Hacking-proofness and Stability in a Model of Information Security Networks

Sunghoon Hong*

Preliminary draft, not for citation.

March 1, 2008

Abstract

We introduce a model of information security networks. A group of agents can form directed information links and use security systems called firewalls. Each agent benefits from direct or indirect information. However, it is costly to maintain an information link or a firewall. In this context, we introduce the probability of being hacked. If an agent is not secured by his firewall and obtains a large enough benefit, then he is attacked with positive probability and loses some fraction of his benefit. We then introduce the notions of stability and hacking-proofness. Stability requires that each agent maximizes his expected utility for a given network, while hacking-proofness requires that no agent is attacked with positive probability. We first present examples of stable networks, and study whether they are hacking-proof. Then we investigate their relation in a general context. As it turns out, under a certain condition, any stable network is hacking-proof. We also provide an upper bound of the probability of being hacked under a stable network.

JEL classification: C72; D85

Keywords: Network formation; Information security networks; Stability; Hacking-proofness.

*E-mail: sunghoon.hong@vanderbilt.edu.

1. Introduction

Information security is a vital part of network economy. So, forming a secure network is not only a technological problem, but also an economic issue. Especially when agents form their links in a decentralized manner, we can apply the theory of network formation to study how to form a secure network. The aim of this paper is to introduce a model of information security networks, and to investigate the relation between network security and its stability. To this end, we assume that there is a security system which perfectly precludes any offensive action to a network, and that agents decide how to organize their own links and whether to use the security system.

We examine the formation of information security networks among a group of agents who can form directed links and use security systems called firewalls. Each agent benefits from direct or indirect information. However, it is costly to maintain an information link or a firewall. In this context, we introduce the probability of being hacked. If an agent is not secured by his firewall, he may be attacked and lose some portion of his information benefit. We assume that each agent knows the distribution over the cost of hacking. Then, the probability of an agent being hacked is defined as the one that the cost of hacking is no more than the amount of the benefit he may lose. So, if an agent is not secured by his firewall and obtains a large enough benefit, then he is attacked with positive probability and loses some fraction of his benefit. Examples of such a situation are abundant in the real world: the Internet, decentralized peer-to-peer networks, etc.

Our model mainly rests on two previous studies. One basis is Bala and Goyal [1]. They propose models of directed networks under which an agent can form links to others by himself. These models are studied by various authors.[1] Most of the studies focus on the existence of stable networks and its relation to efficiency. Alternatively, we can analyze the problem by as-

[1] See for example Bala and Goyal [2], Dutta and Jackson [3], Galeotti, Goyal, and Kamphorst [4], Haller, Kamphorst, and Sarangi [5], Haller and Sarangi [6], Kannan, Ray, and Sarangi [9], Kim and Wong [10].

suming that an agreement between two agents is necessary to form a link, as in Jackson and Wolinsky [7]. The other basis is Jackson and Watts [8]. They introduce a model of networks where an exogenous event may suddenly change the structure of a network. Thus, agents cannot control the uncertain environment in their model.

In this paper, however, agents can control the probability of being hacked by adjusting links or using firewalls. Then, our point of interest is the relation between hacking-proofness and stability. Taking into account the probability of being hacked, stability requires that each agent maximizes his expected utility for a given network. On the other hand, hacking-proofness requires that no agent is attacked with positive probability under a network. We first present examples of stable networks, and study whether they are hacking-proof. Then we investigate their relation in a general context. As it turns out, under a certain condition, any stable network is hacking-proof. We also provide an upper bound of the probability of being hacked under a stable network.

Our main contribution is to incorporate uncertainty into the process of network formation. The probability of an agent being hacked results not only from how the cost of hacking is distributed but also from how an agent forms his network. With the model, we can formalize a topic in the economics of information security: how the price of a security system relates to the security level of a decentralized network.

This paper is organized as follows. Section 2 introduces a model of information security networks, and the notions of stability and hacking-proofness. Section 3 analyzes their relation under certainty. Section 4 studies their relation under uncertainty. Concluding remarks follow in Section 5. All the proofs are in the Appendix.

2. The model

Let $N = \{1, \ldots, n\}$ be the set of agents where $n \geq 3$. Each agent is assumed to have some valuable information which can be accessed by other agents.

To access someone's information, each agent has to form a link to the agent whom he wants to do. We assume one-way flow of information. In other words, if agent i initiates the link to j, i will access j's information but j will not. Each agent benefits from direct or indirect information; it is costly to maintain an information link. In case an agent is attacked by a hacker, he will lose some fraction of his information benefit. We assume that there is a security system called firewall which prevents hacking perfectly. Each agent is able to use his own firewall by paying its price. Therefore, each agent decides how to form links to others and whether to use his firewall.

A concrete example of this model is to manage one's own website. A hyperlink of a website enables the owner to access another website. Furthermore, he can access a website which is not directly linked but indirectly linked through a series of hyperlinks of other websites. It takes resources, time, and effort for the owner to manage his own hyperlinks. It is also costly for him to search for the websites that deserve to be linked in his own. To prevent his website from being hacked, he can use his firewall.

A strategy of agent $i \in N$ is a column vector $g_i \equiv (g_{1i}, \ldots, g_{ni})'$ where $g_{ji} \in \{0, 1\}$ for each $j \in N$. For each $i, j \in N$ with $i \neq j$, we say i has a *link* to j if $g_{ji} = 1$. For each $i \in N$, we say *i is secured by his firewall* if $g_{ii} = 1$. For each $i \in N$, let $\mathcal{G}_i$ be the set of i's pure strategies.

An *information security network*, or in short *network* is defined by a strategy profile $g \equiv (g_i)_{i \in N}$. Note that g is an $n \times n$ matrix where each element is either 0 or 1. Let $\mathcal{G} \equiv \mathcal{G}_1 \times \cdots \times \mathcal{G}_n$ be the set of all networks.

