2008－EIP－42 (7)
2008／11／29

On Analysis of Illegal Content Distribution on the Small-World Network

Pao SRIPRASERTSUK † Wataru KAMEYAMA ††
† Global Information and Telecommunication Institute, Waseda University
†† Global Information and Telecommunication Studies, Waseda University

In this paper, our proposed information distribution model is used to analyze illegal contents distribution by conducting simulations based on both one dimensional and two dimensional small-world networks. Our study also show how to use the model to analyze illegal contents distribution. The simulation results show that two-dimensional small-world network is more appropriate to represent practical social network model than one-dimensional small-world network. In addition, our analytical results can be used to effectively analyze illegal contents distribution.

1 Introduction

In general, there are two kinds of contents flows in an information distribution; one is the primary information distribution, and the other is the secondary information distribution. The primary information distribution is the distribution done by providers or broadcasters to consumers through certain kinds of media such as television, newspaper, etc. The secondary information distribution is the distribution done by users to users. As an example of the secondary information distribution, one student copies rental DVD and gives the copy to his or her friends. Today, the advanced information technologies enable the secondary information distribution to be able to perform by various methods and media.

Nevertheless, the secondary information distribution has been also increasing the power of illegal distribution of contents, and contents industries have been suffering from it, particularly by peer-to-peer file sharing (p2p) networks. The studies from [1] has predicted that global peer-to-peer networks will be effectively stopped by legal means. In some developed countries such as England, there is already a strict law against using p2p software.

In addition, there is another way to distribute contents that is the Small-World Network(SWN)[2] or social network. With the advanced information communication technologies and availability of cheap media, storage devices and high bandwidth network, distributing contents in the small-world network has been becoming easier and more powerful than before. Therefore, it is necessary to analyze and control the power of illegal secondary information distribution.

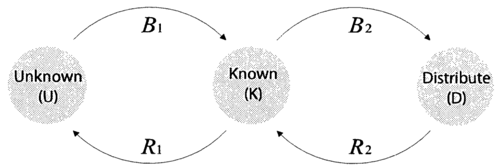

Fig. 1 The Proposed Information Distribution Model.

This paper is organized as follows: in Section 2, summarization of our proposed are described. Section 3 describes the simulation results, and finally Section 4 shows our conclusion and future works.

2 The Proposed Model[3]

2.1 Human Behavior State Model

The illustration of the human behavior state model is shown in Fig. 1 and the definitions are described as below.

- Unknown State (U)
  Individuals in this state do not know information or they do not have contents. They either do not receive information yet or they forget or discard it. In the case of forgetting or discarding information, individuals transit from Known State to this state.

- Known State (K)
  In this state, individuals know information or they have contents but do not have any action to the information distribution.

- Distribute State (D)
  Individuals in this state are active to distribute information or contents. The distribution by their own intentions and other individuals requests are considered as the same.

- Probability of Becoming Known State(B_1)
  This parameter shows the probability of individuals in Unknown State to change into Known State. For example, some individuals in Unknown State are informed with the information or are given with the contents by individuals in Distribute State

- Probability of Becoming Distribute State (B_2)
  This parameter shows the probability of individuals in Known State to change into Distribute State. For example, if individuals in Know State start to distribute the information or contents, they move from Known State to Distribute State.

- Probability of Returning to Unknown State (R_1)
  This parameter shows the probability of individuals in Known State to change into Unknown State. For example, they forget the information or they discard it because they are not interested or the information becomes stale after time passes.

- Probability of Returning to Known State (R_2)
  This parameter shows the probability of individuals in Distribute State to change into Known State. For example, after individuals distributed the information, they may change their mind to stop the action. Thus, their state is changed to Known State.

The use of parameter B_1 is summarized as defined in Eq. (1), where G_i is successful distribution probability of individual i, and n is the number of neighbors of it in Distribute State. For the use of R_1, it is summarized as Eq. (2) where t_p is the time when individual i transit to Know State. In addition, the formula of R_2 and B_2 is shown as Eq. (3).

$$G_i = 1 - (1 - B_1)^n \tag{1}$$

$$R_1(t)_i = 1 - e^{-(\frac{t-t_p}{S})} \tag{2}$$

$$B_2(t), R_2(t) = \begin{cases} \frac{e^{-(t-\mu)^2 \times \lambda_{up}^2}}{\beta} & (t \leq \mu) \\ \frac{e^{-(t-\mu)^2 \times \lambda_{down}^2}}{\beta} & (t > \mu) \end{cases} \tag{3}$$

The detail of each parameters and equations can be found in our previous work[3].

