SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks 2012; **5**:197–210
Published online 24 January 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/sec.307

RESEARCH ARTICLE

Could firewall rules be public – a game theoretical perspective

Qi Liao[1]*, Zhen Li[2] and Aaron Striegel[1]

[1] Department of Computer Science and Engineering, University of Notre Dame, IN, U.S.A.
[2] Department of Economics and Management, Albion College, Albion, MI, U.S.A.

ABSTRACT

Firewalls are among the most important components in network security. Traditionally, the rules of the firewall are kept private under the assumption that privacy of the ruleset makes attacks on the network more difficult. We posit that this assumption is no longer valid in the Internet of today due to two factors: the emergence of botnets reducing probing difficulty and second, the emergence of distributed applications where private rules increase the difficulty of troubleshooting. We argue that the enforcement of the policy is the key, not the secrecy of the policy itself. In this paper, we demonstrate through the application of game theory that *public* firewall rules when coupled with false information (lying) are actually better than keeping firewall rules private, especially when taken in the larger group context of the Internet. Interesting scenarios arise when honest, public firewalls are socially insured by other lying firewalls and networks adopting public firewalls become mutually beneficial to each other. The equilibrium under multiple-network game is socially optimal because the percentage of required lying firewalls in social optimum is much smaller than the percentage in single-network equilibrium and the chance of attacking through firewalls is further reduced to zero. Copyright © 2011 John Wiley & Sons, Ltd.

KEYWORDS

public firewalls; game theory; botnet; productivity efficiency security; network management

***Correspondence**

Qi Liao, Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA
E-mail: qliao@nd.edu

1. INTRODUCTION

Firewalls play a significant role in defending an enterprise network security and have been widely adopted in almost every organization [1]. Many security problems associated with networking can be mitigated by deploying a firewall [2] coupled with other security devices such as Intrusion Detection/Prevention Systems (IDS/IPS) and others [3]. Firewalls are and will continue to be important components in enterprise networks to defend against untrusted network intrusion.

Since the inception of the firewall, the general practice of enterprise network administrators has been to hide the firewall configuration information, to which we refer as *private firewalls*. The conventional wisdom is the less information released to the outsider, the better of the security of the network. For the network environment at the time, such an assumption was not entirely unreasonable. Networks by and large were considerably slower and adversaries tended to only have a limited group of machines from which to clumsily probe to try to infer the open network services. In contrast, the environment of the adversaries of today has shifted considerably. Many techniques have emerged that can carefully craft packets for fooling and penetrating firewalls or to reconstruct firewall rules by probing adaptively based on the firewall response [4,5]. Furthermore, the dramatic increase in scale of the general Internet itself and the emergence of botnets have reduced the "cost" of probing, be it in time or machine exposure, to almost nothing.

The private nature of firewall rules is further complicated by significant increases in the complexity and scale of the applications running on the network [6]. Rather than having relatively simple point-to-point requirements, applications have generally trended toward decentralization and distributed dependencies for optimal operation. As an active firewall can present security risks in and of themselves (redirection, etc.), firewalls more often than not will silently discard packets, appearing to be an ambiguous black hole in the network. The net result is that debugging of connectivity problems becomes an administrative nightmare,

i.e., is the problem the local firewall, the enterprise firewall, your firewall, my firewall, the router, the network link, the application, etc.?

In this paper we ask the relatively innocuous question, how costly is it to make firewall rules public[†], i.e., *public firewalls*? Intuitively, a firewall is a group of systems that enforces an access control policy between two networks [2], which implies the policy and its enforcement is the key, not the secrecy of the policy itself. An equivalent analogy would be that the orderness of society relies on the enforcement of laws, but the laws themselves should be public, namely what behaviors are allowed and what are not allowed. Our contribution is to explore the viability of making firewall rules public and how one might transition from private to public firewalls. When the firewall is given the ability to not only provide *true* public information but interestingly *false* information as well, we demonstrate how one can balance the twin demands of security and productivity on the enterprise. Public information (true or false) can be acquired via either rules/policies querying mechanism [7] or reconstructions of rules through probing [4,5], which is further discussed in Section 5.

To accomplish this analysis, we model the dynamics of interactions between the attacker and the firewall/administrator in a game theoretic framework. While game theories [8] have applications in many areas including information security [9–17], our work focuses on the viability of public versus private firewall rules through game theoretic analysis. Game theory is useful in this case because the network defender wants to know how attackers would respond to the transition from private to public firewalls and what constitutes a good strategy between no information, true information, and false information that can increase productivity, efficiency, and security. With a game theoretical framework, equilibrium strategies for both attackers and administrators can be derived.

As will be shown later, the Nash equilibrium [18] analysis suggests a network would either choose to play the pure strategy of telling truth if it emphasizes productivity or play a mixed strategy (true or false) for self insurance, but would *never* choose null information. In other words, keeping firewall rules public (true or false) is always preferred over private firewalls from the perspective of the administrator, where the attacker's probability of attacking through firewall is reduced compared to the private firewall case.

Furthermore, we extend the game model from one network to *multiple networks* where a *social optimum* solution is possible. Networks can now mutually benefit each other as under *social insurance* and do not have to play strategically (mix of true and false) as under self insurance. At this social optimum, all networks are better off, i.e., only pure strategies are adopted by networks, the percentage of lying firewalls is smaller, and the probability of attacking through firewall can be further reduced to zero.

The rest of the paper is organized as follows: Section 2 introduces the general setting of the firewall game by defining players, strategic spaces, and expected payoff matrices. In Section 3, best responses are discussed and Nash equilibria are derived. The analysis culminates in Section 4, in which an ideal social optimum is reached when the game is extended from a single network to multiple networks. Further discussed in this section is putting together various critical values derived in previous sections to form a global view. Section 5 compares the various ways of acquiring the firewall rule information, suggests implementation details on false information, and discusses the limitation and future work. After discussing related work in Section 6, we conclude the work in Section 7.

2. GAME-THEORETIC FRAMEWORK FOR FIREWALLS

In this section, we formalize a game model to capture the dynamic interactions between defenders of private networks and attackers. A general setting of the firewall game is described, followed by an exploration of the strategic spaces for both players. Following these, the important expected payoff matrices are then defined for both administrators and attackers.

We begin by first defining the two interested parties (*players*):

(1) System/network administrators[‡] who represent the interest of organization networks protected behind firewalls.
(2) Attackers who try to compromise machines behind firewalls and conduct malicious activities.

For simplicity of the game setting, we consider only the administrator and the attacker as in a two-player game. The administrator and normal users are considered as the same interested party because the goal of the system and network administrator is to ensure smooth user experience (e.g., increase user productivity) and strong security of an enterprise network.

Without loss of generality, we assume both attackers and network administrators make decisions based upon intelligent considerations of the possible consequences. Therefore, the interaction between the attacker and the administrator can be modeled as a two-player non-cooperative and general-sum game for which the best-response strategies (Nash Equilibria) are computed.

The interaction between the attacker and the administrator is considered as non-cooperative and competitive in the game. In the non-cooperation theory, the attacker and the administrator are unable to communicate before decisions are made. The main non-cooperative solution concept is the

[†] It is important to note firewalls are still enforcing those rules no matter if the rules are public or private information.

[‡] Administrators, networks, and firewalls are sometimes used interchangeably but all refer to the defense side.

DOI: 10.1002/sec

Table I. Strategy space.

(a) Administrator (S_d)	
S_d^{ϕ}	No information
S_d^{T}	True information
S_d^{F}	False information
(b) Attacker (S_a)	
S_a^{f}	Attack (through firewall)
S_a^{ϕ}	Skip attack (through firewall)

strategic equilibrium. The game is not zero-sum because one party's gains are not necessarily equal to the other party's losses. From the perspective of either market value or loss of confidence, the damage to an enterprise network tends to be a noticeably larger hit by a successful attack compared to the gains received by the attacker.

