KYUNGPOOK Math. J. 56(2016), 1191-1205
http://dx.doi.org/10.5666/KMJ.2016.56.4.1191
pISSN 1225-6951 eISSN 0454-8124

Internet Worm Propagation Model Using Centrality Theory

Su-Kyung Kwon and Yoon-Ho Choi*
Department of Electrical and Computer Engineering, Pusan National University, Pusan, Korea
e-mail : ksk3579@pusan.ac.kr *and* yhchoi@pusan.ac.kr

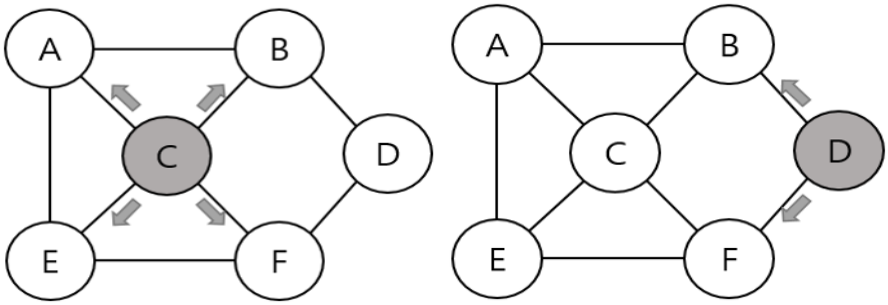
Hunki Baek
Department of Mathematics Education, Catholic University of Daegu, Daegu, Korea
e-mail : hkbaek@cu.ac.kr

Abstract. The emergence of various Internet worms, including the stand-alone Code Red worm that caused a distributed denial of service (DDoS), has prompted many studies on their propagation speed to minimize potential damages. Many studies, however, assume the same probabilities for initially infected nodes to infect each node during their propagation, which do not reflect accurate Internet worm propagation modelling. Thus, this paper analyzes how Internet worm propagation speed varies according to the number of vulnerable hosts directly connected to infected hosts as well as the link costs between infected and vulnerable hosts. A mathematical model based on centrality theory is proposed to analyze and simulate the effects of degree centrality values and closeness centrality values representing the connectivity of nodes in a large-scale network environment on Internet worm propagation speed.

1. Introduction

The Internet worm, a self-replicating computer program, is analogous to a computer virus because it spreads malicious programs. However, contrary to viruses, which are programs that when executed, replicate themselves inside other programs, Internet worms are stand-alone programs that replicate themselves without other

* Corresponding Author.
Received November 2, 2016; accepted November 14, 2016.
2010 Mathematics Subject Classification: 05C82.
Key words and phrases: centrality, degree centrality, closeness centrality, worm propagation, complex network.
This work was supported by a 2-year Research Grant of Pusan National University.



(a) (b)

Figure 1: Worm propagation : (a) from Node C, (b) from Node D

programs [1].

With recent growth in the scale and speed of networks, the widespread damages caused by Internet worms have prompted researchers to model, and analyze the speed at which Internet worms propagate and propose defense strategies against them. In particular, mathematical modelling has been used to predict worm propagation speed and block the spread of damage in early stages [3]-[8].

However, there have been drastic changes in node characteristics in modern networks compared to those from the past. For example, advances in smart devices and technologies have led to widespread Internet use, which has increased the number of access points (APs) available to users for easy access to networks. This increase in APs has elucidated the need for studies of Internet worm propagation modelling that considers node connectivity in a network environment. Previous studies posited that Internet worms spread to the whole nodes in a network. In practice, the spread of Internet worms through a network indicates the existence of connectivity between infected and vulnerable nodes (called target nodes). In other words, target nodes infected during worm propagation are affected by the connectivity of infected nodes.

According to graph theory, infected nodes with high connectivity can infect more target nodes in the same period of time than nodes with low connectivity. For example, Figure 1 shows that with the same graphical topology, the total number of directly connected nodes is four for Node C and two for Node D. Assuming a simultaneous Internet worm propagation from Nodes C and D, count the number of newly infected target nodes for each node. The result shows that the maximum number of infected nodes during worm propagation is two for Node D and four for Node C. This indicates Node C has high connectivity and can infect more neighboring nodes in a shorter time period than other nodes.

