# Mathematical epidemiological analysis of dynamics of delay attacks on pull-based competitive information diffusion

Masahiro Sasabe*

*Division of Information Science, Nara Institute of Science and Technology,*
*8916-5 Takayama-cho, Ikoma, Nara 630-0192, Japan*

**Abstract**

In recent years, competitive information diffusion plays a key role in specific network systems. For example, in the cryptocurrency systems, e.g., Bitcoin, some special nodes, called miners, compete with each other in diffusion races to acquire the system rewards. If the information for diffusion is large, pull-based dissemination is often used to alleviate the communication overhead, where a node notifies a small inventory message to its neighbor(s) before transferring the information itself. On receiving the inventory message, the neighboring node requests the corresponding information from the node. Since the information transfer may require much time due to temporal network/node trouble, the pull-based dissemination generally involves a timeout mechanism. If a timeout occurs, the node retries the information retrieval. A malicious node can conduct a delay attack on the information propagation by exploiting this timeout mechanism, i.e., interrupting the information transfer after receiving a request. In this paper, a risk of information diffusion with interruption is considered, under which adversaries colluding with a specific source node simultaneously conduct the delay attacks to slow down the information diffusion from competitive source nodes. With the help of the deterministic nonlinear model for the propagation of infectious diseases in mathematical epidemiology, a scalar SIRA model is first developed, which captures the information diffusion with interruption under a well-mixed population. Furthermore, a network SIRA model is proposed by extending the scalar SIRA model to support the information diffusion with interruption over arbitrary networks. Through numerical evaluations, the risk of information diffusion with interruption is quantitatively revealed from the viewpoint of attack scale, attack rate, recovery rate, and network structure.

*Keywords:* Pull-based competitive information diffusion, dynamics of delay attack, mathematical epidemiology.

## 1. Introduction

Information diffusion is one of the essential functions to achieve information sharing and consensus formation among nodes in most distributed network systems, e.g., cryptocurrency systems [1], Peer-to-Peer (P2P) live streaming [2], time synchronization [3], and social networking [4]. Speediness of the information diffusion affects not only the Quality of Service (QoS), e.g., initial play-out delay in P2P streaming services, but also determines the winner in the competitive information diffusion, e.g., cryptocurrency systems and social networks. For example, in the cryptocurrency systems, e.g., Bitcoin [5], each of some special nodes, called miners, competes with others to acquire a global consensus on its generating information (i.e., a minted block), so as to acquire rewards from the system.

The way of the information diffusion is roughly categorized into two classes: *push-based* and *pull-based*. In the push-based information diffusion, a sending node pushes (sends) the information to a receiving (neighboring) node without confirming the information possession of the receiver. As a result, more than one copy of identical information may reach the same node from different nodes, which consumes much bandwidth with the increase of the information size. The pull-based information diffusion is often used to alleviate such communication overhead, where a sending node first notifies meta information, which only describes the existence of new information and is much smaller than the actual information, to a neighboring node. On receiving the meta information, the neighboring node first checks whether it already has the information or not. If it does not have the information, it requests the corresponding information from the node.

Since the information transfer might require much time due to temporal network/node trouble, a timeout mechanism for the information transfer is generally involved in the pull-based information diffusion. When a node requests information from a neighboring node, it starts a timer with a timeout. If a timeout occurs, the node retries the information retrieval. A malicious node can intentionally delay the information transfer between neighboring nodes by exploiting this timeout mechanism, i.e., interrupting the information transfer after receiving a request [6].

In this paper, a risk of information diffusion with interruption is considered, under which multiple adversary nodes colluding with a specific source node simultaneously conduct the delay attacks on the information diffusion from competitive source

*Corresponding author
*Email address:* m-sasabe@ieee.org (Masahiro Sasabe)

nodes. With the help of the deterministic nonlinear model for the propagation of infectious diseases in mathematical epidemiology, a scalar SIRA model is first developed, which captures the information diffusion with interruption under a well-mixed population. Furthermore, the scalar model is extended to a network SIRA model to support the information diffusion with interruption over arbitrary networks. Through numerical evaluations, the risk of information diffusion with interruption is quantitatively revealed.

The main contributions of this paper are as follows:

1. With the help of mathematical epidemiology, the pull-based competitive information diffusion with interruption is modeled as a continuous-time Markov chain, SIRA model, which captures the state transitions of honest nodes (i.e., initial stand-by, interrupted, recovered stand-by, or information retrieved states) and those of adversary nodes (i.e., initial stand-by or attackable states). The system dynamics can be written by the ordinary differential equations.
2. Since the system dynamics highly depends on the contact patterns among nodes, two kinds of SIRA models are proposed: the scalar SIRA model over a well-mixed population and the network SIRA model over arbitrary networks.
3. The interruption risk of the competitive information diffusion is comprehensively evaluated through numerical evaluations in terms of the following aspects: 1) impact of attack scale, 2) impact of attack rate, 3) impact of recovery rate, 4) impact of contact patterns, and 5) impact of source/adversaries locations in the network.

The rest of the manuscript is organized as follows. Section 2 describes related work. In Section 3, the competitive information diffusion with interruption is introduced. Section 4 provides two kinds of SIRA model: scalar SIRA model and network SIRA model. Numerical results of the scalar SIRA model and network SIRA model are given in Section 5. Finally, Section 6 gives the conclusion.

## 2. Related Work

### *2.1. Competitive Information Diffusion over Networks*

Competitive information diffusion has played a key role in several network systems: cryptocurrency systems [1, 7], social networks [4, 8, 9], and P2P content distribution [2, 10].

In the Bitcoin system, which is one of the cryptocurrency systems, a distributed ledger is maintained by each node in the system, without relying on a central authority, e.g., a government and a bank. The distributed ledger is a chain of blocks (i.e., blockchain), each of which is a collection of transactions and is generated (minted) through competitive computation among miners. To prevent malicious users (miners) from tampering with the blockchain, the Bitcoin system introduces a novel mechanism, called a proof of work, where each miner is required to solve a cryptographic puzzle to generate (mint) a block. The difficulty of the puzzle is automatically adjusted depending on the total computing power of miners in the system such that the average interval of block generation becomes a target value, e.g., 10 [min]. If a miner succeeds in generating a block and adding it to the blockchains maintained by other nodes, it can acquire new coins as a reward from the system. Since the blocks (and transactions) are shared among nodes through broadcasting over the Bitcoin network, which is a P2P network composed of the participating nodes, each miner has to win both the competitive computation and competitive information diffusion.

In social networks, e.g., Facebook and Twitter, information is exchanged or diffused between individuals through the social networks. Since each link between two individuals in the social network arises from the corresponding social relationship, e.g., friendship, interest, and favor, the social networks has been expected to be tools for market predicting, opinion monitoring, and rumor controlling [11]. There may be several competitors aiming to spread their influences over the social networks.