2.1. Strategic space for the administrator

In general, firewalls controlled by the defender or administrator can have several strategies. The network administrator is the informed party with the full knowledge of firewall rules. Regarding the amount of feedback information that firewalls should reveal, firewalls have three strategies to choose from:

(1) *No information* (S_d^{ϕ}): keep firewall rules hidden.
(2) *True information* (S_d^{T}): tell the truth.[§]
(3) False information (S_d^{F}): lie and give false firewall rules.[‖]

The strategy space for the administrator is denoted as $S_d = \{S_d^{\phi}, S_d^{T}, S_d^{F}\}$ as summarized in Table I(a). The last two choices by the administrator are considered public information. An interesting question arises: what if the firewall lies? The false information means all the original rules are inverted, swapped, and/or falsified. For example, a firewall can lie about non-existent services, a lower version number that is known for vulnerabilities, or even wrong operating system (OS) of end hosts by forging returned probing packets (further discussed in Section 5.4). The consequence (gain and loss) of transition from private to public (true or false) firewall rules is discussed in Section 2.4. We briefly mention here that one immediate benefit derived from lying is that the administrator can now track and identify the attacker who is trying to exploit non-existing services or wrong OS that the administrator wrongfully gives out on purpose (e.g., honeypot [19], which may be required only under the False Information). Since the attacker targets at non-existing services, virtual machines, or wrong OS, there is no chance the attack will be successful in compromising the real hosts. Note that we focus on analyzing the consequence of making firewall rules public and its impact on both attackers and defenders in a formal game-theoretic model, and consider the support for false information to be beyond the scope of the paper.

2.2. Strategic space for the attacker

Although attackers' ultimate goals may vary from stealing sensitive information to performing denial of service attacks to sending spams, the first step of such attacks is usually to compromise machines running vulnerable services behind firewalls. Since attackers need to find ways to traverse the firewall and reach target applications/nodes, attackers want to find out what the firewall allows and disallows, i.e., firewall rules.

Attackers have several strategies in response to administrators' strategies discussed in the previous section. Attackers can either choose to attack through the firewall of an organization or simply choose not to attack at all and move on to next target network on the list. Attacking through the firewall means attackers seek to compromise the end hosts through exploiting vulnerabilities of various services. Because this type of attacks depends on the rules/policies of firewalls, it is categorized as attacking through the firewall. There are other attacks, however, that do not depend on the rule sets of firewalls but through other methods such as social engineering, which usually requires user interaction. For example, attackers can send phishing emails to users and trick them to either run attached executable programs or click on a fraud link which will download and run malicious softwares (malware) on users' machines. Although physically that traffic still goes through the firewall, since emails and web traffic are allowed in almost every organization, the above activities can be categorized as "firewall bypassing attacks," a practice not dependent on the various strategies of firewall rules chosen by the administrator and are thus *not* considered in the modeling. Therefore, in face of various and uncertain firewall rule strategies adopted by the administrator, two options are considered for the attacker: to attack through firewall (S_a^{f}) or not to attack through firewall (S_a^{ϕ}), as summarized in Table I(b).

2.3. The attacker's payoff matrix

Before deriving individual payoffs, we first list a generic payoff matrix of the game in Table II. Given a pair of strategies chosen by the attacker and the administrator, μ is the utility function to compute expected payoffs for both parties. The goal of the game is each player chooses a strategy that maximizes his or her expected payoff by taking into account the opponent's decision.

For the attacker, two parameters of reward and cost factors are considered. The attacker considers not only rewards (R) received from a successful attack but also costs and potential risks of the action (C). The cost function of the attacker, $C = c_1 + c_2$, has two components:

[§] Firewall queries may be answered in an efficient way through SFQL and decision trees like data structure [7].

[‖] The false rules can be returned upon querying in the same mechanism as true rules, or forged packets can be returned upon probing.

Table II. The generic payoff matrix of the game.

Attacker, Administrators	No information (S_d^ϕ)	True information (S_d^T)	False information (S_d^F)
Attack (S_a^f)	$\mu(S_a^f, S_d^\phi), \mu(S_a^f, S_d^\phi)$	$\mu(S_a^f, S_d^T), \mu(S_a^f, S_d^T)$	$\mu(S_a^f, S_d^F), \mu(S_a^f, S_d^F)$
Skip attack (S_a^ϕ)	$\mu(S_a^\phi, S_d^\phi), \mu(S_a^\phi, S_d^\phi)$	$\mu(S_a^\phi, S_d^T), \mu(S_a^\phi, S_d^T)$	$\mu(S_a^\phi, S_d^F), \mu(S_a^\phi, S_d^F)$

Table III. Payoff matrix of the game.

Attacker, Administrator	No information (S_d^ϕ)	True information (S_d^T)	False information (S_d^F)
Attack (S_a^f)	$E_a^\phi, P_0 + E_0 + S_0$	$E_a^T, P_0^+ + E_0^+ + S_0^-$	$-c_2, P_0 + E_0^+ + S_0^{++}$
Skip attack (S_a^ϕ)	$-c_1, P_0 + E_0 + S_0^+$	$0, P_0^+ + E_0^+ + S_0^+$	$0, P_0 + E_0^+ + S_0^+$

(1) Preparation stage cost (c_1): most of time the attacker would research and study the target network and try to find way to get in and compromise hosts. This cost includes port scanning, inferring firewall rules, and probing for potential vulnerability of systems.
(2) Contingent cost (c_2): costs related to potential risks of an attack for the attacker to be detected, traced-back, identified, possibly arrested and punished.

After an attack is initiated, four possible consequences may occur: {*succeed & undetected, succeed & detected, fail & undetected, fail & detected*}. Clearly, the consequence "*succeed & undetected*" is most favored by the attacker while the consequence "*fail & detected*" is the least desirable. The ranking of the other two consequences is ambiguous because two opposing effects are in place. On one hand, a successful attack is more advantageous than a failed attack in the view of the attacker. On the other hand, being detected can make gains from attack temporary and short-lived, e.g., the administrator can remove the attacker from the system, recover damage done by the attacker, reinstall the system, or optionally trace back the attacker.

In Table III, E_a^ϕ is the expected payoff for the attacker under the benchmark strategy $\{S_a^f, S_d^\phi\}$, i.e., the attacker attacks through firewall while the administrator provides no information. Considering the likelihood of each consequence for any pair of strategies ($\{S_a, S_d\}$) by the attacker and the administrator, the attacker's expected payoff is the weighted average of the four possible consequences. Let α_i be the probability of each attack consequence i, i.e., in the order of {succeed & undetected, succeed & detected, fail & undetected, and fail & detected}. Under this benchmark strategy, $E_a^\phi = \alpha_1(R-c_1) + \alpha_2(R-c_1-c_2)-\alpha_3 c_1-\alpha_4(c_1 + c_2)$, where $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$. The first two terms consist of both rewards and costs for successful cases while the last two terms are involved with only costs due to failed attacks. E_a^ϕ can be rewritten as:

$$E_a^\phi = (\alpha_1 + \alpha_2)R-(\alpha_2 + \alpha_4)c_2-c_1 \quad (1)$$

E_a^T in Table III is the attacker's expected payoff under the strategy pair $\{S_a^f, S_d^T\}$, i.e., attacker attacks through firewall while the administrator provides true information. Since the probabilities of attack consequences under this strategy pair can differ from the above benchmark strategy pair, a different set of probabilities β_i is used. Therefore, $E_a^T = \beta_1 R + \beta_2(R-c_2) + \beta_3 0-\beta_4 c_2$, where $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$. Since true information is provided, the c_1 cost equals zero. E_a^T can be rewritten as:

$$E_a^T = (\beta_1 + \beta_2)R-(\beta_2 + \beta_4)c_2 \quad (2)$$

The attacker's expected payoff under the pair of strategies $\{S_a^f, S_d^F\}$ is $-c_2$ due to the fact that only consequence "*fail & detected*" would occur if the attacker attacks upon receiving false information. Should the attacker choose to skip the attack (S_a^ϕ), there occurs only the preparation cost $-c_1$ to the attacker with no rewards under the private firewall case. For public firewalls (true or false), the attacker has no gain or loss as no port scanning costs occur. Therefore, the expected payoffs for the attacker under strategies $\{S_a^\phi, S_d^\phi\}$, $\{S_a^\phi, S_d^T\}$, and $\{S_a^\phi, S_d^F\}$ are $-c_1$, 0, and 0, respectively.