This study proposes a model of Internet worm propagation speed that considers the values of degree centrality (DC) and closeness centrality (CC) to represent the

connectivity of nodes, and we investigate this model simulation.

This paper is organized as follows: In section 2, we describe theoretical background of Internet worm and centrality theory. In section 3, we show the related works and then, in section 4, we show a mathematical model for Internet worm propagation speed depending on node connectivity. After we show the simulation results which show the influence of node connectivity on Internet worm propagation speed in section 5, we conclude this paper in section 6.

2. Background

2.1 Epidemic model

In the case that the total number of the whole of the whole nodes in a network remains constant, Internet worms spreading through a network can be divided into three groups: a susceptible (S) group consisting of a set of venerable nodes; an infected (I) group made up of nodes infected Internet worms; and a removed (R) group consisting of nodes infected with Internet worms, but treated with vaccine and immunized after that [9]. The susceptible, infected (SI) model consists of nodes belonging to either the S or I group that move from S to I group (S→I). Here, infected nodes remain untreated, thereby maintaining the infection and spreading the worm to nodes within the S group.

The susceptible, infected, susceptible (SIS) model is made up of nodes in the I group that have not been immunized after treatment and, therefore, stay in the S group. Nodes move from S through I to the other S group (S→I→S). Treated nodes remain in S group and are subject to infection by Internet worms.

Conversely, the susceptible, infected, removed (SIR) model comprises immunized nodes after treatment. Nodes move from the S through I to the R group (S→I→R). The treated nodes in this model get immunized and are not subject to infection, unlike treated nodes in the SIS model that stay in the S group.

To investigate the effects of node connectivity on Internet worm propagation speed in early stages, this paper focuses on the SI model with fastest propagation speed. In particular, we have designed an Internet worm propagation model that uses centrality theory.

2.2 Centrality theory

Centrality is a value for the position of a given node in graph $G=(v, e)$, with node set v and edge set e. There are several ways to quantify centrality, including degree centrality (DC) and closeness centrality (CC), each of which directly considers node connectivity [2].

2.2.1 Degree centrality

Degree Centrality (DC) considers only node connectivity to quantify centrality. Specifically, the number of edges directly connected to a given node (the degree of

the node) is a value of DC. $DC(v)$ refers to the number of nodes directly connected to a given node, v (the linked number to node v), and is calculated using the above formula (2.1). Figure 2-(a) shows the DC value for each node in Figure 1.

$$DC(v) = deg(v) \tag{2.1}$$

2.2.2 Closeness centrality

Closeness centrality (CC) indicates how close one node is to another. There is a direct relationship between the closeness of to other nodes, and the importance of that node to the network. $CC(v)$, CC for a given node v, is calculated using the following formula: (2.2)

$$CC(v) = \left[\sum_{v \neq j}^{N} d(v,j)\right]^{-1} \tag{2.2}$$

where N is the total number of nodes; j represents the total number of nodes included in N, (v excluded); $d(v,j)$ refers to the shortest distance between two nodes. Accordingly, $CC(v)$ is the reciprocal of the sum of shortest distances from node v to node j. For instance, in the topology of Figure 1, the shortest distance between Node C and Nodes A, B, E, F is 1, and the shortest distance between Node C and Node D is 2. The CC for Node C is calculated as follows:

$$CC(C) = (1+1+1+1+2)^{-1} = 0.2 \tag{2.3}$$

Figure 2-(b) shows the CC for each node in the topology of Figure 1.

