ATLANTIS PRESS

Advances in Computer Science Research, (ACSR), volume 52

2016 International Conference on Computer Engineering and Information Systems (CEIS-16)

Analysis of Virus Propagations on Community-based Wireless Sensor Networks

Qiao Li, Jian-Gao Chen, Xiao-Feng Wang, Cheng-Wang Zhao
Nanjing Artillery Academy,
Nanjing, China
E-mail: liqq007@bit.edu.cn, mrchen1972@sina.com, Londonwxf@163.com, zhaochengwang1976@163.com

Bin Wu
Science and Technology on Space Physics Laboratory,
Beijing, China
E-mail: wubin_005@163.com

***Abstract*-Self-organizing dense sensor networks are expected to resolve the challenges of spectral efficiency, energy equilibrium, and device management. Because of cost and resource constraints, the sensor node is fragile to outside virus attacks. In this paper, we propose a community-based network model to depict complex network structures. Virus attacking behaviors and the virus propagation are theoretically analyzed in three cases when it spreads over the network. The mathematical analysis shows that the total number of infected nodes increases exponentially with the epidemic time. Structures of networks, the infection probability and the shortcut probability affect the rate and extent of spreading of viruses simultaneously. The conclusions are validated through simulations and evaluations, and are envisioned to have future applications in the immunization of wireless sensor networks.**

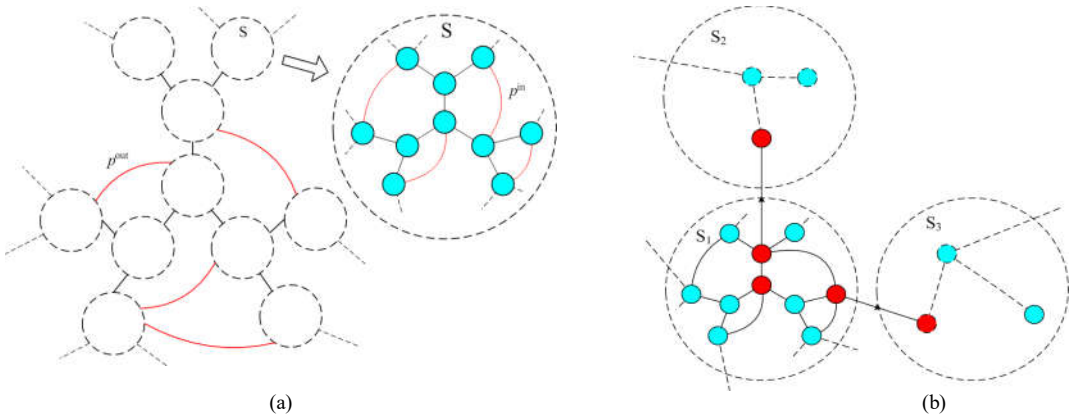
Keywords-community; mathematical analysis; network structure; virus propagation

I. INTRODUCTION

With the rapid development of the Internet of Things (IoT) and the Web of Things (WoT), wireless sensor networks (WSNs) have been garnering increased attention in recent years for both consumers and industrial related applications, such as agricultural monitoring, healthcare service and industrial communication [1]. Numerous tiny nodes cooperate together to achieve complicated application requirements in a certain network which is often partitioned due to geographical separations or node movements. The network structure depends on the software architecture and the deployment of tiny nodes. There is now a trend of proposing different software protocols, topology control algorithms and node deployments for various application requirements to organize sensor nodes into a powerful wireless network [2-4]. Because of cost and resource constraints, the sensor node which is fragile to outside attacks does not have a complicated hardware architecture or operating system to protect the communication safety [5]. Sensor infections such as viruses and worms spread over networks, with different types of networks being exploited by different types of infections. The sensor virus which is a special data packet may inject a piece of malicious code into a sensor node and destroy the software or the hardware. For example, a wireless sensor network which consists of sensor nodes using the Von Neumann architecture is vulnerable to sensor worm attacks if the program does not perform careful boundary checks [6]. The structure and the deployment of networks affect the rate and extent of spreading of sensor infections. In a sense, the propagation becomes faster if sensor nodes are connected with each other more closely. The susceptible node may have more infectious neighbors in the deployment, and it is easier for the susceptible one to be infected. Also, a susceptible node may be surrounded by more infectious neighbors in a mesh structure than in a line structure generally. As a result, the analysis of virus propagations on a certain structure attracts attentions in recent years [7-9]. Both traditional and network-based epidemiological models have been applied to the sensor contagion. In tradition, a wireless sensor network with a complex structure may be deemed as a random network in the epidemic analysis [10].

It is a highly nontrivial task to organize a large-scale wireless sensor network. Variable hierarchical structures based on communities [11] have been proposed to optimize resource utilization and to meet multiple application requirements. Qualitatively, a community is defined as a subset of nodes within the graph such that connections between the nodes are denser than connections with the rest of the network. The hierarchical structure is suitable for such a wireless network which may be divided into several parts to take different monitoring or communication tasks. The nodes in each part communicate frequently with each other, and all the parts are constructed to a whole network by some nodes which act as the gateways.