For each $i, j \in N$ with $i \neq j$ and each $g \in \mathcal{G}$, a *path* from i to j under g is a sequence of agents $i_1, \ldots, i_K$ such that $g_{i_{k+1} i_k} = 1$ for each $k \in \{1, \ldots, K-1\}$ with $i = i_1$ and $j = i_K$. For each $i, j \in N$ with $i \neq j$ and each $g \in \mathcal{G}$, the *(geodesic) distance* from i to j under g, $d(i, j; g)$, is the smallest number of links in the path from i to j; if there is no path from i to j under g, we set $d(i, j; g) = \infty$.

We assume that each information of an agent has a common value to others and that its value is depreciated as the distance from one to another increases. Specifically, as information is spread out through a series of hyperlinks, there will be delays of the dissemination and agents are impatient at these delays. Let $b : \mathbb{N} \to \mathbb{R}_+$ be the *benefit function* which associates with each $d(i, j; g)$ a nonnegative real value. With a slight abuse of notation, a positive integer d denotes an argument of the benefit function. We assume that b is decreasing and $\lim_{d\to\infty} b(d) = 0$.

Let $c \in \mathbb{R}_{++}$ be the *cost of a link*, $p \in \mathbb{R}_{++}$ the *price of a firewall*, and $r \in (0, 1]$ the *rate of a benefit loss.* If an agent gets hacked, he will lose this fraction of the sum of his benefit.

For each $i \in N$ and each $g \in \mathcal{G}$, *i is attacked with probability $h_i(g)$* under g, while he is not attacked with probability $1 - h_i(g)$ under g. We will explain how to determine $h_i(g)$ later. If agent i is not attacked with probability $1 - h_i(g)$ under g, then he earns the sum of the benefits which equals to $\sum_{j\neq i} b(d(i, j; g))$. Otherwise, he loses some portion of his benefits, and obtains the sum of the remaining benefits which equals to $(1 - r)\sum_{j\neq i} b(d(i, j; g))$ with probability $h_i(g)$. So, for each $i \in N$ and each $g \in \mathcal{G}$, the *expected utility* of i under g is,

$$Eu_i(g) \equiv (1 - r \cdot h_i(g)) \sum_{j\neq i} b(d(i, j; g)) - \sum_{j\neq i} g_{ji} \cdot c - g_{ii} \cdot p.$$

Before defining $h_i(g)$ the *probability of agent i being hacked*, we will elaborate on the nature of the hacker in our model. There is a hacker outside of the agents. Hacking incurs some cost to him. The hacker mechanically decides whom he will attack; if the cost of hacking an agent is no more than the amount of the benefit he can take from the agent, he will attack the agent. The benefit appropriated by the hacker from agent i is assumed to have the same value as i's benefit loss.

In general, we assume that each agent knows only the distribution over the cost of hacking. Let $x \in \mathbb{R}_+$ be the *cost of hacking* and $H : \mathbb{R}_+ \to [0, 1]$ the *distribution function* over the cost of hacking. We assume that the cost of hacking is always positive: $H(0) = 0$. Let $x_0 \equiv \inf\{x \in \mathbb{R}_+ | H(x) > 0\}$.

Given $i \in N$, $g \in \mathcal{G}$, b, $r \in (0, 1]$, and H, let

$$h_i(g; b, r, H) \equiv H\left(r(1 - g_{ii}) \sum_{j \neq i} b(d(i, j; g))\right)$$

be the probability of agent i being hacked where $r(1 - g_{ii}) \sum_{j \neq i} b(d(i, j; g))$ is the amount of the benefit the hacker can take from i. Put it another way, it is worthwhile for the hacker to attack agent i if the cost of hacking is no more than the amount of the benefit taken from i, which is equal to i's benefit loss. Note that the probability of agent i being hacked is equal to zero if i is secured by his firewall or the amount of the benefit he may lose is small enough. Let $h(g; b, r, H) \equiv (h_1(g; b, r, H), \ldots, h_n(g; b, r, H))'$. When there is no ambiguity, we write $h_i(g)$ and $h(g)$ instead of $h_i(g; b, r, H)$ and $h(g; b, r, H)$, respectively.

Next we introduce two notions, which will be the main interests of this paper. For each $i \in N$, each $g \in \mathcal{G}$, and each $g'_i \in \mathcal{G}_i$, let (g'_i, g_{-i}) be the matrix substituting g'_i for g_i in g. *Stability* requires that each agent maximizes his expected utility for a given network.

Definition: A network $g \in \mathcal{G}$ is *stable* if for each $i \in N$ and each $g'_i \in \mathcal{G}_i$, $Eu_i(g_i, g_{-i}) \geq Eu_i(g'_i, g_{-i})$.

Let $\mathbf{0} \in \mathbb{R}^n$ and $\mathbf{1} \in \mathbb{R}^n$ be the vectors of 0 and 1, respectively. *Hacking-proofness* requires that no agent is attacked with positive probability.

Definition: A network $g \in \mathcal{G}$ is *hacking-proof* if $h(g) = \mathbf{0}$.

At a glance, we can find an example of a network which is hacking-proof but not stable. An *empty* network is one where no agent has links to others and is secured by his firewall. Formally, a network $g \in \mathcal{G}$ is empty if for each $i, j \in N$ with $i \neq j$, $g_{ij} = 0$, and for each $i \in N$, $g_{ii} = 0$. The following remark shows that a hacking-proof network may not be stable.

Remark: The empty network g is always hacking-proof since $\sum_{j \neq i} b(d(i, j; g)) = 0$ for each $i \in N$ and $H(0) = 0$. However, it is not stable if $c < b(1)$ and $rb(1) < x_0$.

Thus, from now on, we focus on the question whether a stable network is hacking-proof.

