

A SEIR Model Epidemic of Virus on the Online Social Network

Yang Min, Li Qianmu, Song Yaoliang
School of Electronic and Optical Engineering
Nanjing University of Science and Engineering. China
ym8670435@126.com



ABSTRACT: Vigorous development of computer technology has made computers play an integral role in people's lives, but at the same time propagation of the virus on a computer network also cost a great loss to the people. So the research on the mechanism of computer virus propagation and pointing out the key factors in the propagation of the virus has a significant influence on the prevention of computer virus and can also promote the development of relevant policies. The paper researched the influence of the user re-login frequency, the average number of friends on the user's friend list and the initial spread rate of the virus on the social networks during the epidemic of the propagation of the computer virus. By the means of mathematical analysis, an SEIR model to describe the epidemic of the virus on the online social network is established in this paper. The research shows that the re-login frequency of the user and the average friend number of the user would make a significant influence on the virus propagation. The two factors enhance the risk of a virus outbreak in the social network, at the same time.

Subject Categories and Descriptors

I.2.10 [Information Security]: Virus Analysis; **I.4.10 [Model Analysis]**

General Terms: Social Networks, Computer Virus

Keywords: Complex Network, Social Network, Virus Propagation, Epidemic Spreading, Transmission Factor

Received: 11 November 2013, Revised 13 December 2013, Accepted 19 January 2014

1. Introduction

There are a lot of ways that the spread of computer viruses on the Internet, such as email or the sharing of Peer - to - Peer file etc. In recent years, as the number of the service (SNS, hereinafter referred to as the social network) users are growing at a very fast rate; it also provides a new route of spread of the virus. Facebook, the number of Twitter and domestic renren and weibo or other social network 'users are increasing; social network is gradually changing people's life. People can communicate and share files on social networks, even interacting with friends through other platforms that the social networks provide. These convenient services, however, also give the virus more opportunities to infect users. Once a user is infected with the virus, the virus will be automatically sent to the other users who are on this user's friends list. Virus on the Social networks can be dangerous links or videos, etc. As a result, the speed of the spread of virus on the social network is very fast. At the same time, most of the social network users upload their personal information, so the viruses of the social network are more likely to affect the user's privacy and personal safety.

In view of the virus propagation on social networks, some scholars have already carried on the related research. Reference [1] studied virus propagation model based on the email, assuming that the viruses spread when the users click the E-mail attachments. Experiments have proved that if the users see their email in irregular time, the virus will spread faster. Kominos considered that user's behavior will change over time, such as the probability of opening old email attachments is small, as a result he

proposed IM and P2P worm propagation model which spread via email. After studying the propagation law of viruses on Facebook, Fan proposed two models. One is based on the application platform on Facebook and another one is based on sending messages to friends.

But the models above have some limitations. First of all, the viruses that spread on the online social network are more complicated than the viruses that spread by email and IM Network. Secondly, users have their own behaviors, such as online time. Most users are not always online. As a result, user's online status has a crucial effect on the spread of the virus. So the online social network virus propagation model must consider a key factor —user's login interval.

Anderson and May add status *E* to the SIR virus propagation model, and put forward a SEIR model [2]. SEIR model is more detailed and true when it simulate the spread of the virus.

In this paper, the main works are as follows.

- (1) Explained the propagation theory of virus spreading on the social networking [3].
- (2) Put forward SEIR virus propagation model based on online social networks.

After have analyzed the interval of logging, the number of friends in the users' friends list and the influence of the virus's initial spreading rate during the propagation process of the virus, the simulation experiment has proved the effectiveness of this model.

2. Description of the model

The structure of XWLB is firstly analyzed first and then story type is concluded in this section.

2.1 Virus Propagation Mechanisms of the Model

Researchers have found that the topology of the famous social networking such as Facebook is undirected graph structure. The topology of undirected networks can be shown by forms. Every node in the form stands for a user in the Internet. If node *i* and node *j* have a connection side, they have each other in their list of friends [4]. The model defines that every node's degree is the number of users in the list of friends of the node.

