An Epidemic Based Model for the Predictions of OOFI in an IoT Platform

Mohammed Ibrahim, Mohd Taufik Abdullah, Azizol Abdullah, Thinagaran Perumal

Faculty of Computer Science & Information Technology, Universiti Putra Malaysia, Selangor, Malaysia

ABSTRACT

Based on the notion that when a particular node is maliciously infected, there is probability of propagating such infections to other susceptible nodes in a network. This lead to the development of malware spreading models to predict the transmission rate, transmission parameters and the number of infected nodes per unit time. However, the emergence of Internet of Things (IoT) with strong base in both wired and wireless sensor network(WSN), predicting the spreading of malware infections is not the only source of concern for forensic analysis. Considering the heterogeneity and data volatility of IoT nodes, predicting the object of forensic interest (OOFI) in resource-constraint devices like sensor nodes as well as the diffusion of data among the neighboring nodes remain a critical issue for forensic analysis. From the concept of epidemic theory, a novel model is proposed called Susceptible-Infective-Recovered with Forensic (SIR-F) that can predict and isolate OOFI among various nodes in IoT network. The essence of introducing forensic mechanism is to ascertain the OOFI by predicting the responsible nodes holding the data of forensic interest. As such, SIR-F can timely enhance the process of identifying OOFI of the collection phase of digital forensic standard operating procedure (SOP).

Keywords: *Forensic, Internet of Things, Malware, Sensor.*

I. INTRODUCTION

Internet of Things (IoT) as widely known combined the functions of both wired and wireless network technology that utilized sensor capability in exchange of data. Sensor nodes like other nodes on a network are prone to malware attacks, however, based on the notion that when a particular node is maliciously infected, there is a probability of propagating such infections to the neighborhood susceptible nodes. In this regard, malware propagation and malware spreading models are required to predict the transmission rate, transmission parameters and the number of infected sensor nodes per unit time in a given wireless sensor network (WSN). WSN as a fundamental network background of IoT was designed to transmit their observational values to processing/control center as well as a sink node that works as a user interface [1]. Nevertheless, as a result of narrow transmission range, sensor data that are generated distant away from the sink node must be pass along with the intermediate nodes [1]. This shows that source node transmits own data to their neighboring nodes, the neighboring nodes also transmit the data to their various neighboring nodes [1].

In addition, sensor nodes are resource-constrained, they are associated with low-power and limited-memory capability. As such, ascertaining the actual sensor nodes holding the data of forensic interest at the point of investigation remain critical. Therefore, embedding sensor nodes into IoT technology to capture and transmit data autonomously can further complicate the process of outsourcing and predicting object of forensic interest(OOFI). Moreover, in IoT, unlike in other network platforms,OOFI may not always be available or accessible at the point of investigation[2].

Similarly, evidence volatility in IoT remain an issue, data can be locally stored by a thing(device or node) but can later be overwritten/compressed using a lossy techniques [3]. Also, data from a thing can be transferred and consumed by another thing or a local ad-hoc network of things[3]. In some instances, it is likely to acquire the necessary data from the connected devices than from the primary embedded device [4]. Consequently, predicting the device(s) or node(s) holding the data of forensic interest is paramount for forensic analysis. Hence, we argued that besides developing model for malware propagation in sensor based IoT network, there is need for model of predicting OOFI which will likewise turn to predict devices or nodes holding the data of forensic interest.

In recent time, various epidemic models [5], [6]were adopted and modify to predict the spreads of malware attacks in wireless sensor network. However, despite the weakness of the models in handling the spreading of malware attacks in IoT network, the epidemic models does not consider the prediction of OOFI in WSN for the

purpose of investigation.

