Abuse Reporting and the Fight Against Cybercrime

MOHAMMAD HANIF JHAVERI, Southern Methodist University
ORCUN CETIN and CARLOS GAÑÁN, Delft University of Technology
TYLER MOORE, The University of Tulsa
MICHEL VAN EETEN, Delft University of Technology

Cybercriminal activity has exploded in the past decade, with diverse threats ranging from phishing attacks to botnets and drive-by-downloads afflicting millions of computers worldwide. In response, a volunteer defense has emerged, led by security companies, infrastructure operators, and vigilantes. This reactionary force does not concern itself with making proactive upgrades to the cyber infrastructure. Instead, it operates on the front lines by remediating infections as they appear. We construct a model of the abuse reporting infrastructure in order to explain how voluntary action against cybercrime functions today, in hopes of improving our understanding of what works and how to make remediation more effective in the future. We examine the incentives to participate among data contributors, affected resource owners, and intermediaries. Finally, we present a series of key attributes that differ among voluntary actions to investigate further through experimentation, pointing toward a research agenda that could establish causality between interventions and outcomes.

Categories and Subject Descriptors: K.4.1 [**Public Policy Issues**]: Abuse and Crime Involving Computers

General Terms: Measurement, Security, Economics

Additional Key Words and Phrases: Cybercrime, abuse reporting, internet security, security economics

ACM Reference Format:
Mohammad Hanif Jhaveri, Orcun Cetin, Carlos Gañán, Tyler Moore, and Michel Van Eeten. 2017. Abuse reporting and the fight against cybercrime. ACM Comput. Surv. 49, 4, Article 68 (January 2017), 27 pages.
DOI: http://dx.doi.org/10.1145/3003147

1. INTRODUCTION

Every day, millions of resources on the Internet are abused in criminal activities, ranging from infected end-user PCs recruited into botnets sending spam to compromised webservers that surreptitiously infect unsuspecting visitors. While a lot of effort has been put into identifying these compromised resources, the question of what happens after discovery has received a lot less attention. How are criminal resources shut down? Who is supposed to act on the abuse data? And what determines if this entity does indeed act?

This publication was supported by a subcontract from Rutgers University, DIMACS, under Award No. 2009-ST-061-CCI002-06 from the U.S. Department of Homeland Security and by a grant from the Netherlands Organisation for Scientific Research (NWO) under project number 628.001.022.
Authors' addresses: M. H. Jhaveri, 5601 Rock Valley Dr, Fort Worth, TX 76244, USA; email: mjhaveri@alumni.smu.edu; O. Cetin, C. Gañán, and M. Van Eeten, TU Delft, Faculty TBM, Jaffalaan 5, 2628 BX Delft, Netherlands; emails: {F.O.Cetin, C.H.G.hernandezganan, M.J.G.vanEeten}@tudelft.nl; T. Moore, Tandy School of Computer Science, The University of Tulsa, 800 S. Tucker Dr., Tulsa, OK 74114, USA; email: tylermoore@utulsa.edu.

DOI: http://dx.doi.org/10.1145/3003147

The Internet's decentralized, trans-boundary architecture and its highly fragmented ownership structure dictate that voluntary action is at the heart of any response to cybercrime. Incidents evolve rapidly and so do the necessary actions of the players involved. Defenders have to collaborate to be effective. For example, the Conficker Working Group coordinated the preemptive shutdown by registries of domain names used to control a large botnet [Conficker Working Group 2016].

Voluntary action typically takes the form of one party notifying another about potential abuse and asking it to act against it. An average-sized Internet Service Provider (ISP) or hosting provider can easily receive thousands of such abuse reports each day. They cover not only security issues as commonly understood but also alleged infringements of intellectual property and issues of content regulation, such as child pornography. The same abuse reporting mechanisms are also increasingly used to prevent abuse by disseminating information about vulnerable resources that have been discovered to the relevant owners or associated providers [Durumeric et al. 2014; Rossow 2014].

There is a tendency among some in the security community to equate the discovery of abuse with the imperative to act against it by a certain actor who is in a position to act. The reality is that these actors—such as resources owners and network operators—face difficult tradeoffs tied to the competing incentives and objectives under which they operate. These incentives often determine the extent of voluntary action against abuse, yet they remain poorly understood.

Given its key role in fighting cybercrime, the abuse reporting infrastructure is remarkably under-researched. Some parts have been analyzed, but the findings are scattered across more than 50 different studies. This article systematizes the understanding of abuse reporting and voluntary action against cybercrime as follows:

—We construct a framework model of the abuse reporting infrastructure that drives voluntary action against cybercrime (Section 2). This framework helps to explain the roles of various stakeholders and the different actions they take.
—We review key operational aspects of voluntary action against cybercrime, ranging from how abuse data are contributed to the nature of abuse reports to the steps intermediaries take in getting compromised resources cleaned up (Section 3). We report on relevant research where applicable, but supplement this with descriptions of technical operations that have not previously been discussed in the academic literature.
—We examine the incentives for contributors to collect abuse data, for intermediaries to act on abuse reports, and for affected resource owners to act on cleanup requests (Section 4). We review all relevant research on these behavioral aspects and point out gaps in the literature.
—We identify the key attributes of the model that affect the success or failure of cleanup efforts, with an eye towards identifying opportunities for future research that could establish causality between abuse discovery, reporting, and outcomes (Section 5).
—While most action is undertaken voluntarily by private actors, governments do have a key role to play. We discuss how governments can assist in the fight against cybercrime, both by participating directly but especially by supporting existing efforts by aligning the incentives of participants (Section 6).

2. ABUSE REPORTING INFRASTRUCTURE

We have devised a model of how security incidents are reported, as illustrated in Figure 1. The flow reflects the process by which resource owners and third parties can take action to clean up or protect themselves or otherwise be asked to do so. Table I describes the actors (vertices in Figure 1) and actions (edges in Figure 1). Within this process flow, three possible interventions strategies arise: direct remediation, intermediary remediation, and third-party protection.

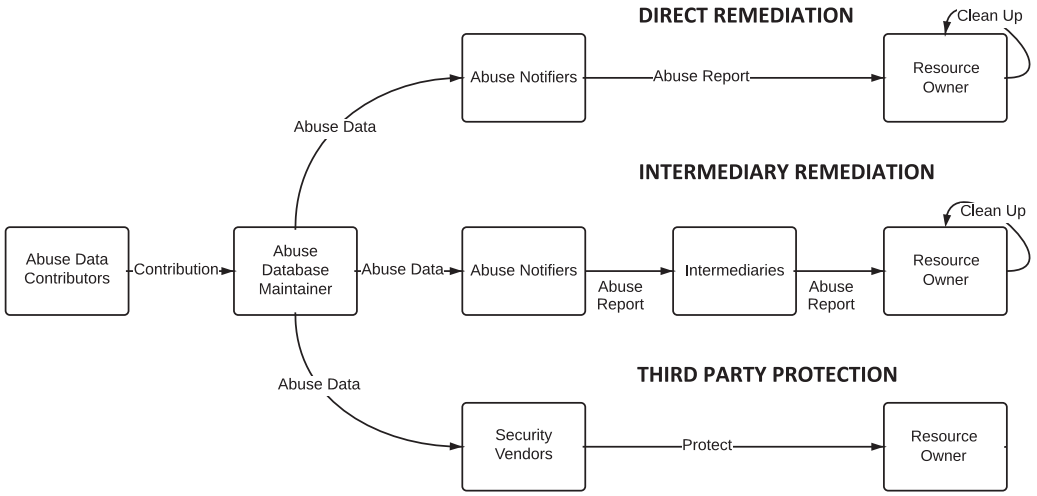

Fig. 1. Abuse reporting infrastructure overview.

Table I. Actors and Actions in the Abuse Reporting Infrastructure Model

Actors	Role
Abuse Data Contributor (ADC)	Identify instances of abuse and contribute this data to some aggregation mechanism
Abuse Database Maintainer (ADM)	Recipients and aggregators of abuse data
Abuse Notifier (AN)	Uses abuse data to send abuse reports to the respective intermediaries or resource owners to facilitate cleanup
Intermediary (INT)	Entity having a business relationship with the affected resource owner, may help facilitate cleanup
Resource Owner (RO)	Individual or entity responsible for the compromised resource identified in the abuse data
Security Vendor (SV)	Uses abuse data to protect third parties against harm
Third Party	Individual or entity using a security service to protect itself from being harmed by the insecurity of others
Actions	**Description**
Contribute Data	Raw indicators of abuse that are provided to ADMs
Share Abuse Data	Send transformed data to abuse notifiers
Send Abuse Report	Send notification and cleanup request to affected resource owner or intermediary.
Protect	Construct defenses for third parties based on abuse data

Direct remediation occurs when abuse notifiers directly report abuse to the Resource Owner with the goal of remediating the situation through a cleanup. For example, a website owner might notice that their website no longer appears in Google Search results. Google allows webmasters to check their site via the Search Console [Google 2016b] and then take action based on what their automated detection system determines is wrong with the site in question. Another example is `BreachAlarm.com`, which informs end-users if their account information has been compromised and whether they need to change their password based on information received about the compromise.

Intermediary remediation occurs when the abuse notifier reports to an intermediary with the goal of instigating a cleanup by the resource owner. For example, an ISP can be notified by Shadowserver when one of its customers is infected [Shadowserver 2016]. Depending on the ISP, it must decide whether to pass this information on to its customers as part of a cleanup strategy. If the ISP does provide this information

to the resource owner who then fails to respond, then the ISP can try a number of remedies, ranging from filtering the port being used for the bot's activity to even suspending the customer's network access. Similarly, when web hosting providers learn that their customers have been compromised, they can first inform the resource owner to request cleanup. Failing that, they could attempt to remediate the infection by either suspending access to the site or cleaning and patching the vulnerability themselves. While filtering network traffic or suspending access does not fundamentally address the root causes of these compromises, they nevertheless are important in limiting their damage and getting the cooperation of resource owners.

Third-party protection occurs when an abuse database maintainer sends abuse data to a security vendor who leverages it in order to protect a third party, usually a customer of a security service. The goal here is to protect the resources of the customer against damage caused by compromised or malicious resources elsewhere rather than remediating the actual abuse itself. One example of third-party protection is Google Safe Browsing, which is an Application Program Interface (API) anyone can use to test whether a requested URL is on Google's phishing or drive-by-download blacklists. SpamHaus provides another example: ISPs query their spam-sending lists to decide whether to accept incoming email from a certain Internet Protocol (IP) address. While not fostering cleanup, we include it in the framework because third-party protection provides a strong incentive to collect abuse data. The primary incentives are to sell services based on the data and to provide a timely response before cleanup can take place.

3. OPERATIONAL ASPECTS OF ABUSE REPORTING

Having established a framework for how cyber incidents are reported, we next investigate key operational aspects of the model. In this section, we describe in greater detail how cybercrime mitigation operates in practice, reviewing relevant research when applicable. We organize the discussion around the following questions arising from actions in the model:

A) How are abuse data contributed?
B) How do abuse data maintainers send data to others?
C) How do abuse notifiers report abuse?
D) How do intermediaries get resource owners to remediate abuse?
E) How do security vendors protect third parties?

3.1. How Are Abuse Data Contributed?

Roughly speaking, abuse data can be contributed either manually or automatically. In the first method, user reports are manually entered and sent to abuse database maintainers by general members of the Internet community, legal practitioners, corporations, and security professionals, varying by database [Vasek and Moore 2012].