In this passage, the model divides all the nodes into 4 statuses: Susceptible node *S*, infected node *I*, immune node *R*, and exposure node *E*. Among them, *S* status shows the node is healthy, but is likely to be infected by a virus [5]. *I* status shows that the node has been infected and it has infectiousness. *E* status shows that the node has been infected, but it doesn't have infectiousness. [6] *R* status shows that the node is immune to the virus. The laws of virus spread in the model are as follows:

- (1) If a node *S* contact with an infected node *I*, it will turn

into status *E* with probability *P*.

- (2) If a node of status *E* doesn't touch any other nodes, it will turn into status *I* with probability ϵ .

- (3) A node of status *I* will not spread the virus all the time. A node *I* will turn into a node of status *R* with probability γ under the condition of it not touching other nodes.

The transition of the node' status is as follows:

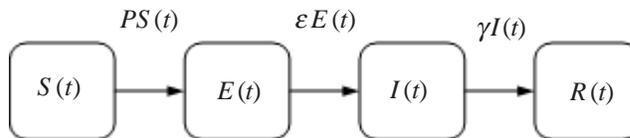


Figure 1. State transition diagram of Node State during virus propagation process

2.2 Mathematical Model of the Virus during Propagation Process of the Model

In the Model, we assume that the number of network nodes is $N(t)$ at moment *t*, that the number of nodes in four states is $S(t)$, $E(t)$, $I(t)$ and $R(t)$ respectively, we can get:

$$S(t) + E(t) + I(t) + R(t) = N(t) \quad (1)$$

Assuming that node *j* at moment is status *S*, P_{ss}^j is the probability for node *j* keeps status *s* at moment $[t, t + \Delta t]$ P_{se}^j is the probability for node *j* turn into status *E* from status, we can get: $P_{ss}^j + P_{se}^j = 1$.

As we have defined the spread law of the virus above, the probability for node in status *S* turning into status *E* is *P*, we get $P_{se}^j = \Delta t P(t)$

Thereinto, $P(t) = \alpha\beta \frac{I(t)}{N(t)}$, α represents the probability for a node in status *I* turning into status *S* (α is approximately equal to 1 in the model), β represents the node's degree, $\frac{I(t)}{N(t)}$ represents Infected nodes account for the proportion of all nodes in the network at moment *t*. We can obtain by formula 1 and 2:

$$P_{se}^j = \Delta t \beta \frac{I(t)}{N(t)} \quad (3)$$

By formula 3, we can get:

$$P_{ss}^j = 1 - \Delta t \beta \frac{I(t)}{N(t)} \quad (4)$$

If node *j* is in status *E* at moment *E*, we define that P_{ee}^j is the probability for node *j* keeping status *E* at moment $[t, t + \Delta t]$ that P_{ei}^j represents the probability for node *j* turning into status

I at moment $[t + \Delta t]$. Similarly, formulas can be obtained as follows:

$$P_{ei}^j = \Delta t \epsilon \quad (5)$$

$$P_{ee}^j = 1 - P_{ei}^j = 1 - \Delta t \epsilon \quad (6)$$

Similarly, if the node j is in state l at moment t , we define that P_{ee}^j is the probability for node j keeping status E at $[t, t + \Delta t]$ moment, that P_{ei}^j is the probability for node j turning into status E at moment $[t + \Delta t]$. Similarly, formulas can be obtained as follows:

$$P_{ir}^j = \Delta t \gamma \quad (7)$$

$$P_{ii}^j = 1 - P_{ir}^j = 1 - \Delta t \gamma \quad (8)$$

From the discussion above, we can get the formula for node in status E at moment;

$$S [t + \Delta t] = S (t) - S (t) P_{se}^j = S (t) - S (t) \Delta t \beta \frac{I(t)}{N(t)} \quad (9)$$