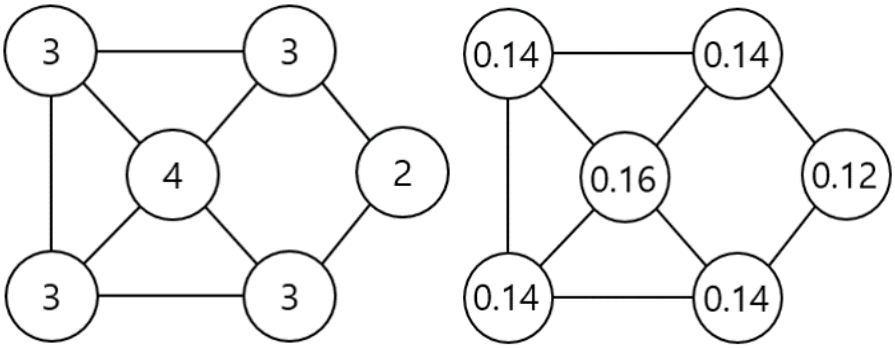
3. Related work

The term Internet worm was first used in John Bruners 1975 novel. The Shockwave Rider and then was adopted by computer researchers to honor the software described in the novel. In 1978, the first Internet worm was implemented by researchers [1]. Since then it creased.

Advancement in network capacity and connectivity has increased the danger of stand-alone Internet worms, thereby promoting many recent studies on Internet worms. Initial researches on Internet worms investigated the SIR model, looking specifically at declines in the nodes of the I group; whereas the spread of vaccine programs resulted in an increase in the nodes of R group [3]. Over time, researchers have increasingly focused on changes in the number of nodes in the I group by installing security programs such as Firework [4]. Recent studies have focused on the vulnerability of nodes in the R group that get re-categorized into the S group over time, due to the adaptations in Internet worm [5].

Ref.	Explanation	Propagation Model
[3]	Two-factor model that considers patching or upgrading susceptible computers, or even disconnecting the susceptible computers from Internet.	$dI(t)/dt = [\beta(t)S(t) - \gamma]\, I(t)$
[4]	Two-factor model based the SIR model, which considers network defence mechanism and Internet worm propagation procedure	$\frac{dIn}{d\frac{S_n}{N_I}} = \beta\frac{S_n}{N_I} - r\frac{N_I - S_n}{N_I}$
[5]	SIRS model based on the Kermack-Mckendrick model, which considers the situation the patched hosts become susceptible again due to user's carelessness	$dI(t)/dt = \beta(t)SI - \gamma I(t)$
[6]	Worm propagation model that considers the node's movements in VANET(Vehicular Ad-hoc Network) by taking into account the full topology of the ad-hoc networks	$S(V) = (S_0 + VT)\left[1 - \left(\frac{V}{V_0}\right)^{\delta}\right]^{-1/2}$
[7]	Worm propagation model that modifies epidemic propagation models using internet worm's propagation pattern	$I(t) = \delta \int\int dxdy$
[8]	Comparison the performance of the SIS model with the performance of the modified SIR model. From the simulation results on the different network topologies, it is shown that the elapsed time for infecting a large portion of the network very significantly depends on where the infection begins.	$\frac{dI(t)}{dt} = \beta I(t)S(t) - \lambda I(t)$

Table 1: Well-known Internet worm propagation models



(a) (b)

Figure 2: Centrality Values: (a) for DC, (b) for CC

Terms	Notation
$S(t)$	Number of susceptible nodes at time t
$I(t)$	Number of infected nodes at time t
$R(t)$	Number of removed node at time t
$S(V)$	The equilibrium gap S between two adjacent cars, the mean vehicular velocity V.
β	Infection rate
N	Number of node
λ, γ	Removed rate
δ	Cured rate

Table 2: Parameters used in Table 1

In addition, an increasingly complex network environment, driven by rapid developments in smart devices, has led to studies on the spread of worms in vehicular ad hoc networks [6] and the spread of information in wireless sensor networks [7]. However, few studies have proposed worm propagation models that consider the connectivity of each node in a network.

4. Internet worm propagation model depending on node connectivity

For any Internet worm that spreads through a network, an infected node usually scans its neighboring nodes, selects target nodes, and spreads the worm to its target nodes.

In turn, the infected node spreads Internet worms to already infected nodes,

Terms	Notation
$I_d(v)$	Number of infected nodes directly connected with node v
$DC(v)$	Degree centrality of node v
$CC(v)$	Closeness centrality of node v

Table 3: Parameters used in the proposed model

as well as nodes that are directly connected to the infected node that have been selected in the scanning process. This suggests a linkage between infected and scanned nodes.