A community construction technology was proposed by F. Wei et al. to transfer emergency information along its individual route in the dynamic network, whose requirements could not be satisfied by conventional static and centralized management [12]. A community-based routing protocol, which balanced the energy consumption and extended the network lifetime, was presented by J. Y. Wu et al. to divide the wireless sensor network into densely connected subgroups by the algorithm of detecting community structure in the network [13]. Combining the community structure and the data recovery algorithm, a self-organizing management scheme was developed by T. Y. Chuang and K. C. Chen to enhance the energy efficiency of wireless sensor networks [14]. An information-centric processing methodology was proposed to tackle the challenges of spectral efficiency, energy equilibrium, and device management. Considering the sensor node is fragile to infections, the virus propagation on community-based wireless sensor networks is in essence complicated.



(a) (b)

Figure 1. (a) The proposed community-based model. (b) Virus attacking behaviors on the community-based network.

In common, the virus propagates fast in one community for dense connections among the nodes, and it extends into other communities after the gateway which is an ordinary node is infected. Two interacting viruses spreading in community-based networks were explored by B. Wang et al., and a random clique model composed of different orders of cliques was proposed [15]. For dynamics of two interacting viruses, the community effect was hypersensitive to the cohesiveness and concentration of cliques. Based on the Susceptible – Infectious – Recovered – Susceptible (SIRS) model with birth and death rates, the spreading of infections in complex heterogeneous networks was analyzed by C. H. Li et al. to find that the dynamics of the model was completely determined by a threshold value [16].

In this paper, we propose a novel community-based model which consists of a regular bottom structure and random bonds for the analysis of virus propagations on the hierarchical network. We study virus attacking behaviors starting from an infectious source on the hypothesis that the virus extends within the community and to other communities until the entire network is attacked. The proposed model has some inner relationship with a small world network [17], which is also a basic property of the complex network [18].

II. COMMUNITY-BASED NETWORK MODEL

Fig. 1(a) is the proposed community-based network model, in which the blue circle is the sensor node and the dotted circle is the community. The black line gives the regular bond between two nodes or two communities. The red line gives the random bond between two nodes or two communities. Communities are abstracted as "blocks" of the network in our study. The nodes are connected densely with each other in every block, which means frequent contacts between two nodes or close bonds among the nodes. The underlying deployment of nodes within the block is simplified as the Cayley tree [19], and random bonds between two nodes are added to the Cayley tree with a probability p^{in} to make the nodes linked closely. The Cayley tree, where every node i has the same degree $k_i=z+1$ (except for the leaf node on the boundary which has the degree $k=1$), is a regular graph with no density fluctuation. Its tree structure starts from a root node whose generation g is 0, and the number of the root's child nodes is $z+1$. Every child node is repeatedly allocated z children, and this process continues for a fixed number of generations g to construct a tree-based network structure. The Cayley tree can grow either in "width" (via z) or in "depth" (via g). The number of nodes in generation $g>0$ is $n(g)=(z+1)z^{g-1}$. The total number of nodes is $N(g)=1+\sum_{g'=1}^{g} n(g')$, and the total number of bonds is $M(g)=N(g)-1=(z+1)(z^{g-1})/(z-1)$. To organize the blocks together and construct a whole network, regular bonds or random bonds are added to the communities. The underlying architecture of the network is also abstracted as a Cayley tree to depict the community-based network and make the analysis simple. Random bonds between two communities are added to the underlying architecture with the probability p^{out} under the condition $p^{in} > p^{out}$. The spatial-temporal dynamics of the prevalence will be analyzed on the hybrid model which depicts the detail structure characteristics of the hierarchical community-based network.

The proposed community-based model has some inner relationship with the small world network, both of which have random bonds besides the regular network structure. In the latter network, every node may be connected with another one which is chosen randomly over the network to construct a symmetrical network. The infectious node would infect its susceptible neighbor with the identical infection probability. In a sense, every susceptible node has the same number of infectious neighbors in the symmetrical network. Compared with a small world network, there exist two different random probabilities within the community and between two communities in our proposed model. The spread of the propagation is faster within the community than that outside of the community for the unbalanced deployment of the hierarchical network. The susceptible node has more infectious neighbors within the community than that outside of the community for denser bonds among the nodes. Fig. 1(b) shows virus attacking behaviors on the proposed community-based model, in which S_1, S_2 and S_3 are three communities of the network. The virus spreads from infectious nodes (red ones) to susceptible neighbors

(blue ones) in the same community and susceptible ones in other communities along external bonds.

III. ANALYSIS OF VIRUS SPREADING

Considering attacking behaviors of the sensor virus, the epidemic starts from one infectious center and the infected area is enlarged gradually. The regular underlying structure and random bonds of the community-based network affect attacking behaviors simultaneously, which means that infected nodes have two sources: the regular underlying structure and random bonds. Multiple infectious sources arise from random bonds which connect a node randomly with another in the same community or outside of the community. Spatial-temporal characteristics of the virus propagation are analyzed on the hypothesis that it is a percolation process [20, 21]. A susceptible node will be infected by the sensor virus coming from its infectious neighbor with the infection probability h. Based on the percolation theory [22], the epidemic will die naturally when the infection probability h is smaller than the percolation threshold h_c. In the case, the virus propagation stops off halfway because the infection behavior cannot continue. We analyze the epidemic supposing that the infection probability h is larger than the percolation threshold h_c in this paper. The sensor virus will attack the network from one side to another until the entire network is attacked as a big infected cluster arises over the network. In the research, every node has two states: the susceptible state and the infectious state. If a susceptible node is connected to an infectious neighbor, it could be infected under attacks of the sensor virus. The new infected node could infect its susceptible neighbors which are located in the same community or outside of the community with the same infection probability as well. The process goes on until the propagation stops off. As a result, a part of susceptible nodes in the network will experience the transformation from the susceptible state to the infectious state, and this propagation is in fact a Susceptible → Infectious (SI) process [23]. The number of infected nodes increases as the sensor virus spreads on the hierarchical network. Considering the unbalanced deployment of the community-based network, the epidemic speed in the community is higher than that between two communities. A susceptible node has more infectious neighbors in the community for denser bonds, which make the epidemic faster.

