A Social Mechanism for Supporting Home Computer Security*

Rick Wash
School of Information
University of Michigan
rwash@umich.edu

Jeffrey K. MacKie-Mason
School of Information
University of Michigan
jmm@umich.edu

Abstract

Hackers have learned to leverage the enormous number of poorly protected home computers by turning them into a large distributed system (known as a botnet), making home computers an important frontier for security research. They present special problems: owners are unsophisticated, and usage profiles are varied making one-size-fits-all firewall policies ineffective. We propose a social firewall that collects security decisions and both user and usage characteristics, and provides users with personalized information to assist with allow/deny recommendations. To succeed, a social firewall must deal with at least three user behavior issues: why contribute private information? why make effort to provide quality information? and, how to prevent manipulation by adversaries? We sketch an *incentive-centered design* approach to each problem. We provide an economic model and some analytic results for a solution to the fundamental problem: why contribute? We show that an excludable public goods mechanism can achieve a better outcome than a system without social motivators.

1 Home Computer Security

For many years, hackers targeted computers approximately in proportion to the information accessible on or from them. Consequently, security research focused on attacks against large enterprise systems, protected by sophisticated human and technical resources. More recently, hackers developed the botnet, hosted by large numbers of insecure computers. These machines are targeted because they rarely are protected by either sophisticated human or technical resources.

To build a botnet, an attacker hacks (say, through a virus or worm) into many computers and installs "control" software. The controlled client (zombie) listens for commands from a "master control" computer. A single command from the hacker to the master can then be carried out by all active zombies. Botnets enable crimes such as spam, click fraud, and distributed denial of service [17]. Observed botnets range in size from a couple hundred zombies to 50,000 or more zombies [1].

Since any computer with an Internet connection can be an effective zombie, hackers logically turned to attacking the most vulnerable population: home computers. Home computer users are usually untrained and have few technical security skills. While some software has improved the average level of security of this class of computers, home computers still represent the largest population of vulnerable computers with decent Internet connections. Existing security research has not made substantial progress on solving the problems of attacks against machines maintained by unsophisticated users.

Botnets are a serious problem. Nine of out ten email messages are spam [16], and 80% of those messages are being sent through botnets [15]. Botnets also are used to steal personal information, and to conduct multiple outbound crimes such as such as click fraud and trust fraud [15], and extortion of "protection" payments under threat of denial of service attacks [10]. As of 28 March 2007, the Shadow Server website[1] was tracking over 1.5 million active zombies attached to over 1300 distinct botnets.

2 A Social Firewall

A recent HoneyNet study of botnet activity reports that botnets are much more active during the day than at night [21]. Over 50% of the methods that the botnets used to spread themselves were exploits of three well-known security holes: ASN1, DCOM, and LSASS.[2] Patches have

*This material is based upon work supported by the National Science Foundation under Grant No. CNS 0716196.

[1] http://www.shadowserver.com

[2] See Microsoft security bulletins MS04-007, MS03-026, MS-04-011 respectively.

existed for these vulnerabilities for months, yet many computers remain vulnerable. From these facts we infer botnets are succeeding largely by infecting unpatched and home computers.

Most home users, lacking the expertise and inclination for security, attempt to delegate their security concerns. They choose to delegate decisions to a technology (e.g., a firewall), a person (e.g., a knowledgable colleague), a local support group, or an institution (e.g., a bank) [5]. We focus in this paper on firewalling, fully aware that this alone is unlikely be a complete solution.

Firewalls are *intrusion detection and prevention systems* (IDPS) [14]. Their design goal is to prevent unauthorized users from taking unwanted actions on a computer. They monitor computer activities and can stop certain technical actions in order to implement a *policy* stating which activities are acceptable and which are not. For example, a firewall might monitor network packets and can block those directed to particular TCP ports.

Firewalls have several technical design challenges: accurately identifying behaviors, offering a policy language flexible enough to map owner preferences over behaviors into actions, and preventing compromise of the firewall itself, among others. One especially difficult challenge for unsophisticated users is to configure the firewall policy appropriately. Yet security experts know what types of machines are being compromised, and how: why are we not stopping the problem?

We believe an important part of the problem is that, particularly for home computers, we try to use one-size-fits-all solutions. The software has the difficult job of detecting malware while letting through all legitimate uses of the computer; but this distinction varies across users. For example, most home computers do not need to be listening for external DCOM connections. However, some do, so it is important that standardized anti-virus tools not flag this activity as a problem. In trying to find a lowest-common-denominator security policy, standardized policies are too weak to stop malicious activity.

A common approach to the customization problem is *interactive policy generation* (see, e.g., ZoneAlarm). For example, a firewall policy may specify one of three actions for any given event (including a default for not otherwise specified events): allow, deny, or ask the user for a decision in real time. Interactivity offers flexibility and convenience for users who may want to postpone decisions until a event is encountered. However, it does not solve the underlying problem: the unsophisticated user may not know the risks or the benefits of permitting certain requests, and may face high costs of acquiring that knowledge, thus preventing effective policy decisions in the face of potential security risks.