2.4. The administrator's payoff matrix

In contrary to the attacker's rewards and costs, the administrator's payoff is three dimensional: *user productivity* (P), *network efficiency* (E), and *system security* (S), by taking consideration of the tradeoff and balance of convenience and security. *Productivity* measures the convenience and easiness for both the enterprise and outside users to collaborate easily so that most of their time can be spent on the actual work, not on the troubleshooting. *Efficiency* regards the effective allocation of network resources to legitimate use only. Port scanning [20] is an important problem for administrators. During scanning, the attacker makes many connection attempts. Consequently, many flows are generated and/or a large number of connections remain open on the end hosts, which in turns waste precious CPU cycles, memory, and bandwidth of the target network [21]. The efficiency metric suggests that most of the network bandwidth or computing resources of an organization should be allocated to only legitimate business, but should not be allocated to illegal, unwanted attacking traffic (such as malicious probing) that affects other legitimate users' quality of service (QoS). The efficiency requirement takes a narrow form of definition: specifically, higher level of malicious probing/scanning implies lower level of network efficiency.

Security refers to the strength and solidity of any distributed system, e.g., sensitive data should not leak out of the network. This means the enterprise network should not be compromised by the attacker, and if there is any malicious and abnormal traffic, there should be a mechanism to detect, identify, and ideally trace back the attacking source. For simplification, the above productivity, efficiency, and security are considered as independent metrics and a total sum in terms of a linear combination is used.

While the above metrics are suitable and important components of administrators, it may be hard to assign specific numeric values to each of the criteria even with empirical study. There has been effort from research community trying to quantify security, for instance, using return on security investment (ROI), which is not the focus of this approach. We show that even without knowing the actual numbers, it is possible to study the *relative* gain/loss[¶] by comparing with the benchmark case. Table III summarizes the expected payoffs for the network administrator. The summation of P_0, E_0, and S_0 is the payoff in the benchmark case (i.e., $\{S_a^f, S_d^\phi\}$), which is the administrator's expected payoff under the current practice of keeping firewall rules hidden.

Compared with the benchmark case, *productivity* of the network will be improved (denoted by the +sign) if true firewall rules become public (S_d^T column) since open rules facilitate debugging by legitimate users. The legitimate users are those users belonging to their organization's network as well as potential collaborators and coworkers outside the organization domain. These users (in contrary to malicious users or attackers) have legitimate reasons to access the content and/or services offered by the organization's network. The key question around the debugging of the connectivity issue among distributed applications is where the packet was dropped. In private firewall cases, since firewalls silently drop packets, essentially acting like *blackholes* in network, troubleshooting becomes an administrative nightmare with moderately large distributed systems.

Generally speaking, the amount of true information released by the firewall is positively related to the productivity of users. Being able to query for firewall rules rather than guessing can be helpful. For example, legitimate users of two research groups want to establish video conferences, which are known legitimate services offered by the enterprise, but the connection fails for unknown reason. Since there are multiple network components between the end-to-end connection, e.g., switches/routers and nested firewalls/NATs, to avoid blaming each other, users can simply query firewalls of both sides to figure out what blocks the connection requests. Being able to quickly troubleshoot and debug greatly increases the overall productivity of enterprise users. In case of false firewall rules (S_d^F column), there is no productivity gain accounted in the payoff for the administrator. Unlike the attacker, legitimate users are supposed to know what legitimate services are available to them in the first place. They should always connect to known services and should never ever try to access those false services. Another fundamental difference between legitimate users and attackers is that the minute users are unable to resolve, they know who to contact and escalate to the administrator. The attacker, on the contrary, does not have the luxury for contacting the administrator. As discussed later, when mixed with "false information", the "true information" should have proportional gain in user productivity. However, to accommodate the worst case, productivity is modeled as no gain over the "no information" case.

Efficiency of the network is also enhanced under public firewall rules (both S_d^T and S_d^F columns) because no matter true or false, the results returned by firewalls nullify the need for malicious port scanning, which saves precious network resources to be solely allocated to legitimate requests. Depending how sophisticated the honeypots can be and to what extent should they trace the attackers, the bandwidth or the hardware required for such systems can be very limited. A few dedicated, inexpensive desktop computers may be sufficient to redirect and handle the attacker's request for false services advertised by the firewall. Therefore, the cost of these minimally featured honeypot-like systems in terms of bandwidth and computing hardware is so drastically low compared with the massive scanning. We argue that the minor efficiency cost associated with the supporting system for the false information is negligible compared with the gain in the bandwidth and computing resources saved from malicious port scanning and probing. As a result, the net gain of efficiency is still considered positive in the case of public false information. Hence, if the firewall tells truth about its rules, legitimate users gain productivity (P) and the network gains efficiency (E), denoted by the plus sign P_0^+ and E_0^+. In contrary, under private firewall rules, an attacker may decide either to attack or to skip based on the port scanning results. Even if he/she decides to skip, network resources are already wasted by the malicious probing. The wasted bandwidth and computing resources for handling those malicious probing requests could otherwise have been used for legitimate requests, hence no efficiency (E) gain from the network perspective even if the attacker decides to move on to the next target organization, i.e., E_0 for $\{S_a^\phi, S_d^\phi\}$.

Security of the network depends on the vulnerability of systems and whether the attacker decides to attack the network. For example, if firewalls always tell truth and the attacker chooses to attack ($\{S_a^f, S_d^T\}$), there is a cost of decreased security denoted by the minus sign S_0^-. However, if the attacker chooses to skip attack and move to other targets because the attacker cannot find any vulnerability even if firewalls always tell truth, then the organization's security will be the same as the skip attack case in the no-information scenario ($\{S_a^\phi, S_d^\phi\}$). Therefore, if the attacker chooses not to attack anyway (S_a^ϕ row), security of the network remains the same in all cases. It is interesting to notice that security is greatly increased (denoted by the ++sign) if the attacker chooses to attack upon given wrong information ($\{S_a^\phi, S_d^F\}$).

[¶] The relative gain or loss compared with the benchmark case is denoted by the superscripts of + and − signs.

When the wrong firewall rules are deliberately released by the administrator, they are passed around bots (much like poison apples/cookies). And when a bot comes back with this wrong information and attacks a non-existent service or a wrong target, not only the attacker will fail, he/she will also be traced and identified, the worst of the four possible attack consequences. Finally, we note that the impacts of different strategies on the relative changes of productivity, efficiency, and security are intuitive and valid assumptions and are in line with generally accepted practices of network administration.

In summarizing the above analysis, Table III lists the expected payoff matrix for both the attacker and the administrator. For $\{S_a^{\phi}, S_d^T\}$, i.e., firewalls tell truth while attackers do not attack, all components of payoff increase, resulting $P_0^+ + E_0^+ + S_0^+$, which is the ideal case for the network administrator.

3. EQUILIBRIUM STRATEGIES

The goal of the game is for each player to choose a strategy that maximizes his or her expected payoff by taking into account the opponent's decision. In this section, best responses for both the attacker and the administrator are analyzed for each case based on the choice made by the other player. We then solve for both pure-strategy and mixed-strategy equilibria of the game.