Consider the example of phishing URLs. Financial institutions discover impersonating websites from user reports, spam filtering, and log inspection. The victim institution may then decide to pass along the information by submitting reports to abuse database maintainers such as PhishTank [2016] and the APWG [2015]. But the victim institutions are not the only ones who identify phishing. General members of the public also have the ability to submit suspected URLs to databases such as PhishTank. Users might have come across these phishing URL in their email inbox or spam folder. Their submission is then voted on by other PhishTank users before being pushed out to the authoritative blacklist. Complementary mechanisms have reduced the barriers for user contribution via one-click reports via browser toolbars. For example, Norton's browser security toolbar has an in-built option to submit one-click reports to SafeWeb [Symantec 2016b], their internal blacklist.

Reports can also be filed by security professionals and volunteers interested in promoting Internet safety. This includes vigilantes and baiters who seek out conflict with the goal of stopping abuse. Such groups have formed particularly around scams that do not attract the attention of established businesses who have a direct economic interest in collecting abuse data but where such scams have consequences for the general public. One example of such a group is the Artists Against 419 (AA419), who combat advanced-fee frauds by identifying fraudulent banking websites that are established as part of the scam [AA419 2016]. Their initial strategy was to deploy bandwidth-hogging tools in hopes that it would trigger hosting providers to shut the scam websites down. Recognizing that the scam sites re-emerged quickly and that their efforts imposed collateral damage, they now maintain a database of scam websites and share their findings with law enforcement and infrastructure providers.

While user reports of abuse can be valuable, attempts to crowd-source the discovery and verification of abuse data have had mixed results. By relying on users to verify submitted phishing URLs, PhishTank's curated blacklist remained persistently outdated [Moore and Clayton 2008b]. However, in contexts where timeliness is less crucial, such as establishing the reputation of websites, crowd-reporting platforms such as Web of Trust (WOT) have been found to be accurate and offer more complete coverage [Chia and Knapskog 2012]. Systems such as WOT distinguish credible contributors based on prior ratings and give stronger weight to their assessments.

A different type of manual contribution is not based on large numbers of participants but on smaller teams that come across some ad hoc, albeit perhaps large, sets of abuse data. For example, in early 2014, German researchers and the prosecutor's office discovered a collection of 16 million email account credentials [BBC 2014]. They handed these over to a government organization, which then set up a website for German citizens to check whether their account was present in the dataset. A similar case occurred around the controversial claim of a small security company where it had uncovered a trove of over 1 billion stolen passwords [Perlroth and Gelles 2014]. These data were later shared with some national Computer Emergency Response Teams (CERTs), who then contacted the relevant intermediaries or resource owners.

The sheer scale of certain types of abuse, or a lack of visibility of the abuse to end-users, may render manual contributions less effective. Here is where the second method comes into play: automated tools for abuse discovery and aggregation. For example, honeypots, web crawlers, or port-scanning systems can detect and generate abuse data by recognizing known abuse patterns.

Typically, machine-generated reports are created by constructing a general-purpose detector for the malicious activity in question. For example, botnet command-and-control servers can be identified by inspecting passive Domain Name Server (DNS) logs [Choi and Lee 2012]. Here, outgoing connections to several newly registered domains from a single IP address are flagged as suspicious. Some researchers have reverse-engineered the algorithms used by the botnet to generate new domains for Command and Control (C&C), enabling preemptive detection of abusive domains [Perdisci et al. 2009; Yadav et al. 2010; Antonakakis et al. 2012]. Bilge et al. described a mechanism for using passive DNS data to identify many additional types of abusive activity, including phishing [Bilge et al. 2011]. Gao et al. built a classifier to recognize the DNS patterns of known maliciously registered domains and then applied it to a dataset of 26 billion DNS queries to identify many previously unidentified malicious websites.

A number of articles have investigate abuse in search-engine results. Provos et al. presented a mechanism for identifying drive-by-downloads in web search results [Mavrommatis and Monrose 2008]. This method has formed the basis of the malware blacklist used in Google Safe Browsing. John et al. [2011] reported on the poisoning of trending search terms to distribute malware. Leontiadis et al. identified

websites that have been hacked to sell counterfeit pharmaceuticals [Leontiadis et al. 2011, 2014]. More generally, a number of articles have proposed classifiers to identify malicious web content. Abu Nimeh et al. compared several methods for classifying phishing websites [Abu-Nimeh et al. 2007], while many others have constructed features for classifying malicious web pages based on website content or behavior [Canali et al. 2011; Ntoulas et al. 2006; Webb et al. 2008; Wang et al. 2011; Bannur et al. 2011]. In all cases, while the articles only collected data during the course of research, they nonetheless explain how mechanisms that identify abuse data work.

It is worth noting that researchers have repeatedly observed that individual sources of abuse data are usually incomplete. Pitsillidis et al. compared 10 feeds of spam-advertised domains, finding that while each has good coverage of broadly advertised campaigns, each list misses many domains promoted by lower-volume spam that are picked up by only one or two of the other lists [Pitsillidis et al. 2012]. Metcalf and Spring took the idea a step further, comparing 25 commonly used blacklists. They found that more than 99% of all abused resources within the lists were unique to one list and not cross-referenced [Metcalf and Spring 2013]. Along the same lines, in an examination of 15 public malware blacklists, Kührer et al. estimated that, collectively, they captured less than 20% of the websites they were supposed to detect [Kührer et al. 2014]. These findings suggest that achieving comprehensive coverage of abuse is difficult, if not impossible, and furthermore that sharing data is an essential prerequisite for dealing effectively with abuse.

3.2. How Do Abuse Database Maintainers Send Data to Others?

Abuse database maintainers amass lists of malicious content, be it IP addresses of machines in botnets or URLs that distribute malware. These are usually organized into a blacklist. Then the challenge for the maintainer is to get the data out into the hands of notifiers to facilitate cleanup or security vendors to protect their customers.

The roles of maintainer, protector, and notifier may be carried out by distinct entities or by the same party. An example of a pure maintainer is the APWG [2015], which maintains a blacklist of active phishing URLs contributed by its members. Many banks and takedown companies subscribe to the list and then craft abuse reports to intermediaries and resource owners as notifiers. Others, such as M3AAWG [Messaging Anti-Abuse Working Group 2007], provide a general mailing list feed where vetted and reputable researchers can share lists of malware-infected sites directly with their respective community.

Symantec's Norton Safeweb is an example an of integrated maintainer-protector. Symantec has built an extensive blacklist of malicious websites. It then incorporates the blacklist into a web browser add-on product to prevent customers from visiting websites it deems dangerous.

Google's Safe Browsing initiative [Google 2016a], by contrast, operates as both distinct maintainer and integrated maintainer-protector. As an example of the former, Mozilla and Apple integrate the Safe Browsing API into their browsers, so every website requested by users is first checked against Google's database. As an example of the latter, Google's Chrome Browser and Google Search also check blacklists before rendering results.

Regardless of whether the roles are integrated or distinct, the maintainer faces choices in how to make the data accessible to others, as well as how they share the data with other maintainers. The simplest choice is to make the entire list public, as PhishTank, Malware Domain List [MDL 2016], and several botnet C&C trackers do [ZeusTracker 2016; Cybercrime tracker 2016]. Others make the raw data available only to vetted members, as the Anti-Phishing Working Group (APWG) does. Still others,

such as Google Safe Browsing and Norton, forbid access to the raw data but allow anyone to query the database to check if a requested resource is on the blacklist.

Of course, the ability to query all data is not the same as publishing a complete list for all to see. Publicizing all abuse data is a form of shaming that its proponents believe will help incentivize better behavior among affected resource owners. Indeed, research suggests that compromised websites made public on PhishTank were less likely to be subsequently recompromised than websites only appearing in private feeds [Moore and Clayton 2011]. Nonetheless, there are legitimate reasons why some data are not made public. One concern is that public reports of compromised resources could aid criminals in finding vulnerable resources for repeated exploitation. After all, a website that can be hacked to host phishing pages can often be used to distribute malware or, at the very least, host more phishing pages. In the case of phishing, the evidence suggests that publishing websites on PhishTank is actually associated with a *reduced* likelihood of recompromise. However, more research is needed to identify cases in which publishing abuse data can itself trigger further abuse.

Another objection to publishing abuse data is that doing so might discourage some contributors from sharing. They might view sharing as counterproductive or worry that if they make a mistake, then they could be exposed to liability for making defamatory claims. This can be a valid concern, especially when automated techniques are deciding what is abuse and the methods have a non-zero false positive rate. When printers receive takedown requests [Piatek et al. 2008], the result can be embarrassing for the sender. Keeping blacklists private is viewed by some as a way to reduce the risk of such embarrassments occurring. The role of public shaming in incentivizing cleanup is discussed further in Section 4.2.

Additionally, maintainers such as Facebook's Threat Intelligence [Facebook 2016], Stop Badware Data Sharing Program [StopBadware 2016], and the APWG eCrime Exchange [APWG 2016] have been established to coordinate contributions from other maintainers. As explained in the prior section, data fragmentation among maintainers is prevalent. Thus, by coordinating across multiple lists, these "meta-maintainers" who set up reciprocal data feeds are attempting to maximize the impact of each submitted abuse report and reduce fragmentation.

3.3. How Do Abuse Notifiers Report Abuse?

Abuse notifiers must decide who to notify and how. Regarding who to notify, the choice is usually between going directly to the resource owner or to an intermediary (such as an ISP or hosting provider) instead. Abuse notifiers who contact resource owners directly are often services that are used by the resource owners themselves, since these services have up-to-date contact information for the resource owner. For example, the Google Search Console [Google 2016b] lets webmasters query whether IP addresses or URLs that the resource owner has demonstrated she controls have malware. Additionally, automated notifications are sent out via email to registered users when their resources are added to one of Google's blacklists. For webmasters who have not signed up for Google's Search Console, they frequently learn that their website has been placed on a blacklist when they or their customers encounter a malware warning when trying to visit the website via Google or using a browser that queries the Safe Browsing API.

However, in many cases, directly informing resource owners is impractical. The resource owner may either be unreachable or lack the necessary expertise to remediate a compromise [Cetin et al. 2015; van Eeten et al. 2010]. In cases where they are willfully perpetrating the compromise (such as in the case of financial frauds or counterfeit goods) and the actual abuser is unknown, direct notification would not help. In these circumstances, informing an intermediary is an important strategic option since the

intermediary both enables the malicious activity and generally has the power to curtail it by denying the resource owner the right to use its resources.

In choosing an intermediary to inform, notifiers select the party who has a direct relationship with the resource owner, provided that the resource has been compromised and has not been maliciously registered. For example, for phishing and web-based malware, the notifier will contact the web hosting provider if the website has been hacked. If the website appears to have been set up by the criminals directly, then the registrar will be notified and requested to suspend the domain. For an abusive hosting provider, they will contact the company providing network capacity and bandwidth. And for botnet-infected computers, they will contact the ISPs, who can map IP addresses to customer contact information.

Regarding the method of notification, abuse reports are transmitted in two ways: push or pull. Using a push mechanism, the notifier periodically sends out unsolicited abuse reports, such as by sending emails to `abuse@domain.com`. Other examples of push mechanisms include ISPs notifying customers that they are in a botnet (more on that in the next section) and disseminating data on public websites.