We draw these threads together to propose a new type of home computer IDPS: a *social firewall*. For our purposes, a social firewall can be conceived as an application-layer service running on top of a standard kernel-level firewall. The idea is simple: provide an automated, interactive way for a user to apply the processed, collective knowledge of a community of trusted users to interactive policy decisions. For example, when presented with an interactive allow/deny query, the social firewall might also offer information about other users who were presented with similar policy decisions:a statistical summary of the allow/deny decisions made by others, and/or contributed information characterizing the reasons for those decisions. This information can help the user determine which decision best fits *her* needs, allowing her to customize her policy appropriately.

Getting participants to *contribute* information to the server is only one step towards an effective social firewall. If the only information needed were the allow/deny decision, then the firewall software could be configured to send that bit, and we would be done (with this set of behavioral issues). However, as we explained above, one-size-fits-all policies are not effective, and thus the social firewall users need additional information to assist in personalization of policy.[3] For example, users could contribute their own reasoning for making the allow/deny decision, which other users can evaluate to see if it applies to them. Additionally, users could contribution something about her security sophistication, risk tolerance, or typical computing activities, all of which are relevant to policy decisions. By enabling users to share this type of information, a social firewall encourages better and more personalized policies.

A social firewall builds on the observed preference of home users to delegate security solutions to technology, but also addresses the need for customized policies. However, there are critical design issues to address before the potential benefits will be realized. In the next section we describe several of these issues, and our method for approaching them.

3 Incentive Centered Design

There are numerous challenges in designing a social firewall, including interface and software engineering problems. However, the novel feature of our social firewall is the harnessing of *knowledge-sharing effort by humans*, so we focus on behavioral issues critical to this effort. Knowledge-sharing depends on human behavioral choices concerning participation, effort, quality-control, and trustworthiness, among others.

[3] In addition, when people rely on aggregating past binary decisions, the problem of *information cascading* can lead to disastrously bad outcomes where everyone follows the crowd even when it is wrong; binary predictions must be supplemented with outcome data or useful contextual data to avoid this [3].

In [18, 19] we argued that "humans are 'smart components' in a system, but cannot be directly programmed to perform; rather, their autonomy must be respected as a design constraint and incentives provided to induce desired behavior." A social firewall is a system whose performance depends unavoidably on human choices about how to interact with the system, and how to interact with other humans using the system. Therefore, we apply the methods of *incentive-centered design* (ICD). In system design we directly include humans as smart, distributed and — crucially — autonomous components, with their own information sets and motivations. We draw primarily on microeconomics, game theory, social and cognitive psychology to model motivations, individual responses to them, and inter-individual strategic awareness and behavior.

We now discuss three fundamental behavioral issues for the design of a social firewall system: "getting stuff in" (contribution), "getting good stuff" (quality), and "keeping bad stuff out" (manipulation).

3.1 Getting stuff in

Sharing knowledge among social firewall participants requires that autonomous, self-motivated individuals voluntarily contribute their knowledge. Knowledge is *nonrivalrous*: its use by one person does not materially reduce its value for use by another person. To use a familiar example, once National Public Radio broadcasts a program over electromagnetic spectrum, consumption by one listener does not crowd out consumption by another listener. For information, nonrivalry is generally true because the incremental costs of (digital) reproduction and distribution are approximately zero, and thus multiple instances of the information can be "consumed" without "using it up". Nonrivalry is a defining characteristic of a *public good* [2].

Thus, knowledge contribution to a social firewalling system is a problem in the *private provision of public goods*, which generally results in underprovision [13, 2]. (The social psychology literature refers to roughly the same problem as "social loafing".) Given the priority of contribution (quality is of little consequence if an insufficient number are contributing), this is the problem on which we focus in this paper.

Theorists propose several approaches to raise the level of private contribution [7, for example]. Each faces practical problems in any application, but in particular the solution for social firewall knowledge-sharing is not straightforward, for three reasons. First, barring external incentives such as monetary payments, those who already have the knowledge may already enjoy most or all of its benefits without incurring the cost of sharing.[4] A user already knows what she knows, and so adding her knowledge to the database does not benefit her.

Second, botnet software has become rather sophisticated. One common feature is that it watches the system state, and waits to put a significant load on the system until most resources are idle. Thus, the local user bears little of the cost of having a zombie, since her processes are not starved for resources. She may not even notice her machine has been infected, and if she does, she may not have much incentive to clean it. Thus, the botnet is a classic externality problem: the costs are borne by others (who suffer from spam, denial of service attacks, etc.), not by the person who is "causing" the problem.

Third, nearly all proposed contribution mechanisms are for contributions made in a numeraire — such as money — that is homogeneous and additive, neither of which are characteristics of information.[5] Therefore, we focus on mechanisms that rely on non-monetary but extrinsic incentives to motivate knowledge-sharing. In particular, we propose and analyze an *excludable public good* mechanism, in which users are excluded from obtaining the benefits of the social firewall unless they make sufficient contributions of their own.[6]

3.2 Getting good stuff

As we mentioned earlier, different types of information might be useful to users of the social firewall, such as policy decisions, reasoning for those decisions, and information about similar users. However, rants, conspiracy theories, and random musings, common on many discussion boards, are less useful. Usefulness also depends on accuracy and clarity. The problem of getting people to not only contribute *quality* as well as quantity activates at least two problems: evaluation (what is the quality of a contribution?), and a behavioral problem known as *hidden action* (or *moral hazard*): how do we induce participants to provide desirable quality when we cannot observe or directly command their effort?