3.1. Best responses of the attacker

Based on the administrator's strategy, the attacker decides whether to attack or not by comparing expected payoffs of each option (as in Table III). If the administrator plays private information (S_d^{ϕ}), the attacker's "best response" (denoted as b_a) is

$$b_a(S_d^{\phi}) = \begin{cases} S_a^f, & \text{if } E_a^{\phi} > -c_1 \\ S_a^{\phi}, & \text{if } E_a^{\phi} \leq -c_1 \end{cases} \tag{3}$$

From Equation (1), $E_a^{\phi} > -c_1$ implies $(\alpha_1 + \alpha_2)R > (\alpha_2 + \alpha_4)c_2$, suggesting the attacker would choose to attack through firewall when the expected rewards overweigh the expected cost of being detected and traced. Similarly, if the administrator plays public true information (S_d^T), we have

$$b_a(S_d^T) = \begin{cases} S_a^f, & \text{if } E_a^T > 0 \\ S_a^{\phi}, & \text{if } E_a^T \leq 0 \end{cases} \tag{4}$$

meaning when the attacker is fully informed of firewall rules, the attacker's best response regarding whether to attack or to skip depends on his/her expected payoff of launching an attack. Since the net payoff of skipping attack is zero, if the expected payoff of an attack is positive, then the attacker's choice would be to attack. Otherwise, he/she will skip the attack.

If the firewall lies by playing public false information (S_d^F), the attacker's best response is always to skip attack due to non-negative cost, or

$$b_a(S_d^F) = S_a^{\phi} \tag{5}$$

Since the probability for the attacker to launch a successful attack under the public true firewall scenario cannot be smaller than hidden firewall rules, i.e. $(\beta_1 + \beta_2) \geq (\alpha_1 + \alpha_2)$, and there is no need to perform port scanning thus avoiding possible detection by IDS, the probability of being detected and traced may be reduced: $(\beta_2 + \beta_4) \leq (\alpha_2 + \alpha_4)$. Since cost ($c_1$) is non-negative, $E_a^{\phi} \leq E_a^T$ based on Equations (1) and (2), which suggests intuitively the attacker is at least equally well regardless of attacking or not if the administrator always tells truth about firewall rules rather than keeping them as hidden information. On the other hand, although the attacker may be better off with expected payoff changing from E_a^{ϕ} to E_a^T, it does not necessarily mean the administrator is worse off (a win–win situation) since the game is not zero sum (i.e., one party's gain equals the other party's loss). First, even when the attacker chooses to attack under public true information when $E_a^T > 0$, $P_0^+ + E_0^+ + S_0^-$ is not necessarily smaller than $P_0 + E_0 + S_0$ depending on how much tradeoff of productivity, efficiency and security the administrator puts on his/her network. Second, when the attacker chooses not to attack when $E_a^T \leq 0$, possible when the true information implies no vulnerability, the administrator's expected payoff increases from $P_0 + E_0 + S_0^+$ to $P_0^+ + E_0^+ + S_0^+$.

3.2. Best responses of the administrator

In strategic analysis, dominance occurs when one strategy is better than another strategy for one player, no matter how that player's opponents may play. In other words, a dominant strategy always does at least as good as the strategies it dominates. For the administrator, S_d^F is the dominant strategy and S_d^{ϕ} is the dominated strategy since the expected payoff of S_d^F is always greater than the expected payoff of S_d^{ϕ}, i.e., $(P_0 + E_0^+ + S_0^{++}) > (P_0 + E_0 + S_0)$ and $P_0 + E_0^+ + S_0^+ > P_0 + E_0 + S_0^+$, as shown in Table III. Thus, *regardless of the actions of the attacker, the dominant strategy S_d^F is always a better choice for the administrator than the dominated strategy S_d^{ϕ}*. Hence, the current practice of hidden rules is not optimal for the administrator who can at least be better off by switching from hidden rules to always lying.

Table IV describes the degenerated game payoff matrix by removing the administrator's dominated strategy S_d^{ϕ}. An interesting question remains: how would the administrator choose between playing honest or dishonest?

If the attacker plays S_a^f, the administrator's "best response" (denoted as b_d) depends on the tradeoff between productivity and security. Note efficiency is no longer a determinant of the administrator's choice of best response in the degenerated game since E_0^+ is same. Hence if productivity is evaluated no less than security (i.e., $P = (P_0^+ - P_0) \geq S = (S_0^{++} - S_0^-)$), S_d^T is the best response; otherwise, S_d^F is

Table IV. Degenerated payoff matrix of the game.

Attacker, Administrator	True information (S_d^T), 1- p	False information (S_d^F), p
Attack (S_a^f), q	E_a^T, $P_0^+ + E_0^+ + S_0^-$	$-c_2$, $P_0 + E_0^+ + S_0^{++}$
Skip attack (S_a^ϕ), $1-q$	0, $P_0^+ + E_0^+ + S_0^+$	0, $P_0 + E_0^+ + S_0^+$

the best response:

$$b_d(S_a^f) = \begin{cases} S_d^T, & \text{if } P \geq S \\ S_d^F, & \text{if } S > P \end{cases} \tag{6}$$

The best response for the administrator when the attacker plays no attack is certainly always tell the truth, i.e.,

$$b_d(S_a^\phi) = S_d^T \tag{7}$$

3.3. Nash equilibrium

Finding equilibria of the game is important because at the equilibrium, there is no incentive for either the attacker or the administrator to deviate from equilibrium strategies because these are the best payoff they can get. Unlike our proof in Section 4, where the optimal equilibrium can only be pure strategies, in this section the equilibria can be either pure or mixed strategies, i.e., administrators are likely to take a portfolio of true and false information sets while attackers would choose a probability of attacks.

In game theory [8], a *pure* Nash equilibrium [18] is a pair of strategies (S_a^*, S_d^*) for the attacker and the administrator that satisfies

$$\begin{aligned} \mu_a(S_a^*, S_d^*) \geq \mu_a(S_i, S_d^*), \forall S_i \in S_a \\ \mu_d(S_a^*, S_d^*) \geq \mu_d(S_a^*, S_j), \forall S_j \in S_d \end{aligned} \tag{8}$$

while a *mixed* strategy is a *probability distribution* that assigns to each available action a likelihood of being selected. In our degenerated 2×2 payoff matrix (Table IV), given that the mixed strategy Nash equilibrium is defined over a discrete support of just two elements (the two pure strategies), each of the players' mixed strategies can be described by a single number:

(1) $p \in [0, 1]$ as the probability for the administrator to play S_d^F.
(2) $q \in [0, 1]$ as the probability for the attacker to play S_a^f.

A mixed-strategy profile for the game is thus an ordered pair $(p, q) \in [0, 1] \times [0, 1]$, which would be a Nash equilibrium if and only if p is a best response by the administrator to the attacker's choice q and q is a best response by the attacker to the administrator's choice p. Therefore (p, q) is a Nash equilibrium if and only if it belongs to the intersection of the graphs of the best-response correspondence p^* and q^*, i.e., $\{(p, q) \in [0, 1] \times [0, 1] : p \in p^*(q), q \in q^*(p)\}$ (Figure 1).

While rules should be all public under the public firewall scheme, to address concerns that there might be "super secret" rules used *only* by the administrator for management purposes, which are not to be publicized, a quick and easy remedy could be to increase the number of false rules to maintain an equivalent p value. For example, the administrator can release a rule set $\{T-S\} + \{F\}$ upon each firewall query/probe, where $\{T\}$ is the true firewall rule set, $\{S\}$ is the true "super secret" rule set unpublicized, and $\{F\}$ is the false rule set with size $|F| = (|T| \cdot p)/(1-p)$, for $p < 1$. Also since attackers know in advance there exist untruthful firewalls, collusion will not help attackers because once the p value is determined, the actual content of true and false rule sets remains consistent among queries from different attackers or consequent queries from the same attacker.

The attacker's best-response correspondence specifies, for each mixed strategy p played by the administrator, the set of mixed strategies q which are best responses for the attacker. The graph of $q^*(p)$ is hence the set of points $\{(p, q) : p \in [0, 1], q \in q^*(p)\}$.