P2P content distribution systems can also be regarded as one of the competitive information diffusion. In the P2P content distribution systems, content files, e.g., software and its security patches and video/audio media, are distributed from distribution servers to clients, called peers, through their cooperative relays over P2P networks [2]. Speedy information diffusion is directly connected to the QoS, e.g., initial play-out delay in the streaming and download time in the file distribution. As a result, some distribution service may have motivation for interrupting/delaying the information distribution of other competing services to degrade their QoS [10].

In what follows, for clarity of explanation, the Bitcoin system is mainly considered as an example of network systems applying the competitive pull-based information diffusion, but the system model in Section 3 and the proposed model in Section 4 are not limited to the case.

### *2.2. Security Risks in Competitive Information Diffusion*

The competitive information diffusion can be cheated by malicious users in both the information generation phase and its diffusion phase. In case of the Bitcoin system, much research effort has been put into the security risks of the first phase based on competitive computation, e.g., 51% attack [12], double spending [13], selfish mining [14]. The technical survey on Bitcoin and its security and privacy issues are summarized in [1, 7], respectively. Since the Bitcoin system relies on the P2P network, which is logically constructed among participating nodes, network aspects of the Bitcoin system has also attracted many researchers [15, 16, 17, 18, 6].

In [15], the authors indicated two vulnerabilities of the Bitcoin network, 1) deanonymization by linking IP addresses to application data and 2) interruption of information flow between nodes. The second risk is highly related to the competitive information diffusion and the following risks were studied: Eclipse attack [17], BGP hijacking [18], and delaying information delivery between neighboring nodes [6].

In Eclipse attack [17], an adversary tries to monopolize all connections to and from a victim node by exploiting the Bitcoin networking protocol and controlling a sufficient number of IP addresses. BGP hijacking [18] focuses on the fact that the

Bitcoin network relies on the underlying IP network, i.e., the Internet, and manipulates (hijacks) the routing data (BGP announcements) of the IP network to attack the information flow over the Bitcoin network. These two attacks require the manipulation of the underling IP network, which may lead to the difficulty of realization. In [6], the authors pointed out the risk of delaying information delivery between neighboring nodes by exploiting the regular timeout mechanism in the Bitcoin protocol.

In this paper, a risk of information diffusion with interruption is considered, where multiple adversaries colluding with a specific source node (miner) simultaneously conduct the delaying attacks to disturb the diffusion of information (block) generated by the competitive source.

### 2.3. Mathematical Epidemiology and Its Application

Information diffusion over networks are similar to infectious (communicable) disease epidemics, which have always been threatening human society. Mathematical epidemiology has been developed to understand the cause of a disease, predict the spread, and control the epidemic [19, 20, 21]. Various types of diseases have been modeled such as SI model, SIS model, and SIR model [19]. In these models, the process of infection (and recovery) is modeled as a continuous-time Markov chain that describes the transitions of the states of the individuals, e.g., susceptible, infected, and removed. As a result, the system dynamics can be written by the ordinary differential equations. There have also been studied on discrete-time models [22] and stochastic models [23]. Since the contact patterns among individuals also affect the infection, the impact of the spatial structure among individuals has also been attracting many researchers [21]. In recent years, the mathematical epidemiological approach has also been applied to the computer networks, e.g., vulnerability analysis of broadcast protocols in wireless sensor networks [24], propagation models of smartphone malware [25], and malware dissemination prevention [26].

The proposed SIRA model, which will be introduced in Section 4, includes the basic diffusion (infection) process as in the conventional models but also has different characteristics. For example, in the infectious disease epidemics and computer virus/malware dissemination, infected individuals/computers become new threats of the infection/attack. On the other hand, in the pull-based information diffusion, the interrupted (infected) nodes are just victims and do not become new sources of infection. Someone may assume that the interruption of the pull-based information diffusion is similar to the prevention of disease infection and malware dissemination. Introducing vaccine against a disease (resp. anti-virus software) can protect the target individual (resp. computer) itself while an adversary in the pull-based information diffusion not only stops its own propagation but also prevents its neighbors from retrieving the information (See Section 3.). As far as I know, there is no mathematical epidemiological model that can deal with the pull-based information diffusion with interruption.

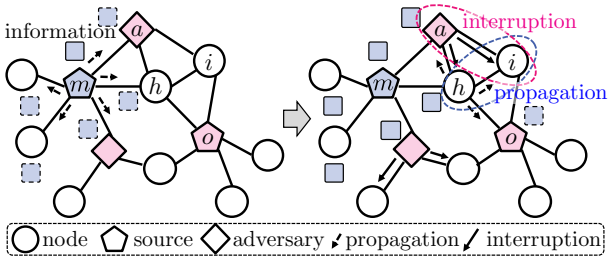
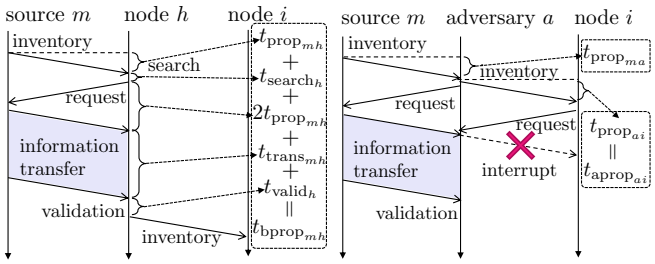



Figure 1: Competitive information diffusion with interruption.




(a) Normal propagation case. (b) Interrupted case.

Figure 2: Pull-based information propagation and its interruption between neighboring nodes.

## 3. Competitive Information Diffusion with Interruption

Fig. 1 illustrates an example of the competitive information diffusion with interruption over a network, where two competitive information source nodes (i.e., sources), two adversary nodes (i.e., adversaries) colluding with source $o$, and eight regular nodes (i.e., nodes) exist. When the source $m$ generates new information, e.g., a block in the Bitcoin system, it first sends the information to its neighboring nodes, as shown in the left side of Fig. 1. Note that two nodes are neighbors when they have a logical link, e.g., TCP connection, between them in case of wired networks and they are in their transmission ranges in case of wireless networks. The information diffusion over the network is achieved by repeating this hop-by-hop information propagation.

If the information size is large, e.g., the block size can be 1 [MB] in case of the Bitcoin system, the pull-based information propagation is applied to alleviate the bandwidth consumption. Fig. 2a illustrates the pull-based information propagation from source $m$ to node $h$. After generating new information, source $m$ first sends an inventory (*inv*) message to its neighbors, e.g., node $h$ and node $a$. The inv message only contains the meta information (e.g., hash value) of the generated information, and thus the size of the inv message is assumed to be negligible. Each receiving neighbor (e.g., node $h$) checks whether it already has the corresponding information. If node $h$ does not have the information, it requests the information from source $m$. After finishing the information retrieval, node $h$ validates the information and further transfers an inv message to its neighbors. Fig. 2a also presents the time required for the one-hop information propagation between source $m$ and node $h$, which is given

by

$$t_{\mathrm{bprop}_{mh}} = 3t_{\mathrm{prop}_{mh}} + t_{\mathrm{search}_h} + t_{\mathrm{trans}_{mh}} + t_{\mathrm{valid}_h}. \quad (1)$$

$t_{\mathrm{prop}_{mh}}$ (resp. $t_{\mathrm{trans}_{mh}}$) is the one-way propagation delay (resp. information transfer delay) between source $m$ and node $h$ while $t_{\mathrm{search}_h}$ (resp. $t_{\mathrm{valid}_h}$) is the local information search time (resp. information validation time) of node $h$.