However, ignoring the above characteristics of connectivity, most previous researches on Internet worm propagation assumed that the infected node scanned the whole nodes during worm propagation. This assumption was utilized to generate formulas analogous to (3.1) for the SI model of worm propagation.

$$I(t+1) = I(t) + [N - I(t)]\,\beta \tag{4.1}$$

$$\beta = \frac{\eta}{N} \tag{4.2}$$

Here, $I(t)$is the number of nodes infected in real time; N is the total number of nodes; β is the rate of infection, calculated using the formula (3.2), where N divided by η signifies the number of nodes scanned per hour, which is the rate of nodes infected per hour.

4.1 Effects of DC on internet worm propagation speed

In the SI model, the infected node scans the whole uninfected nodes [N-$I(t)$]. In practice, the infected node only scans directly connected nodes during worm propagation. Therefore, the total number of nodes to which any infected node, v, spreads an Internet worm is expressed as $DC(v)$.

However, any infected node directly connected with an infected node, v, reduces the total number of nodes to which node v spreads the Internet worm. In Figure 3, $DC(C)$ that the total number of nodes connected to Node C is 4, which suggests that Node C has spread a worm to a total number of 4 other nodes. Given the Node B is already infected, the actual number of nodes that can be infected with the Internet worm is 3. In other words, the total number of nodes to which Node C can spread a worm is dependent on C in $DC(C)$, defined as $[DC(C) - I_d(C)]$, where $Id(C)$ is the number of infected nodes but is excluded from this calculation. In addition, if a limited number of nodes is scanned per hour, the actual number of nodes that can be infected by Node C in time t, is defined as$\beta\,[DC(C) - I_d(C)]$. This expression is an propagation model in which worms are spread only to Node C. In case every infected node spreads worms, the number of infected nodes in $(t+1)$, is defined by (3.3).

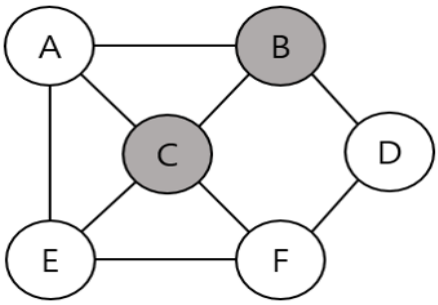


Figure 3: Example where the infected nodes, B and C, are located at one-hop distance

$$I(t+1) = I(t) + \sum_{i \in N}^{I(t)} [\beta(DC(i) - I_d(i))] \tag{4.3}$$

The number of nodes infected over time in $(t+1)$ is calculated as the sum of $\beta\,[DC(C) - I_d(C)]$ which represents the infected node i, at time t.

4.2 Effects of CC on internet worm propagation speed

Closeness centrality (CC) represents the distance nodes. A high DC signifies a higher CC. In Figure 1, while Node C with a high DC can reach Node A directly, Node D with a low DC can go to Node A by travelling around other nodes with one hop. This indicates that a node with high CC takes less time to reach other nodes. The use of CC spreads worms one hop more than using DC alone, within the same time t. The total number of nodes to which the infected node i can spread Internet worms is the same as that of the Internet worm propagation model using DC. However, under the assumption that CC will spread worms within the same time using one more hop, $I(t+1)$, representing the number of nodes infected in $(t+1)$, can be expressed as (3.4).

$$I(t+1) = I(t) + \sum_{i \in N}^{I(t)} [\beta(DC(i) - I_d(i) + \sum_{i \neq k, d(i,k)=1}^{DC(i)-I_d(i)} [\beta(DC(k) - I_d(k))]] \tag{4.4}$$

Node k in which Internet worms are propagated by the current infected node i between time t and $(t+1)$ also propagates Internet worms. For any node k spreading worms with one more hop, the number of nodes infected between t and $(t+1)$ can

be calculated as (3.5).