3.3 Keeping bad stuff out

Any security method needs to be resistant to manipulation. For example the bot-herder will want to masquer-

[4]The private benefit of having a database tool to store the history of one's own decisions and reasons for them may be valuable enough to motivate contributions; cf. our findings on private motivations to store bookmarks in `del.icio.us` [20]. Since storing notes is not a difficult operation for unsophisticated users, we assume that this motivation is not enough to induce efficient contributions to the social firewall database.

[5]*Mechanism* is a term of art referring to a set of rules about messages agents can pass to a mediator, and the mapping of messages to actions taken by the mediator. An auction's bidding and allocation rules are a familiar example.

[6]Many will be familiar with a popular recent application of this approach: the tit-for-tat throttling mechanism in BitTorrent [4].

ade as a trusted member in order to bias security recommendations towards allowing its bots to infect machines.

Manipulation is an instance of a more general class of *hidden information* problems that we refer to as pollution. A participant has personal goals that are unrelated to the community's goals, and in the course of pursuing her goals, she causes harm to others.[7] The underlying problem is that the agent requesting the computer to take actions knows her actions are undesirable, but the computer's owner does not: this private knowledge is the hidden information ("I am a hacker"). Keeping bad stuff out is a common problem in social computing settings because contribution platforms tend to be open. Email, for example, is a rather open system and spammers regularly use it to distribute unsolicited bulk advertisements.

We have shown that many security techniques are instances of a *screening mechanism*, which is an incentive-centered design to sort good participants from bad [18, 19]. For example, passwords, CAPTCHAs and other challenge-response systems are screens, and their success depends on designing them so that the incentive compatibility conditions are satisfied. There are other types of incentive mechanism to consider when designing for manipulation resistance.

4 Inducing Contribution

In this paper, we focus on the first major behavioral problem for a social firewall: encouraging users to contribute their information. We propose an *excludable public goods* mechanism. If participants do not meet or exceed a threshold for contribution activity, they are denied access to the information in the social firewall database. While the proposal is simple, the associated behaviors are not. For example, if the threshold is increased, some users will contribute more, but others may stop participating, so the net effect is not obvious. We formalize this mechanism and analyze its use with a mathematical model that makes explicit the strategic interplay between the participants. Our modeling is intentionally stylized to allow us to capture the main ideas yet obtain analytic results.

4.1 Model

The fundamental novel feature of a social firewall is a repository of information about policy decisions and participant characteristics. When a user is asked to make a policy decision, she is presented with this information (or a useful summary). After making a decision she can report her decision, and some information about it.

[7]This is another example of a negative externality, similar to the problem of the home computer user who allows her machine to be colonized because she is not seriously harmed.

We assume there is a continuum of users indexed by $i \in [0, N]$. Each user i chooses an amount of information to contribute, x_i.[8] Each user receives some value from using this repository $v(X)$, where $X = \int_0^N x_i \, di$ is the total information in the repository.

Contributing is not costless, as it requires time and effort. We assume each user's cost can be represented by a standard cost function $c_i(x)$ which is increasing, convex ($\frac{\partial c_i}{\partial x} > 0$, $\frac{\partial^2 c_i}{\partial x^2} > 0$), and has $c_i(0) = 0$ for everyone. We assume individual costs are ordered such that for all x, either $c_i(x) > c_j(x)$ or $c_i(x) < c_j(x)$ if $i \neq j$. Without loss of generality, we assume that users with low numbers have the highest cost functions: $c_i(x) > c_j(x) \forall x$ iff $i < j$, or equivalently that $\frac{\partial c_i}{\partial i} < 0$.

This is an excludable system: the designer sets a threshold t of contributions such that if, and only if, users contribute greater than t information, they are permitted to use the repository. The total benefit (utility) a user receives from the repository if she contributes x_i information, when everyone else contributes $x_{-i} = \int_{j \neq i} x_j \, dj$, is given by:

$$u_i(x_i; x_{-i}) = \begin{cases} v(X) - c_i(x_i) & \text{if } x_i \geq t \\ -c_i(x_i) & \text{if } x_i < t \end{cases} \quad (1)$$

4.2 Underprovision

We are modeling a standard public good: the consumption of X is nonrivalrous: everyone benefits from all of the information, with no decrease in its value if the number of participating consumers increases. Therefore, we can establish that with voluntary provision, the public good will be *underprovided*: the equilibrium size of the database will be smaller than the socially optimal size.

To find the benchmark social optimum, we imagine that an omniscient social planner can assign contribution levels to each participant to maximize unweighted aggregate social welfare: $\{x_i^+, t^+\} = \text{argmax}_{\{x_i, t\}} U \equiv \int_i u_i(x_i, x_{-i})$. It is trivial to show the optimal threshold is $t^+ = 0$ (everyone consumes), since there is no social cost from consuming once the repository is created.

To find the voluntary non-exclusionary equilibrium, again set $t = 0$, but now let each individual choose her own contribution, x_i^0, to maximize her own utility function. She contributes only to the point at which her marginal benefit equals her marginal contribution cost. In contrast, in the social optimum each individual contributes as long as the sum of *everyone's* marginal benefit equals the individual's marginal cost, which yields

[8]To implement, we need a metric for information quantity. As long is there is some reasonable measure that is correlated with an improvement in policy decisions, our qualitative results go through. For example, x might be the number of policy decisions about which the user reports.

(weakly) more provision from each individual.