2.2 Analytical Method and Simulation Condition

In the present, the most popularly used small-world network models are WS model[2] and NW model[4]. In this paper, we use the NW model instead of the WS small-world network model, because there is probability for the WS model to be broken into unconnected cluster and the average distance between pairs of nodes on the graph is poorly defined due to the rewiring connection[4].

In our simulation, we apply our model and proposed equations to the one-dimension ring lattice small-world network. We deploy the following two steps to construct a small-world network which has the average degree equal to four because real social networks usually have average coordination numbers significantly higher than two[5]. First, we arrange and connect all nodes, then the network is formed as a regular one-dimension ring lattice model where each node has four connections to the nearest neighbor nodes. Second, we create shortcuts by repeatedly connecting two nodes chosen randomly according to the NW model. Table 1 is the list of parameters used in the simulation.

In our simulation, the initial Known State nodes at the beginning of information distribution ($K(0)$), shortcuts are randomly chosen from all nodes in the network. During the process of simulation, all nodes that are in Distribute State distribute information to their neighbors. And, the probability of nodes in Unknown State

Table 1 Parameters Used in the Simulation.

Parameter	Explanation
$K(0)$	Number of initial Known State individuals at the beginning of information distribution. This parameter can be also considered as the result of the initial primary information distribution done in advance.
k	Average number of connections of each node in the network
SC	Number of shortcuts.

becoming Known State depends on numbers of neighbor that are in Distribute State as shown in Eq. (1). For the process of becoming Distribute State and returning to Unknown State, nodes in the Known State are checked to change their state with the probability of $B_2(t)$ and $R_1(t)_i$, respectively. In the process of returning to Known State, nodes in Distribute State are checked to change their state with the probability of $R_2(t)$.

For $B_2(t)$, we use the values shown in Table 2, and these values produce a graph in which the probability increases rapidly at the beginning of the distribution time and gradually decreases after it reaches the maximum value of probability. For $R_2(t)$, the values in Table 2 generate a graph in which the probability increases gradually. In other words, we assume that, the motivation to distribute this information by the nodes in Known State is rapidly created, but after the motivation reaches to a peak, it gradually decreases. And, the lack of motivation to distribute this information by the nodes in Distribute State gradually increases and becomes stagnant later because there is no stimulation. Graphs of $B_2(t)$ and $R_2(t)$ generated by values in Table 2 are shown in Fig. 2. Other parameters for the simulation are also shown in Table 2. The parameter values in this table are a set as default values. As a note, all nodes and shortcuts are randomly created in the simulation as mentioned above. Therefore, in order to make the result data be reliable, every simulation is conducted 5 times.

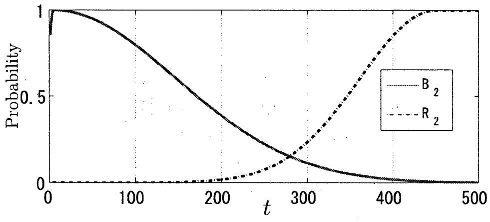

Fig. 2 Graph of $B_2(t)$ and $R_2(t)$ Generated by Values in Table 2.

Table 2 Default Parameter Values.