$$\begin{aligned} I(t+1) =& I(t) + \sum_{i \in I}^{I(t)} [\beta(DC(i) - I_d(i) + \\ & \sum_{i \neq k, d(i,k)=1}^{DC(i)-I_d(i)} [\beta(DC(k) - I_d(k)) + \\ & \sum_{i \neq k \neq l, d(i,k)=1, d(i,l)=2, d(k,l)=1}^{DC(k)-I_d(k)} [\beta(DC(l) - I_d(l))]]] \end{aligned} \tag{4.5}$$

In the case that Internet worm continues to spread to node i and H_N, N hops far of CC on Internet worm propagation can be expressed using (3.6).

$$\begin{aligned} I(t+1) = & I(t) + \sum_{i \in I}^{I(t)} [\beta(DC(i) - I_d(i) \\ & + \sum_{i \neq k, d(i,k)=1}^{DC(i)-I_d(i)} [\beta(DC(k) - I_d(k)) + ... \\ & + \sum_{i \neq k ... \neq H_N, d(i,k)=1, ... d(H_N-1, H_N)=1}^{DC(H_N-1)-I_d(H_N-1)} [\beta(DC(H_N) - I_d(H_N))]]....]] \end{aligned} \tag{4.6}$$

5. Performance evaluation

5.1 Comparison of theoretical performance

This section outlines a comparison of changes in $I(t)$ over time between the SI model described in section 4 and a SI model employing DC, using the values: $I(0) = 5$, $N = 10,000$, and $\beta = 0.4$. According to Figure 4, the SI model using DC spreads Internet worms to the whole nodes at $t = 2$. However, the SI model, with a high initial $I(t)$, actually takes longer to spread Internet worms to the whole nodes. Here, the $I(t)$ is lower than that in the SI model at $t = 1$, but at $t = 2$, all nodes infected by $I(1)$ spread worms simultaneously, thereby infecting the whole nodes much faster. Conversely, the $(N - I(t))$ decreases over time in SI model, thereby slowly increasing $I(t)$. According to the mathematical model, worms are propagated to the whole nodes over a shorter time period in the SI model employing DC. The following section uses an actual simulator to investigate changes in $I(t)$ over time in a variety of environments. We further examine the relative utility of the SI model using DC and the SI model using CC.

5.2 Comparison of experimental performance

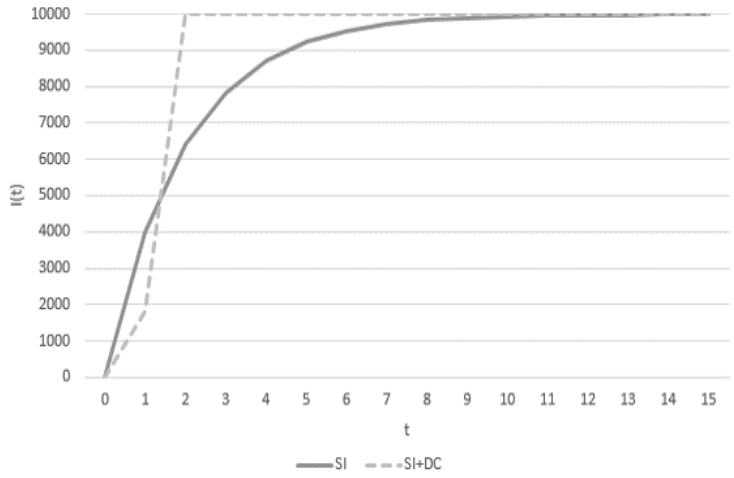


Figure 4: Comparison of $I(t)$ under SI and SI+DC models

This section outlines an experiment employing a simulator on the two Internet worm propagation models described in section 4. The simulator was implemented in eclipse MARS 2 using a Java 1.7. Network topology was also constructed using *GraphStream*, a graph-related library. In addition, the SI model employing CC was simulated to assess the connectivity of network nodes.

5.2.1 Experiment of internet worm propagation

This section provides a comparison of Internet worm propagation speeds between three SI models: a SI model using DC, considering node connectivity; a SI model using CC, considering node connectivity; and a SI model scanning whole nodes, not considering node connectivity by centrality. The experiment used the following values: $I(0) = 5$, $N = 10,000$, and $\beta = 0$.