Proposition 1 *The optimal size of the database $X^+ = \int_i x_i^+ \, di$ created by a social planner is larger than the size of the database $X^0 = \int_i x_i^0 \, di$ that would be voluntarily provided: $X^+ > X^0$.*

Proof: Mathematical proofs are provided in the Appendix.

4.3 An Exclusion Equilibrium

With voluntary contributions and no exclusion from use, the repository will be smaller than ideal. If we set a minimum contribution level, can we increase social welfare by inducing a larger repository? The answer is not obvious, because some users will forgo the repository rather than contribute more.

Suppose that a user expects the total contribution of other users to be fixed at $\bar{x}_{-i}$, regardless of her contribution level. For expository purposes, define $f(x_i, \bar{x}_{-i}) = v(\int_j x_j \, dj) - c_i(x_i)$ to be the total utility to user i from using the repository Then we can define conjectural unconstrained contribution levels, $\hat{x_i} = \text{argmax}_{x_i} f(x_i, \bar{x}_{-i})$. Now we can obtain a user's best response (to $\bar{x}_{-i}$) contribution:

Lemma 1 *User i's optimal contribution x_i^* will be:*

$$x_i^* = \begin{cases} \hat{x_i} & \text{if } \hat{x_i} \geq t \\ t & \text{if } \hat{x_i} < t \text{ and } v(\int_j x_j \, dj) - c_i(t) \geq 0 \\ 0 & \text{if } \hat{x_i} < t \text{ and } v(\int_j x_j \, dj) - c_i(t) < 0 \end{cases}$$

If the user wants to contribute more than the threshold, she can and will do so. Otherwise, she chooses between contributing the minimum to gain access, and contributing nothing (and receiving nothing). If her total utility from contributing enough to get repository access is greater than from not participating (which we earlier normalized to 0), she contributes the minimum, even though the marginal cost of contributing t is higher than the marginal benefit, and she would prefer to contribute less.

Now we characterize a Nash equilibrium: with users ordered by costs, there is a natural partition into three groups, corresponding to the possible contribution levels.

Lemma 2 *For any threshold level t there exists a Nash equilibrium of contributions characterized by an i^0 and an i^* such that*

$$\begin{aligned} x_i^* &= 0 && \text{if } i \leq i^0 \\ x_i^* &= t && \text{if } i^0 < i < i^* \\ x_i^* &= \hat{x_i} && \text{if } i > i^* \end{aligned}$$

4.4 Optimal Exclusion Rule

In Lemmas 1 and 2 we characterize behavior given a specified exclusion level, t. Is using such an exclusion rule worthwhile? That is, is there an exclusion level $t > 0$ such that social welfare (the sum of everybody's utility) is higher than at $t = 0$ (as it is under non-exclusionary voluntary provision)?

Proposition 2 *If there is at least one free-rider who, under voluntary provision, strictly prefers to receive access to the repository, yet contributes nothing, then there exists a $t > 0$ that increases social welfare relative to $t = 0$.*

Now we know that an excludable public goods mechanism can outperform non-exclusionary voluntary provision. We next characterize the optimal threshold :

Proposition 3 *At the socially optimal t^* the sum of marginal gains to everyone who retains access are equal to the sum of marginal costs incurred by all of those who contribute just the threshold amount, t^*.*

Even though it is costless to let everyone access the repository and benefit from its security policy recommendations, we are better off to use an excludable public goods rule that imposes a minimum contribution. The reason is simple: without the incentive ("pay to play"), participants will undercontribute, and fully or partially free-ride on the contributions of others. The social firewall will be more socially valuable if it is at least potentially exclusive. We can provide guidance on setting the exclusion level: set aggregate marginal benefits equal to aggregate marginal cost. That is, at the optimal t^*, a small increase in t would cause the cost increase to those people who contribution the minimum to be equal to the increase in benefits to everyone who still receives benefits. The critical issue is that marginal benefits are only experienced by those wiling to contribute at least the minimum, and finding the optimal threshold in practice requires balancing the incremental gains from higher contribution levels against the loss of contributions from those who decline to participate at higher threshold levels.

5 Discussion

A social firewall can improve home computer security. It leverages two facts: home users prefer to delegate some security decision making, and home firewall policies should be customized. However, a working social firewall requires the solution of several design problems. We focus on a subset of these problems: those that arise because users are autonomous and motivated humans.

We identify problems of contribution, quality, and manipulation.

In this paper we particularly focus on the first, and fundamental problem: how to design a social firewall so that people will contribute. We proposed an excludable public goods mechanism that requires no monetary transactions. We showed that when available content is sufficiently valuable, some participants contribute at least a threshold amount rather than free-ride. Further, the optimal exclusion threshold is greater than zero, and we characterized the trade-off that determines how it should be set.

For quality, we are developing a method that draws on social psychology and computer science. The first component is a reputation system to provide both quality evaluation and a social comparison benchmark [6]. A long literature characterizes ways in which upward (e.g., "leader board") and downward comparisons can be structured to motivate attention to quality (see, for a recent example, [8]). Second, we can reduce the burden on the user to provide textual data to support personalization by mining system usage data. For example, given a user's history of allow/deny decisions, the server could generate recommendations based on the behavior of a cluster of similar others (see, e.g., [11]).