Parameter	Value
B_1	0.5
$R_1(t)$	Eq. (2) with $S = 100$
$B_2(t)$	Eq. (3) with $\mu = 5$, $\lambda_{up} = 0.1$, $\lambda_{down} = 0.005$, $\beta = 1$
$R_2(t)$	Eq. (3) with $\mu = 450$, $\lambda_{up} = 0.008$, $\lambda_{down} = 0$, $\beta = 1$
$K(0)$	10
k	4
N	10,000
range of t	1 to 500
SC	50

3 Simulation and Consideration of Illegal Contents Distribution

3.1 Distribution Effectiveness Rate

In order to observe the effectiveness of information distribution result, we define a new parameter called Distribution Effectiveness Rate(DER), and its equation is shown in Eq. (4) whereas T is rang of the simulation time. This parameter represents the effectiveness of the distribution result, and the maximum value of DER is one.

$$DER = \frac{1}{T \cdot N} \sum_{t=1}^{T} K(t) + D(t) \tag{4}$$

We define the nodes in the network which have the ability to copy and redistribute contents as Bad nodes, and nodes which don' have the ability to redistribute contents as Constraint nodes. Firstly, we suppose that all nodes in the network are Bad nodes. Subsequently, we conduct simulations by changing Bad nodes to Constraint nodes at a constant number(200) and observing impact

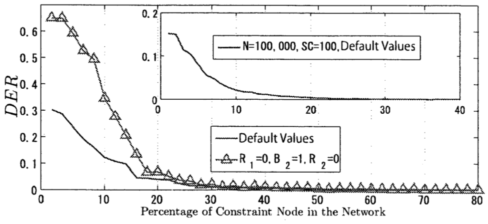

Fig. 3 The Impact of Number of Constraint Nodes to DER.

to DER. For the other necessary parameter values, the values from Table 2 are used. Having conducted the simulations, the result is shown in Fig. 3 whereas y-axis and x-axis are DER and number of Constraint nodes, respectively.

3.2 Impact of Number of Bad Nodes

As seen in Fig. 3, the line graph shows the impact when we use $R_1(t)$,$B_1(t)$,$R_2(t)$ from Table 2 while the triangle line graph is generated by using $R_1(t) = 0$,$B_2(t) = 1$,$R_2(t) = 0$ as static values. In other words, these static values mean all nodes in the network have maximum motivation to distribute the contents. The results show DER is rapidly decreasing until Constraint nodes approximately reach 20%, and then DER is becoming nearly zero. Moreover, we change N to 100,000, $K(0)$ to 100 and SC to 100, and the simulation result is shown in the upper inset of this figure. According to this result, DER is rapidly decreasing until Constraint nodes approximately reach 20%, and then DER is becoming nearly zero. Surprisingly, the result implies that illegal contents distribution in the network which has 4 average degree distribution with a few of shortcuts and 0.1 percent of N as $K(0)$ can be controlled by suppressing the number of Bad nodes less than 80 percent, even though all nodes in the network have maximum motivation to redistribute the contents(triangle ling graph).

3.3 Other Parameters' Impacts

We also conduct the simulation to investigate the impact of SC to illegal contents distribution, and the result is shown in Fig. 4. As seen in this Figure, there is a range of numbers of Constraint nodes that DER is rapidly decreasing in every graphs. We call this range as Quarantine Zone(QZ). For examples, QZ of $SC = 300$, $SC = 1000$, $SC = 2000$, $SC = 3000$ and $SC = 4000$ are estimate has 1000-3000, 2000-4000, 3000-5000, 4000-6000 and 4500-6000 of Constraint Nodes, respectively. The results also imply that the QZ and maximum of DER are slightly changing among the networks which contain very high number of shortcuts. Furthermore, all graphs are decreasing until certain number of Constraint nodes, then DER becomes very small. We call this number as Secure Point(SP). The result also shows SP of the networks which have high SC are not much different. The im-

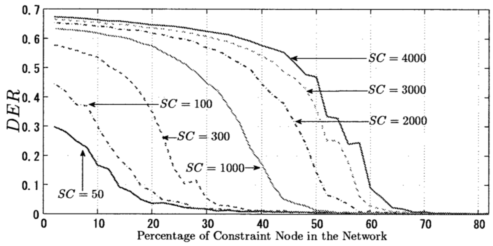

Fig. 4 The Impact of Shortcut(SC).

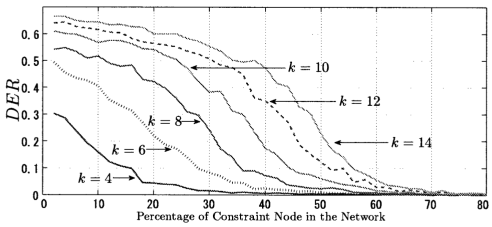

Fig. 5 The Impact of k.

