventive immunity and mutation," in *Proc. 6th Int. Conf. Wireless Commun. Netw. Mobile Comput. (WiCOM)*, Sep. 2010, pp. 1–5.
- [30] S. Peng and G. Wang, "Worm propagation modeling using 2D cellular automata in Bluetooth networks," in *Proc. IEEE 10th Int. Conf. Trust, Secur. Privacy Comput. Commun.*, Nov. 2011, pp. 282–287.
- [31] S. Peng, G. Wang, and S. Yu, "Modeling the dynamics of worm propagation using two-dimensional cellular automata in smartphones," *J. Comput. Syst. Sci.*, vol. 79, no. 5, pp. 586–595, Aug. 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022000012001754>
- [32] 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. 1071–1076, Apr. 2009.
- [33] S. Peng, M. Wu, G. Wang, and S. Yu, "Propagation model of smartphone worms based on semi-Markov process and social relationship graph," *Comput. Secur.*, vol. 44, pp. 92–103, Jul. 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167404814000583>
- [34] W. Liu, C. Liu, X. Liu, S. Cui, and X. Huang, "Modeling the spread of malware with the influence of heterogeneous immunization," *Appl. Math. Model.*, vol. 40, no. 4, pp. 3141–3152, Feb. 2016.
- [35] W. Liu, C. Liu, Z. Yang, X. Liu, Y. Zhang, and Z. Wei, "Modeling the propagation of mobile malware on complex networks," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 37, pp. 249–264, Aug. 2016.
- [36] M. T. Signes-Pont, A. Cortés-Castillo, H. Mora-Mora, and J. Szymanski, "Modelling the malware propagation in mobile computer devices," *Comput. Secur.*, vol. 79, pp. 80–93, Nov. 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167404818302773>
- [37] S. Shen, H. Zhou, S. Feng, J. Liu, and Q. Cao, "SNIRD: Disclosing rules of malware spread in heterogeneous wireless sensor networks," *IEEE Access*, vol. 7, pp. 92881–92892, 2019.
- [38] S. Shen, H. Zhou, S. Feng, J. Liu, H. Zhang, and Q. Cao, "An epidemiology-based model for disclosing dynamics of malware propagation in heterogeneous and mobile WSNs," *IEEE Access*, vol. 8, pp. 43876–43887, 2020.
- [39] G. Meng, M. Patrick, Y. Xue, Y. Liu, and J. Zhang, "Securing Android app markets via modeling and predicting malware spread between markets," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 7, pp. 1944–1959, Jul. 2019.
- [40] M. S. Pour, A. Mangino, K. Friday, M. Rathbun, E. Bou-Harb, F. Iqbal, S. Samtani, J. Crichigno, and N. Ghani, "On data-driven curation, learning, and analysis for inferring evolving Internet-of-Things (IoT) botnets in the wild," *Comput. Secur.*, vol. 91, Apr. 2020, Art. no. 101707.
- [41] M. H. R. Khouzani, S. Sarkar, and E. Altman, "Saddle-point strategies in malware attack," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 31–43, Jan. 2012.
- [42] O. N. Bjørnstad, K. Shea, M. Krzywinski, and N. Altman, "The SEIRS model for infectious disease dynamics," *Nature Methods*, vol. 17, no. 6, pp. 557–558, Jun. 2020.



ARASH MAHBOUBI received the B.E. degree (Hons.) in computer science specializing in computer security from Staffordshire University, Kuala Lumpur, Malaysia, in 2012, the master's degree in information security from University Technology Malaysia, Johor Bahru, Malaysia, in 2013, and the Ph.D. degree in computer science from the Queensland University of Technology (QUT), Brisbane, QLD, Australia, in 2018. From 2016 to 2019, he was a Sessional Academic with the School of Electrical Engineering and Computer Science, QUT. Since 2019, he has been a Lecturer with the School of Computing and Mathematics, Charles Sturt University, Port Macquarie, NSW, Australia. His research interests include computer/mobile malware, ransomware, malware analysis, modeling, and malware epidemic.



SEYIT CAMTEPE (Senior Member, IEEE) received the Ph.D. degree in computer science from the Rensselaer Polytechnic Institute, New York, NY, USA, in 2007. From 2007 to 2013, he was with Technische Universitaet Berlin, Germany, as a Senior Researcher, and a Research Group Leader in security. From 2013 to 2017, he was a Lecturer with the Queensland University of Technology, Brisbane, QLD, Australia. He is currently a Principal Research Scientist and a Team Leader with the CSIRO Data61. His research interests include autonomous and application security covering topics in cyber security and machine learning and applied and malicious cryptography. He has authored more than 90 papers which have received 4460 citations (H-index 26) so far.



KEYVAN ANSARI received the B.E. degree in computer engineering from the Sadjad University of Technology, Mashhad, Iran, in 2006, the master's degree (Hons.) in information technology from the University of Newcastle, Callaghan, NSW, Australia, in 2009, and the Ph.D. degree in computer science from the Queensland University of Technology (QUT), Brisbane, QLD, Australia, in 2014. From 2014 to 2016, he was an Associate Lecturer with the School of Electrical Engineering and Computer Science, QUT. In 2016, he joined the University of the Sunshine Coast, where he currently undertakes research and lectures in Computer Science and ICT with the School of Science, Technology and Engineering. His research interests include pervasive communications, integrated and distributed IoT systems, and cloud computing. He has co-chaired and has been a member of technical program committees of several international conferences and workshops and has been an Editorial Board Member and a Reviewer of several international journals and magazines.

...