Following the description above, the number of infected nodes $I(t)$ will be calculated by the equation

$$I(t)=\sum_{t'=0}^{t} ha(t')[1+2\xi^{-1}I(t-t')], \quad (1)$$

where $a(t')$ is the number of infected neighbors on the underlying architecture of the community-based model at the time t'. ξ is the average distance between the ends of the shortcut, which varies within the community and outside of the community. The above equation consists of two parts because the infected neighbors come from two sources: underlying regular bonds and random bonds. The number of infected nodes increases with the propagation time, and the virus expands from multiple epidemic sources due to the random bonds. Three cases of virus propagations are analyzed in our research.

A. Virus Spreading in Case 1

The network G is split into m subgraphs G_0, G_1, $G_2 \ldots G_m$ in terms of the unbalanced deployment of the hierarchical community-based network. N_0, N_1, $N_2 \ldots N_m$ are the numbers of nodes within each subgraph ($N=N_0+N_1+N_2+\ldots+N_m$). If the network G is split into N_c subgraphs of the same size, each subgraph consists of $N_0=N/N_c$ nodes. The underlying architecture of the network is abstracted as the Cayley tree to make the analysis simple. Every community i has the same degree $k_i=z+1$ which denotes the number of child communities. The number of communities of generation $g_c>0$ is $(z+1)z^{g_c-1}$, and the total number of communities is $N_c=1+\sum_{g'=1}^{g_c}(z+1)z^{g_c-1}$. The total number of external regular bonds is calculated by $N_c-1=(z+1)(z^{g_c}-1)/(z-1)$ in the network. If the virus spreads in the community and attacks its neighbors within other communities along external random bonds only, the local infected clusters are linked together to construct a giant infected cluster over the network. The average distance ξ can be approximately calculated by $\xi=\dfrac{(z-1)N}{(z+1)(z^{g_c}-1)p}$, and $p=p^{\text{out}}$.

Supposing that the size of the community is large enough and the infected parent node will attack all its children nodes in each time unit with $a(t')=(z+1)z^{t'}$, from (1) the total number of infected nodes at any time t can be calculated as follows:

$$I(t)=\sum_{t'=0}^{t} h(z+1)z^{t'}[1+2p\frac{(z+1)(z^{g_c}-1)}{(z-1)N}I(t-t')]. \quad (2)$$

$I(t)$ can be approximately calculated by

$$I(t)=\int_0^t h(z+1)z^{t'}[1+2p\frac{(z+1)(z^{g_c}-1)}{(z-1)N}I(t-t')]dt'. \quad (3)$$

Let $F(t)=z^t\int_0^t z^{-t'}I(t')dt'$, and both sides are differentiated with respect to t, it is obtained that

$$F'(t)=\ln z\cdot F(t)+I(t). \quad (4)$$

Equation (3) is rewritten as

$$I(t)=h(z+1)\int_0^t z^{t'}dt'+\frac{2ph(z+1)^2(z^{g_c}-1)}{(z-1)N}F(t). \quad (5)$$

$F(t)$ is given by

$$F(t)=\frac{(z-1)N}{2ph(z+1)^2(z^{g_c}-1)}I(t)-\frac{(z-1)N}{2p(z+1)(z^{g_c}-1)}\int_0^t z^{t'}dt' . \quad (6)$$

The first-order derivative of $F(t)$ with respect to t is gotten by

$$F'(t)=\frac{(z-1)N}{2ph(z+1)^2(z^{g_c}-1)}I'(t)-\frac{(z-1)N}{2p(z+1)(z^{g_c}-1)}z^t . \quad (7)$$

From (4), (6) and (7), the following equation is obtained as

$$I'(t)-h(z+1)z^t=\ln z[I(t)-h(z+1)\int_0^t z^{t'}dt']$$

$$+\frac{2ph(z+1)^2(z^{g_c}-1)}{(z-1)N}I(t), \quad (8)$$

where g_c is the fixed number of generations of communities, $N=N_cN_0$ and $N_0=(z+1)(z^{g_0}-1)/(z-1)+1$ (g_0 is the number of generations of nodes in the community). The second-order derivative of (8) with respect to t can be calculated by

$$I''(t)=\{\ln z+\frac{2ph(z-1)(z+1)^2(z^{g_c}-1)}{[(z+1)(z^{g_c}-1)+z-1][(z+1)(z^{g_0}-1)+z-1]}\}I'(t) . \quad (9)$$

It can be solved as

$$I(t)=C_1e^{\{\ln z+\frac{2ph(z-1)(z+1)^2(z^{g_c}-1)}{[(z+1)(z^{g_c}-1)+z-1][(z+1)(z^{g_0}-1)+z-1]}\}t}+C_2, \quad (10)$$

where C_1 and C_2 are two constants. There is z+1 infected nodes at t=0, and we can get

$$C_1+C_2=z+1 . \quad (11)$$

If the size of the community is large enough, (10) is suitable for the description of the virus prevalence on the network as the virus attacks senor nodes. If we consider the extreme case of $g_c\to\infty$ ($N\to\infty$), (10) can be rewritten as

$$I(t)=C_1e^{[\ln z+\frac{2ph(z+1)(z-1)}{(z+1)(z^{g_0}-1)+z-1}]t}+C_2 . \quad (12)$$

We consider the extreme case of $g_0\to\infty$ ($N\to\infty$), (10) can be rewritten as

$$I(t)=C_1e^{\ln zt}+C_2 . \quad (13)$$

In (12) and (13), $C_1+C_2=z+1$.