For manipulation, we could pursue a screening mechanism such as we described in [18, 19]. However, we think a more promising approach for a social firewall may be an *influence limiter* [12]. This is an algorithm that is provably resistant to a broad class of attack strategies. In particular, an influence limiter resists Sybil attacks by limiting how much influence each Sybil has on the recommendation.

When we finish developing these components, we will have a social firewall that can be layered on an existing interactive firewall. It will address the core behavioral problems (motivations to contribute both quantity and quality, and to discourage manipulation) through a combination of incentive-centered designs (excludable public goods mechanism; reputation system; collaborative filtering; and an influence limiter). We plan to test a prototype in our human-subjects experiment lab, and a refined version in the field. We do not claim that a social firewall will eliminate the botnet problem by itself, but by matching the design to the pivotal human behavior problems, we hope to contribute a valuable part of the solution.

References

[1] BARFORD, P., AND YEGNESWARAN, V. An inside look at botnets. In *Special Workshop on Malware Detection,* (2006), Advances in Information Security, Springer-Verlag.

[2] BERGSTROM, T., BLUME, L. E., AND VARIAN, H. On the private provision of public goods. *Journal of Public Economics 29*, 1 (January 1986), 25–49.

[3] BIKHCHANDANI, S., HIRSHLEIFER, D., AND WELCH, I. A theory of fads, fashion, and cultural change as information cascades. *Journal of Political Economy 100*, 5 (October 1992), 992–1026.

[4] COHEN, B. Incentives build robustness in bittorrent. In *Workshop on the Economics of Peer-to-Peer Systems* (June 2003).

[5] DOURISH, P., GRINTER, R., DE LA FLOR, J. D., AND JOSEPH, M. Security in the wild: User strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing 8*, 6 (November 2004), 391–401.

[6] FESTINGER, L. A theory of social comparison. *Human Relations* (1954).

[7] GROVES, T., AND LEDYARD, J. Optimal allocation of public goods: A solution to the "free rider" problem. *Econometrica 45*, 4 (May 1977), 783–809.

[8] LING, K., BEENEN, G., LUDFORD, P., WANG, X., CHANG, K., COSLEY, D., FRANKOWSKI, D., TERVEEN, L., RASHID, A. M., RESNICK, P., AND KRAUT, R. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication 10*, 4 (2005).

[9] MILGROM, P., AND SHANNON, C. Monotone comparative statics. *Econometrica 62*, 1 (January 1994), 157–180.

[10] RATLIFF, E. The zombie hunters. *The New Yorker* (October 10 2006).

[11] RESNICK, P., IACOVOU, N., SUSHAK, M., BERGSTROM, P., AND REIDL, J. Grouplens: An open architecture for collaborative filtering. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work* (1994), pp. 175–186.

[12] RESNICK, P., AND SAMI, R. The influence-limiter: Provably manipulation-resistant recommender systems. In *Recommender Systems* (Minneapolis, MN, October 2007), ACM.

[13] SAMUELSON, P. A. The pure theory of public expenditure. *Review of Economics and Statistics 36*, 4 (November 1954), 387–389.

[14] SCARFONE, K., AND MELL, P. Guide to intrusion detection and prevention systems (idps). Tech. Rep. SP 800-94, National Institute of Standards and Technology, February 2007.

[15] SIEBERG, D. Experts: Botnets no. 1 emerging internet threat. *CNN.com, Technology Section* (January 2006).

[16] STONE, B. Spam doubles, finding new ways to deliver itself. *New York Times* (December 2006).

[17] TREND MICRO. Taxonomy of botnet threats. Whitepaper, November 2006.

[18] WASH, R., AND MACKIE-MASON, J. K. Incentive-centered design for information security. In *USENIX Hot Topics in Security (HotSec 06)* (Vancouver, BC, 2006), USENIX.

[19] WASH, R., AND MACKIE-MASON, J. K. Security when people matter: Structuring incentives for user behavior. In *Ninth International Conference on Electronic Commerce (ICEC-07)* (New York, NY, USA, 2007), ACM, pp. 7–14.

[20] WASH, R., AND RADER, E. Public bookmarks and private benefits: An analysis of incentives in social computing. In *American Society for Information Science & Technology* (October 2007).

[21] ZHUGE, J., HOLZ, T., HAN, X., GUO, J., AND ZOU, W. Characterizing the irc-based botnet phenomenon. Tech. Rep. TR-2007-010, The Honeynet Project, 2007.

Appendices

A Proof of Proposition 1

Proof: A social planner would choose the levels of contribution x_i to maximize:

$$\max_{\{x_i\}} \left[\int_i u_i(x_i)\, di \right] = \max_{\{x_i\}} \left[\int_i v(\int_j x_j\, dj) - c_i(x_i)\, di \right]$$

Let us define $\{x_i^+\}$ to be the set of x_i that maximizes this equation. Then we can define $X^+ = \int_i x_i^+ di$. The solution to this would satisfy the first order conditions (FOC) that the first derivative of this function with respect to each choice variable is equal to 0. There are an infinite number of choice variables, so there are an infinite number of first order conditions, one for each person i. The first order conditions are all of the form:

$$\int_j v'(X^+)\, dj - c_i'(x_i^+) \leq 0 \quad \text{(with equality if } x_i^+ > 0) \tag{2}$$

where the notation $v'(X)$ means the first derivative of v with respect to its argument, evaluated at X. Each x_i will be chosen such that the marginal cost to person i is equal to the marginal benefit to society, which is the sum (integral) of the marginal benefits to each person.