In contrast, pull mechanisms enable interested intermediary or resource owners to subscribe to ongoing updates of abuse reports as they are detected. For example, intermediaries may subscribe to a blacklist of web URLs that contain phishing links or IP addresses that are part of a known botnet. Another example of a pull mechanism is the SpamHaus DNSBL Datafeed service [Spamhaus 2016], which compiles a list of IP addresses sending unsolicited email. When requested by the ISP, the service sends all abusive IP addresses that fall within their leased range to trigger cleanup.

The SpamHaus example illustrates another distinction in how abuse reports are provided. SpamHaus does not actually share its global list of spam-sending IPs. Instead, the ISP pulls data specific to their compromised resources and nothing more. This approach can be attractive to both the providers (who do not necessarily want to publish sensitive information willy-nilly) and to the recipients (who are only interested in actionable data).

3.4. How Do Intermediaries Clean Up Resource Owners?

As explained above, abuse notifiers inform intermediaries about affected resources under their purview in the hope that they will fix the problem. Some intermediaries choose not to react to abuse reports [Canali et al. 2013; van Eeten et al. 2010; Liu et al. 2011; Nappa et al. 2013]. Nevertheless, many intermediaries do respond, and they play a crucial role in cleaning up malicious activity. It has been argued that intermediaries are more effective control points for remediation compared to resource owners and end-users because intermediaries possess expertise and visibility into threats that resource owners and end-users often lack [van Eeten et al. 2010]. We now discuss remediation efforts in two broad categories: botnets and website takedown.

Intermediaries may first try to remediate infected sources by passing along abuse reports to resource owners. If that does not fix the problem, then some intermediaries are able to take immediate infection mitigation actions such as preventing access to the resource or taking down the content of the infected source. In botnet takedowns, domains used as C&C servers, drop zones, or malware droppers are usually shut down by intermediaries to prevent bot herders from using them to spread the infection or control the infected machines [Dittrich 2012]. Following the intervention, infected resources cannot be remotely commanded by botnet herders. Additionally, domains used in botnet communication can be sinkholed to point botnet traffic to dedicated servers [Leder et al. 2009]. Thus, botnet traffic can be analyzed to identify the infected machines to start-up end-user cleanup process.

Depending on the type of C&C infrastructure used by the botnet, sinkholing may also require the collaboration of registrars, who have to redirect the domain names used by the botnet to the sinkhole server's IP address. After an ISP receives an abuse report from a sinkhole-maintainer such as Shadowserver, they may then contact infected end-users using a range of methods, including email, telephone call, instant message, Short Message Service (SMS), or web browser notifications [Livingood et al. 2012].

Comcast's web notification system shares critical information with subscribers about the infection and how to clean it up [Mody et al. 2011]. However, the user's Internet connection is not restricted in any way. A more aggressive approach is to place infected customers into a "walled garden" with restrictions on Internet access [Livingood et al. 2012]. Using walled gardens, notified subscribers are also protected from malicious executables and commands from bot herders. The walled-garden approach is more expensive than notification alone, both in terms of infrastructure required and elevated customer service costs. As a baseline, best practices outlined by M3AAWG recommend that ISPs should provide documentation, tutorials, videos on how to prevent future infections, and tools to remediate infected resources in residential networks [Messaging Anti-Abuse Working Group 2007].

We now discuss voluntary efforts by intermediaries in five well-known botnet takedowns. In the case of the Mega-D takedown, coordination with registrars effectively shut down and sinkholed the command and control infrastructure of the botnet. Thus, infected parties were identified and notified by concerned ISPs [Lin 2009]. Similarly, domain names used to control the Conficker botnet were proactively barred by top-level domain registries. Unlike Mega-D, Conficker introduced methods to overwhelm the efforts of registries by increasing the number of daily generated domains, but the takedown efforts succeeded nevertheless via an unprecedented coordinated effort among registries in 110 top-level domains [Conficker Working Group 2011]. The sinkholes of these botnets were used to identify the infected end-users and notify ISPs about the infections in their networks. Similarly to other takedown efforts, in the ZeroAccess takedown, Microsoft and its partners seized servers and sinkholed domain names associated with fraudulent schemes in order to disrupt the botnet's operation [Hiller 2014]. Furthermore, Microsoft continued its mitigation efforts by notifying infected end-users through ISPs [Microsoft News Center 2013].

Moreover, international law enforcement agency efforts have disrupted the activities of sophisticated banking malware called Gameover Zeus. Unlike earlier Zeus variants, Gameover Zeus was more difficult to shut down due to the presence of decentralized, peer-to-peer, and domain generation algorithm– (DGA) based command and control infrastructure. Law enforcement agencies used the design flaws in Gameover's peer-to-peer network to redirect traffic to nodes under its control [Molloy 2014]. Likewise, Gameover's DGA domains were taken down and sinkholed by law enforcement agencies. Furthermore, ShadowServer actively shares infection notification with ISPs and CERTs in order to help identify and notify infected end-users in ongoing cleanup efforts [The Shadowserver Foundation 2014].

In the Bredolab takedown, the Dutch High Tech Crime Unit collaborated with the Dutch hosting provider LeaseWeb in order to gather extensive intelligence on the botnet's command and control infrastructure, which was located inside LeaseWeb's network [de Graaf et al. 2013]. When this phase was concluded and the criminal was identified, all 143 command and control servers were seized [StopBadware 2011].

For websites that have been compromised or set up by criminals to harm others, takedown is frequently used. Note that, unlike for the botnet actions just described, takedown typically occurs at the individual site level, and far less coordination among defenders is required. While resource owners have control over a compromised website, intermediaries such as hosting providers can often exercise control over the content

when needed. Moore and Clayton measured the response times for removing large numbers of phishing websites, finding a skewed lognormal distribution where most infections are remediated quickly but with a long tail of persistent compromises [Moore and Clayton 2007]. When the website itself has been set up by criminals, as in the case of botnet C&C and some phishing, domain name registrars are often in the position to facilitate takedown. Unfortunately, domain-oriented takedowns are not very well established in every country. Moore and Clayton demonstrated that phishing gangs who registered domains iterated through previously untargeted jurisdictions, experiencing a honeymoon period of very slow removal before the authorities got a clue. Liu et al. studied the impact of the registrar-level interventions on countering fraudsters [Liu et al. 2011], finding that criminals can evade detection faster than defenders can shut them down. Canali et al. studied the ability of shared hosting providers to detect vulnerable and compromised websites. The researchers concluded most shared hosting providers did not perform antivirus scans and did too little to detect malicious activity and compromises to help their customers [Canali et al. 2013].

Typically, an intermediary's willingness to act in a timely manner plays a crucial role in successful remediation efforts. For instance, any delayed remediation attempt could result in recovery of a botnet or abusive hosts to victimize more Internet users. To make matters worse, some small intermediaries dedicate themselves to providing bullet-proof hosting, where cybercriminals are safe from abuse notifier pressure and influence. Generally, these networks are associated with a wide variety of malicious activity, especially botnet C&C. Nonetheless, long-lived malicious activity combined with inattention to abuse reports can trigger public attention. When this happens, even bulletproof hosting services could be blacklisted or disconnected by upstream providers, such as McColo [Krebs 2008]. Following McColo's shutdown, spam transmission fell significantly, though the spammers quickly moved elsewhere [Clayton 2009]. In another case, TROYAK-AS, a fraud-friendly ISP associated with many Zeus C&C servers and Zeus spam campaigns, was taken down in 2010 by upstream providers. This triggered a dramatic fall in Zeus communication servers presented in Zeus Tracker [Mansfield-Devine 2010]. Google also maintains lists of such malicious networks and providers, the result of which has been greater attention to the problem. As a result, providers such as RBN and Atrivo have shared a similar fate as McColo when attention was drawn to their activities.

Nonetheless, the decision to immediately engage in a takedown effort is not always straightforward. As the McColo situation illustrates, criminals are adept at transitioning their infrastructure to other hosts when their operations are partially, rather than completely, disrupted. Thus, partial takedowns may actually prolong botnets' ability to continue operating core infrastructure. Moreover, there is the risk of collateral damage stemming from larger efforts to cut off networked criminal activity. Consider the ZeroAccessP2P botnet that monetized itself through click fraud on popular advertising networks [Wyke 2012]. While the Microsoft Digital Crimes Unit (MSDCU) put out a press release [Microsoft News Center 2013] stating they had disrupted nearly 32% of the infected machines within the network, security professionals remained skeptical [Antonakakis and Nadji 2013]. Researchers ultimately determined that the extensive legal work carried out by the FBI, MSDCU, and Europol had a minimal impact in terms of reducing the overall profitability of the botnet. Worse, other botnet takedowns coordinated by such entities obfuscated visibility into the operations of those networks and hampered important data collection efforts by security specialists who had set up sinkholes to collect intelligence.

Furthermore, takedowns can impose unintended collateral damage, particularly when shared resources are involved and the requesting parties do not fully appreciate the nature of the infrastructure they are seizing. Suspending domain names can

impact innocent subdomains, while removing a server that hosts multiple websites is also problematic. For instance, one hosting provider whose customers were affected by a takedown reported that shortly after the 2014 intervention, "some 1.8 million customers are still offline, which translates to more than 4 million domains" [Vixie 2015].

To summarize, despite noble intentions, taking down sophisticated and decentralized criminal networks such as P2P botnets has proven difficult in practice [Rossow et al. 2013]. As a consequence, the decision to engage in a takedown rests on striking a balance between risk and reward. Nadji et al. even quantified these tradeoffs by balancing the cost of collateral damage, visibility into criminal activity, and long-term economic benefit [Nadji et al. 2013].

3.5. How Do Security Vendors Protect Third Parties?

In addition to facilitating cleanup, abuse reports can be leveraged by security vendors to protect their customers from harm, even without fixing the problem for everyone. This step is important because remediation is inevitably slower than protection.

Security vendors operate on a wide range, from Software-as-a-Service (SaaS) providers who sit between users and their content requests, for example, as a DNS service, to offering downloadable software for user devices. Generally, they may engage in one of three actions: warning users when they try to access malicious content, outright blocking users from accessing content, or guiding users to pre-checked content (for example, by integrating a "safe" logo into a list of search results). Typically, they build a corpus of infected sites based on public and private blacklists of compromised content. However, certain vendors will also identify suspicious behaviors such as untrusted embedded iFrames in sites prior to rendering them to the user.

Aside from traditional antivirus and antimalware software, referenced previously, the best-known example of warning users or blocking them outright when requesting malicious content is the integration of Google's Safe Browsing API into web browsers. Additionally, vendors such as Forcepoint, Blue Coat Systems, and FireEye offer blacklists for enterprises and end-users, from filtering DNS lookup requests to browser plug-ins that notify users when dangerous requests are made [Raytheon 2016; Symantec 2016a; FireEye 2016]. The main advantage of blacklists is that they enable third parties to quickly defend against malicious websites outside their control. Because there is a small chance of false positives, most blacklist implementations permit the the user to bypass the protection. False negatives are an even bigger concern, since blacklists are not always up to date or comprehensive.

Empirical evaluations of blacklist-based toolbars are mixed. Several studies have shown them to be at least partially effective in preventing users from accessing and subsequently having their machines infected by harmful websites, despite relying on information that might be outdated or incorrect at times [Angai et al. 2010; Sheng et al. 2009].