Figure 5 shows a graph of increases in $I(t)$ over time t. The SI model rapidly increases in $I(t)$ during the early stage of infection, compared to its behavior in the other two cases; but it progressively declines. SI models using DC and CC are initially lower in $I(t)$ than the SI model, but over time, are higher in $I(t)$ than the SI model.

The SI model, not restricting the scanned object, initially infects a greater number of nodes than the SI model employing DC and restricting the scanned object formula. Yet, given the initial spread of Internet worms in a number of nodes, the number of $(N - I(t))$ declines and the increase in $I(t)$ gradually declines over time. SI model using DC and SI model using CC, initially restricting the object of scan, have a low $I(t)$. But as the growing number of infected nodes spread worms as much as their DC over time, $I(t)$ increase drastically increases, compared to SI

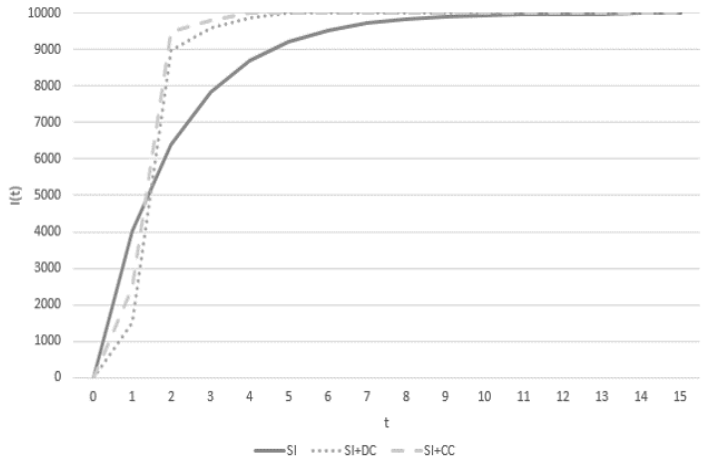


Figure 5: Influence of DC and CC on $I(t)$

model.

The SI model using CC spreads worms once more than the SI model using DC, which is expected to have a much higher $I(t)$ than the latter. However, an increase in $I(t)$ suggests increases in $Id(i)$. Namely, a number of already infected nodes is excluded from worm propagation between t and $(t+1)$, which makes little differences in $I(t)$.

5.2.2 Investigating changes in network architecture

This section discusses the effects on $I(t)$ in two Internet worm propagation models that change the network environment. One such model is an Internet worm propagation model using DC, and the other is an Internet worm propagation model using CC.

The current widespread network environment allows a variety of terminals accessible to the nodes of an access point (AP) for network connectivity structure in which the connectivity of all nodes in a network is not equal or similar, to a high connectivity for AP nodes and low connectivity for terminal nodes.

Any infra-structured network infected by worms that follow the SI model employing DC and SI model using CC will increase the number of nodes that get scanned, leading to an increase in $I(t)$.

Therefore, this experiment examines the effect on $I(t)$ caused by changes in network architecture by comparing Internet worm propagation speed in an infra-structured network topology with AP and ad-hoc network topology without AP. Both networks have an average DC of 70.

Figure 6 shows the effects on $I(t)$ caused by Internet worm propagation in different network environments. One being an infra-structured network topology