In contrast, in the voluntary provision equilibrium, each person will separately choose their own x_i to satisfy

$$\max_{x_i} \left[v(\int_j x_j\, dj) - c_i(x_i) \right]$$

Define x_i^0 to be the x_i that maximizes this expression. Then we can define $X^0 = \int_i x_i^0 di$. Each x_i^0 that is chosen will satisfy its first order condition:

$$v'(X^0) - c_i'(x_i^0) \leq 0 \quad \text{(with equality if } x_i^0 > 0) \tag{3}$$

Under voluntary provision, each person i will keep increasing their contribution x_i until their marginal cost of contribution is equal to their own personal marginal benefit.

Remember that by definition, $v''(\cdot) < 0$ (v is concave) and $c_i''(\cdot) > 0$ (c is convex). This means that the marginal value $v'(\cdot)$ is decreasing in x_i and the marginal cost $c_i'(\cdot)$ is increasing in x_i. Comparing equations 2 and 3, we can see that the marginal cost will be lower in the voluntary provision equilibrium than the social optimal. Since c is concave, the only way this can happen is when x_i is smaller. Since this is true for all i, the total X^0 must be smaller than X^+. *Q.E.D.*

B Proof of Lemma 1

Proof: First note that in the definition of individual i's utility (Eq 1), for any given x_i the first line is always greater than the second line since $v(\cdot) > 0$ by assumption. As such, if the unconstrained maximum $\hat{x_i} > t$ then it is also the maximum of $u_i(\cdot)$. Therefore, if $\hat{x_i} > t$ then the user's optimal choice is to choose $\hat{x_i}$.

Next, notice that $f(x, \bar{x}_{-i})$ is concave in x everywhere by assumption ($v(\cdot)$ is concave, and $-c_i(\dot{})$ is concave since $c_i(\cdot)$ is convex). Therefore, for any $x > \hat{x_i}$, $f(x, \bar{x}_{-i})$ is non-increasing in x. So, if $\hat{x_i} < t$ then the contribution x_i that is sufficient to access the repository and maximizes utility is $x_i = t$. Also, notice that the second line (where the user does not receive access to the repository) is also non-increasing in x. The utility from this option is maximized when contributing nothing ($x_i = 0$). So, if the user's ideal contribution $\hat{x_i} < t$ then they have to choose between contributing the minimum to receive access ($x_i = t$) or not contributing anything ($x_i = 0$). It is straightforward to see that the user will choose to receive access whenever the benefit $v(X)$ with $x_i = t$ is greater than the increase in cost $c_i(t)$. *Q.E.D.*

C Proof of Lemma 2

First, we must step aside, repeat a famous result from Milgrom and Shannon, and prove a simple fact about our utility function.

Milgrom and Shannon [9] define a function $f(x, i)$ to have *increasing differences* (ID) if for all $x' > x'', i' > i''$, $f(x', i') - f(x'', i') > f(x', i'') - f(x'', i'')$. Another way of saying this is that for $x > y$, $f(x, i) - f(y, i)$ is increasing in i. For continuous and differentiable functions, this is similar and related to the property that the cross derivative is positive. Milgrom and Shannon [9] were then able to prove the following theorem:

Theorem 1 (Milgrom and Shannon, 1994 [9]) *If $f(x, i)$ is supermodular in x, and $f(x, i)$ has increasing differences in (x, i) then $\hat{x} = argmax_x f(x, i)$ is non-decreasing in i.*

Now, let us define the two functions. First, let us define a simple function representing the total amount of public good: Let

$$X(x_i, \bar{x}_{-i}) = \int_j y(j) dj$$

where

$$y(j) = \begin{cases} \bar{x}_j & \text{if } j \neq i \\ x_i & \text{if } j = i \end{cases}$$

Now we can define a function that specifies the total utility of person i contributing amount x:

$$g(x, i) = v(X(x, \bar{x}_{-i})) - c_i(x_i)$$

Note that $g(x,i) = f(x, \bar{x}_{-i})$, and that both $\frac{\partial X}{\partial x_i} > 0$ and $\frac{\partial X}{\partial \bar{x}_{-i}} > 0$.

In order to use the theorem from Milgrom and Shannon, we must first show when $g(x,i)$ has increasing differences:

Lemma 3 *If users expect that $\bar{x}_{-j} \geq \bar{x}_{-i}$ for all $i > j$, then $g(x,i)$ as increasing differences in (x,i).*

Proof: Let $x' > x''$, and $i > j$. We need to show that $g(x',i) - g(x'',i) > g(x',j) - g(x'',j)$.

$$\begin{aligned} & g(x',i) - g(x'',i) \\ &= v(X(x', \bar{x}_{-i})) - v(X(x'', \bar{x}_{-i})) - (c_i(x') - c_i(x'')) \\ &\geq v(X(x', \bar{x}_{-i})) - v(X(x'', \bar{x}_{-i})) - (c_j(x') - c_j(x'')) \\ &\geq v(X(x', \bar{x}_{-j})) - v(X(x'', \bar{x}_{-j})) - (c_j(x') - c_j(x'')) \\ &= f(x',j) - f(x'',j) \end{aligned}$$