The attacker's expected payoff for an arbitrary mixed-strategy profile (p, q) is the weighting of each of the attacker's pure-strategy profile payoffs by the probability of that profile's occurrence as determined in Table IV, i.e., $\mu_a(q; p) = (1-p)qE_a^T - pqc_2$. The attacker's best-response correspondence can be found by solving his/her utility

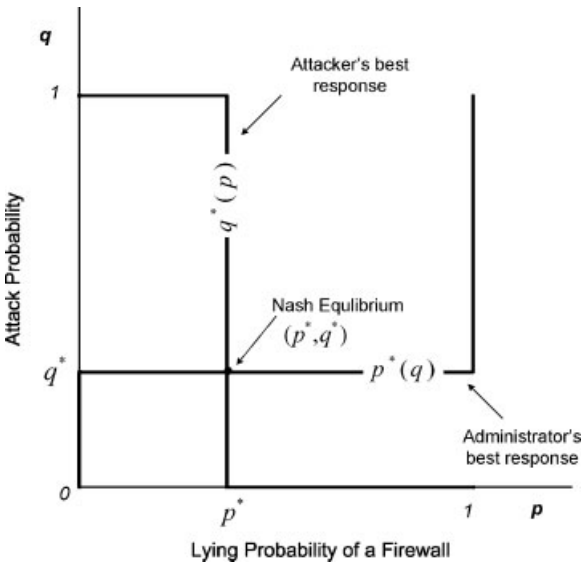

Figure 1. Nash equilibria can be found at the intersection of the best-response correspondence for both the administrator and the attacker. $p^*(q)$ is the administrator's best response when the attacker plays strategically, and $q^*(p)$ is the attacker's best response when the administrator plays strategically.

maximization problem as

$$\max_{q\in[0,1]} \mu_a(q; p) = \delta(p)q \tag{9}$$

where $\delta(p) = E_a^T - (E_a^T + c_2)p$, which vanishes at

$$p^* = \frac{E_a^T}{E_a^T + c_2} \tag{10}$$

Since $\delta(p)$ is decreasing in p, the attacker will choose the pure strategy S_a^f (i.e., $q = 1$) against p's on the interval $[0, p^*)$ and the pure strategy S_a^ϕ (i.e., $q = 0$) against p's on the interval $(p^*, 1]$. Against $p = p^*$, the attacker is indifferent to playing any of his two pure strategies (or any convex combination of them) since they both lead to an expected payoff of 0.

Similarly, the administrator's best-response correspondence can be found by maximizing the administrator's expected payoff for an arbitrary mixed-strategy profile (p, q):

$$\max_{p\in[0,1]} \mu_d(p; q) = \zeta(q)p + \chi(q) \tag{11}$$

where $\zeta(q) = (S_0^{++} - S_0^-)q - (P_0^+ - P_0)$, and $\chi(q) = (P_0^+ + E_0^+ + S_0^+) - (S_0^+ - S_0^-)q$. $\zeta(q)$ vanishes at

$$q^* = \frac{P_0^+ - P_0}{S_0^{++} - S_0^-} \tag{12}$$

according to which, two cases may occur.

3.3.1. Case I: the administrator has a dominant strategy (i.e., pure strategy of *public true* information).

As a trade-off, if the administrator values productivity (P) gains more than security (S), then there exists a Nash equilibrium in which the administrator always plays public true information. This means the administrator will have the same best response for every attack probability q, and the administrator's lying probability $p = 0$ since $\zeta(q) \geq 0$ on [0,1], and the administrator has a strongly dominant pure strategy of S_d^T. Since the attacker's best response to S_d^T is $b_a(S_d^T)$ as in Equation (4), the *pure-strategy* Nash equilibrium of the game is

$$(S_a^*, S_d^*) = \begin{cases} (S_a^f, S_d^T), & \text{if } E_a^T > 0 \wedge P \geq S \\ (S_a^\phi, S_d^T), & \text{if } E_a^T \leq 0 \wedge P \geq S \end{cases} \tag{13}$$

Such a pure-strategy Nash Equilibrium would only exist if S_d^T is the dominant strategy. Notice that false information S_d^F cannot be a dominant strategy for the administrator over true information S_d^T since $q^* > 0$. From Table IV, obviously S_d^T would be a better strategy for the administrator if the attacker chooses not to attack, thus S_d^F cannot dominate S_d^T. To prove by contradiction, if the administrator did choose S_d^F, the attacker would respond by playing S_a^ϕ. If the attacker continued playing the no-attack strategy, the administrator would be better off by switching from S_d^F to S_d^T. The attacker would further respond by playing S_a^f, and so on. Hence no equilibrium solution exists if S_d^F were chosen as a dominant strategy, and S_d^F cannot be a dominant strategy in one network. In next section (Section 4), we will discuss how the situation changes when multiple networks play the game rather than just one network. When the entire Internet is considered, the vast majority of networks can choose public true information while a small percentage of networks play public false information as their dominant strategies, respectively.

3.3.2. Case II: the administrator plays strategically (i.e., a probability distribution of mixed strategies of *public* information, true or false).

Unlike Case I, if an administrator has no preference on productivity over security, it is more likely he/she will create a portfolio of public firewall rules (most true information with some deliberate false information). Similarly, attackers will have their portfolio of attacking targets within the network (with varying attacking probabilities q). Since $\zeta(q)$ is increasing in q, the administrator chooses the pure strategy $p = 0$ against q's on the interval $[0, q^*)$ and the pure strategy $p = 1$ against q's on the interval $(q^*, 1]$. Against $q = q^*$, the administrator is free to choose any mixing probability. In this case, there is a unique *mixed-strategy* Nash equilibrium, i.e., a strategy profile:

$$\left(p^* = \frac{E_a^T}{E_a^T + c_2}, q^* = \frac{P_0^+ - P_0}{S_0^{++} - S_0^-} \right) \tag{14}$$

The Nash equilibrium analysis suggests that unless a network's preference is biased toward productivity (Case I) thus telling truth would be the administrator's dominant choice, the administrator must play strategically (i.e., a mix of true and false information) when facing a tradeoff between productivity and network security. By providing false information at an equilibrium probability of p^*, the network is essentially self insured. On the other hand, the attacker must also play strategically (i.e., a mix of attack and no attack) by having an equilibrium attack probability of q^*, which is smaller than the attack probability under the private firewalls. If either player deviates from the mixed-strategy Nash Equilibrium unilaterally, the deviating party would be worse off with a lower expected payoff. It is interesting to note that a network will choose to either play the pure strategy S_d^T or play a mixed strategy of S_d^T and S_d^F, but will never choose the strategy S_a^ϕ. The equilibria suggest that keeping firewall rules public (true or false) is preferred to keeping them hidden from the perspective of the administrator.

4. SOCIAL OPTIMUM: PUBLIC FIREWALLS AT A LARGER SCALE

While the Nash Equilibrium analysis in Section 3 focuses on a single network, in this section we extend our analysis in an important and interesting direction where more than one network adopt public firewalls. How can strategies chosen by each individual network be mutually beneficial to all participating networks? How would the situation change if *multiple* networks' firewalls go from private to public? How would the *uncertainty* introduced by the mixture of practices across the entire Internet communities change attackers' behaviors? The resulting advantages, as shown in this section, are intriguing. Most importantly, the social equilibrium now becomes the optimal solution. By optimal, it means three things. First, under multi-network game, each individual network does not have to play a mixed strategy of true or false but simply chooses a *pure* public information set to maximize its best interests. Second, the *optimal* equilibrium percentage of lying firewalls P^* in multiple networks is much *smaller* than the p^* in one-network game. Lastly, the optimal equilibrium attack probability drops to *zero* meaning under social optimum attackers will not choose to attack through firewalls.

4.1. Uncertainty matters

To identify the impacts of public firewalls on attackers' decision-making, let us briefly review how the attacker makes decisions under the current practice of hidden information. Under the private information scenario, the attacker will choose to attack if $E_a^{\phi} > (-c_1)$ (Table III). The only uncertainty for the attacker arises from the unknown likelihood of the four possible consequences of an attack regarding success and detection.