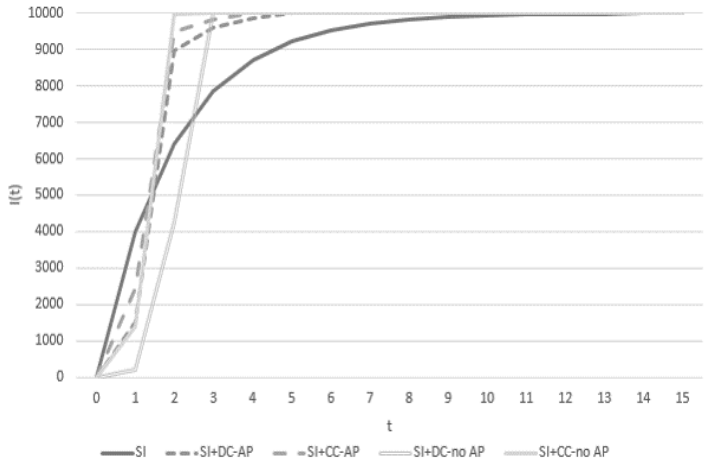


Figure 6: Comparison of $I(t)$ under various network environments

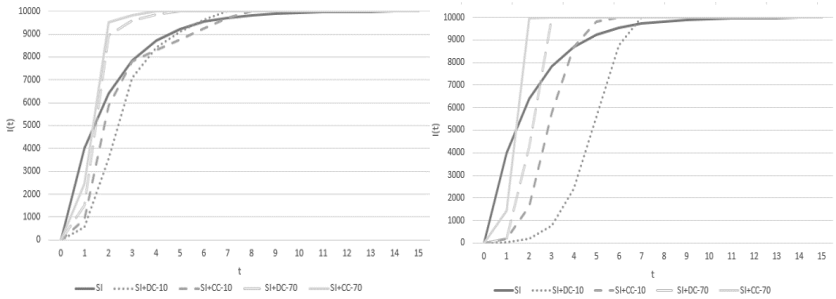
with AP, the other is an ad-hoc network topology without AP, both of which employ the SI model and either DC or CC.

If infected, the infra-structured network topology with AP spreads Internet worms through connected terminals, with a rapid increase in $I(t)$ during the early spreading; yet, once worm propagation is complete in the AP and begins in terminals, increases in $I(t)$ significantly decline, as most terminals with low connectivity spread worms to fewer nodes. In addition, with many nodes already infected by worms from AP, the increases in $I(t)$ drastically decline over time.

Conversely, ad-hoc network topology without AP initially has a lower $I(t)$ than the infra-structured network topology with AP, but for each, the DC of its nodes is almost equally constructed. In an ad-hoc network topology, the number of uninfected nodes connected to one node is similar to those of other nodes. This node does not experience a downturn in increases in $I(t)$ over time that appears in the infra-structured network topology. This suggests that an ad-hoc network topology propagates Internet worms to the whole nodes over shorter period of time.

5.2.3 Investigating the effects of average DC

In the SI model, regardless of the respective connectivity of individual nodes, propagation speed is not affected by topology. However, models that use DC and CC consider the respective connectivity of nodes, and are affected by changes in network connectivity. For instance, an increase in the number of terminals connected to an AP indicates an increase in the number of nodes propagating Internet worms in the AP. In an infra-structured network topology in which an increase in the average number of terminals connected to the AP leads to an increase in the average DC of the network, SI models employing DC and CC are expected to have



(a) (b)

Figure 7: Influence of DC or CC on $I(t)$ under different network topologies: (a) infra-structured network topology with AP and (b) ad-hoc network topology without AP

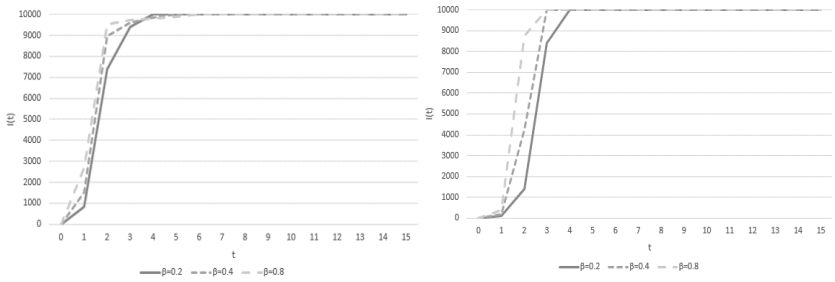
great effects on $I(t)$.

Figure 7-(a) shows a comparison between infra-structured network topologies with an average DC of 10 and 70, respectively. Here, $I(t)$ is largely affected by changes in DC. This indicates a clear difference in the initial propagation of Internet worms between the two. Thus, changes in average DC within a network have a significant effect on $I(t)$ in ad-hoc network topology Figure 7-(b). As seen in Figure 10, unlike the SI model that remains unaffected by changes in the network, a SI model employing DC and the SI model using CC exhibits great changes in values depending on changes in the network environment. This indicates that the results of SI models employing DC and SI model using CC are closer to those in a real world environment.