The first equality is by definition. The first inequality comes from the definition of c_i: We know that $c_i(x') > c_i(x'')$ since $c_i(\cdot)$ is increasing. We also know that $c_j(x') < c_i(x')$ by definition. We make an addition technical assumption that $-c_i(x)$ has increasing differences in (x,i). The second inequality results from the fact that $v(\cdot)$ is convex. If the users expect that $\bar{x}_{-j} \geq \bar{x}_{-i}$ then $X(x, \bar{x}_{-j}) \geq X(x, \bar{x}_{-i})$ since $\frac{\partial X}{\partial \bar{x}_{-i}} > 0$. $v(\cdot)$ is increasing and concave, so the second inequality holds. *Q.E.D.*

A simple correlary of the Milgrom and Roberts theorem and Lemma 3 follows:

Corollary 1 *If users expect $\bar{x}_{-j} > \bar{x}_{-i}$ for $i > j$, then $\hat{x_i}$ is non-decreasing in i.*

We use this corollary to finally prove Lemma 2. As a reminder, the lemma states that there exists a i^0 and a i^* such that the following is a Nash equilibrium:

$$x_i^* = \quad 0 \qquad \text{if } i \leq i^0 \tag{4}$$
$$x_i^* = \quad t \qquad \text{if } i^0 < i < i^* \tag{5}$$
$$x_i^* = \quad \hat{x_i} \qquad \text{if } i > i^* \tag{6}$$

Proof: Here we prove that contributions characterized by lines 4-6 constitute a fulfilled expectations Nash equilibrium, in that if the users expect everyone else to make contributions according to this schedule, then they do not want to deviate. First note that this schedule of contributions is non-decreasing in i: no user i contributes less than any user numbered less than i. If users expect each other to contribute according to this schedule of contributions, then the precondition for Lemma 3 is fulfilled.

Let us begin with line 6. Assume that for some i, $x_i^* = \hat{x_i}$, meaning that user i chose to contribute his optimal amount, which is greater than the threshold t by Lemma 1. Then all users $j > i$ will also want to contribute their optimal amount $\hat{x_j}$, since by Corollary 1, $\hat{x_j} > \hat{x_i}$ and the user's optimal choice according to Lemma 1 is to contribute $\hat{x_j}$. Define i^* to be the smallest i that contributes $\hat{x_i}$.

Next we move to line 4. Assume that for some j, $x_j^* = 0$. By Lemma 1, we know that $\hat{x_j} < t$ and $v(X(t, x_{-j})) < c_j(t)$. This last statement is equivalent to saying $g(t,j) < 0$. Then all users $i < j$ will also want to contribute 0. We know that $\hat{x_i} \leq \hat{x_j}$ by Corollary 1, so $\hat{x_i} < t$. $c_i(t) > c_j(t)$ by assumption on $c(\cdot)$. By Lemma 3, we know that $g(x,i)$ has increasing differences. This means that in particular, $g(t,i) - g(0,i) < g(t,j) - g(0,j)$. Since we are looking at a Nash equilibrium here, person i chooses expecting person j to contribute nothing, and person j chooses expecting person i to contribute nothing. Therefore, $x_{-i} = x_{-j}$, $c_i(0) = c_j(0) = 0$ (by assumption) and

$$\begin{aligned} & g(0,i) \\ &= v(X(0, x_{-i})) - c_i(0) = v(X(0, x_{-j})) - c_j(0) \\ &= g(0,j) \end{aligned}$$

By Lemma 3, $g(t,i) < g(t,j) < 0$. Therefore, person i would also choose to contribute $x_i^* = 0$. Define i^0 to be the maximum i such that $x_i^* = 0$.

Line 5 is all that is left, and is fairly straightforward now. Choose an i such that $i^0 < i < i^*$. We know that $\hat{x_i} < t$ since $i < i^*$. We know that $g(t,i) > 0$ since $i > i^0$. Therefore, by Lemma 1, person i will choose to contribute t.

Finally, we note that in this equilibrium, $x_i^* < x_j^*$ for all $i < j$. This means that in equilibrium, the precondition for Lemma 3 holds. *Q.E.D.*

D Proof of Propositions 2 and 3

Both Propositions 2 and 3 depend on the first derivative of the social welfware with respect to t. As such, we first prove a lemma about this derivative, and then use this lemma to prove the two propositions.

Lemma 4

$$\frac{\partial}{\partial t} \int_i [v(X^*) - c_i(x_i^*)]\, di =$$
$$(N - i^0) v'(X^*) \frac{\partial X}{\partial t} - \int_{i^0}^{i^*} c_i'(t) di - \int_{i^*}^{N} c_i'(\hat{x_i}) \frac{\partial \hat{x_i}}{\partial t} di$$

Proof: To prove this, we first note that Lemma 1 and Lemma 2 together imply that:

$$v(X(t, x_{-i^*}^*)) - c_{i^*}(t) = v(X(x_{i^*}^*, x_{-i^*}^*)) - c_{i^*}(x_{i^*}^*) \tag{7}$$
$$v(X(t, x_{-i^0}^*)) - c_{i^0}(t) = 0 \tag{8}$$

Now, first we split up the initial integral to its four separate parts (based on Lemma 2):

$$\frac{\partial}{\partial t}\int_i [v(X^*) - c_i(x_i^*)]\,di$$
$$= \frac{\partial}{\partial t}\left[\int_{i^0}^{i^*} v(X^*)di - \int_{i^0}^{i^*} c_i(t)di + \int_{i^*}^{N} v(X^*)di - \int_{i^*}^{N} c_i(\hat{x_i})di\right]$$