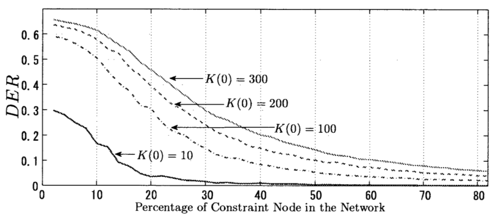

Fig. 6 The Impact of K_0.

pact of k to illegal contents distribution is shown

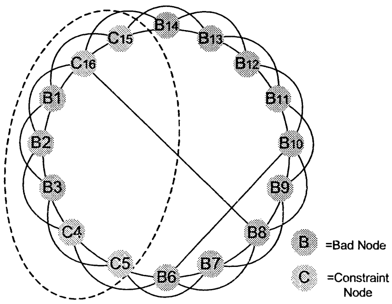

Fig. 7 The Illustration of the Distribution is Confined by Constraint Nodes.

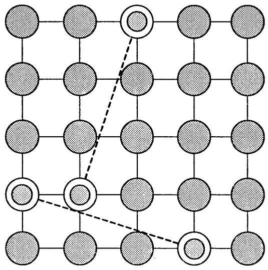

Fig. 8 The Illustration of Two-Dimensional Small World Network.

in Fig. 5. As seen in this figure, the QZ , SP and DER are slightly changing among the networks which have high k. In other words, the gap of SP and DER between $k = 6$ and $k = 8$ are bigger than $k = 12$ and $k = 14$. Furthermore, the impact of $K(0)$ is investigated and the result is shown in Fig. 6. In this figure, all graphs have similar pattern but the maximum value of DER is different. However, these maximum values when $K(0)$ are high are not much different. The results also show that slope of graphs in QZ are decreasing and minimum value of DER is increasing when $K(0)$ is increasing. In other words, if there are more initial illegal distributions, it is more difficult to make DER become very low.

3.4 Consideration on Simulation Results

As seen in Fig. 3, DER is considerably low when number of Constraint nodes is more than 20 percent. This is due to the fact that, the distribution is effectively confined by Constraint nodes when less than 80 percent of Bad nodes are randomly suppressed. The illustration of this phenomenon is shown in Fig. 7. According to this figure, the distribution from "B2" cannot reach to "B6" and "B14" because it is blocked by "C4", "C5" and "C16","C15", respectively. On the other hand, if the network contains many SC or k, it would be difficult to confine the distribution as seen in and Fig. 4 and Fig. 5.

4 Two-Dimension Small World Network and Analytical Results

The models that represent the small world properties are not only limited to one-dimension but also two-dimension such as [6]. Furthermore, it is still questioned that practical human structure network should be represented in one-dimensional or two-dimensional model. Therefore, we also conduct simulation on two-dimensional small world network by construct 100×100 two-dimensional square lattice and also randomly created shortcuts. The illustration of the network is simply shown in Fig. 8.

Having conducted the simulations, the result is shown in Fig. 9. According to this figure, the line graph shows the impact when we use $R_1(t)$,$B_1(t)$,$R_2(t)$ from Table 2 while and triangle line graph is generated by using $R_1(t) = 0$,$B_2(t) = 1$,$R_2(t) = 0$ as the static values. The results show DER is rapidly decreasing until Constraint nodes approximately reach 50%, and then DER is becoming nearly zero. The result implies that illegal contents distribution in two-dimensional small-world network which has 4 average degree distribution with a few of shortcuts and 0.1 percentage of N as $K(0)$ can be controlled by suppressing the number of Bad nodes less than 50 percent, even though all nodes in the network have maximum motivation to redistribute the contents(triangle ling graph). We also conduct the simulation to investigate the impact of SC, k and K_0, and the results are shown in Fig. 10, Fig. 11 and Fig. 12, respectively. As shown in Fig. 10, the patterns of SC impact are similar to Fig. 4 in which the QZ and DER are slightly changing among the networks which contain very high number of shortcuts. For the impact of k shown in Fig. 11, the illegal distribution

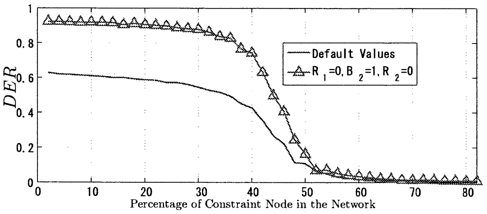

Fig. 9 The Impact of Number of Constraint Nodes to DER in Two-Dimensional Small-World Network.