From (10) we can see that the total number of infected nodes of the network $I(t)$ increases exponentially with the epidemic time t as the virus propagates on the community-based model. If the size of the community is fixed, the prevalence will be affected by the underlying architecture, the shortcut probability, the infection probability and the size of the community in the extreme case of $N\to\infty$. If the number of the communities is fixed, the prevalence will only be affected by the underlying architecture in the extreme case of $N\to\infty$.

B. Virus Spreading in Case 2

Providing that the boundary of the community expands to the entire network, the difference between the shortcut probability within the community and that outside of the community will be ignored in this case. The network is transformed into a homogeneous small world network with $p=p^{in}=p^{out}$ and $N=N_0$. ξ can be approximately calculated by $\xi=\frac{N}{(N-1)p}$. Supposing that the infectious parent node may attack all its child nodes with $a(t')=(z+1)z^{t'}$ in each time unit, from (1) the total number of infected nodes of the network can be calculated by

$$I(t)=\sum_{t'=0}^{t}h(z+1)z^{t'}[1+2p\frac{N-1}{N}I(t-t')] . \quad (14)$$

The second-order derivative of (14) with respect to t can be obtained by

$$I''(t)=[\ln z+\frac{2ph(z+1)(N-1)}{N}]I'(t) . \quad (15)$$

It can be solved as

$$I(t)=C_3e^{[\ln z+\frac{2ph(z+1)(N-1)}{N}]t}+C_4, \quad (16)$$

where C_3 and C_4 are two constants. There is z+1 infected nodes in the network at t=0, and we get

$$C_3+C_4=z+1 . \quad (17)$$

If the extreme case of $N\to\infty$ is considered, the following result can be obtained

$$I(t)=C_3e^{[\ln z+2ph(z+1)]t}+C_4, \quad (18)$$

and $C_3+C_4=z+1$.

Based on the analysis of Case 2, we can see that the total number of infected nodes of the network increases exponentially as the infection spreads on the assumption that the infectious parent node may attack all its child nodes on the community-based model in each time unit. In the extreme case of $N\to\infty$, the prevalence will be affected by the underlying architecture, the shortcut probability and the infection probability.

C. *Virus Spreading in Case 3*

If the community will be deemed as a whole, we analyze the virus propagation in Case 3. Case 3 is the expansion of Case 1. The propagation is analyzed in the hypothetical scene that each community can be deemed as a whole for strong internal connections and the highly contagious sensor worm. If one sensor node is infected in the community, all susceptible neighbors will suffer from the virus with the identical infection probability. In the analysis, the epidemic starts from certain communities.

If the size of the community is finite and the infectious parent community may attack all its child communities in each time unit with $a(t') = (z+1)z^{i} N / N_c$, from (1) the total number of infected nodes can be calculated by

$$I(t) = \sum_{t'=0}^{t} \frac{hN(z+1)z^{i}}{N_c}[1+2p\frac{(z+1)(z^{g_c}-1)}{(z-1)N} I(t-t')], \quad (19)$$

where $N_c = (z+1)(z^{g_c}-1)/(z-1)+1$ is the total number of communities. g_c is a fixed number of generations of communities, $N=N_cN_0$, and $p=p^{out}$ .The total number of nodes in the community is calculated as $N_0 = (z+1)(z^{g_0}-1)/(z-1)+1$ (g_0 is the number of generations of nodes in the community). The second-order derivative of (19) with respect to t can be obtained by

$$I''(t) = [\ln z + \frac{2ph(z+1)^2(z^{g_c}-1)}{(z+1)(z^{g_c}-1)+z-1}]I'(t). \quad (20)$$

It can be solved as

$$I(t) = C_5 e^{[\ln z + \frac{2ph(z+1)^2(z^{g_c}-1)}{(z+1)(z^{g_c}-1)+z-1}]t} + C_6, \quad (21)$$

where C_5 and C_6 are two constants. Supposing that there is one infected community on the underlying architecture in the initial stage, the following formula is gotten with

$$C_5 + C_6 = \frac{(z+1)(z^{g_0}-1)}{z-1} + 1. \quad (22)$$

If we consider the extreme case of $g_c \to \infty$ ($N \to \infty$), (21) can be rewritten as

$$I(t) = C_5 e^{[\ln z + 2ph(z+1)]t} + C_6, \quad (23)$$

where $C_5 + C_6 = (z+1)(z^{g_0}-1)/(z-1)+1$. In the extreme case, we get the similar result with (18) except for the difference in the initial stage.

From (21) we can see that the total number of infected nodes of the network increases exponentially as the infection spreads. In the extreme case of $N \to \infty$, the prevalence will be affected by the underlying architecture, the shortcut probability and the infection probability.