Similarly, we can respectively get formulas for the node in status E/I/R at moment $[t + \Delta t]$:

$$E [t + \Delta t] = E (t) + S (t) P_{se} - E (t) P_{ei} = E (t) + S (t) \Delta t \beta \frac{I(t)}{N(t)} - E (t) \Delta t \varepsilon \quad (10)$$

$$I [t + \Delta t] = I (t) + E (t) P_{ei} - I (t) P_{ir} = I (t) + E (t) \Delta t \varepsilon - E (t) \Delta t \gamma \quad (11)$$

$$R [t + \Delta t] = R (t) + I (t) P_{ir} = R (t) + I (t) \Delta t \gamma \quad (12)$$

We can get from formula 9:

$$\frac{S [t + \Delta t] - S (t)}{\Delta t} = -\beta \frac{S (t) I (t)}{N (t)} \quad (13)$$

In the formula 13, when $\Delta t \rightarrow 0$, we can get:

$$S' = -\beta \frac{S (t) I (t)}{N (t)} \quad (14)$$

Similarly, we can get from formula 10, 11 and 12:

$$E' = \beta \frac{I(t)}{N(t)} S(t) - \varepsilon E(t) \quad (15)$$

$$I' = \varepsilon E(t) - \gamma I(t) \quad (16)$$

$$R' = \gamma I(t) \quad (17)$$

From formula 14 to 17, we can get the rate of the number of 4 kinds of SEIR nodes.

3. The design and analysis of simulation experiment

Through the simulation experiment, this article study the time interval of users login social network; the number of users' friends list and initial spread rate of viruses- " λ " these three factors which affect the spread of the virus. Set up the cellular automata model $T(C, Q, V, f)$: the cellular space T take quadrilateral element cell space and cell number is 10000; set cellular automata state set $Q = \{S, E, I, R\}$; cellular neighbor number is the average number of users' friends in the social network- " β ", according to the C Jin [3] and other researches, this experiment under the condition of standard make β value of 16 and cellular type

boundary take cycle; cellular evolution, such as the virus propagation law described in section 2.1, the evolvement of mathematical has been described in section 2.2.

Through the numerical simulation, this section validates this problem of the SEIR virus propagation model we propose. In the simulation process, the following three conditions which affect the spread of the virus are mainly considered:

- (1)The average- " β " of degree in social network and the average number of users' friends in the network
- (2)The frequency of users who access to social network- " ε "
- (3)The initial transmission rates- " λ "

Other parameters selection are as follows: $N = 10000$, $E(0) = R(0) = 0$, at the same time, assume that the outbreak time is short, and there aren't new or disappeared social network nodes.

Experiment 1. The change of the number of network nodes in the standard case

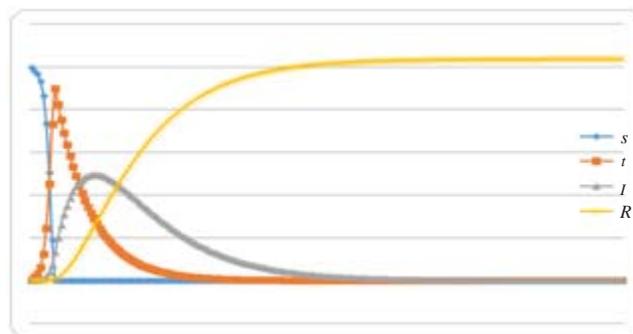


Figure 2. Shows the change graph of the number of nodes of SEIR four states over time. Among them $t \in [0, 200]$, $\varepsilon = 1/12$, $\beta = 16$, $\lambda = I(0) / N(0) = 10 / 10000$.

As we can see from figure 2, the number of nodes of Immune status (R) in the social network grows rapidly. After steady state is reached in the course of virus spread, the number of node R approaches that of node N. The number of node in the infection status (I) grows rapidly when the virus propagation begins. And it slows down when the peak is reached. It approaches 0 in the steady state. The number of the nodes of Exposure status grows very rapidly when the virus propagation begins, and it contracts at once when the peak is reached. Finally, it approaches 0. While the nodes of susceptible status (s) decrease continuously and the number of them approaches 0 at last.