However, this paper consider an attack on both multihop and wireless sensor network with heterogeneous IoT nodes. Using contact tracing strategy, the paper bring about a novel Susceptible-Infective-Recovered with Forensic (SIR-F) to discover the dynamic relationship and data diffusion among heterogeneous IoT nodes to determine the object of forensic interest. The malware attack start by infecting a particular node in a wireless based IoT network, which will spreads the malware along with forensic data to the neighboring nodes. Upon receiving any malicious packet from the infected node, the susceptible neighboring node can get infected. However, due to the resource-constrained of some of the susceptible node, associating or interacting with other unconstrained resource nodes can consume the data of forensic interest. By introducing the forensic mechanism into the wireless based IoT network, our SIR-F can not only isolate infected nodes but generate OOFI at the point of investigation. We obtain analytical solution and outlined some of the practical application of the model. The proposed SIR-F can complement the capability of some of the existing digital forensic tool in enhancing IoT forensic processes.

II. EPIDEMIC MODEL AND IOT NETWORK

In this section, we describe epidemic model based on the concept of contact tracing strategy and the malware propagation as well as the dynamic nature of data dispersion among various devices in IoT network.

A. Epidemic Model Based On Contact Tracing Strategy

As often exploit in social or medical sciences, epidemic model is employed to analyse the outcomes of infection in a given population that contains susceptible factors with regard to the infection [7]. This lead to the development of mathematical modeling of epidemiological phenomenon to analyse the infectious disease data. The general and the most common epidemiological model is commonly referred to as SIR(Susceptible-Infection-Recovered) model. In SIR model, the susceptible individual found to be infective, holds for a given time, recovers, and then turn to be insusceptible to the other infections [1].

Despite its widely application in epidemic model, SIR was extended to take into account the effect of contact tracing control measures on the spread of epidemic [8]. Contact tracing strategy is an active detection mechanism that consist in asking infected individual to name persons with whom they have possible infectious contact, on the basis of information provided, the contacted individual can be subjected to medical examinations or cure at the event of infections[8]. All persons that have sufficient exposure contact with the infected persons will be compile on contact list for further examination [9].

B. IoT Network Model

We study a WSN that contained N stationery heterogeneous IoT nodes uniformly distributed with a node density of Ø over a particular space. Each IoT node is assumed to have a communication range to the neighboring node. Also, generated data from a source node can be communicated to its neighboring nodes using a transmission signal within the communication range. Upon receiving the data, neighboring nodes transmit such data to the other nodes continuously over the IoT network geographical space.

Similarly, malware attack against single node can result to transmission of malware packets alongside with the typical data from the compromised node to the susceptible neighbor nodes through a broadcast protocol [10]. Consequently, the entire IoT nodes are at risk of getting compromised. Compromised nodes usually broadcasts data to their neighbor nodes, as such cumulative dataset may

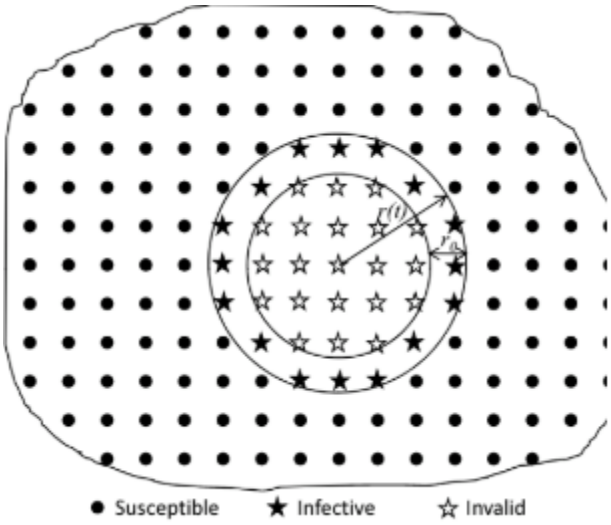

Fig 1. A Model of Virus Spread in a WSN [1]

exist in multiple location [4]. This will require forensic investigators to determine not only the compromised nodes, but the IoT nodes holding the data of forensic interest.