The analysis of three cases above shows that the total number of infected nodes of the network $I(t)$ increases exponentially with the epidemic time t as the virus spreads on the community-based model. Considering the reduction of remaining susceptible nodes in the network, the exponential growth will experience a decline in the real propagation. In fact, the growth of the number of infected nodes will go on after the exponential phase ends. The total number of infected nodes converges towards a fixed value as the time increases on the basis of the percolation theory. In our study, the susceptible nodes will be infected with the infection probability in any case if they are linked with the infectious neighbors. So, the entire network will be impacted at last as the sensor virus attacks it from one side to another.

IV. FOCUS ON DIFFERENT STRUCTURES

In the analysis of epidemics on the community-based network or the hybrid network, the random graph is selected traditionally as the network model. In a hypothetical scene that all nodes of a network move randomly or flood messages under no rule, the random graph is suitable for depicting the network architecture. The classical simple epidemic model is used to analyze the propagation of sensor viruses [23]. Here sensors have two states: the susceptible and the infectious. The overall rate of new infections given by this model is

$$\frac{dI(t)}{dt} = \frac{hI(t)(N-I(t))}{N}, \quad (24)$$

where N is the size of the network and h is the infection probability. By solving this differential function, the number of infected nodes at any time t is obtained as

$$I(t) = \frac{N}{1+e^{-ht}CN}, \quad (25)$$

where C is a constant factor. From (25) we can see that the number of infected nodes converges towards a fixed value as the time increases. In the hypothetical situation, the position of the node is moveable and random, and the attack of the virus happens in a moment. For the mobility and randomness of the network, the result can not indicate structure characteristics of the network. We cannot observe the epidemiological process in detail according to (25), although the formula hints the exponential characteristic of the propagation. On the basis of the percolation theory, there exists a percolation threshold in the network [22]. The virus dies naturally in the prevalence when the infection probability is below the percolation threshold, and the number of the infected nodes stays at a fixed value at last. When the infection probability is above the percolation threshold, the virus will attack the network from one side to another. We only observe the spatial-temporal dynamic of

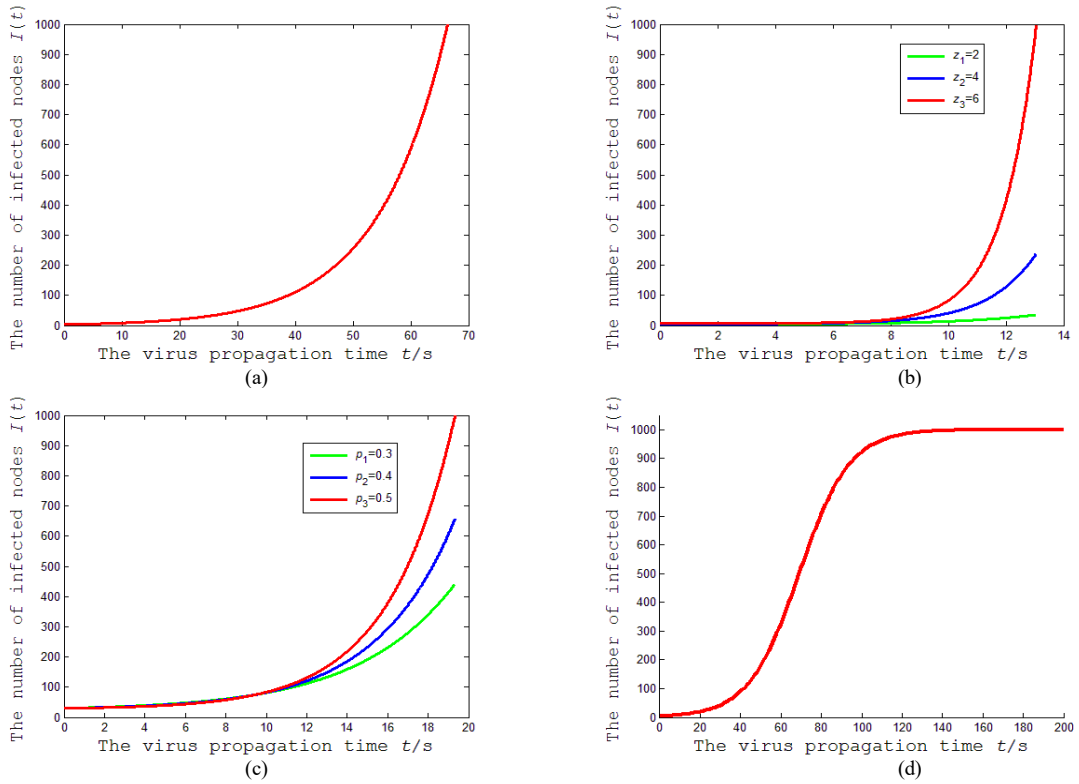


Figure 2. Virus propagations on theoretical models.

the virus prevalence on the assumption that the infection probability is above the percolation threshold.

The proposed community-based model in our research possesses some inner relationship with the small world network. Providing that there is only one node in the community and p is the average number of shortcuts per bond, the proposed community-based model will be transformed into a small world network.