Experiment 2: The change of the number of nodes infected in the immune status under the condition of different people logging at intervals.

Picture 3 Infection nodes in social network as the change of users' access to different frequency curve.

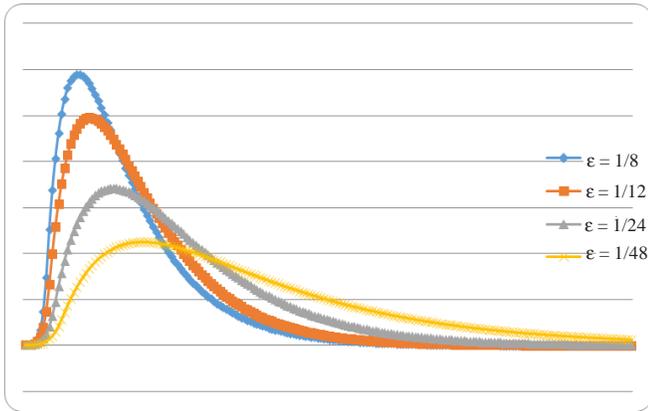


Figure 3. It shows the curve of the number of node I in the social network with different access frequency. $t \in [0, 200)$, $1/\varepsilon = 8, 1/\varepsilon = 12, 1/\varepsilon = 24, 1/\varepsilon = 48, \beta = 48, \beta = 16, \lambda = 0.001$

It comes from the picture obviously when $1/s$ is lesser, infection nodes number can peak earlier, then decreases quickly, and goes to 0 at last; With the decrease of $1/s$, the rate of I nodes number reaching the maximum value and the 0 is greater. Meanwhile, Picture 3 also shows when β remain the same, the spread rate of the virus in social network decreases as the increase of $1/s$, which is the virus in social network spreads faster as the interval of users reenter being shorter [8].

Experiment 3: The change of infection nodes in social network at different friend number Picture 4 shows that with the difference of friend number, the number change curve of I nodes, which is infected, in social network. In it $t \in [0, 200)$, $1/\varepsilon = 12, \lambda = 0.01, \beta = 8, 16, 24, 35, 48$.

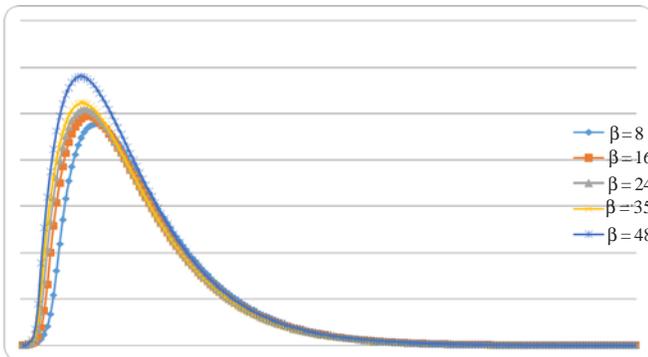


Figure 4. The relation between infection nodes number and friend number in social network

It comes from P4 obviously that the smaller the β is, the longer the time that nodes number in infected status spends in reaching its peak, then the nodes in infected will decrease quickly, and goes to 0 last; meanwhile, with the increase of β , the rate of nodes number in infected reaching peak and almost shrinking to 0 is faster. Combining Ex2, it concludes when $1/s$ remains the same, the virus spreads faster as the increase of β .

Experiment 4: How does the initial spread rate λ of virus affect the infected nodes number in social network.

Figure 5 shows that with the change of initial infection rate λ , the number change curve of I nodes, which is infected, in social network. In it $t \in [0, 200)$, $1/\varepsilon = 12, 1/\varepsilon = 16, \lambda = 0.001, 0.005, 0.01$ and 0.05 .