In this one-hop information propagation, the information transfer may consume much time, due to temporal network/node troubles. To cope with such blocking problems, the pull-based information distribution typically adopts a timeout mechanism. In the timeout mechanism, a node sets a timer with a time $t_{\mathrm{TO}}$ just after requesting the information. When a timeout occurs, the node retries the information retrieval. However, it has been reported that this pull-based information propagation can be delayed by exploiting the timeout mechanism [6].

In Fig. 1, we observe that node $i$ has three neighbors (i.e., node $h$, adversary $a$, and source $o$) and node $h$ and adversary $a$ are the neighbors of source $m$. As a result, node $i$ may receive inv messages from both adversary $a$ and node $h$. In such cases, it is rational for node $i$ to send a request to a node from which it first received the inv message for the speedy information propagation, as in the Bitcoin protocol. In addition, it is also rational for each node to receive information from one node at the same time, so as to save the bandwidth consumption.

Fig. 2b illustrates how adversary $a$ delays (interrupts) the information propagation from source $m$ through node $h$ to node $i$. In the original delay attack [6], the authors considered that an adversary could increase the attack opportunity by skipping the validation process after the information retrieval. In this paper, we focus on the fact that the adversary can further increase the attack opportunity by immediately forwarding the incoming inv message to neighbors, which is known as pipelining information propagation [16]. In Fig. 2b, adversary $a$ immediately forwards the incoming inv message from source $m$ to the neighboring node $i$. The time required for the one-hop inventory propagation from source $m$ to adversary $a$ is equal to the one-hop propagation delay between them, i.e., $t_{\mathrm{prop}_{ma}}$, which is the waiting time for adversary $a$ to start the delay attack to node $i$. On the other hand, the honest node $h$ must wait for the normal information propagation delay $t_{\mathrm{bprop}_{mh}}$ to send the inv message to node $i$ as mentioned above. Since $t_{\mathrm{prop}_{ma}}$ tends to be much smaller than $t_{\mathrm{bprop}_{mh}}$, adversary $a$ can increase the attack opportunity by becoming the first node sending the inv message to node $i$. Therefore, adversary $a$ can absorb the information requests from the neighboring nodes. The time required for adversary $a$ to send an attack, i.e., inv message, to node $i$ is given by

$$t_{\mathrm{aprop}_{ai}} = t_{\mathrm{prop}_{ai}}. \quad (2)$$

In some systems, e.g., the Bitcoin system, the inv message does not include the source of the generated information and adversaries require to retrieve the information itself to judge the source of the generated information. In such cases, adversary $a$ concurrently retrieves the information from source $m$ during the inv forwarding to node $i$, as shown in Fig. 2b. On receiving the request from node $i$, adversary $a$ initially waits for sending the information to node $i$. After retrieving the information, adversary $a$ can recognize the source of the information. If the information is generated by the competitive source, adversary $a$ continues to interrupt the information transfer. Otherwise, it immediately completes the validation process and starts to transfer the information. If adversary $a$ succeeds in the interruption, the victim node $i$ will wait for the information retrieval from adversary $a$ until the timeout occurs. Note that adversary $a$ can also send the information at a low speed to pretend to be an honest node.

In [6], the authors only indicated the risk of the one-hop information propagation with interruption. In this paper, a risk of *information diffusion interruption* is considered, where multiple adversaries colluding with a specific source simultaneously conducts the information delaying attacks over the network.

Table 1: Notations in the model.

| Notation | Definition |
|---|---|
| $S_0$ | The number of nodes at initial standby state $\mathrm{S}_0$ |
| $S_1$ | The number of nodes at recovered standby state $\mathrm{S}_1$ |
| $I$ | The number of nodes at interrupted state I |
| $R$ | The number of nodes at retrieved state R |
| $A$ | The number of adversaries at attackable state A |
| $N$ | The total number of nodes ($N = S_0 + S_1 + I + R + A$) |
| $\alpha$ | The ratio of adversaries to total nodes ($0 \le \alpha \le 1$) |
| $\beta_{0_{ij}}$ | Information retrieval rate from node $i$ to node $j$: $t^{-1}_{\mathrm{bprop}_{ij}}$ |
| $\beta_{1_{ij}}$ | Attack reception rate from node $i$ to node $j$: $t^{-1}_{\mathrm{aprop}_{ij}}$ |
| $\delta$ | Recovery rate: $t^{-1}_{\mathrm{TO}}$ |
| $c$ | Contact rate with other nodes ($0 < c \le 1$) |

## 4. SIRA: Information Propagation with Interruption Model

The information diffusion process given in Section 3 is similar to the epidemic propagation over contact networks. In the field of the mathematical epidemiology, various types of epidemic models have been developed [23, 20, 21]. In this section, with the help of the existing epidemic models, a standby-interrupted-retrieved-attackable (SIRA) model is newly developed to capture the process of the information diffusion with interruption. Table 1 shows notations that will be used.

### *4.1. SIR model*

One of the epidemic models, susceptible-infected-removed (SIR) model [27], is the most similar to the assumed information diffusion. In the SIR model, individuals in a population are divided into three *compartments* (*states*): susceptible (S), infected (I), and removed/recovered (R). Individuals are initially in the state S where they are healthy but susceptible to becoming infected. If the healthy individuals come in contact with infected individuals, they become infected (i.e., move to the state I) with some infection rate. The infected individuals will recover (i.e., move to the state R) with some recovery rate. This state transition is described by a Markov process.

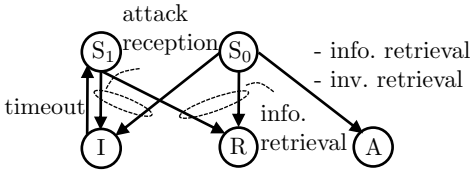



Figure 3: SIRA: Information diffusion with interruption model.

### 4.2. SIRA model

In the competitive information diffusion, each source node has to distribute its generating information as speedy as possible. In this paper, we focus on the situation how the information generated by one source node is distributed over the network under the interruption from adversaries. Suppose a system of $N$ ($N > 0$) nodes consisting of one source node, $(1-\alpha)(N-1)$ honest nodes, and $\alpha(N-1)$ adversary nodes. Fig. 3 illustrates the SIRA model that describes the information diffusion with interruption just after the information generation as a continuous-time Markov chain.