V. PERFORMANCE EVALUATIONS

In this section, we will validate our analytical results with simulations and evaluations. The mathematical evaluations and real architecture evaluations of virus propagations are presented to show that mathematical conclusions reveal virus attacking behaviors in detail. Real architecture evaluations are tested with Zigbee FFDs (Full Function Devices), which construct a community-based network structure over a designated deployment area if some properties of routing layer primitives are modified to achieve application requirements [24].

A. Mathematical Evaluations of Virus Propagations

Mathematical evaluations of virus propagations on the proposed community-based model are presented in this section.

The mathematical evaluation of virus propagations in Case 1 is presented in Fig. 2(a). In this case, the epidemic spreads from a source in the community and the infected clusters are linked together only by external random bonds outside of the community to construct a giant infected cluster over the network. All susceptible child nodes will be attacked by the infectious parent node with $C(t') = (z+1)z^{t'}$ on the underlying tree-based architecture in each time unit. In our evaluation, there are totally 1000 nodes distributed over the deployment area.We may adjust the distribution of nodes by changing g_0, g_c and z. In the simulation, the parameters are set with h=0.8, p=0.3, g_0=2, g_c=5 and z=2. If each time unit includes 10 seconds, C_1 and C_2 of (10) can be approximately calculated at t = 0 and t = 10s. We check how the total number of infected nodes changes with the epidemic time through mathematical simulations. Fig. 2(a) shows that the total number of infected nodes $I(t)$ increases exponentially with the epidemic time t on the basis of (10). Structures of networks affect the rate and extent of virus infections, which can be denoted by changing g_0, g_c and z. In Fig. 2(a), with t being increased to 66.41s, the number of infected nodes increases to 1000 and the entire network is infected.

The mathematical evaluation of virus propagations in Case 2 is presented in Fig. 2(b). In this case, the boundary of the community expands to the entire network which is transformed into a homogeneous small world network, and the difference between the shortcut probability within the community and that outside of the community is ignored. If the virus attacks all susceptible neighbors on the underlying tree-based architecture in each time unit, we set $C(t') = (z+1)z^{t'}$. We check how the total number of

infected nodes changes with the epidemic time through mathematical simulations under h=0.8, p=0.6 and N=1000. If each time unit includes 10 seconds, C_3 and C_4 of (16) can be approximately calculated at t=0 and t=10s. Fig. 2(b) shows that the total number of infected nodes $I(t)$ increases exponentially with the epidemic time t on the basis of (16). We check how structures of networks affect virus attacking behaviors under the condition z_1=2, z_2=4 and z_3=6. In Fig. 2(b), with t being increased to 13.05s, the number of infected nodes increases to 1000 and the entire network is infected at z_3=6. When z_1 is 2, the number of infected nodes increases to 35 at t=13.05s. When z_2 is 4, the number of infected nodes increases to 238 at the same time.

The mathematical evaluation of virus propagations in Case 3 is presented in Fig. 2(c). If each community may be deemed as a whole for strong internal connections, all susceptible neighbors will suffer from the virus with the same infection probability when one node is infected in the community. If the virus attacks all susceptible neighbors on the underlying tree-based architecture in each time unit, we set $a(t') = (z+1)z^{i} N / N_c$. There are totally 1000 nodes distributed over the deployment area. We check how the total number of infected nodes changes with the epidemic time through mathematical simulations under z=2, h=0.8, p=0.3, g_0=2 and g_c=5. From Fig. 2(c) we can see that the total number of infected nodes $I(t)$ increases exponentially with the epidemic time t on the basis of (21). We check how the shortcut probability affects virus attacking behaviors under the condition p_1=0.3, p_2=0.4 and p_3=0.5. In the figure, with t being increased to 19.34s, the number of infected nodes increases to 1000 and the entire network is infected at p_3=0.5. When p_1 is 0.3, the number of infected nodes increases to 441 at t=19.34s. When p_2 is 0.4, the number of infected nodes increases to 657 at the same time.

Fig. 2(d) shows the time evolution of the infected numbers as the virus spreads on the basis of (25). In our evaluation, there are totally 1000 nodes distributed randomly over the deployment area. We check how the total number of infected nodes changes with the epidemic time through mathematical simulations under h=0.8. If each time unit includes 10 seconds, the constant C in (25) can be approximately calculated at t = 0. Fig. 2(d) shows that the virus spreads exponentially in most time if the network is deemed as the random network. In the later stage of the propagation, the exponential growth experiences a decline due to the reduction of remaining susceptible nodes in the network. The infected number keeps increasing until most nodes of the network are infected and the number of the infected nodes stays at a fixed value at last. In Fig. 2(d), with t being increased to 140s, the number of infected nodes increases to 1000 and the entire network is infected.

Mathematical evaluations show that the virus will propagate throughout the entire network as it spreads from the source to other communities. The number of infected nodes increases exponentially with the epidemic time, which means a rapid propagation. The growth of the number of infected nodes will go on after the exponential phase ends. Multiple infectious centers arise due to random bonds, which accelerate the virus propagation. Meanwhile, structures of networks affect the rate and extent of spreading of sensor viruses as well as the infection probability and the shortcut probability. The evaluation results conform to the conclusions based on the classical simple epidemic model, in which the complex architecture of the network is deemed as a random graph.