Because these vendors are usually competing with each other for customers, fragmentation of data feeds remains a problem. Since the early 1990s, antivirus companies have shared virus samples [Moore and Clayton 2008a]. However, this spirit of cooperation has not always extended to other types of security data, including URL blacklists. There have been a few noteworthy exceptions, however. One is a governing framework proposed by Riccardi et al. to coordinate collective efforts to mitigate financial trojans through periodic sharing of intelligence with security vendors and intermediaries [Riccardi et al. 2010]. With a similar goal in mind, vendors providing real-time email spam blacklists have offered lookup access to each others' databases to verify incoming emails and improve their collectively ability to identify and flag spam emails [Ramachandran et al. 2006]. These can then be either rejected

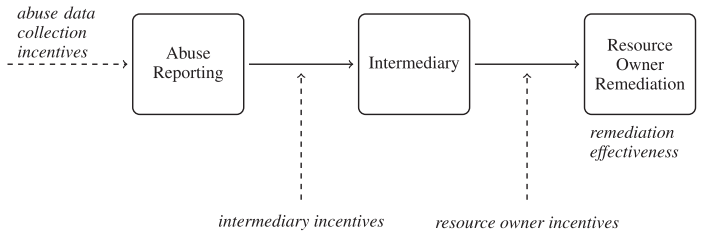

Fig. 2. Causal model.

immediately, tagged as junk, or filtered further using content-matching rules. While imperfect, third-party protection helps to filter much of the bad traffic emanating from compromised resources, even if the problem remains for those not utilizing the services.

4. BEHAVIORAL ASPECTS OF ABUSE REPORTING

Having just described how cybercrime is remediated, we now turn to why actors participate in this system. Because no single entity is responsible for reporting, maintaining, and acting on abuse data, incentives determine why participants take action. The causal model in Figure 2 illustrates the incentives that underpin the abuse reporting and remediation process. Three types of incentives influence the overall effectiveness of cleanup: incentives to collect abuse data, incentives for intermediaries to act on abuse reports, and incentives on resource owners to clean themselves up. We discuss each in turn.

4.1. Incentives to Collect Abuse Data

None of the steps in the process of remediation could take place without people and organizations gathering abuse data in the first place. However, collecting this data involves cost. What motivates these actors to do so?

Among academic researchers, the possibility of being able to publish high-quality research articles that lead to a publication is an incentive to collect abuse data. Among individuals, the intentions can be altruistic or be driven by personal ethics and beliefs. Volunteer groups such as AA419 [2016], Stop Escrow Fraud [2016], and Malware Must Die [2016] cooperate to fight online scourges out of a sense of moral duty. Malware Must Die! even invokes the Crusades as an analogy. The inspiration for participation in these groups is similar to those found in the open-source community. Research on that community has shown that people volunteer for such projects because they value the contribution to the community [Hars and Ou 2001; Lakhani and Wolf 2005]. The added moral dimension of combating cybercrime provides further motivation for some.

A key incentive for firms to collect and contribute abuse data is to protect a brand or service. Firms impersonated by phishing criminals certainly have an incentive to collect timely reports of phishing websites—or to pay a company to do it for them. Phishing attacks can lead to direct losses from compromised accounts and an erosion of trust in the firm's brand or in online banking and e-commerce altogether. Search-engine operators seek out hacked webservers distributing drive-by-downloads in order to provide customers a safer search experience.

One can sometimes separate out the incentive for a firm to protect its reputation and the incentive to minimize financial harm. Frequently in cybersecurity, protecting reputation carries more weight than reducing losses. Moore and Clayton showed that while banks are vigilant in shutting down phishing websites impersonating their brand, they were less active in pursuing websites that recruited money mules to launder the proceeds of phished accounts [Moore and Clayton 2009b]. The discrepancy may stem from

the fact that phishing directly impersonated the brand, whereas the recruitment sites did not. Moreover, direct impersonation directly harms reputation and imposes financial losses, whereas mule recruitment imposes financial losses spread among many banks but no reputational harm.

A related incentive to collect and share abuse data is to receive useful data in return from others. For instance, many firms and ISPs monitor their own networks to mitigate incoming attacks (e.g., scanning activity). After blocking the attack, some operators choose to pass collected information on to the originating network operator or to an abuse data maintainer such as SANS [SANS Institute 2016], which collects firewall logs from contributors in its DShield system. The motivation for such sharing arrangements can be altruistic, but it may also be quid pro quo: The expectation is that they will receive notifications when their own customers are affected. The incentive for maintainers is similar, as sharing data in exchange for other data allows each maintainer to have greater and more relevant data at any given time, increasing the value of their services to their users.

While not an incentive per se, the cost of discovery definitely influences which data, and how much of it, gets collected. Some forms of abuse are intrinsically harder to discover than others. This explains why so much research into abuse and abuse reporting infrastructure has relied on spam data, for example. It is abundantly available, as anyone operating a mail server can generate it, albeit with very different quality levels. Other forms of abuse are much harder to detect: infected machines without a C&C infrastructure, for example, or targeted attacks or the resources used in Distributed Denial of Service (DDoS) attacks using spoofed origin-IP addresses.

When the main business model of the data collector is takedown and remediation, this may provide a disincentive to share data with others. For instance, brand-protection companies involved in website takedown maintain lists of websites to be removed. Moore and Clayton examined the feeds maintained by two large take-down companies, along with their client lists [Moore and Clayton 2008a]. They found that substantial overlap existed in the datasets: Phishing websites impersonating one company's clients appeared also in the other company's list, albeit at different moments in time. Because the companies refused to share their data feeds out of competitive concerns, this measurably undermined their effectiveness in combating phishing.

4.2. What Are The Incentives for Intermediaries to Receive and Act on Abuse Reports?

Intermediaries play a crucial role in finding and notifying resource owners to comply with remediation requests. ISPs, hosting providers, and registrars are routinely criticized for not being vigilant enough, both in the security industry as well as policy debates [Lichtman and Posner 2006]. This is typically blamed on a lack of incentives [STC 2007]. This raises the question of what incentives intermediaries actually have to receive abuse reports and then take action on them.

A number of factors weigh against intermediaries dealing directly with abuse reports. Foremost, the abuse often does not harm the intermediary directly. When a consumer PC is infected with keystroke-logging malware, the ISP is not affected. When websites have been hacked, the hosting provider may not be impacted. Furthermore, when there are negative externalities from insecurity, frequently the intermediary escapes harm. This may explain why Van Eeten et al. found that Dutch ISPs remediating customer infections were frequently unaware of free and publicly available sources of abuse reports that the authors relied on to conduct their study [van Eeten et al. 2011].

While the actual abuse may not harm the intermediary directly, there are still costs associated with inaction. A portion of the infected end-users contact their ISPs for help. This raises the ISP's cost of customer support. Interviews with ISPs [van Eeten and

Bauer 2008] suggests that incoming and outgoing calls to the call center cost between € 8 and € 16.

Some forms of abuse lead to impacts on a network's own operation. Web hosting providers might be blacklisted for refusing to deal with abuse requests promptly. ISPs that do not clamp down on spam senders within their network might end up being blocked partially or completely by blacklist operators or large ISPs.

Even when the harm is recognized by stakeholders, a coordination failure may prevent action from being taken. Most cybercrimes exploit multiple components of Internet infrastructure, and it is all too easy for one actor to decide not to take action in hopes that another will do so. Furthermore, since some cybercrimes such as botnet C&C require coordinated action to succeed, no single actor may feel strongly enough to take the lead. Traditionally, governments have helped solve collective action problems, and they have experimented with interventions to align incentives to deal with abuse.

Countries such as the United States, Canada, France, and Germany have laws that shield providers from liability for what is going on in their networks, as long as they act against abuse when notified—these are sometimes called safe harbor provisions. These are extended through trade agreements such as the Trans-Pacific Partnership [Bridy 2015], which allows such provisions to become de facto laws for participating countries. In light of such legal regimes, these industries have adopted notice and takedown procedures as a form of self-regulation to protect their safe harbor status. To what extent providers really act on such notices is highly contested. Another regulatory incentive stems from the threat of government regulation. The countries with telecommunication regulators that participated in the London Action Plan [Brown and Marsden 2007] consortium, an international anti-spam and malware initiative of public and private actors, were found to have lower infected levels in their ISPs' networks [Asghari et al. 2015]. In a study for the Dutch market, the researchers noticed reputation effects in the botnet mitigation efforts of ISPs. During a closed-door session, the researchers reported back to the ISPs how they were ranked in terms of remediating infections. Afterwards, the worst-performing ISP used the poor showing to get buy-in from their management to invest more resources [van Eeten et al. 2011]. Their performance dutifully reverted to the mean in subsequent quarters. In short, governmental attention to abuse issues has in some cases managed to motivate operators to take self-regulatory actions and become more active in botnet mitigation, for example (for a more detailed discussion, see Section 6).

Finally, public reputation and the possibility of being shamed is a potential incentive for intermediaries, particularly when they can link a decrease in their reputability (for hosting malicious content) to a decrease in their revenues. The incentive is similar to the concept of corporate responsibility/sustainability initiatives that are correlated with higher overall sales growth and subsequent increase of overall firm value [Lo and Sheu 2007]. An intermediary that behaves "responsibly" by remediating compromises and actively preventing the misuse and abuse of their networks may be able to increase their share of the overall market. Unfortunately there is mixed evidence on whether this holds in practice. Work by Tang et al. [2013] indicates that while information disclosure about networks originating spam from SpamRankings.net led to an overall 15.9% reduction in spam, networks originating the greatest amount of spam were indifferent to the rankings, suggesting that they were unconcerned with their public reputation. Similarly Stone-Gross et al. [2009] observes that while public information plays a role in helping to disrupt the operations of bulletproof hosts, they simply reappear shortly thereafter.

4.3. What Are The Incentives for Resource Owners to Act on Cleanup Requests?

Whether or not resource owners have a strong incentive to clean up infections when requested depends largely on whether the abuse affects the owner directly and whether

this harm is visible to the owner. When the harm is direct, say, for example, a malware-infected computer that limits productivity, the user may attempt to engage in cleanup. In certain cases, this ends the infection. In other cases, the user does not always possess the technical skill to remediate an infection, and the cost of bringing in a security professional can be too high relative to the incentive to do so [van Eeten and Bauer 2008].

When the harm is indirect, or borne by third parties, then the incentive to act is weaker than before. This is the case for those willfully engaging in illegal activity. They are generally unlikely to engage in any kind of self-directed cleanup following notification. Some may even utilize "bulletproof" hosting providers that actively advertise that they will not act on abuse reports [Sood and Enbody 2013].

Where the resource owner may not have an incentive to clean up but negative externalities persist, two strategies can be used to mitigate them. The first is to isolate the compromised resource. When compromised sites are blacklisted or removed from Google search results, the decline in traffic and revenue gets the attention of the resource owner [Edwards et al. 2012]. Traffic declines can have large monetary impacts on resource owners. Chachra et al. observes that, when blacklisted, the earnings of spammy domains plateaued within 2h, compared to non-blacklisted spammy domains that continued to rise, concluding that, all things being equal, the quicker to blacklist, the quicker to elicit a response from a resource owner [Chachra et al. 2014]. Similarly, quarantine procedures by ISPs that limit Internet connectivity for bot-infected customers may incentivize the customers to participate in cleanup efforts. The second strategy is to pursue legal action, as we will discuss when we consider the role of governments in Section 6.