5.2.4 Investigating the effects of infection rate

The infection rate, β, is an estimate of the rate of maximum propagation of an Internet worm over time t. Particularly, in a SI model that considers only directly connected nodes, and follows SI model using CC, the propagation of an Internet worm in nodes with low connectivity puts great limitations on the number of nodes that can possibility propagate, depending on the value of β. The experiment conducted in this paper compares the effects of changes in β on $I(t)$ in a SI model employing DC.

Figure 8-(a) shows the propagation speed in an infra-structured network topology with a β value of 0.2, 0.4, and 0.8 respectively. Despite some differences in initial $I(t)$s, depending β, these values are similar in their Internet worm propagation patterns to the whole nodes. A low β value yields a low $I(t)$ at an early stage. As some nodes connected to a given node are already infected, increasing $I(t)$ does not significantly decline over time. Alternatively, a high β leads to high $I(t)$ in early stage, but increase in $I(t)$ significantly decline over time. As seen in Figure 8-(b),



(a) (b)

Figure 8: Influence of infection rate on $I(t)$ under: (a) infra-structured network topology with AP and (b) ad-hoc network topology without AP

the value of $I(t)$ varies depending on β, but the amount of worm propagation to the whole nodes is similar.

6. Conclusion

Previous papers on Internet worm propagation speed focused either on the network environment or the spread of vaccine programs, assuming that the scanning object consisted of uninfected whole nodes during the process for worm propagation. However. node connectivity actually has great effects on Internet worm propagation speed through large-scale networks. This paper focuses on DC and CC models of Internet worm propagation in relation to node centrality. We also mathematically analyze and simulate the effects of node connectivity on Internet worm propagation speed.

References

[1] Computer Worm, *https://en.wikipedia.org/wiki/Computmer_worm*

[2] Centrality, *https://en.wikipedia.org/wiki/Centrality*

[3] C. C. Zou, W. Gong and D. Towsley, *Code red worm propagation modeling and analysis*, the 9th ACM conference on Computer and communications security, (2002), 138-147.

[4] A. Mohammed, S. M. Nor and M. N. Marsono, *Network Worm Propagation Model Based on a Campus Network Topology*, the 2011 International Conference on Internet of Things and 4th International Conference on Cyber, (2011), 653–659.

[5] D. Zhang and Y. Wang, *SIRS: Internet Worm Propagation Model and Application*, Electrical and Control Engineering, 2010 International Conference on, (2010).

[6] M. Nekovee, *Modeling the Spread of Worm Epidemics in Vehicular Ad Hoc Networks*, Vehicular Technology Conference, (2006).

[7] M. S. Haghighi, S. Wen and Y. Xiang, *On the Race of Worms and Patches: Modeling the Spread of Information in Wireless Sensor Networks*, IEEE Transactions on Information Forensics and Security, (2016).

[8] J. Kim, S. Radhakrishnan and S. K. Dhall, *Measurement and Analysis of Worm Propagation on Internet Network Topology*, Computer Communications and Networks, (2004).

[9] C. C. Zou, D. Towsley and W. Gong, *The Study of Network Worm Propagation Simulation*, IEEE Transactions on Dependable and Secure Computing, (2007).

[10] C. C. Zou, D. Towsley and W. Gong, *On the performance of Internet worm scanning strategies*, Performance Evaluation, (2006), 700-723.

[11] A. Wagner, T. Dubendorfer, B. Plattner and R. Hiestand, *Experiences with Worm Propagation Simulations*, the 2003 ACM workshop on Rapid malcode, (2003), 34-41.

[12] I. Gaye, G. Mendy, S. Ouya and D. Seck, *New centrality measure in Social Networks based on Independent Cascade (IC) model*, 2015 3rd International Conference on Future Internet of Things and Cloud, (2015), 675-680.

[13] Y. Wang, S. Wen and Y. Xiang, *Modeling the Propagation of Worms in Networks: A Survey*, IEEE Communications Surveys & Tutorials, (2013), 942-960.