The target space for the attacker is unlimited to any specific network and the attacker has a list of choices for target networks that he wants to compromise. This is especially true in multi-network game where many enterprise networks play together when treating the entire Internet as a whole. The attacker as a rational agent would only choose to attack when the reward is greater than the cost to compromise a network, or simply move onto the next vulnerable target that is easier and more profitable to compromise. With the proposed model, any target network may be adopting either strategy S_d^T or S_d^F. Compared to the current practice of hidden firewall rules, another uncertainty occurs: does the network tell me true or false information?

At the presence of sufficient chance of untruthful firewalls from other networks, which acts like social insurance, self insurance by playing strategically (mix of true and false) is no longer necessary. Therefore, to maximize expected payoffs each network would prefer a pure strategy over the mixed strategy. Although it is straightforward for attackers to tell whether firewall rules are hidden or public, it is not evident for them (even under collusion) to tell if any particular rules or packets returned by a firewall are true or false since each firewall returns answers consistently. Converting firewall rules from private to public information makes firewall rules asymmetric information to attackers, which raises an interesting question as how the attacker's behaviors would be affected by the *uncertainty* introduced by such a mixture of firewall practice?

Let P be the chance of a lying network under an N-network game, i.e., $P = \sum_{i=1}^{N} p_i/N$, where $p_i \in \{0, 1\}$, and $(1-P)$ is the probability for a target network to be truth-telling. If the attacker chooses to attack, the expected payoff[#] is $E_a^f = (1-P)E_a^T(P) - Pc_2(P)$. If the attacker skips attack, the expected payoff is $E_a^{\phi} = 0$. The attacker would choose to attack if $E_a^f > E_a^{\phi}$, or if $(1-P)E_a^T(P) > Pc_2(P)$, meaning the following inequality must be satisfied:

$$R > \eta c_2(P) \tag{15}$$

where $\eta = (\frac{P}{1-P} + \beta_2 + \beta_4)/(\beta_1 + \beta_2)$. Under the current practice of hidden information, the attacker chooses to attack if

$$R > \theta c_2(P) \tag{16}$$

where $\theta = (\alpha_2 + \alpha_4)/(\alpha_1 + \alpha_2)$. Comparing Equations (15) and (16), since the left-hand-side is equal, which inequality has a larger solution region depends on the comparison of η and θ. If the former is greater, Equation (15) would have a smaller solution region, thus the chance of attack would reduce when firewall rules go from private to public, or vice versa. Equations (15) and (16) would have the same solution region if $\eta = \theta$, solving which results in

$$\hat{P} = \frac{(\alpha_2 + \alpha_4)(\beta_1 + \beta_2) - (\alpha_1 + \alpha_2)(\beta_2 + \beta_4)}{(\alpha_1 + \alpha_2)(\beta_1 + \beta_3) + (\alpha_2 + \alpha_4)(\beta_1 + \beta_2)} \tag{17}$$

The critical $\hat{P}$ in Equation (17) is the minimum percentage of networks playing pure strategy S_d^F that would make the chance of attacks equal between *private* and *public* firewalls under multiple-network game. If the actual P value is greater than $\hat{P}$ ($P > \hat{P}$), the probability of attack would decrease; or if the actual P value is smaller than $\hat{P}$ ($P < \hat{P}$), the probability of attack would increase.

4.2. Social optimum

Game theory suggests a Nash equilibrium solution does not have to be optimal. The Nash equilibria derived in Section 3.3 for a single network is therefore not necessarily an optimal solution. Under multiple-network game where many networks are involved, a network would prefer a pure strategy over the mixed strategy to maximize its expected payoff.

[#] Expected payoffs are slightly different from Table IV. Both E_a^T and c_2 are functions of P; the former is decreasing in P and the latter is increasing in P.

Therefore, at the presence of sufficient social insurance, it is possible to reach a social optimum equilibrium.

Ideally, we want to find the critical P that would prevent the attacker from launching an attack at all. We define P^* as a social equilibrium, which is the smallest percentage of networks that need to give false information (i.e., play S_a^F) to prevent attackers from attacking (i.e., before an attacker starts playing S_a^f). As shown later, P^* is also socially optimal solution.

For the attacker to play S_a^ϕ, $E_a^f \le E_a^\phi$ needs to hold. The tolerance is the smallest P for which the condition holds, i.e., $E_a^f = E_a^\phi$. This gives the following social equilibrium:

$$\left(P^* = \frac{E_a^T(P)}{E_a^T(P) + c_2(P)}, Q^* = 0 \right) \tag{18}$$

4.3. Why multiple public firewalls are best

Here we put together the equilibria derived in this and previous section (Section 3) to show the important relationships between those critical values, which insightfully explains further why *public firewalls* are superior to *private firewalls*. As an aggregate view, Figure 2 interestingly captures the relations among the critical values. Functions of $q^*(p)$ and $p^*(q)$ under self insurance are the same as in Figure 1 and are included for comparison. Accordingly, q^* and p^* are the attacker/administrator's equilibrium strategies, respectively under single network scenario where all networks are self insured and both the attacker and the network play strategically. For simplicity of illustration a downward-sloping line function $q^*(P)$ is used to show how the attacker reacts to the percentage of networks playing the pure strategy of false information (S_d^F). Moving along the P axis from right to left, the attack probability q increases, and at $\hat{P}$, q becomes equal to $q_0 = Pr(E_a^\phi > -c_1)$, which is the chance of attack under the current practice of private firewalls. However, q_0 will never occur because once P drops below $\tilde{P}$, which is essentially the boundary between self and social insurance, the social insurance would be insufficient and networks will shift to self insurance by playing strategically (S_d^T and S_d^F). But if $P \ge \tilde{P}$, a sign of mutually beneficial and positive externality, networks can choose freely their optimal pure strategies (S_d^T or S_d^F). Therefore, the actual value range for attack probability q is:

$$q = \begin{cases} q^*, & \text{for } P \in [0, \tilde{P}] \\ (0, q^*), & \text{for } P \in (\tilde{P}, P^*) \\ 0, & \text{for } P \in [P^*, 1] \end{cases}$$

While the equilibria analysis in Section 3 suggests public firewalls[**] are superior to private firewalls, Figure 2 also shows a decreased attack probability caused by public firewalls, i.e., $q^* < q_0$. This section proves that multiple public firewalls are even better than single public firewalls by reaching a social optimum. It is important to note that the social equilibrium (P^*, Q^*) under multiple networks is better than the equilibrium (p^*, q^*) in a single network because $P^* < p^*$ and $Q^* < q^*$ (see Figure 2). This means the percentage of required lying firewalls in social optimum is much smaller than the percentage of lying contents in a single firewall in single-network equilibrium. Also, when playing p^* on a single-network basis, the network has a chance of q^* to be attacked. When social insurance is at present, i.e., the game involving all networks playing as one party against attackers, once the percentage of lying networks reaches the socially optimal solution P^*, the network has no chance to be attacked through firewalls, namely $Q^* = 0$. Therefore, social insurance under multiple networks is advantageous than self insurance in single network by further reducing the equilibrium likelihood of attack through firewall to zero. When there exist multiple public firewalls rather than just one, there can be gains in all three aspects, i.e., productivity, efficiency, and security.