On the other hand, susceptible resource-constrain nodes are busy interacting with the powerful counterpart nodes by means of certain operations such as "push", "publish/subscribe" and "subscribe/notify" [11]. These operations are associating constrain IoT nodes to the

unconstrained once by means of message or data exchanges. In such operations, unconstrained IoT nodes keep information of their respective clients.

As illustrated in Fig 1. [1], all the wireless nodes are considered susceptible nodes until a particular node is infected by malware attack. The infected node is then transmits malicious packet along with the forensic data to the susceptible neighboring nodes over a communication range. The neighboring susceptible nodes will turn to be infective within the communication range. However, infected nodes that are resources constrain can not keep their data due to limited memory or interaction with powerful nodes that consume their data at a particular time. Therefore, the propose SIR-F can enhance the process of identifying OOFI by predicting the infected nodes holding the data of forensic interest.

Next, is to study malware propagation and data spreading in the IoT network by taking into consideration the nature of the IoT nodes and operations associating data movement with respect to IoT nodes over time

III. SIR-F MODELING

Let consider four basic classes of sensor nodes $S(t), I(t), R(t)$ and $F(t)$ stands for susceptible, infection, recovered and OOFI nodes at a given time (t). The SIR-F model can be develop in both WSN and multi-hop IoT network.

A. Modeling in WSN

Given that the sensor nodes are uniformly distributed with density of Ø. Each infected node can contact its neighbor node with the order $\phi\pi r^2$ [1]. At a given time (t), only a fraction of the infected node's neighbor can get infected, this can be mathematically expressed as $\frac{S(t)}{N}$ where N is the total number of nodes in the network. Contrary to the work of Tang and Mark [1] that put into consideration the active and sleep modes in developing SIR-M,in our propose SIR-F, characteristics of infected nodes(constrain or unconstrained) is to be considered. This is due to the interest of the proposed model in predicting nodes holding the forensic data.

As commonly symbolized in the literature, β is the probabilistic rate of transmitting an infection to the susceptible nodes. As such β is largely depends on the infectivity of the virus and the rate of transmission of a protocol since the malware propagate itself by modifying on normal data through regular communication range [1]. Therefore, the modified data on transmission and its destination nodes along the communication range is needed for forensic analysis. Along the communication range, susceptible nodes can turn to be infectious nodes, also infectious nodes I can either belongs to the class of powerful computing resources nodes q_u or low computing resources nodes q_l .

As early stated, susceptible nodes can become infectious through transmission via encounter with infectious node(using malware transmission parameter β). Then, mathematically, the rate of change of susceptible nodes S per unit time can be express as

$$dS(t)/dt = -\beta\pi r^2 \frac{S(t)}{N} I(t) \quad (1)$$

Infectious IoT nodes can either be resource-constraint q_l or resource unconstrained q_u . resource unconstrained infectious nodes usually holds the data of forensic interest. Therefore, resource unconstrained infectious IoT nodes can become object of forensic interest. Also, infectious IoT nodes with strong antivirus can be recovered and become immune using recovery parameter γ .

$$dI(t)/dt = \beta\pi r^2 \frac{S(t)}{N} I(t) - (q_u + \gamma) I(t) + \alpha F(t) \quad (2)$$

Similarly, recovered resource unconstrained IoT nodes can holds data of forensic interest. As such, recovered resource unconstrained IoT nodes can become the object of forensic interest.

$$dR(t)/dt = \gamma I(t) - q_u R(t) \quad (3)$$

Also, data from a thing can be transferred and consumed by another thing or a local ad-hoc network of things [3]. As such, nodes in forensic class that loss their data can return to their infectious without any forensic value. In this case, let α be the rate at which data can be consume from certain number of nodes.