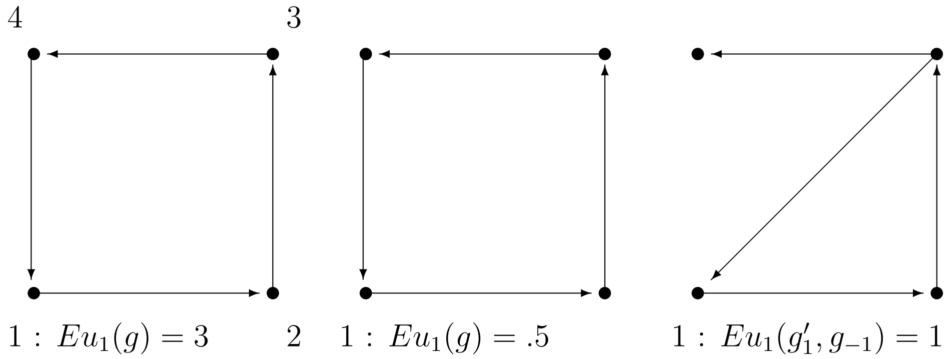
3. Hacking-proofness and stability under certainty

In this section, we assume that $H(x_0) = 1$. In other words, each agent knows the true cost of hacking which equals to x_0. Then, for each $i \in N$, we have

$$h_i(g) = \begin{cases} 1 & \text{if } r(1 - g_{ii}) \sum_{j \neq i} b(d(i, j; g)) \geq x_0, \\ 0 & \text{otherwise.} \end{cases}$$

We first present an example showing that the cost of hacking may affect the hacking-proofness or the stability of a network.

Example 1: Let $n = 4$, and g be such that $g_1 = (0, 0, 0, 1)'$, $g_2 = (1, 0, 0, 0)'$, $g_3 = (0, 1, 0, 0)'$, and $g_4 = (0, 0, 1, 0)'$. Assume that $b(1) = 5$, $b(2) = 3$, $b(3) = 2$, $c = 7$, $p = 3$, and $r = .25$. If $x_0 = 3$, then $h(g) = \mathbf{0}$, since $r(b(1) + b(2) + b(3)) = 2.5 < 3 = x_0$. So, g is hacking-proof. Furthermore, it is stable. See Figure 1 (a). However, if $x_0 = 2.1$, then $h_i(g) = 1$ for each $i \in N$, since $x_0 = 2.1 \leq 2.5 = r(b(1)+b(2)+b(3))$. So, g is not hacking-proof. Now, let $g'_1 = (0, 0, 1, 0)'$. Since $r(b(1)+b(2)) = 2 < 2.1 = x_0$, $h_1(g'_1, g_{-1}) = 0$. So, $Eu_1(g'_1, g_{-1}) = b(1) + b(2) - c = 1 > .5 = (1 - r)(b(1) + b(2) + b(3)) - c = Eu_1(g)$. Thus, g is not stable. See Figure 1 (b). Note that each arrow indicates the flow of information, which is the reverse direction of each link.



(a) $x_0 = 3$ (b) $x_0 = 2.1$

Figure 1 The cost of hacking plays a significant role

Our next example shows that a stable network may not be hacking-proof.

Example 2: Let $n = 4$, and g be such that $g_1 = (1,1,1,1)'$, $g_2 = (1,0,0,0)'$, $g_3 = (1,0,0,0)'$, and $g_4 = (1,0,0,0)'$. Assume that $b(1) = 10$, $b(2) = 9$, $b(3) = 6$, $c = 5$, $p = 3$, $x_0 = 2$, and $r = .1$. Since $x_0 = 2 \leq 2.8 = r(b(1) + 2b(2))$, $h_i(g) = 1$ for $i = 2,3,4$. So, g is not hacking-proof. However, it is stable. See Figure 2. Here the circle around agent 1 denotes his firewall.

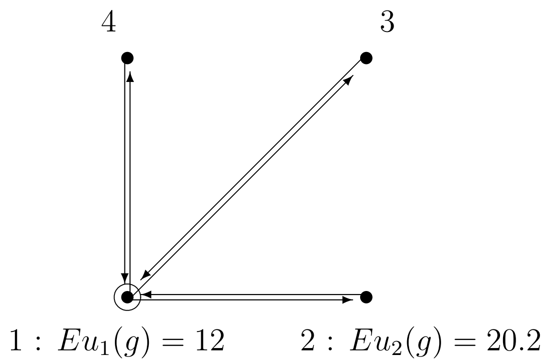


Figure 2 A stable network may not be hacking-proof

Now we present examples of stable networks, and study whether they are hacking-proof. A *complete* network is one where each agent has links to all the other agents. Formally, a network $g \in \mathcal{G}$ is complete if for each $i, j \in N$ with $i \neq j$, $g_{ij} = 1$. A network *with no agent secured* is one where no agent is secured by his firewall. Figure 3 shows a complete network with no agent secured when $n = 3$.

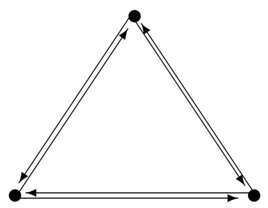


Figure 3 A complete network with no agent secured

Proposition 1: *The complete network with no agent secured is stable if*

(i) $r(n-1)b(1) \leq p$, *and*

(ii) $c \leq (1 + r - rn)b(1) - b(2)$.

In the assumptions of Proposition 1, (i) says that the price of a firewall is no less than the benefit loss of an agent under the complete network with no agent secured, and (ii) says that the cost of a link is no more than the minimum increment of benefit by a link addition.

Remark: The complete network with no agent secured is hacking-proof if and only if $r(n-1)b(1) < x_0$. So, together with Proposition 1, if $x_0 \leq r(n-1)b(1)$, $r(n-1)b(1) \leq p$, and $c \leq (1 + r - rn)b(1) - b(2)$, it is stable but not hacking-proof.

A *wheel* network is one where each agent has exactly one link and each pair of agents has a path from one to another. Formally, a network $g \in \mathcal{G}$ is a wheel if there is a sequence of agents $i_1, \ldots, i_n$ such that (i) for each $k \in \{2, \ldots, n\}$ and each $l \notin \{k-1, k\}$, $g_{i_{k-1}i_k} = 1$ and $g_{i_l i_k} = 0$, and (ii)

$g_{i_n i_1} = 1$, and for each $l \notin \{n, 1\}$, $g_{i_l i_1} = 0$. Figure 4 shows a wheel network with no agent secured when $n = 3$.