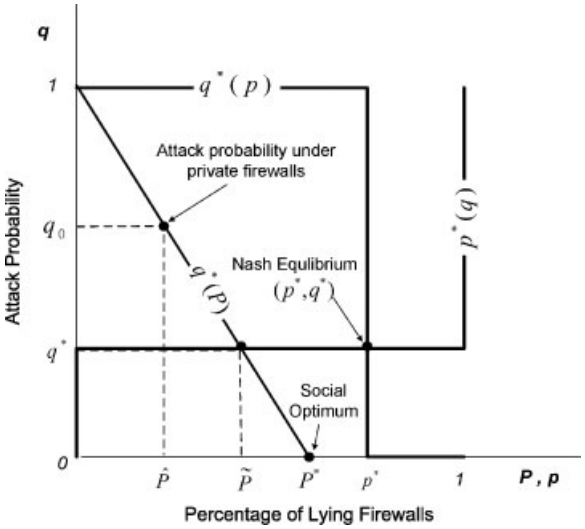

Figure 2. Relationship between important critical values under both single network and multiple network scenarios. See Section 4.3 for explanation.

5. DISCUSSION

In this section, we put theory aside and think critically and expand a few thoughts in practice. How can public firewalls be implemented? For a small percentage of lying firewalls, what type of false information can be used? How about attackers always do probing instead of querying firewalls? Also, a few limitations of theoretical modeling and future work are discussed.

5.1. Probing versus querying

A natural response to public firewalls is that, what if attackers always do probing instead of querying firewalls? We

[**] Throughout the paper, public firewalls refer to public true, public false, or a mixture of both. Public (true only) firewalls are only superior to private firewalls in Case I of single-network games.

want to emphasize that regardless of querying or probing, the returned results purely depend on the truthfulness of firewalls. In other words, attackers have nothing to gain (i.e., the amount of information inferred from probing is less than or equal to that from directly querying for firewall rules) but would only incur extra probing cost.

It is also important to note that querying and probing are just two mechanisms/methods to acquire information. The mechanisms *do not change* two things:

(1) The nature of the firewall, i.e., *private or public* (the public firewalls are still queryable and appear public to the vast majority of the world including legitimate users and administrators).
(2) The nature of firewall information, i.e., *true or false*.

By blindly always choosing to probe, attackers are covering their eyes and ears similar to 'Ostrich Logic' but do not change the nature of public firewalls.

In the context of the model, costless probing is equivalent to having zero preparation component of the cost for the attacker (i.e., $c_1 = 0$). While probing does affect the efficiency component of the administrator's payoff, the efficiency factor becomes irrelevant to the administrator's choice between true or false firewall rules when solving Nash equilibria (Section 3) for the degenerated game (Table IV). Therefore, our entire analysis is unaffected regardless attackers use any mix of querying and probing or just stick to one method only, and all modeling analysis and conclusions remain valid.

5.2. Static versus dynamic

While stateless firewall rules are most commonly deployed due to the efficiency and simplicity, the game theoretic analysis in this paper does not depend on whether firewall rules are stateless or stateful. Therefore, stateful rules should also work for the developed models. To facilitate the modeling and derivation of Nash equilibrium solutions, we abstract the firewall implementation details, i.e., static or dynamic filtering. Rather, we focus on the two fundamental natures of firewall policies, i.e., private or public, and if public, true or false, as described in Section 5.1.

5.3. Limitation and future work

As with all game theoretical analysis, one limitation of our model is rationality assumption, which may not always hold in certain attack models such as state-sponsored attacks. However, since equilibrium is the best attackers can get, irrational attackers not following equilibrium can only have one consequence: being further worse off. Therefore, it is reasonable to assume rational attackers who will decide and behave to the best interests of themselves.

Second, in this paper for simplicity the attacker's strategy space consists of only two strategies, i.e., attack through firewall or not attack through firewall. While the social optimum discussed in Section 4.2 may reduce the probability of attack through firewall down to zero, it does not mean attacks will vanish at all since attackers may shift to other types of attacks such as social engineering. Future work may expand the game model by taking into consideration other types of attacks.

Lastly, for simplification the administrator's expected payoff is an unweighted sum of productivity, efficiency, and security, while in actual practice, different administrators may evaluate the importance of each aspect differently. The balance between productivity and security is a tradeoff that has been widely recognized. On one hand, a fully open system running maximum amount of applications and data provides greatest flexibility for users but at a cost of reduced security. On the other hand, a perfect secure but extremely hard to debug system is hardly of any use. It is reasonable to argue that productivity may be the single most important goal in any organization because ultimately at the end of day all other technical means are just playing supporting roles to increase user productivity so real progress can be made. While productivity should probably have most importance among the three criteria, equal weights are considered in the expected payoff analysis in this study as a starting point. Nevertheless, should productivity have a higher weight, our conclusion will become even stronger that public firewalls are better choices than private firewalls because of the gain of productivity.

5.4. Implementation suggestions

While public firewalls are better than private firewalls in terms of user productivity, network efficiency, and security, they nevertheless pose challenges to security community. One suggestion for implementing public firewalls can be to use a querying mechanism similar to Structured Firewall Query Language (SFQL) structured language described in Ref. [7] for querying firewall rules. While most firewalls in the Internet are truthful, there exist a small percentage of lying firewalls. Although the technical support for false information is considered beyond the scope of the paper, examples of false information that firewalls can release include: lying about non-existing services (ports); lower versions of vulnerable applications (e.g., SSH v1 vs. v2); or even non-existing host IP addresses. The deliberately given wrong information is then circulate within bots controlled by botmasters (similar to poison cookies), and when an attacker comes back trying to contact those "poison" service ports or IPs, it is a strong indication of malicious activities. To support the false information (such as lying about non-existent or lower version of services), consequent service requests may need to be redirected to a dedicated machine similar to a honeypot system [19] to handle the connections. The lying capacity largely depends on the complexity and sophistication of the backend support system. Since firewalls lie *consistently* both in terms of between querying results and probing results and in terms of among different query results over time, collusion between attackers is useless because cooperating and sharing information among bots will not identify the wrong information sets.

At1tackers know from the very beginning that some of the querying and probing results they get from public firewalls will be lies but they do not know which ones are. Finally, we note that while the idea of extra lying capability of public firewalls for insurance purpose is interesting, the main contribution of the paper is not about lying but that private firewall rules are no better than public firewalls.

6. RELATED WORK

Today's practice is to treat firewalls as private information known only to administrators. A firewall may have a large number of conflicting rules [22–24], which can be detected and rectified through policy anomaly detection algorithm [22,23], visualization [25], or firewall queries [7]. Through a SQL-like query language called the SFQL and decision trees as its underlying data structure, firewall queries can be answered in an efficient way and used to assist human users to understand, analyze, maintain and debug firewalls. Although the above firewall querying mechanism can be used in our proposed public firewall framework, the firewall queries described in Ref. [7] are intended primarily for the administrator and neither normal users nor outsiders are able to query the firewalls.

Concerning the validation of false firewall rules, one option may be using the honeypot/honeynet technology [19], which has gained increasing adoption ranging from analyzing malware behavior to tracing botmasters [26–28]. We note that while we focus on private and public rules of firewalls, honeypots *as a mechanism* can be used to support firewalls' lying. On the other hand, the supporting role from the increasing sophistication of honeypot technology further facilities the proposed practice of open firewall rules.

Game theory [8] provides a formal mathematical framework to study complex interactions among interdependent rational players and is therefore useful in the security analysis of computer communication networks. The results from the game are characterized as one or more Nash equilibria [18] that are the best responses for all players. Although the game theory was originally applied primarily to economics, it has been used in many disciplines including recent research activities that apply game theory to model and analyze the security of computer networks [9–17].

For example, game theoretical approaches have been applied to network packet sampling strategy for intrusion detection [12], to graph path and edge model [15], and to infer Attacker Intent, Objectives and Strategies in an incentive-based conceptual framework [13]. Interactions between the attacker and the administrator in a two-player general-sum stochastic game as a nonlinear programming problem are studied in Ref. [14]. However, the administrator's actions consist of only removing compromises done by the attacker. Authors of Ref. [11] model decision-making by homogeneous and heterogeneous users under five game models (total effort, weakest-link, best shot, weakest target with and without mitigation). However, the only players considered are defense agents whose action space consists of either choosing an insurance level or a protection level, which constitutes the utility function to be maximized. Hence the problem is trying to find the equilibrium of insurance and protection levels. Work in Ref. [29] follows a similar track by studying the threshold of buying insurance versus self-protection, and insurance is found to be a powerful incentive mechanism for self-protection.