Therefore, the object of forensic interest can be determined using the expression below:

$$dF(t)/dt = q_u (R(t) + I(t)) - \alpha F(t) \quad (4)$$

Thus the initial condition is given as

$$S(0) = N - 1, I(0) = 1, R(0) = 0, \quad \text{and} \quad F(0) = 0 \quad (5)$$

As indicated in Fig.1 [1], the malware spreads to a radius of r(t) at a time (t), and the nodes within the inner circle contribute no further to the spreading of the malware. Therefore, the number of susceptible and infective nodes can be respectively express as

$$S(t) = N - \phi\pi r(t)^2$$

$$I(t) = \phi\pi r(t)^2 - \phi\pi [r(t) - r_0(t)]^2$$

B. Modeling in Multihop IoT Network

[12] analyse a multihop IoT Mesh network that composed of sink and multiple nodes. By considering the computing resources of the nodes, there are found that some nodes are more powerful than the others. As such, the powerful nodes can be used to identify malicious nodes [12]. however, we argued that besides identifying malicious nodes, powerful nodes can be used to predict OOFI. Unlike in WSN, in multihop IoT network, there would be a multiple paths from a source node S node to a destination node D, each path may contain various number of relay nodes. The relay nodes can contain both resource constrain and non-constrained nodes [12]. Similar to WSN, we consider here that each node has a communication range of radius r with sufficient network connectivity/diversity to transmit data. While transmitting data all over the connected paths, each nodes receive packets from its preceding nodes within r along the path and forwards to all its succeeding nodes within r [12].

In this regard, at the event of an attack, compromised IoT nodes send modified packet along the connected paths, the relay nodes along these paths can receive the modified packets. On receiving the modified packets, the relay nodes can be infected and turn to be compromised IoT nodes. Therefore, the probability of transmitting an infection to the susceptible relay nodes along the connected paths is given as

β = no. of modified packets/total no. Of packets transmitted.

Then for malware spreading , there should be a fractions of susceptible nodes along a specific path that can have contact with the modified packets from a compromised node I. Compromised node I can be associated with multiple paths, then the number of susceptible nodes that can get infected along multiple paths to a destination node is given as

$$dS(p)/dp = -\beta\eta\phi A\frac{S(p)}{N}I(t) \qquad (6)$$

Where η = is a threshold that ensures every node is connected to more than one path in the network.

A = Area covered by the distributed IoT nodes.

ϕ = Node density per unit area

Contrary to WSN in which source node spreads information to all its neighboring nodes base on the coverage of communication antenna, in multihop IoT network, source node send a packet to the next succeeding node along the connected paths. Therefore, spreading a malware attack can also be determine by the nature of the succeeding node. If the next succeeding node is equip with antivirus γ, on getting infected by the modified packet, the succeeding node will recover and remove from the infectious class. Also, multihop IoT network consists of both powerful and resource constrain nodes, as such fraction of the powerful nodes ω associated with infectious nodes can consume data from the resource constrain nodes. Consequently, the fraction of the nodes consuming the forensic data can be move to the forensic class leaving the remaining infective nodes to the infectious group , mathematically, the rate of change of infectious node along the connected path can be express as

$$dI(p)/dp = \beta\eta\phi A\frac{S(p)}{N}I(p) - (\gamma+\omega)I(p) + \alpha F(p) \qquad (7)$$

Similarly, recovered infectious IoT nodes associated with powerful IoT nodes can also holds data of forensic interest. As such, their can be remove and send to the forensic class. Therefore, recovered IoT nodes along the connected path can be expressed mathematically as

$$dR(p)/dp = \gamma I(p) - \omega R(p)$$

(8)

For object of forensic interest, both the infectious and recovered IoT nodes with all the capability of holding data of forensic interest can be considered. However, data from a thing can be transferred and consumed by another thing or a local ad-hoc network of things [3]. As such, those nodes in forensic class that loss their data can return to the infectious class without any forensic value. In this case, let α be the rate at which data can be consume from certain number of nodes.