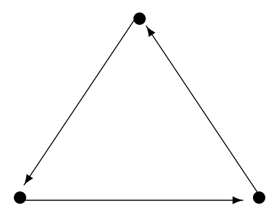


Figure 4 A wheel network with no agent secured

Proposition 2: *Let n be odd. Then the wheel network with no agent secured is stable if*

(i) $r \sum_{d=1}^{n-1} b(d) \leq p$,

(ii) $r \sum_{d=1}^{n-1} b(d) \leq b(n-1)$, *and*

(iii) $(1+r) \sum_{d=1}^{(n-1)/2} b(d) - (1-r) \sum_{d=(n+1)/2}^{n-1} b(d) \leq c \leq (1-r) \sum_{d=1}^{n-1} b(d)$.

In the assumptions of Proposition 2, (i) says that the price of a firewall is no less than the benefit loss of an agent under the wheel network with no agent secured, (ii) says that the value of the information from the most distant agent is no less than the benefit loss of an agent under the network, and (iii) says that the cost of a link is no less than the maximum increment of benefit by a link addition and also no more than the sum of the remaining benefits when an agent is attacked.

Remark: Let n be even. Then the wheel network with no agent secured is stable if (i) $r \sum_{d=1}^{n-1} b(d) \leq p$, (ii) $r \sum_{d=1}^{n-1} b(d) \leq b(n-1)$, and (iii) $(1 + r) \sum_{d=1}^{n/2-1} b(d) - (1-r) \sum_{d=n/2+1}^{n-1} b(d) + rb(n/2) \leq c \leq (1-r) \sum_{d=1}^{n-1} b(d)$.

Remark: The wheel network with no agent secured is hacking-proof if and only if $r \sum_{d=1}^{n-1} b(d) < x_0$. So, together with Proposition 2, it is stable but not hacking-proof if $x_0 \leq r\sum_{d=1}^{n-1} b(d)$, $r\sum_{d=1}^{n-1} b(d) \leq p$, $r\sum_{d=1}^{n-1} b(d) \leq b(n-1)$, and $(1+r)\sum_{d=1}^{(n-1)/2} b(d) - (1-r)\sum_{d=(n+1)/2}^{n-1} b(d) \leq c \leq (1-r)\sum_{d=1}^{n-1} b(d)$, when n is odd.

A *star network with its center secured* is one where the *center* has links to all the other agents and is secured by his firewall, and each of the others has one link to the center and is not secured by his firewall. Formally, a network $g \in \mathcal{G}$ is a star with its center secured if there exists the center $i \in N$ such that (i) for each $j \neq i$, $g_{ij} = g_{ji} = 1$, and for each $j, k \neq i$ with $j \neq k$, $g_{jk} = 0$, and (ii) $g_{ii} = 1$, and for each $j \neq i$, $g_{jj} = 0$. Figure 5 shows the network when $n = 4$.

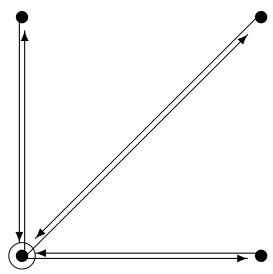


Figure 5 A star network with its center secured

Remark: Let $n = 3$. Then the star network with its center secured is stable if $x_0 \leq 2rb(1)$, $r(b(1)+b(2)) \leq p \leq 2rb(1)$, and $b(1)-b(2) \leq c \leq (1-2r)b(1)$.

Proposition 3: *Let $n \geq 4$. Then the star network with its center secured is stable if*

(i) $x_0 \leq r(n-1)b(1)$,

(ii) $r(b(1)+(n-2)b(2)) \leq (n-3)(b(2)-b(3))$,

(iii) $r(b(1)+(n-2)b(2)) \leq p \leq r(n-1)b(1)$, *and*

(iv) $(1+r)b(1) - (1+2r-rn)b(2) \leq c \leq (1+r-rn)b(1)$.

In the assumptions of Proposition 3, (i) says that the cost of hacking is no more than the benefit loss of the center when he is not secured by his firewall, (ii) says that the benefit loss of a peripheral agent under the star network with its center secured is no more than the minimum decrement of benefit by changing his link to another peripheral agent, (iii) says that the price of a firewall is no less than the benefit loss of a peripheral agent under the network and also no more than that of the center, and (iv) says that the cost of a link is no less than the maximum increment of benefit by a link addition and also no more than the minimum decrement of benefit by a link deletion.

Remark: The star network with its center secured is hacking-proof if and only if $r(b(1)+(n-2)b(2)) < x_0$. So, together with Proposition 3, it is stable but not hacking-proof if $x_0 \leq r(b(1)+(n-2)b(2))$, $r(b(1)+(n-2)b(2)) \leq (n-3)(b(2)-b(3))$, $r(b(1)+(n-2)b(2)) \leq p \leq r(n-1)b(1)$, and $(1+r)b(1)-(1+2r-rn)b(2) \leq c \leq (1+r-rn)b(1)$.

In a nutshell, even though each agent knows the true cost of hacking, stable networks may not be hacking-proof. Thus, our next question is whether we can find a condition under which every stable network is hacking-proof. As shown in the following proposition, we have a positive result: every stable network is hacking-proof if each agent knows the true cost of hacking and it is greater than the price of a firewall.

Proposition 4: *If $H(x_0) = 1$ and $p < x_0$, then any stable network is hacking-proof.*

4. Hacking-proofness and stability under uncertainty

We now assume that each agent does not know the true cost of hacking, but just knows its distribution. Our next question is whether we can generalize

Proposition 4 in this context. We have a positive result in Proposition 5. Before continuing our discussion, we introduce the following notion.

Definition: A network $g \in \mathcal{G}$ is *fully secured down to degree 1* if

(i) for each $i \in N$ with $\sum_{j \neq i} g_{ji} \geq 1$, $g_{ii} = 1$, and

(ii) for each $i \in N$ with $\sum_{j \neq i} g_{ji} < 1$, $g_{ii} = 0$.

A network is fully secured down to degree 1 if (i) an agent is secured by his firewall whenever he has links to 1 or more agents, and (ii) an agent is not secured by his firewall otherwise. We then establish the following lemma.

Lemma 1: *If $p < rb(1) \cdot H(rb(1))$, then any stable network is fully secured down to degree 1.*

Since the sufficient condition of Lemma 1 requires that the price of a firewall is less than the minimum expected loss of an agent who has links to 1 or more agents, he has an incentive to use his firewall. If an agent has no link, then he has no incentive to be secured by his firewall. Therefore, any stable network is fully secured down to degree 1 under the condition. Furthermore, the following proposition shows that given the same condition, stability implies hacking-proofness.