We first focus on the state transitions of honest nodes. Honest nodes start from the initial standby state $S_0$ because they initially do not have the information generated by the source. On retrieving the information from the source or other honest node, they move to the retrieved state R. On the other hand, if they receive attacks (i.e., inv messages) from adversaries, they move to the interrupted state I. The honest nodes at the state I will recover and move to the recovered standby state $S_1$ after a timeout occurs, that is, $t_{\mathrm{TO}}$ passes. The transitions from the state $S_1$ are the same as those from the state $S_0$. Next, we focus on the state transition of adversary nodes. Adversary nodes also start from the initial standby state $S_0$. When they retrieve inv messages from the source or other honest/adversary nodes, they move to the attackable state A where they can start the delay attacks.

In the SIR model, it is assumed that each individual randomly makes $cN$ ($0 < c \leq 1$) contacts with others per unit of time under a well-mixed population and the infection rate (resp. recovery rate) is identical among individuals. In this paper, a scalar SIRA model under the same assumption is first developed. From the viewpoint of network systems, such a random contact scenario may not be realistic. In Section 4.3, the scalar SIRA model is further extended to the network SIRA model where each node can only communicate with its neighboring nodes over a network.

The ordinary differential equations that describe the rate of change of individuals at the states $S_0, S_1, \mathrm{I}, \mathrm{R}$, and A are given as follows:

$$\dot{S}_0 = -S_0 c(\beta_0 R + \beta_1 A), \tag{3}$$
$$\dot{S}_1 = -S_1 c(\beta_0 R + \beta_1 A) + \delta I, \tag{4}$$
$$\dot{I} = ((1-\alpha)S_0 + S_1)\beta_1 cA - \delta I, \tag{5}$$
$$\dot{R} = ((1-\alpha)S_0 + S_1)\beta_0 cR, \tag{6}$$
$$\dot{A} = \alpha S_0 c(\beta_0 R + \beta_1 A). \tag{7}$$

(3) can be interpreted as follows. There are three kinds of transitions from $S_0$: $S_0 \to R, S_0 \to I$, and $S_0 \to A$, as shown in Fig. 3. $S_0 \to R$ occurs per unit of time when $(1-\alpha)S_0$ honest nodes at the state $S_0$ communicate with $cR$ ($= cN \cdot \frac{R}{N}$) honest nodes at the state R and then retrieve the information at the rate $\beta_0$, which is the reciprocal of the normal information propagation delay $t_{\mathrm{bprop}}$:

$$\beta_0 = t_{\mathrm{bprop}}^{-1}. \tag{8}$$

Recall that the information retrieval rate $\beta_0$ is assumed to be identical among each pair of neighbors in the scalar SIRA model. In actual systems, $t_{\mathrm{bprop}}^{(\mathrm{p})}$ may change depending on the pair of neighbors, as in (1). The network SIRA model, which will be presented in Section 4.3, can deal with such complex situations.

Similarly, $S_0 \to I$ happens per unit of time when $(1-\alpha)S_0$ honest nodes at the state $S_0$ communicate with $cA$ adversary nodes at the state A and then are attacked at the rate $\beta_1$, which is the reciprocal of the attack (inv) propagation delay $t_{\mathrm{aprop}}$:

$$\beta_1 = t_{\mathrm{aprop}}^{-1}. \tag{9}$$

Note that $\beta_1$ is also assumed to be identical among each pair of adversary and node (or adversary), and $\beta_1 > \beta_0 > 0$ is satisfied by considering (1), (2), (8), and (9).

$S_0 \to A$ represents the transition of adversaries: It occurs per unit of time when $\alpha S_0$ adversaries at the state $S_0$ communicate with $cR$ honest nodes at the state R or $cA$ adversary nodes at the state A, and then retrieve the information or inv message at the rate of $\beta_0$ or $\beta_1$, respectively. (4)–(7) can be derived similarly. Note that $\delta$ is the recovery rate that is the reciprocal of the timeout value $t_{\mathrm{TO}}$:

$$\delta = t_{\mathrm{TO}}^{-1}. \tag{10}$$

The total number of nodes, $N$, satisfies

$$S_0 + S_1 + I + R + A = N, \tag{11}$$

and it is assumed that the total number of nodes, $N$, does not change. Therefore, four of the above five equations, i.e., (3)–(7), are enough to describe the system dynamics. Note that the scalar SIRA model without adversaries, i.e., $\alpha = 0, \beta_1 = 0$, is equivalent to a model that describes the normal information diffusion where only two states, i.e., $S_0$ and R, and the transition $S_0 \to R$ exist.

Finally, we focus on the convergence property of the scalar SIRA model. From (3) and (7),

$$\frac{dA}{dS_0} = -\alpha$$

is satisfied. Integrating both sides of this equation by $S_0$ yields

$$A = C - \alpha S_0,$$

where $C$ is the integral constant, which can be derived as $C = \alpha(N-1)$ by substituting the initial condition, i.e., $S_0 = N - 1, S_1 = 0, I = 0, R = 1$, and $A = 0$, into this equation. As a result,

$$A = \alpha(N - 1 - S_0) \tag{12}$$

is satisfied. Since the number of nodes at each state is non-negative, (3) indicates that $S_0$ monotonically decreases and eventually converges to zero. From (12), $A$ increases from zero and converges to $\alpha(N-1)$, which is actually equal to the number of adversaries.

From (6), we observe that $R$ monotonically increases. By substituting $S_0 = 0$ into (6), $S_1 = 0$ is required to satisfy $\dot{R} = 0$. In addition, (4) indicates that $S_1 = I = 0$ should be hold to achieve $\dot{S_1} = 0$. As a result, $S_0 = S_1 = I = 0, R = (1-\alpha)(N-1) + 1$, and $A = \alpha(N-1)$ are satisfied at the steady state.

### *4.3. Network SIRA model*

In most network systems, each node has a limited number of neighbors and the structure of the network affects the information diffusion. In this section, the scalar SIRA model is extended to the network SIRA model where each node can only communicate with its neighbors. For example, in Fig. 1, honest node $i$ can only communicate with source $o$, honest node $h$, and adversary $a$. There have been studied on the dynamics of epidemic propagation over networks [28]. With the help of the model in [28], a network SIRA model is proposed as follows. Note that the scalar SIRA model focuses on the population dynamics while the network SIRA model focuses on the state transition of each individual.