Therefore, the object of forensic interest can be determined using the expression below:

mathematically express as

$$dF(p)/dp = \omega(I(p) + R(p)) - \alpha F(p)$$

(9)

And the initial condition is given as

$$S(0) = N - I, I(0) = 1, R(0) = 0, F(0) = 0,$$

Considering the mesh topology diagram[12], based on that diagram, the malware spreads along the connected path(p), the nodes within that area can be affected.

$$S(p) = N - \phi A$$

(10)

By substituting equation (10) into equation (6) and (7), analytical solutions can be obtain to predict the susceptible, infectious, recovery and object of forensic interest at the event of an attack against IoT nodes.

IV. CONCLUSION

Based on the strategy of predicting the spreads of malware and viruses in a network platform. Various

epidemic model were adopted in predicting susceptible, infectious and recovery nodes in a network platform. With the emergence of IoT and its heterogeneous nature, data volatility among various nodes complicate further the process of identifying object of forensic interest. In this regard, SIR-F model was develop to predict not only the susceptible, infectious and recovery IoT nodes at the event of an attack. The model is capable of predicting object of forensic interest at the point of investigation and recovering of digital evidence.

ACKNOWLEDGMENT

We acknowledge that this research received support from the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia and the Fundamental Research Grant Scheme FRGS/1/2019/ICT03/UPM/02/1 awarded by Malaysian Ministry of Education.

REFERENCES

[1] S. Tang, and B.L. Mark. "*Analysis of virus spread in wireless sensor networks: An epidemic model*",. in 2009 7th International Workshop on Design of Reliable Communication Networks. 2009. IEEE.

[2] E. Oriwoh, D. Jazani, G. Epiphaniou, and P. Sant. *"Internet of things forensics: Challenges and approaches*,". in Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), 2013 9th International Conference Conference on. 2013. IEEE.

[3] R. Hegarty, D.J. Lamb, and A. Attwood. "*Digital Evidence Challenges in the Internet of Things*",. in INC. 2014.

[4] S. Watson and A. Dehghantanha, "*Digital forensics: the missing piece of the internet of things promise*",. Computer Fraud & Security, 2016. 2016(6): p. 5-8.

[5] A. Dadlani, M.S. Kumar, S. Murugan, and K. Kim, "*System dynamics of a refined epidemic model for infection propagation over complex networks*",. IEEE Systems Journal, 2014. 10(4): p. 1316-1325.

[6] J.C. Wierman and D.J. Marchette, "*Modeling computer virus prevalence with a susceptible-infected-susceptible model with reintroduction*". Computational statistics & data analysis, 2004. 45(1): p. 3-23.

[7] R.M. Anderson, B. Anderson, and R.M. May, "*Infectious diseases of humans: dynamics and control*", 1991: Oxford university press.

[8] S.Clémençon, V. Chi Tran, and H. De Arazoza, "*A stochastic SIR model with contact-tracing: large population limits and statistical inference*", Journal of Biological Dynamics, 2008. 2(4): p. 392-414.

[9] A. Pandey, K.E. Atkins, J. Medlock, N. Wenzel, J.P. Townsend, J.E. Childs, T.G. Nyenswah, M.L. Ndeffo-Mbah, and A.P. Galvani, "*Strategies for containing Ebola in west Africa*", Science, 2014. 346(6212): p. 991-995.

[10] J.W. Hui and D. Culler. "*The dynamic behavior of a data dissemination protocol for network programming at scale*", in Proceedings of the 2nd international conference on Embedded networked sensor systems. 2004. ACM.

[11] F. Carrez, M. Bauer, M. Boussard, and N. Bui, "*Final architectural reference model for the IoT v3. 0*", EC FP7 IoT-A Deliverable, 2013. 1.

[12] X. Liu, M. Abdelhakim, P. Krishnamurthy, and D. Tipper. "*Identifying malicious nodes in multihop iot networks using diversity and unsupervised learning*", in 2018 IEEE International Conference on Communications (ICC). 2018. IEEE.