Proposition 5: *If $p < rb(1) \cdot H(rb(1))$, then any stable network is hacking-proof.*

Next we show that the probability of being hacked has an upper bound under a stable network. For each $p \in \mathbb{R}_{++}$, let $x_p \equiv \min\{x \in \mathbb{R}_+ | p \leq x \cdot H(x)\}$.

Proposition 6: *If g is stable, then $h(g) \leq H(x_p) \cdot \mathbf{1}$.*

By Proposition 6, we show that if the cost of hacking is uniformly distributed over a closed interval and the price of a firewall is no more than the maximum hacking cost, then the upper bound can be expressed as a function of three variables: the minimum hacking cost, the maximum hacking cost, and the price of a firewall.

Corollary 1: *Let H be the uniform distribution function over $[x_0, x_1]$. If $p \leq x_1$ and g is stable, then*

$$h(g) \leq \frac{-x_0 + \sqrt{x_0^2 + 4p(x_1 - x_0)}}{2(x_1 - x_0)} \cdot \mathbf{1}.$$

5. Concluding remarks

In this paper, we investigate the implications of hacking-proofness and stability for an information security network model. An underlying assumption of our approach is that each agent knows the entire structure of a network. However, if we study complex networks whose structures cannot be identified clearly, we need to find an alternative framework. As in Lopez-Pintado [11], we can analyze the problem by taking a random graph theoretic approach: first, assume that each agent knows the degree (the number of links an agent maintains) distribution of a network, and then define the expected utility of an agent with respect to the distribution.

Another interesting question is to consider the externality of hacking. In our model, we assume that an agent loses some fraction of his benefit only if he is attacked with positive probability. However, if hacking is contagious through a network, it may be reasonable to assume that other agents, who are not attacked but have links to the agents attacked with positive probability, also lose some portion of their benefits. We hope to address the issue in our future research.

Appendix

Now we present the proofs for all propositions.

Proof of Proposition 1: Let g be the complete network with no agent secured. Then, for each $i \in N$, $Eu_i(g) = (n-1)b(1) - (n-1)c$ if $r(n-1)b(1) <$

x_0, or $Eu_i(g) = (1-r)(n-1)b(1) - (n-1)c$ if $x_0 \leq r(n-1)b(1)$. In both cases, $Eu_i(g) \geq (1-r)(n-1)b(1) - (n-1)c$. First, let g_i' be the strategy such that i is secured by his firewall and has links to k agents. Note that $k \leq n-1$. Then, for each $i \in N$, $Eu_i(g_i', g_{-i}) = kb(1) + (n-1-k)b(2) - kc - p$. So, for each $i \in N$,

$$\begin{aligned}
&Eu_i(g) - Eu_i(g_i', g_{-i}) \\
&\quad \geq (1-r)(n-1)b(1) - kb(1) - (n-1-k)(b(2)+c) + p \\
&\quad \geq (1-r)(n-1)b(1) - kb(1) - (n-1-k)(b(2)+c) + r(n-1)b(1) \\
&\quad \geq (n-1-k)((1-r)b(1) - b(2) - c) \\
&\quad \geq 0,
\end{aligned}$$

by (i) and (ii).

Next, let g_i' be the strategy such that i is not secured by his firewall and has links to k agents. If $k = n-1$, $g_i' = g_i$. So, for each $i \in N$, $Eu_i(g_i', g_{-i}) = Eu_i(g)$. If $k \leq n-2$, for each $i \in N$, $Eu_i(g_i', g_{-i}) \leq kb(1) + (n-1-k)b(2) - kc$. So, for each $i \in N$,

$$\begin{aligned}
&Eu_i(g) - Eu_i(g_i', g_{-i}) \\
&\quad \geq (1-r)(n-1)b(1) - kb(1) - (n-1-k)(b(2)+c) \\
&\quad = (n-1-k-r(n-1))b(1) - (n-1-k)(b(2)+c) \\
&\quad = (n-1-k)((1 - \frac{r(n-1)}{n-1-k})b(1) - b(2) - c) \\
&\quad \geq (n-1-k)((1-r(n-1))b(1) - b(2) - c) \\
&\quad \geq 0,
\end{aligned}$$

by (ii). □

Proof of Proposition 2: Let n be odd, and g the wheel network with no agent secured. Then, for each $i \in N$, $Eu_i(g) = \sum_{d=1}^{n-1} b(d) - c$ if $r \sum_{d=1}^{n-1} b(d) < x_0$, or $Eu_i(g) = (1-r) \sum_{d=1}^{n-1} b(d) - c$ if $x_0 \leq r \sum_{d=1}^{n-1} b(d)$. In both cases, $Eu_i(g) \geq (1-r) \sum_{d=1}^{n-1} b(d) - c$. First, let g_i' be the strategy such that i is secured by his firewall and has links to k agents. If $k = 0$, then $Eu_i(g_i', g_{-i}) =$

$-p \leq 0 \leq Eu_i(g)$, by (iii). If $k = 1$, then $Eu_i(g_i', g_{-i}) \leq \sum_{d=1}^{n-1} b(d) - c - p$. This implies $Eu_i(g_i', g_{-i}) \leq Eu_i(g)$, by (i). If $2 \leq k \leq n-1$, then $Eu_i(g_i', g_{-i}) \leq k \sum_{d=1}^{(n-1)/2} b(d) - (k-2) \sum_{d=(n+1)/2}^{n-1} b(d) - kc - p$. So, for each $i \in N$,

$$
\begin{aligned}
& Eu_i(g) - Eu_i(g_i', g_{-i}) \\
\geq\ & (1-r) \sum_{d=1}^{n-1} b(d) - c - k \sum_{d=1}^{(n-1)/2} b(d) + (k-2) \sum_{d=(n+1)/2}^{n-1} b(d) + kc + p \\
\geq\ & \sum_{d=1}^{n-1} b(d) - k \sum_{d=1}^{(n-1)/2} b(d) + (k-2) \sum_{d=(n+1)/2}^{n-1} b(d) + (k-1)c \\
=\ & (k-1) \left(c - \sum_{d=1}^{(n-1)/2} b(d) + \sum_{d=(n+1)/2}^{n-1} b(d) \right) \\
\geq\ & 0,
\end{aligned}
$$

by (i) and (iii).