Applying the Fundamental Theorem of Calculus

$$= (i^* - i^0)v'(X^*)\frac{\partial X}{\partial t} + v(X^*)\left(\frac{\partial i^*}{\partial t} - \frac{\partial i^0}{\partial t}\right)$$
$$+ (N - i^*)v'(X^*)\frac{\partial X}{\partial t} + v(X^*)\left(-\frac{\partial i^*}{\partial t}\right)$$
$$- \int_{i^0}^{i^*} c_i'(t)di - \left(c_{i^*}(t)\frac{\partial i^*}{\partial t} - c_{i^0}(t)\frac{\partial i^0}{\partial t}\right)$$
$$- \int_{i^*}^{N} c_i'(\hat{x_i})\frac{\partial \hat{x_i}}{\partial t}di - \left(-c_{i^*}(\hat{x_i})\frac{\partial i^*}{\partial t}\right)$$

and grouping terms

$$= \frac{\partial i^*}{\partial t}\left(v(X^*) - c_{i^*}(t) - (v(X^*) - c_{i^*}(\hat{x_i}))\right)$$
$$- \frac{\partial i^0}{\partial t}\left(v(X^*) - c_{i^0}(t)\right)$$
$$+ (N - i^0)v'(X^*)\frac{\partial X}{\partial t}$$
$$- \int_{i^0}^{i^*} c_i'(t)di - \int_{i^*}^{N} c_i'(\hat{x_i})\frac{\partial \hat{x_i}}{\partial t}di$$

The first two lines can be eliminated by equations 7 and 8, leaving the intended conclusion. Eliminating these two lines is equivalent to saying that as t changes, changes in i^* don't matter because person i^* wants to voluntarily contribute exactly t, and changes in i^0 don't matter because person i^0 doesn't receive any net value from access (and contributing t). *Q.E.D.*

We need one more simple lemma before we can prove the propositions:

Lemma 5 *As t increases, everyone who is voluntarily contributing greater than t will alter their contribution in exactly the opposite direction as the overall change in database size.*

Proof: The FOC for $\hat{x_i}$ states that

$$v'(X) - c'(x_i) = 0$$

Using the implicit function theorem, we find that

$$\frac{\partial x_i}{\partial x_{-i}} = -\frac{v''(X)}{v''(X) - c''(x_i)}$$

This derivative is always negative (by the concavity and convexity assumptions on $v(\cdot)$ and $c(\cdot)$), and furthermore has the same sign for all $i \geq i^*$. Therefore, as X (and, consequently, x_{-i}) increases, the voluntary contributors decrease their contribution slightly, but not enough to change the overall direction of change in X. *Q.E.D.*

Now that we have proved these lemmas, it is easy to derive the two propositions. For proposition 2, we assume that there is at least one free rider. What his means is that in the voluntary provision equilibrium, at least one person (person number 0) has $v(X^0) < c_i'(0)$, or his marginal benefit of contributing anything is below his marginal cost. It must also be true that $v(X^0) > 0$, meaning he receives positive value from accessing the database. He would not voluntarily contribute anything, and would free-ride on the contributions of others.

For this person, if we increase t from 0 to some small amount $t^0 > 0$, he will have to pay a cost $c_0(t^0)$ if he still wants access to the database. But, since the value of the database is strictly positive and everything is continuous, there must exist a small enough t^0 such that $v(X(t^0, X^0)) > c_0(t^0)$. This will cause the total size of the database to increase. Everyone else will decrease their contribution, but the total decrease from everyone else will be less than t^0 (by Lemma 5), so the total size of the database will increase. By Lemma 4, this must result in a positive increase in welfare. Thus, there is a positive t that leads to a welfare improvement from the voluntary contribution equilibrium, proving Proposition 2.

The socially optimal t^* will maximize social welfare. The first order condition of this maximization sets the derivative of social welfare with respect to t (the derivative in Lemma 4) equal to 0. By the envelope theorem, we know that

$$\frac{\partial v}{\partial x_i} - \frac{\partial c}{\partial x_i} = 0$$

We can re-arrange the conclusion of Lemma 4 such that:

$$(N - i^0)v'(X^*)\frac{\partial X}{\partial t}$$
$$- \int_{i^0}^{i^*} c_i'(t)di - \int_{i^*}^{N} c_i'(\hat{x_i})\frac{\partial \hat{x_i}}{\partial t}di$$
$$= (i^* - i^0)v'(X^*)\frac{\partial X}{\partial t} - \int_{i^0}^{i^*} c_i'(t)di$$
$$\int_{i^*}^{N}\left(\frac{\partial v}{\partial x_i} - \frac{\partial c}{\partial x_i}\right)\frac{\partial x_i}{\partial t} + \frac{\partial v}{\partial x_{-i}}\frac{\partial x_{-i}}{\partial t}di$$
$$= \int_{i^0}^{N} v'(X^*)\frac{\partial x_{-i}}{\partial t}di - \int_{i^0}^{i^*} c_i'(t)di$$

The first order conditions say that the optimal t^* is the one that makes this expression equal to 0. This is the statement of the conclusion of Proposition 3.