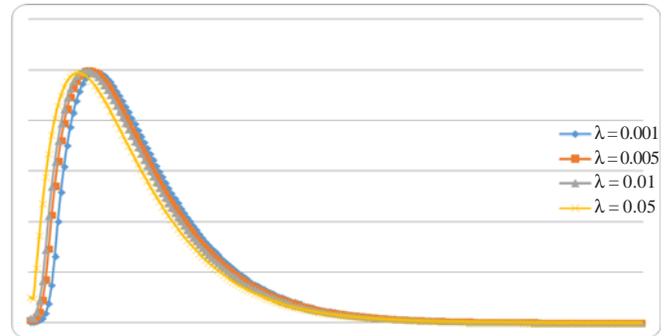


Figure 5. The time change regular pattern of infected nodes in social and the change of virus' initial spread rate λ

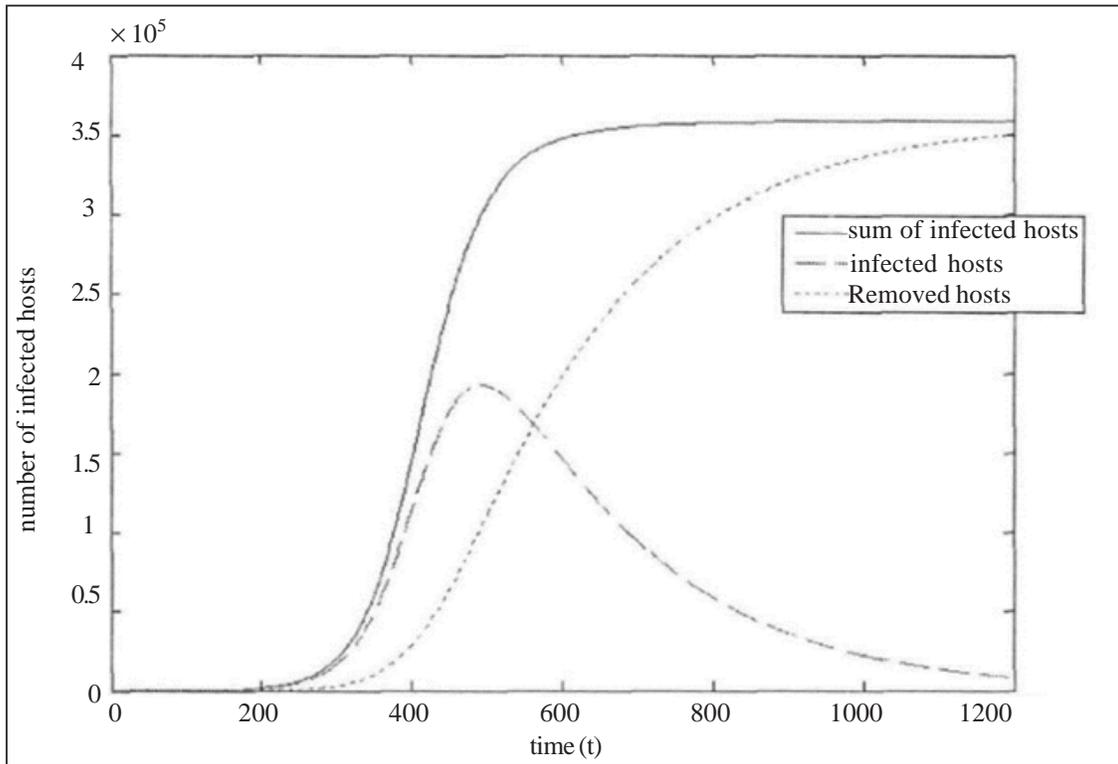
It comes from the P5 with the initial infection rate λ of the virus in social network increasing gradually, the spread rate of virus in initial stage is faster, and the time of reaching its peak is short, and after downing back quickly almost goes to 0 gradually; Meanwhile we can see it that for the smaller average of initial infection rate λ , their peak in the course of spread is very similar.