Next, let g_i' be the strategy such that i is not secured by his firewall and has links to k agents. If $k = 0$, then $Eu_i(g_i', g_{-i}) = 0 \leq Eu_i(g)$, by (iii). If $k = 1$ and $g_i' = g_i$, then $Eu_i(g_i', g_{-i}) = Eu_i(g)$. If $k = 1$ and $g_i' \neq g_i$, then $Eu_i(g_i', g_{-i}) \leq \sum_{d=1}^{n-2} b(d) - c \leq Eu_i(g)$, by (ii). If $2 \leq k \leq n-1$, then $Eu_i(g_i', g_{-i}) \leq k \sum_{d=1}^{(n-1)/2} b(d) - (k-2) \sum_{d=(n+1)/2}^{n-1} b(d) - kc$. So, for each

$i \in N$,

$$
\begin{aligned}
& Eu_i(g) - Eu_i(g_i', g_{-i}) \\
\geq\ & (1-r)\sum_{d=1}^{n-1} b(d) - c - k\sum_{d=1}^{(n-1)/2} b(d) + (k-2)\sum_{d=(n+1)/2}^{n-1} b(d) + kc \\
=\ & \sum_{d=1}^{n-1} b(d) - k\sum_{d=1}^{(n-1)/2} b(d) + (k-2)\sum_{d=(n+1)/2}^{n-1} b(d) + (k-1)c - r\sum_{d=1}^{n-1} b(d) \\
=\ & (k-1)\left(c - \sum_{d=1}^{(n-1)/2} b(d) + \sum_{d=(n+1)/2}^{n-1} b(d) - \frac{r}{k-1}\sum_{d=1}^{n-1} b(d)\right) \\
\geq\ & (k-1)\left(c - \sum_{d=1}^{(n-1)/2} b(d) + \sum_{d=(n+1)/2}^{n-1} b(d) - r\sum_{d=1}^{n-1} b(d)\right) \\
\geq\ & 0,
\end{aligned}
$$

by (iii). □

Proof of Proposition 3: Let g be the star network with its center secured, and $i \in N$ the center. Then, $Eu_i(g) = (n-1)b(1) - (n-1)c - p$. First, let g_i' be the strategy such that i is secured by his firewall and has links to k agents. Then, $Eu_i(g_i', g_{-i}) = kb(1) - kc - p$. So,

$$
\begin{aligned}
& Eu_i(g) - Eu_i(g_i', g_{-i}) \\
=\ & (n-1)b(1) - (n-1)c - p - kb(1) + kc + p \\
=\ & (n-1-k)(b(1) - c) \\
\geq\ & 0,
\end{aligned}
$$

by (iv).

Next, let g_i' be the strategy such that i is not secured by his firewall and has links to k agents. If $k = n-1$, then $h_i(g_i', g_{-i}) = 1$, by (i). Thus,

$Eu_i(g_i', g_{-i}) = (1-r)(n-1)b(1) - (n-1)c$. So,

$$\begin{aligned}
&Eu_i(g) - Eu_i(g_i', g_{-i}) \\
&= (n-1)b(1) - (n-1)c - p - (1-r)(n-1)b(1) + (n-1)c \\
&= r(n-1)b(1) - p \\
&\geq 0,
\end{aligned}$$

by (iii). If $0 \leq k \leq n-2$, then $Eu_i(g_i', g_{-i}) \leq kb(1) - kc$. So,

$$\begin{aligned}
&Eu_i(g) - Eu_i(g_i', g_{-i}) \\
&\geq (n-1)b(1) - (n-1)c - p - kb(1) + kc \\
&\geq (n-1)b(1) - (n-1)c - r(n-1)b(1) - kb(1) + kc \\
&= (n-1-k)((1 - \frac{r(n-1)}{n-1-k})b(1) - c) \\
&\geq (n-1-k)((1 - r(n-1))b(1) - c) \\
&\geq 0,
\end{aligned}$$

by (iii) and (iv).

For each $j \neq i$, $Eu_j(g) = b(1) + (n-2)b(2) - c$ if $r(b(1) + (n-2)b(2)) < x_0$, or $Eu_j(g) = (1-r)(b(1) + (n-2)b(2)) - c$ if $x_0 \leq r(b(1) + (n-2)b(2))$. In both cases, $Eu_j(g) \geq (1-r)(b(1) + (n-2)b(2)) - c$. First, let g_j' be the strategy such that j is secured by his firewall and has links to k agents. If $k = 0$, then $Eu_j(g_j', g_{-j}) = -p \leq 0 \leq Eu_j(g)$, by (iv). If $k = 1$ and j has the link to the center, then $Eu_j(g_j', g_{-j}) = b(1) + (n-2)b(2) - c - p$. So, for each $j \neq i$,

$$\begin{aligned}
&Eu_j(g) - Eu_j(g_j', g_{-j}) \\
&\geq (1-r)(b(1) + (n-2)b(2)) - c - b(1) - (n-2)b(2) + c + p \\
&= p - r(b(1) + (n-2)b(2)) \\
&\geq 0,
\end{aligned}$$

by (iii). If $k = 1$ and j has a link to an agent other than the center, then

$Eu_j(g'_j, g_{-j}) = b(1) + b(2) + (n-3)b(3) - c - p$. So, for each $j \neq i$,

$$
\begin{aligned}
& Eu_j(g) - Eu_j(g'_j, g_{-j}) \\
\geq\ & (1-r)(b(1) + (n-2)b(2)) - c - b(1) - b(2) - (n-3)b(3) + c + p \\
\geq\ & (n-3)(b(2) - b(3)) - r(b(1) + (n-2)b(2)) \\
\geq\ & 0,
\end{aligned}
$$

by (ii). If $2 \leq k \leq n-1$, then $Eu_j(g'_j, g_{-j}) \leq kb(1) + (n-1-k)b(2) - kc - p$. So, for each $j \neq i$,

$$
\begin{aligned}
& Eu_j(g) - Eu_j(g'_j, g_{-j}) \\
\geq\ & (1-r)(b(1) + (n-2)b(2)) - c - kb(1) - (n-1-k)b(2) + kc + p \\
\geq\ & (1-r)(b(1) + (n-2)b(2)) - c - kb(1) - (n-1-k)b(2) + kc + r(b(1) + (n-2)b(2)) \\
=\ & (k-1)(c - b(1) + b(2)) \\
\geq\ & 0,
\end{aligned}
$$

by (iii) and (iv).