B. Real Architecture Evaluations of Virus Propagations

Real architecture evaluations of virus propagations are presented in this section. The evaluations are tested with Zigbee FFDs, which construct a community-based network structure over a designated deployment area if some properties of routing layer primitives are modified to achieve application requirements. In evaluations, by default, 1000 nodes are randomly distributed in a deployment area of X×Y=1000m×1000m. The transmission range of the node is 200m. A community-based network structure, in which the inner connection probability δ_{in} of the community is 0.9 and the external connection probability δ_{out} is 0.3, is constructed for the unbalanced deployment of the nodes. We define that maximum 15 nodes are distributed in a community to reduce the energy burden of the gateway and balance the energy consumption of the network. Providing that node i is susceptible, it will be infected with the infection probability h if it is connected to an infectious neighbor in the experiments. The tested virus is a sensor worm which may inject a piece of malicious code into a sensor node with the Von Neumann architecture and destroy the software or the hardware if the program does not perform careful boundary checks. The node will die if the buffer overflow problem occurs. We check how the total number of infected nodes $I(t)$ changes with the epidemic time t and the infection probability h.

We check how the total number of infected nodes $I(t)$ changes with the epidemic time t at h=0.6, and the result is presented in Fig. 3(a). As shown in the figure, $I(t)$ in the evaluation increases exponentially with the epidemic time t as the virus spreads on the community-based network. The exponential propagation slows down due to the reduction of remaining susceptible nodes in the later phase, and the exponential curve comes to a decline at t=120s. In this case, the sensor virus can propagate throughout the entire network of 1000 nodes within 150s. The total number of infected nodes increases until all nodes are infected as the virus attacks the network from one side to another. Structures of networks affect the rate and extent of spreading of the virus. The virus spreads from one community to another along regular bonds and shortcuts, and multiple virus sources which speed up the propagation arise in the hybrid network.

The effect of the infection probability on the propagation is tested in the further evaluation. We check how the total number of infected nodes $I(t)$ changes with the epidemic time t under the condition h_1=0.5, h_2=0.7 and h_3=0.9. The evaluation result is presented in Fig. 3(b). As shown in the figure, the disparity of infection probabilities does not change the exponential form of the propagation. $I(t)$ in the evaluation increases exponentially with the epidemic time t as the virus spreads on the community-based network.

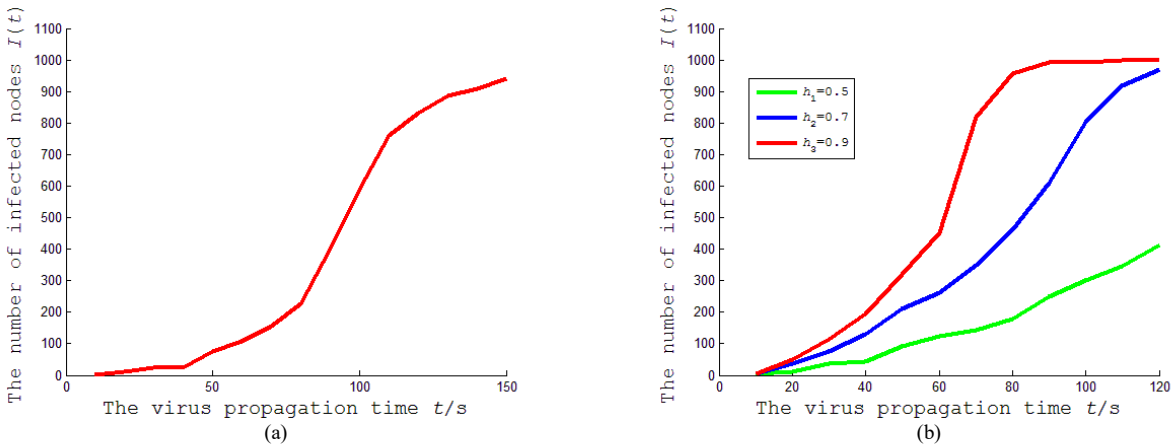


Figure 3. Virus propagations on the community of Zigbee FFDs.

Fig. 3(b) shows that the enhancement of the infection probability speeds up the propagation. The larger the infection probability is, the higher the speed of prevalence is. A larger infection probability means more infectious neighbors around the susceptible node. The total number of infected nodes $I(t)$ increases rapidly from 1 to 1000 under h_3=0.9, and all nodes are infected at t=100s. The number of infected nodes increases more slowly under h_2=0.7, and 805 nodes are infected at t=100s. The number of infected nodes increases most slowly under h_1=0.5, and 312 nodes are infected at t=100s.

VI. CONCLUSIONS

The sensor node which is fragile to virus attacks does not have a complicated hardware architecture or operating system to protect the communication safety because of cost and resource constraints. In this paper we not only propose a novel community-based model which consists of a regular bottom structure and random bonds, but also present the mathematical analysis of virus attacking behaviors based on the proposed model. Our analytical results show that total number of infected nodes of the network $I(t)$ increases exponentially with the epidemic time t when the virus spreads over the network. The growth of the number of infected nodes will go on after the exponential phase ends. The total number of infected nodes converges towards a fixed value as the time increases. We also analyze the impact of structures of networks, the infection probability and the shortcut probability on the rate and extent of spreading of viruses. The virus spreads from one community to another along regular bonds and shortcuts, and multiple virus sources which speed up the propagation arise in the hybrid network. Mathematical evaluations and real architecture evaluations are presented to validate our analytical results. The conclusions provide a guidance for the survivability of wireless sensor networks under virus attacks. In the future, we will investigate the immunization of sensor networks.