4. The proving and explain of model's validity

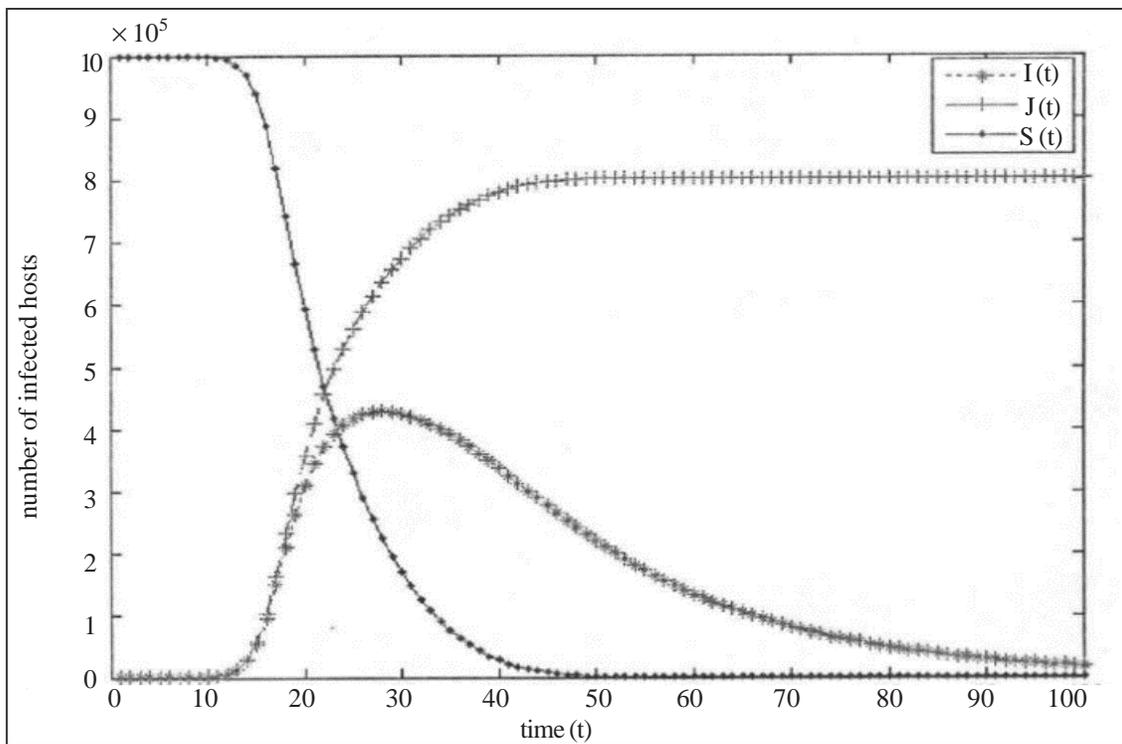
We mostly analysis the presenting SEIR virus spread model in online social network from two sides: ① as the validity analysis of complex network virus spread model; ② as the validity analysis of social network virus spread model.

(1) In the experiment of the third section, it comes the spread course of SEIR model, which this article presents, in social network in standard situation. Meanwhile, in these models this article presenting, selected node degree is put forward by scholars of average degree of each node in the network. Considering from complex network's topology structure side, the nodes of this model have same the node degrees and clustering coefficient, which is Homogeneous network [9]. There are many scholars presenting the classic model of virus in Homogeneous network such as SEM model, KM model and double factor model. Figure 6 is the picture of classic KM spread model and double factor spread model.

It's worth noting that the SIR model's R of the picture (a) is not the Recover status of the models in this article, but is Remove status, which means to remove the nodes from network. We can see from sciences' classic conclusions on Homogeneous network when virus is in spread course, the nodes number in S status is steady in initial stage of spread, then decrease quickly; meanwhile I nodes number of network gently increases in the initial stage, and infects the nodes in network when it reaches the epidemic threshold, and backs down after reaching the peak; When I nodes reach the epidemic threshold, S nodes number



(a)



(b)

Figure 6. KM model and double factor model spread picture

decreases quickly, while R nodes is the status most nodes will be when network becomes balanced status. Comparing the classic experiment's conclusions with simulation experiment 1, we can find that experiment 1 can greatly describe virus' propagation law in rule network, thus verifies the validity of the model.