Suppose that the information is distributed over a network $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ are the set of nodes, denoted by $\mathcal{V} = \{1, \ldots, N\}$, and that of links, respectively. In this paper, it is assumed that $G$ is undirected and strongly connected. $\mathcal{V}$ consists of one source node, $(1-\alpha)(N-1)$ honest nodes, and $\alpha(N-1)$ adversaries. $\alpha_i$ is introduced to distinguish the node type: If node $i \in \mathcal{V}$ is the source or honest node, $\alpha_i = 0$, and otherwise, $\alpha_i = 1$. Each honest node can be at one of the four states, i.e., initial standby state $\mathrm{S}_0$, recovered standby state $\mathrm{S}_1$, interrupted state I, and retrieved state R. On the other hand, each adversary node can be at one of the two states, i.e., initial standby state $\mathrm{S}_0$ and attackable state A. Let $S_{0_i}$ ($S_{1_i}, I_i, R_i$, and $A_i$ resp.) be the probability that node $i$ is in the initial standby (recovered standby, interrupted, retrieved, and attackable resp.) state at time $t$. Note that $S_{0_i} + S_{1_i} + I_i + R_i + A_i = 1$ ($i \in \mathcal{V}$). As a result, the expected number of nodes at each state in the system can be obtained by the sum of the probabilities of nodes at the corresponding state, i.e., $S_0 = \sum_{i\in\mathcal{V}} S_{0_i}, S_1 = \sum_{i\in\mathcal{V}} S_{1_i}, I = \sum_{i\in\mathcal{V}} I_i, R = \sum_{i\in\mathcal{V}} R_i$, and $A = \sum_{i\in\mathcal{V}} A_i$.

As mentioned in Section 4.2, the network SIRA model can support different information/inventory/attack propagation delay between neighboring nodes $i$ and $j$, depending on their network conditions and computation power. As a result, the information retrieval rate $\beta_{0_{ij}}$ and attack reception rate $\beta_{1_{ij}}$ from node $i$ to node $j$ are defined as follows.

$$\beta_{0_{ij}} = t^{-1}_{\mathrm{bprop}_{ij}}, \quad \beta_{1_{ij}} = t^{-1}_{\mathrm{aprop}_{ij}}, \tag{13}$$

where $t_{\mathrm{bprop}_{ij}}$ and $t_{\mathrm{aprop}_{ij}}$ are given by (1) and (2), respectively. Note that both $\beta_{0_{ij}}$ and $\beta_{1_{ij}}$ become zero if node $i$ and node $j$ do not have the neighboring relationship.

The network SIRA model on the networks $G$ is given by

$$\dot{S}_{0_i} = -S_{0_i}\Big(\sum_{j\in\mathcal{V}} \beta_{0_{ji}} R_j + \sum_{j\in\mathcal{V}} \beta_{1_{ji}} A_j\Big), \tag{14}$$

$$\dot{S}_{1_i} = -S_{1_i}\Big(\sum_{j\in\mathcal{V}} \beta_{0_{ji}} R_j + \sum_{j\in\mathcal{V}} \beta_{1_{ji}} A_j\Big) + \delta I_i, \tag{15}$$

$$\dot{I}_i = ((1-\alpha_i)S_{0_i} + S_{1_i}) \sum_{j\in\mathcal{V}} \beta_{1_{ji}} A_j - \delta I_i, \tag{16}$$

$$\dot{R}_i = ((1-\alpha_i)S_{0_i} + S_{1_i}) \sum_{j\in\mathcal{V}} \beta_{0_{ji}} R_j, \tag{17}$$

$$\dot{A}_i = \alpha_i S_{0_i}\Big(\sum_{j\in\mathcal{V}} \beta_{0_{ji}} R_j + \sum_{j\in\mathcal{V}} \beta_{1_{ji}} A_j\Big). \tag{18}$$

With the definition of $\alpha_i$, the state transitions of honest node $i$ ($\alpha_i = 0$) are given by (14)–(17) while those of adversary nodes $i$ ($\alpha_i = 1$) are given by (14) and (18). (14)–(18) can be derived similarly as in case of the scalar SIRA model. For example, (14) presents the decrease rate of $S_{0_i}$ and is the network version of (3). In the scalar SIRA model, each node can homogeneously make contacts with others, and thus the product of information retrieval rate and contact probability with information holders is given by $c\beta_0 R$ in (3). In the network SIRA model, node $i$ can only communicate with its neighbors. Therefore, node $i$ at the state $\mathrm{S}_0$ can retrieve the information at the rate of $\sum_{j\in\mathcal{V}} \beta_{0_{ji}} R_j$ in (14). Recall that $\beta_{0_{ji}}$ becomes zero if node $j$ is not the neighbor of node $i$.

Finally, we focus on the convergence property of the network SIRA model. The convergence property for adversaries can be derived in a manner similar to that in the scalar SIRA model (Section 3). From (14) and (18),

$$\frac{dA_i}{dS_{0_i}} = -\alpha_i$$

is satisfied. Integrating both sides of this equation by $S_{0_i}$ yields

$$A_i = C_i - \alpha_i S_{0_i}$$

where $C_i$ is the integral constant, which can be obtained as $C_i = \alpha_i$ by substituting the initial condition, i.e., $S_{0_i} = 1$ and $A_i = 0$, into this equation. As a result, we obtain

$$A_i = \alpha_i(1 - S_{0_i}). \tag{19}$$

From (14), $S_{0_i}$ monotonically decreases from one and eventually converges to zero. Since adversary node $i$ has $\alpha_i = 1$, $A_{0_i}$ increases from zero and converges to one from (19). From the viewpoint of the whole system,

$$A = \sum_{i\in\mathcal{V}} A_i = \alpha(N-1) - \sum_{i\in\mathcal{V}} \alpha_i S_{0_i}$$

is satisfied, and thus $A$ starts from zero and finally converges to the number of adversaries, i.e., $\alpha(N-1)$, as in the scalar SIRA model.

On the other hand, the convergence property for honest nodes is affected by the network structure. (17) indicates that honest node $i$ ($\alpha_i = 0$) at the state $\mathrm{S}_{0_i}$ or state $\mathrm{S}_{1_i}$ can retrieve the information from the neighbors at the rate $\sum_{j\in\mathcal{V}} \beta_{0_{ji}} R_j$. As a result,

honest nodes with many neighbors tend to retrieve the information speedily. There may also exist the situation where all the neighbors of honest node $i$ are adversaries, which means that honest node $i$ cannot retrieve the information. Note that the no retrieval risk of honest node $i$ tends to decrease with increase of the neighbors. For example, in case of a $d$–regular connected graph ($d \geq 2$) where each node has exact $d$ neighbors, each of the $d$ neighbors can be an adversary with the probability of $\alpha(N-1)/N$, and thus the probability that all the neighbors of a certain node are adversaries can be expressed by $(\alpha(N-1)/N)^d \simeq \alpha^d$. Above-mentioned heterogeneous retrieval speed among nodes and no retrieval risk only exist in the network SIRA model.

## 5. Numerical Results

In this section, we reveal how the risk of the information diffusion with interruption changes according to the settings of the parameters ($\alpha, \beta_0, \beta_1$, and $\delta$) and the communication patterns among nodes.