5. MEASURING EFFECTIVENESS OF ABUSE REPORTING

We have described a complex ecosystem of actors who face different incentives and take a range of approaches in combating cybercrime. As researchers, our ultimate goal is to understand how effective abuse reporting infrastructure is and to identify the circumstances that best promote remediation.

The effectiveness of abuse reporting is a controversial topic. Many in the security field see it as too little, too late. They have pointed to its reactive nature [Brenner 2004], to lax intermediaries or resource owners [Lichtman and Posner 2006], and to the rapid replenishment of criminal resources even after extensive takedown and remediation campaigns [Klimburg 2011]. Abuse reporting efforts can be construed as treating the symptoms rather than going after the root causes.

All of these complaints are justified to some extent, but they overlook evidence to the contrary. Some notification campaigns or shaming efforts have achieved remarkable results, even though the abuse issue did not directly burden the intermediary that was asked to act or the resource owner [Vasek and Moore 2012; Durumeric et al. 2014; Tang et al. 2013]. To a large extent, this is matter of expectations. Skeptics could point out that 2 months after the disclosure of Heartbleed, 3% of HTTPS sites in the Alexa Top 1 Million remained vulnerable. Although the aspiration might be to reach 100% success [Durumeric et al. 2014], the fact that 97% of potentially vulnerable systems were patched in a reasonable timeframe suggests significant improvement in the overall security of the Internet, even if notifications cannot exclusively be credited for all patching in that period of time. Of course, the positive experience patching Heartbleed may not generalize, given that it was so broadly publicized.

It is widely accepted that many good security tools and practices are only partially effective. Abuse reporting is no different. But how partial is its success? The truth of the matter is that the actual effectiveness has never been systematically assessed. We will summarize the existing literature, en route to a more systematic analysis. We then review key attributes of the model that might impact how successful abuse reporting

efforts will be. We believe that this will enable future research to establish causality between the efforts to discover and report abuse and what is ultimately remediated.

5.1. Existing Research

A few researchers have set out to measure the effectiveness of abuse reporting, usually with the goal of improving the overall response to cleanup efforts. Vasek and Moore sent out notifications to remediate websites distributing malware [Vasek and Moore 2012]. They randomly assigned incoming reports from StopBadware's community malware feed to three groups: minimal notifications, detailed notifications that included evidence for why the website was distributing malware, and a control group receiving no notifications. Abuse reports were sent by email to public abuse contacts at hosting providers and resource owners (if the websites were hacked) and to registrars (if the websites were maliciously registered). They found that 62% of websites receiving detailed notices were cleaned within 16 days, compared to 45% of those not receiving notifications. Notably, they found no difference in response rates between receiving a minimal report and not receiving any report at all. This research established the importance of providing infection details when reporting to intermediaries.

Following up on this study, Cetin et al. [2015] notified operators of websites infected with the Asprox botnet. In order to assess the impact of sender reputation, they devised three treatment groups with varied email addresses: a webmail account, a University account, and an anti-malware non-profit. The authors found no statistically significant difference between the treatment groups, suggesting that the sender email address does not matter greatly when responding to abuse reports. Nonetheless, the study reaffirmed Vasek and Moore's findings that detailed notices work, because all three treatment groups outperformed the control group where no abuse reports were sent. Based on a 16-day observation period, the first wave showed a 35% rate of cleanup for the control group and between 65% and 80% for the treatments, while the second wave showed 26% for the control group and between 44% and 49% rate for the treatments. The median cleanup times were also substantially reduced with the control groups at 8 and 14 days for the two waves, respectively, whereas every treatment group ranged between 1.5 and 4 days.

A recent observational study using data collected by Google offers the largest-scale data analysis on the effectiveness of abuse reporting to date [Li et al. 2016]. In examining over 700,000 website compromises over the course of 1 year, the reserachers found that abuse reports sent via the Google Search Console triggered a 50% increase in the probability of cleanup and, furthermore, that the durations of infections fall by 62%. They also provided some evidence that website operators who have not signed up for alerts nonetheless clean their site up more often and faster if users are blocked from visiting their website by interstitial warnings.

Two additional studies have experimented with notifications involving web-based malware. Canali et al. set up vulnerable webservers on 22 hosting services [Canali et al. 2013]. They then compromised the webservers and sent out notifications to all hosting providers after 25 days had passed. Half the providers did not respond; among those who did, the most common response was to suspend the compromised webserver. The authors also sent false abuse reports to gauge the response to false positives. They found that 3 of the 22 notified hosting providers either suspended the account or warned the user of being compromised, despite a lack of evidence. Meanwhile, Nappa et al. sent abuse reports to providers hosting 19 long-lived exploit servers [Nappa et al. 2013]. Only 7 (39%) responded to the report, taking an average of 4.3 days to disconnect the server.

Two recent studies described experiments to notify affected webserver operators running vulnerable software. Following the high-profile disclosure of the Heartbleed bug in

OpenSSL, many webservers were patched. After a few weeks, though, around 2.4% of all HTTPS hosts remained vulnerable. Durumeric et al. ran an experiment to automatically notify 150,000 vulnerable hosts identified through IP WHOIS lookups [Durumeric et al. 2014]. They found that the rate of patching increased by 47%, an impressive feat given the automatic and unsolicited nature of the notification. Another highly publicized issue involved DDoS amplification vulnerabilities in several User Datagram Protocol (UDP)-based protocols [Rossow 2014]. Kührer et al. engaged CERT/CC and Cisco in a campaign to notify the owners of equipment running vulnerable Network Time Protocol (NTP) servers. They observed a 92% reduction in vulnerable servers, from 1.6 million to 126,000 in under 3 months. These studies show that, for vulnerabilities at least, notifications can really work.

Instead of sending notifications to intermediaries or end-users, simply contributing to influential abuse database maintainers can sometimes have a positive impact on cleanup. In a study of Zeus botnet C&C domains, Ganan et al. found that domains reported to online trackers were remediated more quickly than domains identified by a private company that did not report findings beyond its own customers [Gañán et al. 2015].

Mentioned earlier in Section 4, Tang et al. [2013] aggregated information on spam activity taking place at ASes and published data for a subset of the ASes (the treatment group) while reporting nothing for others (the control group). They then monitored resulting spam activity, reporting preliminary evidence that the publicly shamed treatment group experienced a 15.9% reduction in spam activity.

In some cases, research not directly focused on notification experiments has offered insight on the effectiveness of certain attributes. For example, Stone-Gross et al. successfully leveraged personal relationships with intermediaries to remediate 16 of 20 hosts carrying spambot activity [Stone-Gross et al. 2011b]. In another article, the same team observed that cleanups were somewhat ineffective when webmasters failed to upgrade vulnerable software, observing recompromise in 476 sites [Stone-Gross et al. 2011a].

Studies such as these represent the proverbial tip of the iceberg when it comes to measuring the effectiveness of voluntary remediation. We next set out to describe key attributes that might influence cleanup, in hopes of guiding future research in this area.

5.2. Key Attributes That Influence Abuse Reporting Effectiveness

In addition to the incentives discussed in Section 4, attributes of the abuse reporting infrastructure can influence the overall effectiveness of cleanup efforts.

Table II summarizes some of these attributes, along with the possible values that each could take, in an attempt to help establish a potential research agenda. This list should not be considered exhaustive. The first group involves attributes of actions, followed by characteristics of data and the actors involved.

Action attributes cover characteristics of sharing abuse data, sending abuse reports and cleaning up infections. Consider the delivery mechanism, which can influence whether the intermediaries accepts the incoming reports and can process them at scale and efficiently pass the information onwards to resource owners or not. For example, if the email comes from a personal contact, the notifier is more likely to get a response. On the other hand, if the abuse reports are provided as a dump or an unsolicited email, then the intermediary may ignore it. The delivery mechanism for abuse reports to resource owners is also critical, because they are likely not expecting to be contacted. In this case, phone calls or even postal letters may be warranted. Also, the escalation process, in the case of non-response, can be automated to some extent, where certain

Table II. Key Attributes of the Abuse Reporting Infrastructure That May Influence the Effectiveness of Remediation Efforts. Refer Back to Table I to Review the Roles of Actors

Attribute	Description of Attribute	Source	Sink	Action(s)	Possible Values
Action Attributes					
Delivery Mechanism	How is information shared?	ADM	AN	Share Abuse Data	Solicited Email, Unsolicited Email, Feed, Query, Dump, Personal Contact
Delivery Mechanism	How is information shared?	AN	INT, RO	Send Abuse Report	Unsolicited Email, Phone Call, SMSes, Quarantine, Query, Public Post
Recipient Addressability	Who receives the abuse report?	AN	INT, RO	Send Abuse Report	Public Contact, Private Contact, General Contact
Follow-up/ Judgment	How does the intermediary evaluate abuse reports?	AN	INT	Send Abuse Report	Manual, Automated
Added Value to the abuse report	How does the AN transform the abuse report?	AN	INT, RO	Send Abuse Report	Legalese, None, Education, Simplification, Explanation, Threats
Blocking Level	Any connectivity restrictions before cleanup completes?	INT, RO	RO	Cleanup	Partial, Complete, None
Treatment	Does the cleanup treat the root cause or just symptoms?	RO, INT	–	Cleanup	Deleted, Repaired (but not upgraded), Repaired (and upgraded)
Data Attributes					
Reliability	What is the dependability of the data (e.g., false-positive rate)?	AC, ADM, AN	–		High, Medium, Low, Anonymous
Coverage	What is the geographic scope of the data?	AC, ADM, AN	–		International, National, Regional, Select Group, Random
Cost	Does the data cost money?	AC, AN, ADM	–		Free, Freemium, Paid
Type	What is the type of abuse?	AC, ADM	–		Malware, Spam, Phishing, . . .
Freshness	How "fresh" are the data?	AC, ADM, AN	–		High, Medium, Low
Regularity	Is the data one-off or ongoing?	AC, ADM, AN	–		One-off, Ongoing
Parseability	Is the format parsable/designed for the receiving system?	AC, ADM, AN	–		Standard (Automatable), Non Standard (Manual)
Shareability	Are there any limitations on what can be done with the data?	ADM, AN	–		Yes, No
Accessibility	Who can access the data and for how long?	ADM	–		Fully public, Oracle, Private (All), Private (Tailored)
Actor Attributes					
Longevity	How old is the organization?	AC, ADM, AN, SV	–		# of years
Reputation	How well known is the organization?	AC, ADM, AN, SV	–		High, Medium, Low, Anonymous
Institutional Affiliation	What association does the participant have? (influences credibility)	AC, ADM, AN, SV	–		Nonprofit, Corporation, Government, Education

conditions trigger filtering or quarantining, for example. Automation is important to determine the scale at which an intermediary can process and act on abuse reports.

The addressability of the recipient of an abuse report also matters. When contacts are publicly available, then it is more likely that the abuse report can be delivered. On the other hand, the public address may not be responded to as expeditiously. This is linked to the follow-up and judgment attribute, which can either be automated or manual, and further helps to explain action or inaction.

Furthermore, the abuse notifier may transform the report in such a way that encourages greater compliance. For example, providing additional contextual information to the original abuse report could increase response rates, such as details about the compromise [Vasek and Moore 2012]. Educational information such as instructions on how to remediate the compromise or explaining the harm imposed on others might help. Even attaching legalese to the report conveying the earnestness of the message might elicit greater compliance.