(2) Due to the spread of the virus in social networks with a high concealment and a long incubation period, therefore it's very difficult to obtain the empirical data on the spread of the virus on the network. Internet data analyses cooperation organization (CAIDA) use the characteristic of Nyxem worm virus successfully infecting after access

to the DNS server to empirically study the spread of the virus on the social network firstly. But by the influence of normal network traffic and other malicious information flow, the statistical result is not absolute accuracy, and only reflects the social network worm's propagation characteristics from trend; Meanwhile, CAIDA's research report points out that the spread of social network virus is the key to the user's safety consciousness and constant attention, so whether can reflect user behavior's influence on social network the spread of the virus can be the important index for evaluation of simulation model is whether in line with the fact [10].

In experiment 2 and experiment 3, we respectively analyze the frequency of users entering social network site and users' friend number these two users behavior's influence on the spread of virus. We can see very clearly from the result of the experiment, the more frequenter frequency users enter their social network, the faster virus spread, and the higher the peak number of infected users is; The more users' friend number in infected social network, the faster virus reach the peak in spread course and the higher the peak number of infected users is. Experiment can well reflect user behavior on the influence of social network the spread of the virus, which is not different from CAIDA's report conclusion, besides it point out the factors that affect the spread of the virus more intuitively and more specifically.

5. Conclusions

This article analyzes the related factor of virus spread in social network by means of building math model, which can actively reflect the effect of network and users in social network spread, meanwhile it also can offer some thought for against virus. It concludes following conclusions from simulation experiment : (1) in social network, the shorter interval time the users enter their social network ID, and the more frequent the frequency is, the faster rate the virus spread is; (2) in social network, the more average friend number the users have, the shorter time the virus spends in reaching peak and backing down in spread course; (3) when the initial infection rate of virus in social network is smaller, the effect of the virus initial infection's rate is greater, the effect of infected nodes' peak in network is smaller.

References

- [1] Zou Changchun, Gong Weibo, Towsley Don. (2007). Modeling and Simulation Study of the Propagation and Defence of internet e-mail worms. *IEEE Transactions on Dependable and Secure Computing*, 4 (2) 105-118.
- [2] Anderson, R., May, R M. (2002). *Infectious Diseases in Humans*. Oxford: Oxford University Press.
- [3] Jin, C., Liu, J. (2009). Network virus propagation model based on effects of removing time and user vigilance. *International Journal of Network Security*, 9 (2) 156-163.
- [4] Li Qianmu, Hou Jun, Qi Yong. (2013). A classification matching and conflict resolution method on meteorological disaster monitoring information. *Disaster Advances*. 6 (1) 415-421.
- [5] Li Qianmu, Zhang Hong. (2012). Information Security Risk Assessment Technology of Cyberspace: a Review. *Information- an International Interdisciplinary Journal*, 15 (11) 4677-4684.
- [6] Li Qianmu. (2011). Multiple QoS Constraints Finding Paths Algorithm in TMN. *Information- an International Interdisciplinary Journal*. 14 (3) 731-738.
- [7] Qianmu Li, Jun Hou, Yong Qi, Hong Zhang. (2012). The Rule Engineer Model on the high-speed processing of Disaster Warning Information. *Disaster Advances*. 5 (4) 432-437.
- [8] Li Qianmu, Li Jia. (2012). Rough Outlier Detection Based Security Risk Analysis Methodology. *China Communications*. 5 (7) 14-21.
- [9] Aldeer, M. M. N. (2010). Performance Comparison of Packet-Level Multiplexing Algorithms with Bursty Traffic. *Journal of Engineering Science and Technology Review*. 3 (9) 46-52.
- [10] Kopitov, R. (2012). Formalization of a Reliable Enterprise Design. *Computer Modelling and New Technologies*. 16 (1) 7-15.