### 5.1. Evaluation Settings

We consider that a network consists of one thousand nodes ($N = 1000$), which include one source node, $(1-\alpha)(N-1)$ honest nodes, and $\alpha(N-1)$ adversary nodes. At $t = 0$, the source node generates new information and the information diffusion with interruption starts. In the succeeding evaluations, the default parameters are set as follows. The fraction of adversaries to the total nodes, $\alpha$, is set to be 0.01 because it is important to figure out whether a limited number of adversaries can interrupt the information diffusion.

In the scalar SIRA model, the number of contacts per unit of time, $cN$, is set to be eight. In the network SIRA model, we use two kinds of networks: regular network and scale-free network generated by BA model [29]. The regular (resp. BA-based scale-free) network is used to evaluate how the limitation of communications between nodes affects the degree of the risk under the homogeneous (resp. heterogeneous) network structure. For comparison purpose, the average number of neighbors, i.e., degree, is set to be eight in both network topologies.

To focus on the impact of the topological structure, the one-way propagation delay $t_{\text{prop}_{ij}}$ of a link between neighboring nodes $i$ and $j$, $e_{ij}$ ($i, j \in \mathcal{V}, e_{ij} \in \mathcal{E}$), is identical and set to be 0.1 [s]. From (2) and (13), $\beta_{1_i} = \beta_1 = 10$ [1/s] ($i \in \mathcal{V}$). To focus on how the balance among the one-way propagation delay $t_{\text{prop}_{ij}}$, information transfer delay $t_{\text{trans}_{ij}}$, and timeout $t_{\text{TO}}$ affects the interruption risk, the local information search time $t_{\text{search}_i}$ and the information validation time $t_{\text{valid}_i}$ of node $i$ ($i \in \mathcal{V}$) are assumed to be negligible, i.e., $t_{\text{search}_i} = t_{\text{valid}_i} = 0$. The information transfer delay between neighboring nodes $i$ and node $j$, $t_{\text{trans}_{ij}}$, is obtained by $B/D_{ij}$ where $B$ and $D_{ij}$ are the information size and the transmission rate between them, respectively. For example, the Bitcoin system supports the maximum of 1 [MB] block size, which results in one second transfer delay under 8 [Mbps] transmission rate. To focus on the relative relationship between the information retrieval rate $\beta_0$ and the attack rate $\beta_1$,

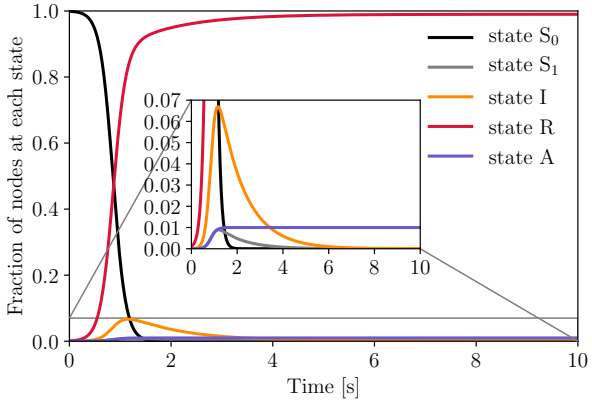



Figure 4: Evolution of scalar SIRA model ($N = 1000, \alpha = 0.01, \beta_0 = 1, \beta_1 = 10, \delta = 1.0$).

$\beta_{0_{ij}} = \beta_0 = 1$ [1/s] ($i, j \in \mathcal{V}$) is used as the default value where $t_{\text{prop}_{ij}} = 0.1$ [s] and $t_{\text{trans}_{ij}} = 0.7$ [s] are substituted into (1) and (13). The timeout value $t_{\text{TO}}$ should be larger than the normal information propagation delay, $t_{\text{bprop}} = 1.0$. The recovery rate $\delta = t_{\text{TO}}^{-1}$ is set to be the upper bound, i.e., 1.0 [1/s].

To solve both the scalar and network SIRA models composed of the ordinary differential equations, scipy.integrate.odeint function of Scipy [30] is used.

### 5.2. Scalar SIRA model

We first focus on the fundamental characteristics of the information diffusion with interruption through numerical results based on the scalar SIRA model.

#### 5.2.1. Fundamental Dynamics

Fig. 4 illustrates the evolution of the scalar SIRA model, i.e., the transition of the fraction of nodes at each state, when $t \in [0, 10]$. To focus on the transition of the fraction of nodes at the state A, the vertically enlarged graph is also given as the inner graph. We observe that the fraction of nodes that retrieve the information, i.e., $R/N$, steeply increases to a certain level, i.e., 0.9, and then gradually converges to the situation where all honest nodes complete the information retrieval. On the other hand, the fraction of the attackable nodes, $A/N$, also shows the steep increase and converges to $\alpha = 0.01$. The fraction of the interrupted nodes, $I/N$, initially increases in response to the increase of $A/N$, and then gradually decreases with the recovery.

#### 5.2.2. Impact of Attack Scale

Fig. 5 depicts the transition of the fraction of nodes at state R when the attack scale $\alpha$ ranges $[0, 0.05]$. If there is no adversary in the system, i.e., $\alpha = 0.00$, the normal information diffusion requires only about two seconds to reach the consensus. We observe that the initial increase speed of $R$ does not almost change, regardless of $\alpha$. However, after the initial increase of $R$, the diffusion speed (resp. ratio) becomes slower (resp. smaller) with the increase of $\alpha$.

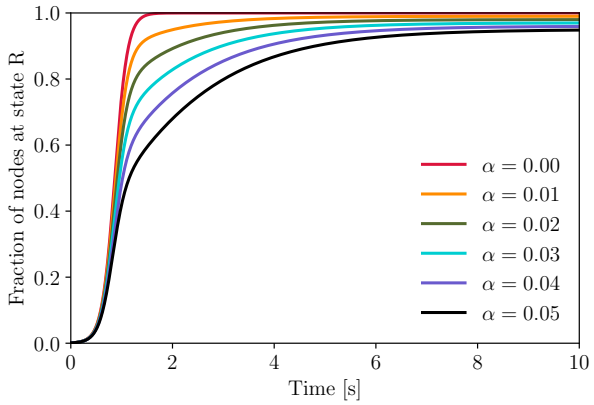



Figure 5: Impact of attack scale $\alpha$ (scalar SIRA model, $N = 1000, \beta_0 = 1, \beta_1 = 10, \delta = 1.0$).

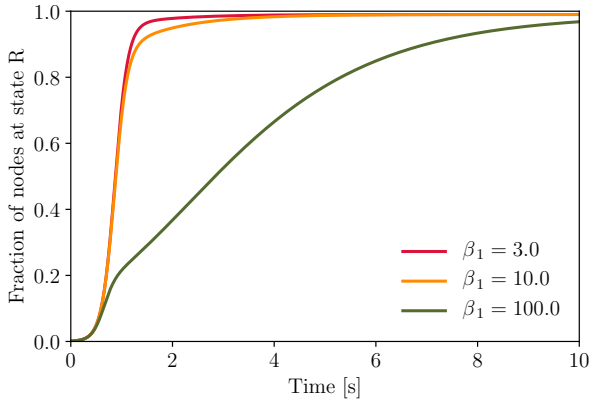



Figure 6: Impact of attack rate $\beta_1$ (scalar SIRA model, $N = 1000, \alpha = 0.01, \beta_0 = 1, \delta = 1.0$).