Remediation attributes focus on the actual act of cleanup that an intermediary or a resource owner undertakes. These include whether the resource is blocked and ultimately the treatment it receives. Blocking access to a resource is a step taken to prevent access to the compromised resource and can be partial (filtering connections to ports associated with a botnet), complete (denial of network access for an illegal site), or none. While some intermediaries limit their response to notification of abuse to these options, others use this as a first step towards treatment, including deleting the content, repairing the content, or repairing and protecting the content from recompromise. The important distinction here is that, when the treatment leads to an upgraded installation (as opposed to a simple repair), the likelihood of recompromise [Moore and Clayton 2009a] should go down substantially, as the original vulnerability is generally patched.

Characteristics of the abuse data itself can play a big role in how effective remediation becomes. Foremost among these is the reliability of the data itself. It is crucial that abuse reports have very low false positives. The reason is that the volume of abuse reports frequently necessitates an automated response. Without a human in the loop to sanity-check the data, even a small number of false positives can overwhelm an abuse team. Aware of this, active attempts to poison data feeds and reduce their reliability and credibility have to be considered (and potentially countered) in any collection and dissemination attempts. Otherwise, intermediaries may ignore such data sources or, worse, stop trying to remediate abusive sites altogether.

Actor attributes highlight participants and intrinsic characteristics that affect their ability to successfully participate in the abuse reporting infrastructure. These include longevity, reputation, and institutional affiliation. For example, the remarkable response rate reported by Kührer et al. (92% reduction in 3 months) may be related to the reputation of CERT/CC and Cisco, who were part of the notification campaign.

Notifiers with a longer history might get a better response relative to notifier's that were more recently started, other things being equal. One way to measure this is by looking at technically similar abuse reports issued by groups with longer histories and comparing them with control groups created internally that lack any operational history, in order to understand how longevity influences response from intermediaries and ultimately, cleanup rates. However, one must be careful in not conflating longevity and reputation.

Reputation as an attribute would measure some external perception of the organization, based on a combination of quality, reliability, and trustworthiness of the organization. While it is highly likely an organization with high reputation would also have been around longer, it is also possible that new and emerging industry-affiliated notifiers might have the reputation needed to get sufficient responses from intermediaries,

thus affecting cleanup rates. This links into the third attribute: institutional affiliation, which is divided between non-profit, government, corporate, and educational.

6. GOVERNMENT AND ABUSE REPORTING

At the heart of the abuse reporting infrastructure is voluntary action. This voluntary nature does not imply that the decisions of intermediaries or resource owners to act on abuse reports are made in absolute freedom and are only driven by their good (or not-so-good) intentions. As presented in Section 4, these actors operate under incentives that are shaped by other actors, including governments. Increasing the effectiveness of abuse reporting should therefore explore ways in which governments can support better remediation.

When market incentives for security need to be changed, people typically look to governments for assistance [Anderson et al. 2008]. The reality is that governments already shape the incentives of intermediaries and resource owners. Think of intellectual property regimes and safe harbor provisions, which have led to the emergence of notice and takedown procedures in many countries. Intellectual property law was also instrumental in Microsoft's botnet takedown actions, using civil law provisions to seize domains and resources [Goguen 2014].

Another area where governments shape incentives is privacy law. These regulations are mentioned by some intermediaries as an excuse to avoid monitoring for abuse in their own networks. Privacy considerations also put constraints on data sharing among actors and jurisdictions. For example, a European Union– (EU) funded anti-botnet project that aimed to collect and share infection data across the EU in order to protect citizens got bogged down in the legal complexities of sharing these data across jurisdictions that all have their own specific privacy regulations [ACDC 2016].

How could governments—or transnational governance regimes like the London Action Plan [Brown and Marsden 2007]—make abuse reporting more effective? They can do so in two roles: as participants and as incentivizers.

6.1. Governments as Participants

With respect to its role as an actor, the FBI's Cyber Crimes Investigative Unit [FBI 2016] or its European equivalent, the European Cybercrime Centre [Europol 2016], are both known to play the roles of abuse data contributor (as they have internal teams that may make contributions to databases), maintainer (as they collect user-filed reports and store and share abuse reports with the public and private sectors), and notifier (as they reach out to intermediaries and resource owners who expect that they would remediate abuse successfully [Vixie 2014]). At the national level, CERTs and regulators typically collect abuse data that are relevant to their constituencies.

The research in this case focuses on if and how the government should play a more direct and active role as an actor. Mulligan and Schneider make the case for a government agency that coordinates all abuse reports and reduces the coordination failures involved with the existing fragmented nature of vendors and databases [Mulligan and Schneider 2011]. Such coordination failures are reinforced by literature [Metcalf and Spring 2013] indicating that more than 99% of blacklist entries are unique, with hundreds of blacklists available. National CERTS, many of which are government organizations or government funded, play a key role here. They are a natural point of contact for abuse data relevant to their jurisdiction, which they then pass on to the relevant network operators. Another way in which governments participate is by supporting Information Sharing and Analysis Centers (ISACs). These non-profit organizations bring together participants by sector to share intelligence regarding threats and vulnerabilities affecting critical infrastructure with the hope of increasing overall security. In the United States, these organizations emerged following the release of

Presidential Decision Directive-63 (PDD-63) in 1998. Today, individual ISACs exist encompassing everything from aviation to finance to water, while a National Council of ISACs coordinates bamong them [National Council of ISACs 2016]. Effectiveness and participation rates have ranged considerably, with the financial services ISAC (FS-ISAC) widely seen to be the most active. The ISAC model has been exported worldwide, for example, within Europe [isac.eu 2016] and India [ISAC 2016], though the US ISACs are the furthest along at this point.

6.2. Governments as Incentivizers

One key way for governments to change incentives is through subsidies. In Germany [Anti-Botnet-Advisory Centre 2016], the Netherlands [Abuse Information Exchange 2016], and Japan [Krebs 2010], the government has subsidized public-private initiatives to collect data on malware infections to then notify ISPs and end-users. The idea is that these centralized clearinghouses and support centers reduce the cost of ISPs and users to counteract infections, thereby strengthening their incentives for remediation.

Such a strategy could be taken further. For example, work by Clayton [2010] compares the problem of malware infection to public health, as both potentially carry negative externalities. By proposing the government subsidize companies that directly engage in the cleanup of end-users, the cost for users to access a remediation service is reduced. This stimulates greater purchasing of cleanup services, while also potentially giving governments room to justify regulations that mandate the timely cleanup of malware infections. The model proposes a fee for service of $30 for the end-user (to reduce the possibility and/or perception of moral hazard). Government then tenders bids on the additional cost it must bear of providing cleanup services to end-users taking the fee for service into account. Clayton estimates companies could bid as low as $11.05 per cleanup because of the value companies place on an end-user's willingness to purchase other services while being serviced (i.e., anti-virus subscription). Assuming a payout rate of 0.5% of all computers in the United States along with one computer per person, the total cost is effectively $0.66 per person, per year, or less than a number of public health initiatives.

Another way to shape incentives is via law and regulation. Right now, it is frequently unclear which party, if any, is responsible for remediating a compromised device. Intermediaries may be in a position to observe a compromise and take action, and many do so voluntarily. But governments could further clarify responsibility through the law. Lichtman and Posner argued that ISPs should be held responsible for cleaning up infected customer machines on their networks by applying the principle of indirect intermediary liability [Lichtman and Posner 2006].

About a decade ago, the Finnish telecommunication regulator FICORA legally required the main ISPs in the country to clean up known bots. It also deployed automated tools to report abuse incidents to the providers [Kiuru 2016]. In terms of botnet infections, Finland has long been recognized as harboring the cleanest networks in the world [Rains 2014].

Finland is unique with regards to a mandatory botnet cleanup regime. That being said, liability regimes have been tried in other countries to combat other undesirable online activities. For example, the United States regulates the publication of offensive content via the Communications Decency Act, copyrighted material via the Digital Millennium Copyright Act, and online gambling services via the Unlawful Internet Gambling Enforcement Act. In each case, an intermediary such as an ISP or payment network is obliged to assist in removing illegal content. Moore calls for a similar regime to combat end-user malware, in which ISPs must act on abuse reports affecting their customers or assume responsibility for the harm they cause [Moore 2010].

What remains clear is that differences in intermediary liability law across countries do not explain the large variation in security performance. Peterson et al. compared the intermediary liability regimes of high-performing Finland and low-performing Lithuania [Peterson et al. 2014]. They noted that the laws concerning ISP responsibility were similar, even though the outcomes could not differ more.

Approaches via criminal law have been used only sparingly, against a number of high-profile targets. This mostly reflects the high costs of investigating and prosecuting globalized cybercrime. One way in which this approach might be extended is for law enforcement agencies to target intermediaries. In the Netherlands, the public prosecutor's office and the police recently announced a project to go after "bad hosting" based in the Netherlands [van't Hof 2014]. Naming and shaming is explicitly considered a strategy. Even without the end result of a successful criminal conviction, this kind of attention may encourage lax providers to shore up their remediation efforts.

We should note that while a variety of proposals have been circulating for governments to allocate further funding or adopt new rules to generate stronger incentives for intermediaries and resources owners, none of these proposals are anywhere close to being accepted, let alone implemented. For the immediate future, it seems more promising to increase the effectiveness of the abuse reporting infrastructure within the existing incentive structure.

7. CONCLUDING REMARKS

The Internet has developed as a remarkable feat of self-organization. That same approach is being leveraged to arrest the rise of cybercrime. Private actors, ranging from volunteers who meet on online forums to technology titans and experts from security firms, gather abuse data and disseminate reports to relevant intermediaries and resource owners. They shut down websites impersonating banks and botnets infecting many thousands of computers. Most of this is achieved without any mandate from governments or assistance from law enforcement.

That this system works at all in countering cybercrime is impressive, made all the more so by the fact that its operation is so poorly understood. In this article, we have set out to explain how the abuse reporting infrastructure functions. We have outlined incentives for participation among notifiers, intermediaries, and resource owners. Finally, we briefly considered the role governments can play in fostering a more capable response.

We hope that by explaining the important role of abuse reporting, we can spur the research community to more rigorously study what works and how to make future efforts more effective.

REFERENCES

AA419. 2016. Artists Against 419–AA419. Retrieved from https://www.aa419.org.

Saeed Abu-Nimeh, Dario Nappa, Xinlei Wang, and Suku Nair. 2007. A comparison of machine learning techniques for phishing detection. In *Proceedings of the 2nd APWG eCrime Researchers Summit*. ACM, 60–69.

Abuse Information Exchange. 2016. Abuse Information Exchange. Retrieved from https://www.abuse informationexchange.nl/english.

ACDC. 2016. Advanced Cyber Defence Centre. Retrieved from https://www.acdc-project.eu.

Ross Anderson, Rainer Böhme, Richard Clayton, and Tyler Moore. 2008. Security economics and European policy. In *Managing Information Risk and the Economics of Security*, M. E. Johnson (Ed.). Springer, 55–80.

Frankie Angai, Calvin Ching, Isaiah Ng, and Cameron Smith. 2010. *Analysis on the Effectiveness of Safe Browsing Services*. Technical Report. University of British Columbia.

Anti-Botnet-Advisory Centre. 2016. Anti-Botnet Advisory Centre. Retrieved from https://www.botfrei.de.