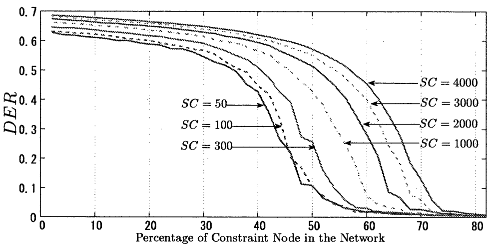

Fig. 10 The Impact of Shortcut(SC)in Two-Dimensional Small-World Network.

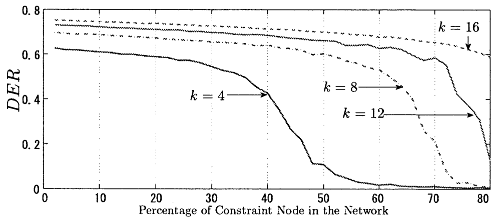

Fig. 11 The Impact of k in Two-Dimensional Small-World Network.

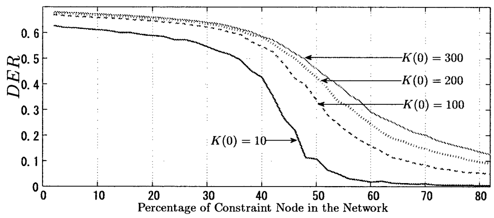

Fig. 12 The Impact of K_0 in Two-Dimensional Small-World Network.

are almost unstoppable when $k \geq 12$ even though the number of constraint nodes are 70%, approximately. We consider that these models can be used to represent the illegal distribution in P2P file sharing networks because the distribution is P2P networks is hardly to stop. The impact of K_0 is shown in Fig. 12. As seen in this figure, obviously slope of graphs in QZ are decreasing and minimum value of DER is increasing when $K(0)$ is increasing.

5 Considering Node Dynamics

In this simulation, we analyze how many number of Bad nodes in the network can effectively give rise to the Illegal Distribution result. In this simulation, we consider there are three kinds of node in the network that are Bad nodes(crackers), Good nodes and Neutral nodes. Bad nodes are nodes which distribute contents illegally, and Good nodes are nodes which distribution contents legally. The distribution which done by Good nodes and Bad nodes are legal distribution and illegal distribution, respectively. For the Neutral nodes, they are nodes which can be transitioned to both Bad nodes and Good nodes. Their transition depends on where they received the contents from. For example, if they received contents from Bad nodes they will become Bad nodes, but if they received content from Good nodes they will become Good nodes.

First, we select randomly initial Good nodes and Bad nodes in the network, and the distribution starts from initial Good nodes as the legal distribution. Except for the selected initial Good nodes and Bad nodes, the remaining nodes in the network are Neutral nodes. Subsequently, we increase number of Bad nodes by changing Neutral nodes to Bad nodes at a constant number(200) and observing impact of Legal Distribution Rate (LDR) and Illegal Distribution Rate (IDR) that are defined as Eq. (5) and Eq. (6), respectively. $G(t)$ is the total number of Good nodes which are in Known and Distribute State at time t, and $B(t)$ is the total number of Bad nodes which are in Known and Distribute at time t. For the other necessary parameter values, the values from Table 2 are used. In this simulation scenario, if the legal distribution reaches to Neutral nodes, these nodes will become Good nodes and redistribute content legally. As the same concept, if the illegal

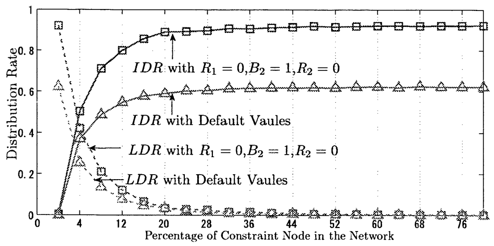

Fig. 13 The Impact of Bad nodes to Legal and Illegal Distribution Results.