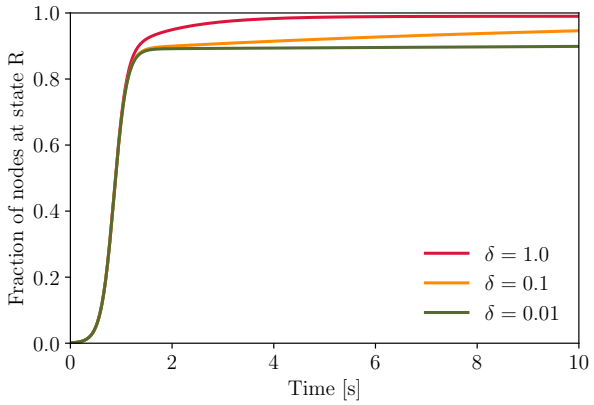



Figure 7: Impact of recovery rate $\delta$ (scalar SIRA model, $N = 1000, \alpha = 0.01, \beta_0 = 1, \beta_1 = 10$).

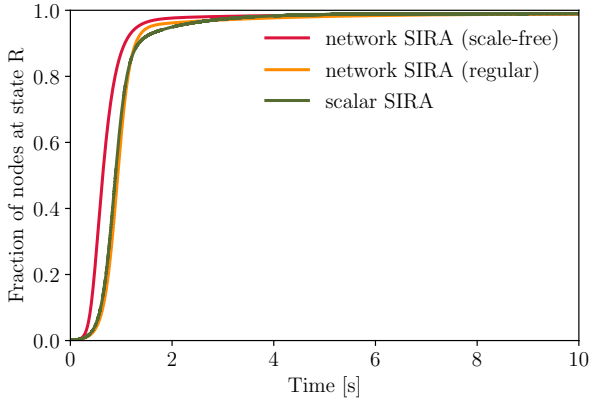



Figure 8: Impact of contact pattern ($N = 1000, \alpha = 0.01, \beta_0 = 1, \beta_1 = 10, \delta = 1.0$).

#### 5.2.3. *Impact of Attack Rate*

Next, we focus on the impact of the attack rate $\beta_1$ on the transition of the fraction $R/N$ of nodes that retrieved the information. Since $\beta_0$ is set to be one, $\beta_1$ can also be regarded as the attack rate normalized by the information retrieval rate. From the definitions of $\beta_0$ and $\beta_1$, which are given by (1), (2), (8), and (9), the normalized attack rate has a lower bound of three, which is achieved when the information size is negligible, i.e., $t_{\text{trans}} = t_{\text{search}} = t_{\text{valid}} = 0$, and increases with the information size.

Fig 6 shows the transition of the fraction of nodes at state R when $\beta_1$ is set to be 3, 10, and 100. As we expected, speed down (resp. speed up) of the attack rate $\beta_1$ decreases (resp. increases) the risk of interruption. Specifically, $\beta_1$ mainly affects the degree of the initial diffusion ratio.

#### 5.2.4. *Impact of Recovery Rate*

The attack rate $\beta_1$ affects how often honest nodes would be interrupted by adversaries. On the other hand, the recovery rate $\delta$ determines how fast the interrupted nodes throw off the interruption. Recall that $\delta$ is the reciprocal of the timeout value $t_{\text{TO}}$ as given by (10). Since the original purpose of the timeout mechanism is disconnecting undesired long information transfer, $t_{\text{TO}}$ should be larger than the information transfer delay $t_{\text{trans}}$, and thus $\delta$ should be less than $\beta_0$.

Fig. 7 illustrates the transition of the fraction of nodes at state R when $\delta$ is set to be 0.01, 0.1, and 1.0. As we expected, we observe that slow (resp. fast) recovery increases (resp. decreases) the risk of interruption. In particular, $\delta$ mainly affects the convergence speed after the initially steep increase.

### 5.3. *Network SIRA model*

Next, we reveal the impact of the network structure on the risk of interruption through the network SIRA model.

#### 5.3.1. *Impact of Contact Pattern among Nodes*

Fig. 8 presents the transition of the fraction of nodes at state R under three kinds of contact patterns among nodes. First one is the random contact pattern under the well-mixed population, which can be analyzed by the scalar SIRA model. Second and third ones are the spatially limited contact patterns over regular networks and BA-based scale-free networks, respectively. These can be analyzed by the network SIRA model. In the second and third ones, the locations of source and adversaries follow a *random* manner where they are randomly located in the networks.

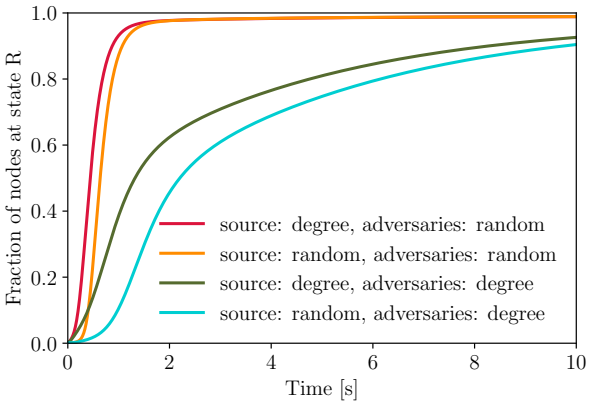



Figure 9: Impact of source/adversaries locations (scale-free graph, $N = 1000, \alpha = 0.01, \beta_0 = 1, \beta_1 = 10, \delta = 1.0$).

We first observe that the network SIRA model over the regular network shows the similar tendency to the scalar SIRA model but can slightly achieve higher initial diffusion ratio, due to the spatial limitation of interruptions from adversaries. We also find that the BA-based scale-free network achieves the fastest information diffusion, with the help of high-degree and honest nodes. However, in case of the scale-free network, the locations of source and adversaries will affect the information diffusion, and thus this point will be deeply analyzed in the next section.

#### 5.3.2. *Impact of Locations of Source and Adversaries in Networks*

Fig. 9 illustrates how the locations of source and adversaries affect the transition of the fraction of nodes at state R when $\alpha = 0.01$. In addition to the random manner, a *degree-based* manner is also used as the location selection of source and adversaries. In the degree method, the source and adversaries are selected from all nodes in descending order of their degrees. In particular, it is rational for malicious users to invade the network so as to maximize the delay effect.

We first focus on the impact of the source location on the information diffusion. As we expected, the degree-based location of the source can improve the initial information diffusion ratio and speed compared with the random location. However, we also confirm that the degree-based locations of the adversaries drastically decrease both the ratio and speed of information diffusion even under the limited fraction of adversaries, i.e., $\alpha = 0.01$. This is because the high-degree become adversaries and they tend to absorb the information requests from their many neighboring nodes, as mentioned in Section 3.