Manos Antonakakis and Yacin Nadji. 2013. Microsoft DCU—strike three. Now what? *Damballa Blog* Retrieved from https://www.damballa.com/microsoft-dcu-strike-three-now-what-2/.

Manos Antonakakis, Roberto Perdisci, Yacin Nadji, Nikolaos Vasiloglou II, Saeed Abu-Nimeh, Wenke Lee, and David Dagon. 2012. From throw-away traffic to bots: Detecting the rise of DGA-based malware. In *Proceedings of the USENIX Security Symposium*. USENIX, 491–506.

APWG. 2015. Anti-Phishing Working Group. Retrieved from http://www.antiphishing.org/.

APWG. 2016. Report Phishing—APWG. Retrieved from https://apwg.org/report-phishing/.

Hadi Asghari, Michel J. G. van Eeten, and Johannes M. Bauer. 2015. Economics of fighting botnets: Lessons from a decade of mitigation. *IEEE Secur. Priv.* 13, 5 (2015), 16–23.

Sushma Nagesh Bannur, Lawrence K. Saul, and Stefan Savage. 2011. Judging a site by its content: Learning the textual, structural, and visual features of malicious web pages. In *Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence*. ACM, 1–10.

BBC. 2014. Millions of German passwords stolen. Retrieved from http://www.bbc.com/news/technology-25825784.

Leyla Bilge, Engin Kirda, Christopher Kruegel, and Marco Balduzzi. 2011. EXPOSURE: Finding malicious domains using passive DNS analysis. In *NDSS*.

Susan W. Brenner. 2004. Distributed security: Moving away from reactive law enforcement. *Int. J. Commun. Law Policy* 9 (2004).

Annemarie Bridy. 2015. A user-focused commentary on the TPP'S ISP safe harbors. *Stanford IP-Watch Blog.* Retrieved from http://cyberlaw.stanford.edu/blog/2015/11/user-focused-commentary-tpp's-isp-safe-harbors.

Ian Brown and Chris Marsden. 2007. Co-regulating internet security: The London action plan. In *Proceedings of the Global Internet Governance Academic Network 2nd Annual Symposium.*

Davide Canali, Davide Balzarotti, and Aurélien Francillon. 2013. The role of web hosting providers in detecting compromised websites. In *Proceedings of the 22nd International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 177–188.

Davide Canali, Marco Cova, Giovanni Vigna, and Christopher Kruegel. 2011. Prophiler: A fast filter for the large-scale detection of malicious web pages. In *Proceedings of the 20th International Conference on World Wide Web*. ACM, 197–206.

Orcun Cetin, Mohammad Hanif Jhaveri, Carlos Gañán, Michel van Eeten, and Tyler Moore. 2015. Understanding the role of sender reputation in abuse reporting and cleanup. In *Proceedings of the 14th Annual Workshop on Economics of Information Security (WEIS'15)*.

Neha Chachra, Damon McCoy, Stefan Savage, and Geoffrey M. Voelker. 2014. Empirically characterizing domain abuse and the revenue impact of blacklisting. In *Proceedings of the Workshop on the Economics of Information Security (WEIS'15)*. Retrieved from http://econinfosec.org/archive/weis2014/papers/Chachra-WEIS2014.pdf.

Pern Hui Chia and Svein Johan Knapskog. 2012. Re-evaluating the wisdom of crowds in assessing web security. In *Financial Cryptography and Data Security*. Springer, 299–314.

Hyunsang Choi and Heejo Lee. 2012. Identifying botnets by capturing group activities in DNS traffic. *Comput. Networks* 56, 1 (2012), 20–33.

Richard Clayton. 2009. How much did shutting down McColo help? In *Proceedings of the 6th Conference on Email and Antispam (CEAS)*.

Richard Clayton. 2010. Might governments clean-up malware? In *Proceedings of the 9th Annual Workshop on the Economics of Information Security (WEIS'10)*. Retrieved from http://weis2010.econinfosec.org/papers/session4/weis2010_clayton.pdf.

Conficker Working Group. 2011. Conficker working group: Lessons learned. Retrieved from http://www.confickerworkinggroup.org/wiki/uploads/Conficker_Working_Group_Lessons_Learned_17_June_2010_final.pdf.

Conficker Working Group. 2016. Conficker Working Group. (2016). Retrieved from http://www.conficker workinggroup.org.

Cybercrime tracker. 2016. Cybercrime tracker. Retrieved from http://cybercrime-tracker.net.

Daan de Graaf, Ahmed F. Shosha, and Pavel Gladyshev. 2013. BREDOLAB: Shopping in the cybercrime underworld. In *Digital Forensics and Cyber Crime*. Springer, 302–313.

David Dittrich. 2012. So you want to take over a botnet. In *Proceedings of the 5th USENIX Conference on Large-Scale Exploits and Emergent Threats*. USENIX Association, 6–6.

Zakir Durumeric, James Kasten, F. Li, Johanna Amann, Jethro Beekman, Mathias Payer, Nicholas Weaver, J. A. Halderman, Vern Paxson, and Michael Bailey. 2014. The matter of heartbleed. In *ACM Internet Measurement Conference (IMC)*.

Benjamin Edwards, Tyler Moore, George Stelle, Steven Hofmeyr, and Stephanie Forrest. 2012. Beyond the blacklist: Modeling malware spread and the effect of interventions. In *Proceedings of the 2012 Workshop on New Security Paradigms*. ACM, 53–66.

Europol. 2016. A Collective European Response to Cybercrime. Retrieved from https://www.europol.europa.eu/ec3.

Facebook. 2016. Threat Exchange—Threat Exchange - Facebook for Developers. Retrieved from https://developers.facebook.com/products/threat-exchange.

FBI. 2016. FBI Cyber Crimes Division. Retrieved from http://www.fbi.gov/about-us/investigate/cyber.

FireEye. 2016. Cyber Security & Malware Protection—FireEye, Inc. Retrieved from http://www.fireeye.com.

Carlos Gañán, Orcun Cetin, and Michel van Eeten. 2015. An empirical analysis of ZeuS C&C lifetime. In *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security*. ACM, 97–108.

Natalie Goguen. 2014. Update: Details on Microsoft Takeover. Retrieved July, 10 2014 from http://www.noip.com/blog/2014/07/10/microsoft-takedown-details-updates/.

Google. 2016a. Google Developers—Safe Browsing API. Retrieved 2016 from https://developers.google.com/safe-browsing.

Google. 2016b. Search Console. Retrieved June 23, 2016 https://www.google.com/webmasters/tools/home?hl=en.

Alexander Hars and Shaosong Ou. 2001. Working for free? Motivations of participating in open source projects. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. IEEE, 9–pp.

Janine S. Hiller. 2014. Civil cyberconflict: Microsoft, cybercrime, and botnets. *Santa Clara Comput. High Tech. LJ* 31 (2014), 163.

ISAC. 2016. About—Information Sharing and Analysis Center. Retrieved from https://certisac.org/about. isac.eu. 2016. isac.eu. Retrieved from http://isac.eu.

John P. John, Fang Yu, Yinglian Xie, Arvind Krishnamurthy, and Martin Abadi. 2011. deSEO: Combating search-result poisoning. In *Proceedings of the USENIX Security Symposium*. USENIX Association.

Antti Kiuru. 2016. Incident Response Made Better by Agile Robots. Retrieved from June 13, 2016 https://www.first.org/resources/papers/conf2016/FIRST-2016-110.pdf.

Alexander Klimburg. 2011. Mobilising cyber power. *Survival* 53, 1 (2011), 41–60.

Brian Krebs. 2008. Host of internet spam groups is cut off. *Washington Post, Nov* 12 (2008). Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2008/11/12/AR2008111200658.html.

Brian Krebs. 2010. Talking bots with Japan's "cyber clean center." *KrebsonSecurity* Retrieved from http://krebsonsecurity.com/2010/03/talking-bots-with-japans-cyber-clean-center.

Marc Kührer, Christian Rossow, and Thorsten Holz. 2014. Paint it black: Evaluating the effectiveness of malware blacklists. In *Research in Attacks, Intrusions and Defenses*. Springer, 1–21.

Karim R. Lakhani and Robert G. Wolf. 2005. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. *Perspect. Free Open Source Softw.* 1 (2005), 3–22.

Felix Leder, Tillmann Werner, and Peter Martini. 2009. Proactive botnet countermeasures–an offensive approach. *Virtual Battlefield: Perspect. Cyber Warf.* 3 (2009), 211–225.

Nektarios Leontiadis, Tyler Moore, and Nicolas Christin. 2011. Measuring and analyzing search-redirection attacks in the illicit online prescription drug trade. In *Proceedings of USENIX Security 2011*. San Francisco, CA.

Nektarios Leontiadis, Tyler Moore, and Nicolas Christin. 2014. A nearly four-year longitudinal study of search-engine poisoning. In *Proceedings of ACM CCS 2014*.

Frank Li, Grant Ho, Eric Kuan, Yuan Niu, Lucas Ballard, Kurt Thomas, Elie Bursztein, and Vern Paxson. 2016. Remedying web hijacking: Notification effectiveness and webmaster comprehension. In *Proceedings of the International World Wide Web Conference*.

Doug Lichtman and Eric Posner. 2006. Holding internet service providers accountable. *Supr. Court Econ. Rev.* (2006), 221–259.

Phillip Lin. 2009. Anatomy of the mega-d takedown. *Network Secur.* 2009, 12 (2009), 4–7.

He Liu, Kirill Levchenko, Márk Félegyházi, Christian Kreibich, Gregor Maier, Geoffrey M. Voelker, and Stefan Savage. 2011. On the effects of registrar-level intervention. In *Proceedings of the 4th USENIX LEET*.

Jason Livingood, Nirmal Mody, and Mike O'Reirdan. 2012. Recommendations for the Remediation of Bots in ISP Networks. RFC 6561 (Informational). Retrieved from http://www.ietf.org/rfc/rfc6561.txt.

Shih-Fang Lo and Her-Jiun Sheu. 2007. Is corporate sustainability a value-increasing strategy for business? *Corp. Gov.: Int. Rev.* 15, 2 (2007), 345–358.

Malware Must Die. 2016. Homepage. Retrieved from http://malwaremustdie.org

Steve Mansfield-Devine. 2010. Battle of the botnets. *Netw. Secur.* 2010, 5 (2010), 4–6.

Niels Provos Panayiotis Mavrommatis and Moheeb Abu Rajab Fabian Monrose. 2008. All your iFrames point to us. In *Proceedings of the 17th USENIX Security Symposium*.

MDL. 2016. Malware domain List (MDL). Retrieved from http://www.malwaredomainlist.com.

Messaging Anti-Abuse Working Group. 2007. M3AAWG best practices for the use of a walled garden. *San Francisco, CA* (2007).

Leigh Metcalf and Jonathan M. Spring. 2013. *Everything You Wanted to Know About Blacklists But Were Afraid to Ask*. Technical Report. Software Engineering Institute—Carnegie Mellon University.

Microsoft News Center. 2013. Microsoft, the FBI, Europol and industry partners disrupt the notorious ZeroAccess botnet. (December 2013). Retrieved December 5, 2013 from https://news.microsoft.com/2013/12/05/microsoft-the-fbi-europol-and-industry-partners-disrupt-the-notorious-zeroaccess-botnet/.