distribution reaches to nodes which are not Good nodes, these nodes will become Bad nodes and redistribute content illegally. However, when the legal distribution reach to Bad nodes, the legal distribution will be stopped. And, if those Bad nodes redistribute the received contents, the distribution is considered as the illegal distribution. These processes are repeated during the distribution time.

$$LDR = \frac{1}{T \cdot N} \sum_{t=1}^{T} G(t) \tag{5}$$

$$IDR = \frac{1}{T \cdot N} \sum_{t=1}^{T} B(t) \tag{6}$$

Having conduct the simulation, the result is shown in Fig. 13 where x-axis and y-axis are number of Bad nodes and Distribution Rate, respectively. According to this figure, the triangle solid line(LDR) and triangle dot graph(IDR) show the impact when we use $R_1(t)$,$B_1(t)$,$R_2(t)$ from Table 2 while and square solid line(LDR) and square dot graph(IDR) are generated by using $R_1(t) = 0$,$B_2(t) = 1$,$R_2(t) = 0$ as static values. As seen in Fig. 13, two of LDR is rapidly decreasing while two of IDR is rapidly increasing untill number of Bad nodes reach approximately 20%, and then LDR is becoming nealy zero. The result shows that with a few of initial Good nodes in the network, legal distribution result of contents is rapidly decreasing by small amount of Bad nodes. Nevertheless, this above simulation scenario is just a simple case. We will investigate more realistic scenario and conduct simulations in our future works.

6 Conclusion and Future Works

In this paper, the proposed information distribution model is simulated and analyzed by applying both the one-dimensional and the two-dimensional small-world network to analyze illegal contents distribution.

The results also show that two-dimensional square lattice SWN is more appropriate to represent and analyze real world content distribution on SWN because only 20 percent of Constraint nodes in one-dimensional SWN can stop effectively illegal content distribution, and this kind of phenomenon is rare in the real world. Furthermore, the analytical results can be used to optimize a number of equipped DRM software in the network in order to protect illegal distribution. For example, 50 percent of DRM software are sufficient to confine illegal distribution in the network which have $k = 4$, 0.1 percent of N as $K(0)$ and a few of shortcuts. Moreover, two-dimensional the small-world network with $k \geq 12$ can represent contents distribution on peer-to-peer network because the distribution in those condition is almost unstoppable. This kind of phenomenon occurred in the real world peer-to-peer network.

Our analytical results also shows that with 0.1 percent of N as $K(0)$ but only 20 percent of Bad nodes in the network reduces extremely the legal distribution result. This result implies that even small number of crackers in the network can threaten contents industries. However this simulation results is conducted based on a simple scenario. Therefore, we will include more realistic scenarios and conduct simulations to analyze the results in our future works.

There are remaining issues in our work. For example, we will estimate the appropriate values of parameters in the proposed model for the real world social network. Furthermore, some influencing factors and their impacts for real-world contents distribution will be investigated and analyzed by using the proposed model to analyze and control the illegal contents distribution.

References

[1] P. Biddle, P. England, M. Peinado, and B.Willman, "The Darknet and the Future of Content Distribution", In Proceedings of the 2002 ACM Workshop on Digital Rights Management, 2002.

[2] D.J. Watts and S.H. Strogatz, "Collective dynamics of small world networks , Nature", vol.393, pp.440-442,June 1998.

[3] S.Pao and K.Wataru, "A Proposal of Contents Distribution Model and Analysis of Illegal Contents Distribution on the Small-World Network", FIT 2008, N-202, September 2008.

[4] M.E.J. Newman and D.J Watts, "Renormalization group analysis of the small-world network model", Physics Letters A, vol.263, pp.341-346, December 1999.

[5] M.E.J. Newman and D.J Watts, "Scaling and percolation in the small-world network model", Physical Review E, vol.60, pp.7332-7442, 1999.

[6] J. M. Kleinberg, "The Small-World Phenomenon: An Algorithmic Perspective", Proc.32nd annual ACM symposium on Theory of computing, pp.163-170, June 2000.