Next, let g'_j be the strategy such that j is not secured by his firewall and has links to k agents. If $k = 0$, then $Eu_j(g'_j, g_{-j}) = 0 \leq Eu_j(g)$, by (iv). If $k = 1$ and j has the link to the center, then $g'_j = g_j$ and $Eu_j(g'_j, g_{-j}) = Eu_j(g)$. If $k = 1$ and j has a link to an agent other than the center, then $Eu_j(g'_j, g_{-j}) \leq b(1) + b(2) + (n-3)b(3) - c$. So, for each $j \neq i$,

$$
\begin{aligned}
& Eu_j(g) - Eu_j(g'_j, g_{-j}) \\
\geq\ & (1-r)(b(1) + (n-2)b(2)) - c - b(1) - b(2) - (n-3)b(3) + c \\
=\ & (n-3)(b(2) - b(3)) - r(b(1) + (n-2)b(2)) \\
\geq\ & 0,
\end{aligned}
$$

by (ii). If $2 \leq k \leq n-1$, then $Eu_j(g'_j, g_{-j}) \leq kb(1) + (n-1-k)b(2) - kc$.

So, for each $j \neq i$,

$$\begin{aligned}
& Eu_j(g) - Eu_j(g'_j, g_{-j}) \\
&\geq (1-r)(b(1) + (n-2)b(2)) - c - kb(1) - (n-1-k)b(2) + kc \\
&\geq (k-1)(c - (1 + \frac{r}{k-1})b(1) + (1 - \frac{r(n-2)}{k-1})b(2)) \\
&\geq (k-1)(c - (1+r)b(1) + (1 - r(n-2))b(2)) \\
&\geq 0,
\end{aligned}$$

by (iv). □

Proof of Proposition 4: Suppose that g is not hacking-proof. Then, there exists $i \in N$ such that $h_i(g) > 0$. Since $h_i(g) = H(r(1-g_{ii}) \sum_{j \neq i} b(d(i,j;g))) > 0$, $r(1-g_{ii}) \sum_{j \neq i} b(d(i,j;g)) \geq x_0$. So, $h_i(g) = 1$. Since $H(0) = 0$ and $H(x_0) = 1$, $x_0 > 0$. Thus, $r(1-g_{ii}) \sum_{j \neq i} b(d(i,j;g)) > 0$. This implies $g_{ii} = 0$ and $r \sum_{j \neq i} b(d(i,j;g)) \geq x_0$. Since $h_i(g) = 1$ and $g_{ii} = 0$, $Eu_i(g) = (1-r) \sum_{j \neq i} b(d(i,j;g)) - \sum_{j \neq i} g_{ji} \cdot c$.

Now, let g'_i be the strategy such that $g'_{ii} = 1$, and for each $j \neq i$, $g'_{ji} = g_{ji}$. Then, $Eu_i(g'_i, g_{-i}) = \sum_{j \neq i} b(d(i,j;g'_i, g_{-i})) - \sum_{j \neq i} g'_{ji} \cdot c - p$. Since $g'_{ji} = g_{ji}$ for each $j \neq i$, $d(i,j;g'_i,g_{-i}) = d(i,j;g)$. So,

$$\begin{aligned}
& Eu_i(g'_i, g_{-i}) - Eu_i(g) \\
&= \sum_{j \neq i} b(d(i,j;g'_i,g_{-i})) - \sum_{j \neq i} g'_{ji} \cdot c - p - (1-r) \sum_{j \neq i} b(d(i,j;g)) + \sum_{j \neq i} g_{ji} \cdot c \\
&= r \sum_{j \neq i} b(d(i,j;g)) - p \\
&> r \sum_{j \neq i} b(d(i,j;g)) - x_0 \\
&\geq 0,
\end{aligned}$$

since $p < x_0$ and $r \sum_{j \neq i} b(d(i,j;g)) \geq x_0$. Thus, g is not stable. □

Proof of Lemma 1: Suppose that g is not fully secured down to degree 1. We consider two cases.

Case 1: For some $i \in N$ with $\sum_{j \neq i} g_{ji} \geq 1$, $g_{ii} = 0$. Since $\sum_{j \neq i} g_{ji} \geq 1$, $\sum_{j \neq i} b(d(i,j;g)) \geq b(1)$. So, $h_i(g) \geq H(rb(1))$, since $g_{ii} = 0$. Note that

$Eu_i(g) = (1 - r \cdot h_i(g)) \sum_{j \neq i} b(d(i,j;g)) - \sum_{j \neq i} g_{ji} \cdot c$. Now, let g'_i be the strategy such that $g'_{ii} = 1$, and for each $j \neq i$, $g'_{ji} = g_{ji}$. Then, $Eu_i(g'_i, g_{-i}) = \sum_{j \neq i} b(d(i,j;g'_i, g_{-i})) - \sum_{j \neq i} g'_{ji} \cdot c - p$. Since $g'_{ji} = g_{ji}$ for each $j \neq i$, $d(i,j;g'_i, g_{-i}) = d(i,j;g)$. Altogether,

$$
\begin{aligned}
&Eu_i(g'_i, g_{-i}) - Eu_i(g) \\
&\quad = \sum_{j \neq i} b(d(i,j;g'_i, g_{-i})) - \sum_{j \neq i} g'_{ji} \cdot c - p - (1 - r \cdot h_i(g)) \sum_{j \neq i} b(d(i,j;g)) + \sum_{j \neq i} g_{ji} \cdot c \\
&\quad = r \cdot h_i(g) \sum_{j \neq i} b(d(i,j;g)) - p \\
&\quad \geq rb(1) \cdot H(rb(1)) - p \\
&\quad > 0,
\end{aligned}
$$

since $p < rb(1) \cdot H(rb(1))$. Thus, g is not stable.
Case 2: For some $i \in N$ with $\sum_{j \neq i} g_{ji} < 1$, $g_{ii} = 1$. Since $\sum_{j \neq i} g_{ji} < 1$, $\sum_{j \neq i} b(d(i,j;g)) = 0$. So, $Eu_i(g) = -p < 0$. Thus, g is not stable. □