Nirmal Mody, Alex Kasyanov, Jason Livingood, Brian Lieu, and Chae Chung. 2011. Comcast's Web Notification System Design. RFC 6108 (Informational). (February 2011). DOI:http://dx.doi.org/10.17487/rfc6108

Meaghan Molloy. 2014. Operation Tovar: The Latest Attempt to Eliminate Key Botnets. Retrieved July 8, 2014 from https://www.fireeye.com/blog/threat-research/2014/07/operation-tovar-the-latest-attempt-to-eliminate-key-botnets.html.

Tyler Moore. 2010. The economics of cybersecurity: Principles and policy options. *Int. J. Crit. Infrastruct. Protect.* 3, 3–4 (2010), 103–117.

Tyler Moore and R. Clayton. 2007. Examining the impact of website take-down on phishing. In *Proceedings of the 2nd APWG eCrime Researcher's Summit*.

Tyler Moore and Richard Clayton. 2008a. The consequence of non-cooperation in the fight against phishing. In *Proceedings of the 3rd APWG eCrime Researchers Summit*.

Tyler Moore and Richard Clayton. 2008b. Evaluating the wisdom of crowds in assessing phishing websites. In *Financial Cryptography and Data Security (Lecture Notes in Computer Science)*, Gene Tsudik (Ed.), Vol. 5143. Springer, 16–30. Retrieved from http://lyle.smu.edu/~tylerm/fc08.pdf.

Tyler Moore and Richard Clayton. 2009a. Evil searching: Compromisel and recompromise of internet hosts for phishing. In *Proceedings of the 13th International Conference on Financial Cryptography and Data Security*.

Tyler Moore and Richard Clayton. 2009b. The impact of incentives on notice and take-down. In *Managing Information Risk and the Economics of Security*, M. E. Johnson (Ed.). Springer, 199–223.

Tyler Moore and Richard Clayton. 2011. The impact of public information on phishing attack and defense. *Commun. Strat.* 1, 81 (2011), 45–68.

Deirdre K. Mulligan and Fred B. Schneider. 2011. Doctrine for cybersecurity. *Daedalus* 140, 4 (2011), 70–92.

Yacin Nadji, Manos Antonakakis, Roberto Perdisci, David Dagon, and Wenke Lee. 2013. Beheading hydras: Performing effective botnet takedowns. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*. ACM, 121–132.

Antonio Nappa, M. Zubair Rafique, and Juan Caballero. 2013. Driving in the cloud: An analysis of drive-by download operations and abuse reporting. In *Proceedings of the 10th Conference on Detection of Intrusions and Malware & Vulnerability Assessment*. Springer, Berlin, 1–20.

National Council of ISACs. 2016. National Council of ISACs—About NCI. Retrieved from http://www.nationalisacs.org/about-nci.

Alexandros Ntoulas, Marc Najork, Mark Manasse, and Dennis Fetterly. 2006. Detecting spam web pages through content analysis. In *Proceedings of the 15th International Conference on World Wide Web*. ACM, 83–92.

Roberto Perdisci, Igino Corona, David Dagon, and Wenke Lee. 2009. Detecting malicious flux service networks through passive analysis of recursive DNS traces. In *Computer Security Applications Conference, 2009. ACSAC'09. Annual*. 311–320.

Nicole Perlroth and David Gelles. 2014. Russian Hackers Amass Over a Billion Internet Passwords. Retrieved from http://nytimes.com/2014/08/06/technology/russian-gang-said-to-amass-more-than-a-billion-stolen-internet-credentials.html.

Jason H. Peterson, Lydia Segal, and Anthony Eonas. 2014. Global cyber intermediary liability: A legal & cultural strategy. *Pace Law Rev.* 34 (2014), 586.

PhishTank. 2016. PhishTank. Retrieved from https://www.phishtank.com/.

Michael Piatek, Tadayoshi Kohno, and Arvind Krishnamurthy. 2008. Challenges and directions for monitoring P2P file sharing networks—or why my printer received a DMCA takedown notice. In *Proceeding of the 3rd USENIX Workshop on Hot Topics in Security (HotSec'08)*. Retrieved from http://www.usenix.org/events/hotsec08/tech/full_papers/piatek/piatek.pdf.

Andreas Pitsillidis, Chris Kanich, Geoffrey M. Voelker, Kirill Levchenko, and Stefan Savage. 2012. Taster's choice: A comparative analysis of spam feeds. In *Proceedings of the ACM SIGCOMM Conference on Internet Measurement*. 427–440.

Tim Rains. 2014. And the Gold Medal Goes to ...Finland! Retrieved Feburary 20, 2014 https://blogs.microsoft.com/cybertrust/2014/02/20/and-the-gold-medal-goes-to-finland/.

Anirudh Ramachandran, David Dagon, and Nick Feamster. 2006. Can DNS-based blacklists keep up with bots? In *CEAS*. Citeseer.

Raytheon. 2016. Forcepoint. Retrieved from https://www.forcepoint.com.

Marco Riccardi, David Oro, Jesus Luna, Marco Cremonini, and Marc Vilanova. 2010. A framework for financial botnet analysis. In *eCrime Researchers Summit (eCrime), 2010*. IEEE, 1–7.

Christian Rossow. 2014. Amplification hell: Revisiting network protocols for DDoS abuse. In *Proceedings of the 2014 Network and Distributed System Security (NDSS) Symposium*.

Christian Rossow, Dennis Andriesse, Tillmann Werner, Brett Stone-Gross, Daniel Plohmann, Christian J. Dietrich, and Herbert Bos. 2013. Sok: P2pwned-modeling and evaluating the resilience of peer-to-peer botnets. In *Proceedings of the 2013 IEEE Symposium onSecurity and Privacy* (SP). IEEE, 97–111.

SANS Institute. 2016. SANS Information Security Training. Retrieved from http://www.sans.org.

Shadowserver. 2016. Shadowserver Foundation. Retrieved from https://www.shadowserver.org.

Steve Sheng, Brad Wardman, Gary Warner, Lorrie Cranor, Jason Hong, and Chengshan Zhang. 2009. An empirical analysis of phishing blacklists. In *Sixth Conference on Email and Anti-Spam (CEAS)*.

Aditya K. Sood and Richard J. Enbody. 2013. Crimeware-as-a-service—a survey of commoditized crimeware in the underground market. *Int. J. Crit. Infrastruct. Protect.* 6, 1 (2013), 28–38.

Spamhaus. 2016. Spamhaus Datafeed. Retrieved from http://www.spamhaus.org/datafeed.

STC 2007. *Personal Internet Security*. Technical Report. Authority of the House of Lords. Retrieved from http://www.publications.parliament.uk/pa/ld200607/ldselect/ldsctech/165/165ii.pdf.

Brett Stone-Gross, Marco Cova, Christopher Kruegel, and Giovanni Vigna. 2011a. Peering through the iframe. In *Proceedings of the 2011 IEEE INFOCOM*. IEEE, 411–415.

Brett Stone-Gross, Thorsten Holz, Gianluca Stringhini, and Giovanni Vigna. 2011b. The underground economy of spam: A botmaster's perspective of coordinating large-scale spam campaigns. In *Proceedings of the 4th USENIX Conference on Large-scale Exploits and Emergent Threats (LEET'11)*. USENIX Association, Berkeley, CA, 4–4.

Brett Stone-Gross, Christopher Kruegel, Kevin Almeroth, Andreas Moser, and Engin Kirda. 2009. Fire: Finding rogue networks. In *Proceedings of the 2009 Computer Security Applications Conference (ACSAC'09)*. IEEE, 231–240.

Stop Escrow Fraud. 2016. Home - Escrow Fraud Prevention. Retrieved from http://www.escrow-fraud.com.

StopBadware. 2011. *The State of Badware*. Technical Report. StopBadware. Retrieved from https://www.stopbadware.org/files/state-of-badware-june-2011.pdf.

StopBadware. 2016. Data Sharing Program. Retreived from https://www.stopbadware.org/data-sharing.

Symantec. 2016a. Blue Coat—Network + Security + Cloud. Retrieved from https://www.bluecoat.com.

Symantec. 2016b. Norton Safe Web. Retrieved from https://safeweb.norton.com.

Qian Tang, L. Linden, J. S. Quarterman, and A. B. Whinston. 2013. Improving internet security through social information and social comparison: A field quasi-experiment. In *WEIS 2013*.

The Shadowserver Foundation. 2014. GameoverZeus & Cryptolocker. Retrieved July 8, 2014 from http://blog.shadowserver.org/2014/06/08/gameover-zeus-cryptolocker/.

Michel van Eeten, Hadi Asghari, Johannes M. Bauer, and Shirin Tabatabaie. 2011. Internet Service Providers and Botnet Mitigation: A fact-finding study on the Dutch market. (2011). Report prepared for the Netherlands Ministry of Economic Affairs, Agriculture and Innovation. Retrieved from http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2011/01/13/internet-service-providers-and-botnet-mitigation.html.

Michel van Eeten, Johannes M. Bauer, Hadi Asghari, Shirin Tabatabaie, and Dave Rand. 2010. *The Role of Internet Service Providers in Botnet Mitigation: An Empirical Analysis Based on Spam Data*. Technical Report. OECD Publishing.

Michel J. G. van Eeten and Johannes M. Bauer. 2008. *Economics of Malware: Security Decisions, Incentives and Externalities*. Technical Report. OECD Publishing.

Chris van't Hof. 2014. How the Dutch police and public prosecutor form smart coalitions against "bad hosting". *TekTok* Retrieved from http://www.tektok.nl/index.php/2014-05-23-11-21-59/147-8-5-how-the-dutch-police-and-public-prosecutor-form-smart-coalitions-against-bad-hosting.

Marie Vasek and Tyler Moore. 2012. Do malware reports expedite cleanup? An experimental study. In *Proceedings of the 5th USENIX Conference on Cyber Security Experimentation and Test (CSET'12)*.

Paul Vixie. 2014. Testimony of Paul Vixie before the subcommitee on crime and terorism united states senate committee on the judiciary - hearing on taking down botnets: Public and private eforts to disrupt and dismantle cybercriminal networks. Retrieved from http://www.judiciary.senate.gov/imo/media/doc/07-15-14VixieTestimony.pdf.

Paul Vixie. 2015. Targeted takedowns: Minimizing collateral damage using passive DNS. In *Black Hat USA 2015*. Retrieved from https://www.blackhat.com/docs/us-15/materials/us-15-Vixie-Targeted-Takedowns-Minimizing-Collateral-Damage-Using-Passive-DNS.pdf.

David Y. Wang, Stefan Savage, and Geoffrey M. Voelker. 2011. Cloak and dagger: Dynamics of web search cloaking. In *Proceedings of the 18th ACM Conference on Computer and Communications Security*. ACM, 477–490.

Steve Webb, James Caverlee, and Calton Pu. 2008. Predicting web spam with HTTP session information. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*. ACM, 339–348.

James Wyke. 2012. The zeroaccess botnet: Mining and fraud for massive financial gain. *Sophos Technical Paper* (2012).

Sandeep Yadav, Ashwath Kumar Krishna Reddy, A. L. Reddy, and Supranamaya Ranjan. 2010. Detecting algorithmically generated malicious domain names. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*. ACM, 48–61.

ZeusTracker. 2016. ZeuS Tracker at abuse.ch. Retrieved from https://zeustracker.abuse.ch.

Received July 2015; revised September 2016; accepted September 2016