Game theoretic concepts have been suggested to develop a formal decision and control framework in a distributed IDS [9,10]. While these works focus on the resource allocation problems in intrusion sensor network, the problem we target is the firewall design and game model is used to prove public rules are superior to private ones. In all above cases, however, game theory can provide deeper understanding of complex network dynamics and lead to better design of efficient and robust networks.

We formally analyze the feasibility of keeping firewall rules open, and study the potential of public firewall rules with either true or false feedback in a game-theoretic model. This study extends our earlier work [30] in one important direction: derivation and proof of the existence of a *socially optimal equilibrium* under multiple-network game. Under the single-network scenario, we found there are two Nash equilibria: a pure strategy of public true information and a mixed strategy. While the Equilibrium analysis in single-network scenario suggests public firewalls are superior to private firewalls, when more than one network adopt public firewalls, the outcome is even better, and is actually *optimal* in terms of three things. First, networks no longer have to be self-insured by playing strategically (a probability distribution of true and false information) but rather choose their own preferred pure strategies to maximize their best interest. Second, the percentage of lying firewalls in the Internet is smaller than that in the single network game. Finally, the chance of attack through firewalls is further reduced to zero. Therefore, the impact of expanding single-network scenario to multi-network game is significant.

7. CONCLUDING REMARKS

To summarize, when managing firewalls, conventional wisdom has held that firewall rules should remain hidden in order to improve security. With the emergence of botnets and distributed applications, we argue that such wisdom is no longer valid. In this paper, we provided arguments to question the benefits of private firewall rules and took initial steps to explore the viability of public firewall rules. Through the application of a game theoretic analysis, we showed that public firewall rules, when coupled with the ability to provide false information, can indeed increase productivity, efficiency, and security. When multiple networks adopt public firewalls, a socially optimal equilibrium can be reached, where vast majority of truthful firewalls are insured by smaller percentage of lying firewalls and the chance of attack through firewalls is further reduced to zero while enjoying all other benefits from going public. The idea of public firewalls poses challenge and opportunity to network

security community and it is our hope that our exploration offers an unconventional yet promising approach to this important problem and future firewall design.

REFERENCES

1. Stiemerling M, Quittek J, Eggert L. NAT and firewall traversal issues of host identity protocol (HIP) communication. *Network Working Group Request for Comments (RFC) 5207*, April 2008.
2. Cobb S. Establishing firewall policy. In Conference Record of Southcon '96, Orlando, FL, June 1996; 198–205.
3. Turner D, Fossi M, Johnson E, *et al*. Symantec global internet security threat report – trends for July--December 2007. *Symantec Enterprise Security XIII*, April 2008.
4. Samak T, El-Atawy A, Al-Shaer E. Firecracker: a framework for inferring firewall policies using smart probing. In IEEE International Conference on Network Protocols, Beijing, China, October 2007; 294–303.
5. Samak T, El-Atawy A, Al-Shaer E, Hong L. Firewall policy reconstruction by active probing: an attacker's view. In *2nd IEEE Workshop on Secure Network Protocols*, November 2006; 20–25.
6. Chapple MJ, D'Arcy J, Striegel A. An analysis of firewall rulebase (mis)management practices. *Journal of Information System Security Association (ISSA)* 2009; **7**: 12–18.
7. Liu AX, Gouda MG, Ma HH, Ngu AH. Firewall queries. In *8th International Conference on Principles of DIstributed Systems (OPODIS'04)*, Springer LNCS (3544) 2005; 197–212.
8. von Neumann J, Morgenstern O. Theory of Games and Economic Behavior. Princeton University Press, 1944.
9. Alpcan T, Basar T. A game theoretic approach to decision and analysis in network intrusion detection. *Proceedings of 42nd IEEE Conference on Decision and Control* 2003; **3**: 2595–2600.
10. Alpcan T, Basar T. A game theoretic analysis of intrusion detection in access control systems. In *Proceedings of 43nd IEEE Conference on Decision and Control* 2004; 1568–1573.
11. Grossklags J, Christin N, Chuang J. Security investment (failures) in five economic environments: a comparison of homogeneous and heterogeneous user agents. In Proceedings of Workshop on the Economics of Information Security (WEIS '08), Hanover, New Hampshire, June 2008.
12. Kodialam M, Lakshman TV. Detecting network intrusions via sampling: a game theoretic approach. *IEEE INFOCOM* 2003; **3**: 1880–1889.
13. Liu P, Zang W. Incentive-based modeling and inference of attacker intent, objectives, and strategies. In *Proceedings of the 10th ACM conference on Computer and communications security, SESSION: Information Warfare* 2003; 179–189.
14. Lye K-W, Wing JM. Game strategies in network security. *International Journal of Information Security* 2005; **4**: 71–86.
15. Mavronicolas M, Papadopoulou V, Philippou A, Spirakis P. A graph-theoretic network security game. In *First International Workshop on Internet and Network Economics (WINE'05), Springer LNCS 3828* 2005; 969--978.
16. Sallhammar K, Knapskog S, Helvik B. Using stochastic game theory to compute the expected behavior of attackers. In *Proceedings of the 2005 Symposium on Applications and the Internet Workshops (SAINT-W'05)* 2005; 102–105.
17. You XZ, Shiyong Z. A kind of network security behavior model based on game theory. In *Proceedings of the Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'03)*, August 2003; 950–954.
18. Nash J. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences* **36**(1): 1950; 48–49.
19. Bächer P, Holz T, Kötter M, Wicherski G. Know your enemy: tracking botnets. In *The Honeynet Project & Research Alliance*, March 2005; www.honeynet.org/papers/bots.
20. Jamieson S. The ethics and legality of port scanning. SANS Institute 2001.
21. Gu Y, McCallum A, Towsley D. Detecting anomalies in network traffic using maximum entropy estimation. In *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement*, 2005; 345–350.
22. Al-Shaer E, Hamed H, Boutaba R, Hasan M. Conflict classification and analysis of distributed firewall policies. *IEEE Journal on Selected Areas in Communications* 2005; **23**, 2069–2084.
23. Ferraresi S, Pesic S, Trazza L, Baiocchi A. Automatic conflict analysis and resolution of traffic filtering policy for firewall and security gateway. *IEEE International Conference on Communications* 2007; 24/28: 1304–1310.
24. Gouda MG, Liu AX. Structured firewall design. Computer networks. *The International Journal of Computer and Telecommunications Networking* 2007; **51**: 1106–1120.
25. Tran T, Al-Shaer E, Boutaba R. Policyvis: firewall security policy visualization and inspection. In *Proceedings of the 21st large Installation System Administration Conference (LISA '07)*, Dallas, TX, November 2007.
26. Grizzard JB, Sharma V, Nunnery C, Kang BB, Dagon D. Peer-to-peer botnets: overview and case study. In

First Workshop on Hot Topics in Understanding Botnets (HotBots07), Cambridge, MA, April 2007; 1–1.
27. Rajab MA, Zarfoss J, Monrose F, Terzin A. A multifaceted approach to understanding the botnet phenomenon. In *6th ACM SIGCOMM Conference on Internet Measurment, SESSION: Security and Privacy* 2006; 41–52.
28. Zou C, Cunningham R. Honeypot-aware advanced botnet construction and maintenance. In International Conference on Dependable Systems and Networks, Philadelphia, PA, June 2006; 199–208.
29. Bolot JC, Lelarge M. A new perspective on internet security using insurance. *IEEE INFOCOM* 2008; 1948--1956.
30. Liao Q, Li Z, Striegel A. Information game of public firewall rules. In the Fifth Workshop on Secure Network Protocols (NPSEC '09) in Conjunction With the 17th IEEE ICNP, Princeton, NJ, October 2009.