Proof of Proposition 5: Let g be a stable network. Since $p < rb(1) \cdot H(rb(1))$, by Lemma 1, g is fully secured down to degree 1. First, for each $i \in N$ with $\sum_{j \neq i} g_{ji} \geq 1$, $g_{ii} = 1$. Since $g_{ii} = 1$, $h_i(g) = H(0) = 0$. Next, for each $i \in N$ with $\sum_{j \neq i} g_{ji} < 1$, $g_{ii} = 0$. Since $\sum_{j \neq i} g_{ji} < 1$, $\sum_{j \neq i} b(d(i,j;g)) = 0$. So, $h_i(g) = H(0) = 0$. Thus, g is hacking-proof. □

Proof of Proposition 6: Suppose that there exists $i \in N$ such that $h_i(g) > H(x_p)$. Since $h_i(g) = H(r(1 - g_{ii}) \sum_{j \neq i} b(d(i,j;g))) > H(x_p)$, $r(1 - g_{ii}) \sum_{j \neq i} b(d(i,j;g)) > x_p$. Since $p > 0$, we have $x_p > 0$. Since $r(1 - g_{ii}) \sum_{j \neq i} b(d(i,j;g)) > 0$, we have $g_{ii} = 0$. Thus, $r \sum_{j \neq i} b(d(i,j;g)) > x_p$. Note that $Eu_i(g) = (1 - r \cdot h_i(g)) \sum_{j \neq i} b(d(i,j;g)) - \sum_{j \neq i} g_{ji} \cdot c$.

Now, let g'_i be the strategy such that $g'_{ii} = 1$, and for each $j \neq i$, $g'_{ji} = g_{ji}$. Then, $Eu_i(g'_i, g_{-i}) = \sum_{j \neq i} b(d(i,j;g'_i, g_{-i})) - \sum_{j \neq i} g'_{ji} \cdot c - p$. Since $g'_{ji} = g_{ji}$

for each $j \neq i$, $d(i,j;g_i',g_{-i}) = d(i,j;g)$. Altogether,

$$\begin{aligned}
& Eu_i(g_i',g_{-i}) - Eu_i(g) \\
&= \sum_{j\neq i} b(d(i,j;g_i',g_{-i})) - \sum_{j\neq i} g_{ji}' \cdot c - p - (1 - r\cdot h_i(g)) \sum_{j\neq i} b(d(i,j;g)) + \sum_{j\neq i} g_{ji}\cdot c \\
&= r\cdot h_i(g) \sum_{j\neq i} b(d(i,j;g)) - p \\
&> r\cdot H(x_p) \sum_{j\neq i} b(d(i,j;g)) - p \\
&\geq r\cdot H(x_p) \sum_{j\neq i} b(d(i,j;g)) - x_p\cdot H(x_p) \\
&= H(x_p)\left(r \sum_{j\neq i} b(d(i,j;g)) - x_p\right) \\
&> 0,
\end{aligned}$$

since $h_i(g) > H(x_p)$, $p \leq x_p \cdot H(x_p)$, and $r\sum_{j\neq i} b(d(i,j;g)) > x_p$. Thus, g is not stable. □

Proof of Corollary 1: Since H is the uniform distribution function over $[x_0, x_1]$,

$$x\cdot H(x) = \begin{cases} 0 & \text{if } 0 \leq x < x_0, \\ \frac{x(x-x_0)}{x_1-x_0} & \text{if } x_0 \leq x \leq x_1, \\ x & \text{if } x_1 < x. \end{cases}$$

Note that $x\cdot H(x)$ is continuous on $\mathbb{R}_+$. Since $p \leq x_1$, we have $p = \frac{x_p(x_p-x_0)}{x_1-x_0}$. So, $x_p^2 - x_0x_p - p(x_1-x_0) = 0$. Thus, $x_p = \frac{x_0+\sqrt{x_0^2+4p(x_1-x_0)}}{2}$. By Proposition 6, $h(g) \leq H(x_p)\cdot \mathbf{1} = \frac{-x_0+\sqrt{x_0^2+4p(x_1-x_0)}}{2(x_1-x_0)} \cdot \mathbf{1}$. □

References

[1] V. Bala and S. Goyal, A non-cooperative model of network formation, *Econometrica* **68** (2000), 1181-1229.

[2] V. Bala and S. Goyal, A strategic analysis of network reliability, *Review of Economic Design* **5** (2000), 205-228.

[3] B. Dutta and M. O. Jackson, The stability and efficiency of directed communication networks, *Review of Economic Design* **5** (2000), 251-272.

[4] A. Galeotti, S. Goyal, and J. Kamphorst, Network formation with heterogeneous players, *Games and Economic Behavior* **54** (2006), 353-372.

[5] H. Haller, J. Kamphorst, S. Sarangi, (Non-)existence and scope of Nash networks *Economic Theory* **31** (2007), 597-604.

[6] H. Haller and S. Sarangi, Nash networks with heterogeneous links, *Mathematical Social Sciences* **50** (2005), 181-201.

[7] M. O. Jackson and A. Wolinsky, A strategic model of social and economic networks, *Journal of Economic Theory* **71** (1996), 44-74.

[8] M. O. Jackson and A. Watts, The evolution of social and economic networks, *Journal of Economic Theory* **106** (2002), 265-295.

[9] R. Kannan, L. Ray, and S. Sarangi, The structure of information networks, *Economic Theory* **30** (2007), 119-134

[10] C. Kim and K-C. Wong, Network formation and stable equilibrium, *Journal of Economic Theory* **133** (2007), 536-549.

[11] D. Lopez-Pintado, Contagion and coordination in random networks, *International Journal of Game Theory* **34** (2006), 371-381.