REFERENCES

[1] Z. G. Sheng, S. S. Yang, Y. F. Yu, A. V. Vasilakos, J. A. Mccann, and K. K. Leung, "A survey on the IETF protocol suite for the Internet of Things: standards, challenges, and opportunities," IEEE Wireless Communications, vol. 13, Dec. 2013, pp. 91-98.

[2] R. A. M. Khan and H. Karl, "MAC protocols for cooperative diversity in wireless LANs and wireless sensor networks," IEEE Communications Surveys and Tutorials, vol. 16, no. 1, Mar. 2014, pp. 46-63.

[3] A. R. Pinto, C. Montez, G. Araújo, F. Vasques, and P. Portugal, "An approach to implement data fusion techniques in wireless sensor networks using genetic machine leaning algorithms," Information Fusion, vol. 15, Jan. 2014, pp. 90-101.

[4] L. Nachabe, M. G. Genet, and B. E. Hassan, "Unified data model for wireless sensor network," IEEE Sensors Journal, vol. 15, no. 7, Jan. 2015, pp. 3657-3667.

[5] S. Roy, M. Conti, S. Setia, and S. Jajodia, "Secure data aggregation in wireless sensor networks: filtering out the attacker's impact," IEEE Transactions on Information Forensics and Security, vol. 9, no. 4, Apr. 2014, pp. 681-694.

[6] Y. Yang, S. C. Zhu, and G. H. Cao, "Improving sensor network immunity under worm attacks: a software diversity approach," Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MobiHoc' 08), May 2008, pp.149-158.

[7] C. Nowzari, V. M. Preciado, and G. J. Pappas, "Analysis and control of epidemics: a survey of spreading process on complex networks," IEEE Control Systems, vol. 36, no. 1, Jan. 2016, pp. 26-46.

[8] X. Wang, W. Ni, K. F. Zheng, R. P. Liu, and X. X. Niu, "Virus propagation modeling and convergence analysis in large-scale networks," IEEE Transactions on Information Forensics and Security, vol. 11, no. 10, Jun. 2016, pp. 2241-2254.

[9] P. S. Romualdo, C. Castellano, P. V. Mieghem, and A. Vespignani, "Epidemic processes in complex networks," Reviews of Modern Physics, vol. 87, no. 3, Jul. 2015, pp. 925-979.

[10] B. K. Mishra and N. Keshri, "Mathematical model on the transmission of worms in wireless sensor network," Applied Mathematical Modelling, vol. 37, no. 6, Mar. 2013, pp. 4103-4111.

[11] M. E. J. Newman and T. P. Peixoto, "Generalized communities in networks," Physical Review Letters, vol. 115, no. 8, Aug. 2015, pp. 088701-1-5.

[12] F. Wei, M. E. Haque, X. D. Lu, and K. Mori, "Autonomous community construction and coordination technology to achieve real-time transmission in multiple emergencies' situation," Telecommunication Systems, vol. 54, no. 1, Sep. 2013, pp. 61-78.

[13] J. Y. Wu, X. Y. Shao, and H. P. Zhu, "A novel clustering routing protocol with community structure detection for wireless sensor networks," Applied Mechanics and Materials, vol. 472, Jan. 2014, pp. 460-465.

[14] T. Y. Chuang and K. C. Chen, "Information centric sensor network management via community structure," IEEE Communications Letters, vol. 19, no. 5, May 2015, pp.767-770.

[15] B. Wang, L. Cao, H. Suzuki, and K. Aihara, "Impacts of clustering on interacting epidemics," Journal of Theoretical Biology, vol. 304, Jul. 2012, pp. 121-130.

[16] C. H. Li, C. C. Tsai, and S. Y. Yang, "Analysis of epidemic spreading of an SIRS model in complex heterogeneous networks," Communications in Nonlinear Science and Numerical Simulation, vol. 19, no. 4, Apr. 2014, pp. 1042-1054.

[17] J. W. Duncan, Small Worlds, the Dynamics of Networks between Order and Randomness. Princeton: Princeton University Press, Sep. 1999.

[18] E. Estrada, "Introduction to complex networks: structure and dynamics," Evolutionary Equations with Applications in Natural Sciences, vol. 2126, Oct. 2014, pp. 93-131.

[19] J. P. Bagrow, "Communities and bottlenecks: trees and treelike networks have high modularity," Physical Review E, vol. 85, no. 6, Jun. 2012, pp. 066118-1-9.

[20] J. Sanz, L. M. Floría, and Y. Moreno, "Spreading of persistent infections in heterogeneous populations," Physical Review E, vol. 81, no. 5, Mar. 2010, pp. 056108-1-9.

[21] S. Pei and H. A Makse, "Spreading dynamics in complex networks," Journal of Statistical Mechanics: Theory and Experiment, vol. 12, Dec. 2013, pp. 12002-1-21.

[22] S. Dietrich and A. Ammon, Introduction to Percolation Theory, 2nd ed., Oxford: Taylor & Francis, Jul. 1994.

[23] T. J. Norman, The Mathematical Theory of Infectious Diseases, 2nd ed., New York: Hafner Press, Sep. 1975.

[24] C. Pham, "Communication performances of IEEE 802.15.4 wireless sensor motes for data-intensive applications: a comparison of WaspMote, Arduino MEGA, TelosB, MicaZ and iMote2 for image surveillance," Journal of Network and Computer Applications, vol. 46, Nov. 2014, pp. 48-59.