## 6. Conclusions

In this paper, the interruption risk of the pull-based competitive information diffusion was analyzed through the mathematical epidemiological approach. First, the information diffusion with interruption in the system was modeled as the SIRA model, which was the continuous-time Markov chain of the five states, i.e., $S_0, S_1, I, R$, and A. The population dynamics was described by the ordinary differential equations. In addition, the scalar SIRA model was extended to the network SIRA model, which could consider the spatial contact patterns among nodes.

Through numerical experiments, the fundamental characteristics of the pull-based competitive information diffusion with interruption were revealed. 1) The initial diffusion speed is not almost affected by the interruption attacks but the increase of attack scale degrades both diffusion speed and ratio. 2) The attack rate mainly affects the initial diffusion ratio while the recovery rate determines the diffusion speed after the initial diffusion. 3) The scale-free structure of the network contributes to the speedy diffusion if adversaries are randomly located in the network. On the other hand, if adversaries are located at high-degree nodes, both diffusion speed and ratio drastically deteriorate.

## Acknowledgment

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI (C) under Grant 19KT0045 and 19K11942, and the Telecommunications Advancement Foundation, Japan.

## References

[1] F. Tschorsch, B. Scheuermann, Bitcoin and Beyond: A Technical Survey on Decentralized Digital Currencies, IEEE Communications Surveys & Tutorials 18 (3) (2016) 2084–2123.
[2] N. Anjum, D. Karamshuk, M. Shikh-Bahaei, N. Sastry, Survey on Peer-Assisted Content Delivery Networks, Computer Networks 116 (2017) 79–95.
[3] M. Lévesque, D. Tipper, A Survey of Clock Synchronization Over Packet-Switched Networks, IEEE Communications Surveys & Tutorials 18 (4) (2016) 2926–2947.
[4] M. Li, X. Wang, K. Gao, S. Zhang, A Survey on Information Diffusion in Online Social Networks: Models and Methods, Information 8 (4) (2017) 1–21.
[5] S. Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System, https://bitcoin.org/bitcoin.pdf (2008).
[6] A. Gervais, H. Ritzdorf, G. O. Karame, S. Capkun, Tampering with the Delivery of Blocks and Transactions in Bitcoin, in: Proc. of ACM SIGSAC CCS'15, 2015, pp. 692–705.
[7] M. Conti, S. K. E, C. Lal, S. Ruj, A Survey on Security and Privacy Issues of Bitcoin, IEEE Communications Surveys & Tutorials (2018) 1–39.
[8] M. Agha Mohammad Ali Kermani, S. F. Fatemi Ardestani, A. Aliahmadi, F. Barzinpour, A Novel Game Theoretic Approach for Modeling Competitive Information Diffusion in Social Networks with Heterogeneous Nodes, Physica A: Statistical Mechanics and its Applications 466 (2017) 570–582.
[9] H. Li, L. Pan, P. Wu, Dominated Competitive Influence Maximization with Time-Critical and Time-Delayed Diffusion in Social Networks, Journal of Computational Science 28 (2018) 318–327.
[10] H. Ismail, Analyzing and Mitigating Security Threats in P2P Systems, Ph.D. Thesis, Technische Universität (2018).
[11] Y. Li, J. Fan, Y. Wang, K. Tan, Influence Maximization on Social Graphs: A Survey, IEEE Transactions on Knowledge and Data Engineering 30 (10) (2018) 1852–1872.
[12] J. A. Kroll, I. C. Davey, E. W. Felten, The Economics of Bitcoin Mining, or Bitcoin in the Presence of Adversaries, in: Proc. of WEIS'13, 2013, pp. 1–21.
[13] M. Rosenfeld, Analysis of Hashrate-Based Double Spending, arXiv:1402.2009 [cs] (2014) 1–13.
[14] I. Eyal, E. G. Sirer, Majority is Not Enough: Bitcoin Mining is Vulnerable, Communications of The ACM 61 (7) (2018) 95–102.

[15] T. Neudecker, H. Hartenstein, Network Layer Aspects of Permissionless Blockchains, IEEE Communications Surveys & Tutorials 21 (1) (2018) 838–857.
[16] C. Decker, R. Wattenhofer, Information Propagation in the Bitcoin Network, in: Proc. of IEEE P2P 2013, 2013, pp. 1–10.
[17] E. Heilman, A. Kendler, A. Zohar, S. Goldberg, Eclipse Attacks on Bitcoins Peer-to-Peer Network, in: Proc. of 24th USENIX Security Symposium (USENIX Security 15), 2015, pp. 129–144.
[18] M. Apostolaki, A. Zohar, L. Vanbever, Hijacking Bitcoin: Routing Attacks on Cryptocurrencies, in: Proc. of 2017 IEEE Symposium on Security and Privacy, 2017, pp. 375–392.
[19] F. Brauer, P. van den Driessche, J. Wu (Eds.), Mathematical Epidemiology, 2008.
[20] M. Martcheva, An Introduction to Mathematical Epidemiology, Springer US, 2015.
[21] F. Brauer, Mathematical Epidemiology: Past, Present, and Future, Infectious Disease Modelling 2 (2) (2017) 113–127.
[22] L. J. S. Allen, Some Discrete-Time SI, SIR, and SIS Epidemic Models, Mathematical Biosciences 124 (1) (1994) 83–105.
[23] T. Britton, Stochastic Epidemic Models: A Survey, Mathematical Biosciences 225 (1) (2010) 24–35.
[24] P. De, Y. Liu, S. K. Das, An Epidemic Theoretic Framework for Vulnerability Analysis of Broadcast Protocols in Wireless Sensor Networks, IEEE Transactions on Mobile Computing 8 (3) (2009) 413–425.
[25] S. Peng, S. Yu, A. Yang, Smartphone Malware and Its Propagation Modeling: A Survey, IEEE Communications Surveys & Tutorials 16 (2) (2014) 925–941.
[26] T. Spyridopoulos, K. Maraslis, A. Mylonas, T. Tryfonas, G. Oikonomou, A Game Theoretical Method for Cost-Benefit Analysis of Malware Dissemination Prevention, Information Security Journal: A Global Perspective 24 (4-6) (2015) 164–176.
[27] W. O. Kermack, A. G. McKendrick, A Contribution to the Mathematical Theory of Epidemics, Proc. of the Royal Society of London 115 (772) (1927) 700–721.
[28] W. Mei, S. Mohagheghi, S. Zampieri, F. Bullo, On the Dynamics of Deterministic Epidemic Propagation over Networks, Annual Reviews in Control 44 (2017) 116–128.
[29] A.-L. Barabási, R. Albert, Emergence of Scaling in Random Networks, Science 286 (5439) (1999) 509–512.
[30] SciPy.org, scipy.integrate.odeint, `https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.odeint.html`.