Ethical Dilemmas in Take-down Research

Tyler Moore[1] and Richard Clayton[2]

[1] Center for Research on Computation and Society, Harvard University, USA
tmoore@seas.harvard.edu
[2] Computer Laboratory, University of Cambridge, UK
richard.clayton@cl.cam.ac.uk

Abstract. We discuss nine ethical dilemmas which have arisen during the investigation of 'notice and take-down' regimes for Internet content. Issues arise when balancing the desire for accurate measurement to advance the security community's understanding with the need to immediately reduce harm that is uncovered in the course of measurement. Research methods demand explanation to be accepted in peer-reviewed publications, yet the dissemination of knowledge may help miscreants improve their operations and avoid detection in the future. Finally, when researchers put forward solutions to problems they have identified, it is important that they ensure that their interventions demonstrably improve the situation and do not cause undue collateral damage.

1 Introduction

This paper is a case study of the ethical dilemmas we have faced in our computer security research. We do not set out any over-arching ethical theories of behavior. Instead, we discuss our personal experiences and those of other researchers, reporting with the benefit of hindsight when we did the right thing and perhaps also when we did not.

Over the past few years we have researched and published a number of papers about 'phishing', where criminals entice people into visiting websites that impersonate the real thing and dupe them into revealing passwords and other credentials, which will later be used for fraud. The main countermeasure to phishing is the removal, or 'take-down', of the fake websites. In some papers, notably [15], we consider take-down of other types of Internet content.

Our research approach might very loosely be described as econometrics. We obtain large numbers of measurements of real-world activity, particularly of website take-down times. From this data, and particularly from variations in this data, we tease out an understanding of the underlying criminality. This approach has been extremely valuable, in that it has allowed us to explain the relative success of some criminal gangs, and to reveal the harm caused by a lack of information sharing between companies offering take-down services.

Our research also lies in the general field of 'security economics', the relatively recent understanding that computer and networking security problems are better explained by economic considerations than by considering the more technical

'computer science' aspects of the situation [2]. Our research can also be seen to be criminological, in that we are looking at crime scenes, gauging the rates of victimization and assessing the effectiveness of crime prevention measures.

Because what we're looking at is criminal activity in the real world, with real victims suffering real monetary losses, we have found ourselves facing a number of ethical dilemmas, and it is those dilemmas upon which we focus in this paper. In Subsection 1.1 we describe our take-down research in more detail to better explain what we've been doing over the past few years that has triggered ethical headaches. In Section 2 we set out four ethical dilemmas that arise in the context of measuring criminal activity. In Section 3 we discuss two further ethical dilemmas that arise when explaining how take-down works, and then in Section 4 we explain three ethical dilemmas to be confronted when writing about 'fixes' for the original problem. Thereafter Section 5 discusses related work and finally in Section 6 we draw some conclusions.

1.1 An Overview of our Phishing Research Papers

In order to avoid continually having to break off from our later discussion of ethical dilemmas to explain details of our research, we will now present a quick overview of relevant aspects of our phishing research papers. Our account here is superficial and incomplete, and – for those concerned more about phishing than ethics – no substitute for consulting the original work.

For our first paper in 2007 [14] we measured phishing website lifetimes. We found that some websites, operated by the 'rock-phish' gang, were using a technically innovative scheme whereby the website hostname resolved to a different set of intermediary machines every few minutes. These intermediaries relayed the website traffic to a hidden 'mothership'.

In order to calculate the harm done by phishing, we fetched 'world-readable' log-summary files (created by The Webalizer[1]) from a subset of compromised machines. This gave us data from which to estimate the number of visitors that a phishing website received, a figure that was previously unknown. We were surprised to find that some visitors turned up weeks after it was first reported.

We also found that a small number of websites were storing the credentials they had stolen in files on the websites themselves. We inferred the locations of these credentials and found that about half appeared genuine and the rest took the form of messages, mainly abusive, directed at the criminals. Determining the proportion of visitors who were actually fooled into divulging real credentials allowed us to estimate the total harm that phishing was causing.

In 2008 we wrote [15], in which we considered a range of different types of content for which a 'take-down' regime exists: defamation, copyright violations, child sexual abuse images, phishing, and various types of fraudulent website. Our conclusion was that lifetimes were determined more by the incentives of those trying to remove the content, than by the nature of the content or the technical arrangements used to host it.

[1] `http://mrunix.net/webalizer/`

Also in 2008, we revisited our website lifetime statistics using 'feeds' of phishing website URLs from competing take-down companies. We were able to show in [16] that, because these feeds were not being shared with competitors, websites which were not universally known about were not removed.

In 2009 we took a further look at Webalizer data in [17]. In particular we considered what was revealed about the search terms that had been used to locate the phishing website. We were able to demonstrate that search was an important way for criminals to locate websites that they were able to compromise, and that being findable in this way was a contributory factor to recompromise rates.

2 Measuring Take-down

A natural tension exists between conducting accurate, reproducible research and reducing the harm caused by the content that is being removed. We explore some of the issues which arise in the following four dilemmas.

Dilemma 1: Should researchers notify affected parties in order to expedite take-down? In our research investigating phishing website take-down [14], we observed that many fake websites remained online for several weeks. Furthermore, our measurements suggested that consumers continued to visit such long-lived websites, indicating that their continued presences caused direct harm. Consequently, we discussed whether we should bring these websites to the attention of the banks being impersonated.

On one hand, it seems like a no-brainer: passing the information along might reduce the harm caused by phishing. However, there are several compelling reasons why we might prefer not to share this information. First, doing so could taint our measurements. One aim of our research was to independently measure the lifetime of phishing websites. Security companies had reported short take-down times, yet we found high variation that fitted a lognormal distribution. If we had notified banks and firms immediately, we could never have accurately measured the slow take-down speed.

Second, had we chosen to notify others, who should we tell? In some cases the relevant bank contact details could be easily inferred, but in other cases not. Significant additional effort would have been required to identify the appropriate points of contact for several hundred banks. Many banks hire specialist take-down companies to take down phishing websites on their behalf. Without knowing the arrangements a bank had put in place, we could not readily determine who we should be sharing our findings with.

Third, even if we had wanted to share, our arrangements with data sources precluded this. We negotiated real-time 'feeds' of phishing reports from two large take-down providers, but signed non-disclosure agreements to secure access. Notifying banks about long-lived phishing websites would have violated these agreements and could have caused financial harm to the feed providers.

In the end, we decided to mainly keep the reports to ourselves. After conducting the initial research and publishing our first paper, we began sharing reports

with individual banks and take-down companies on a one-off basis as requested, but we chose not to attempt to notify all banks. We later discovered that the primary explanation for long-lived phishing websites is that take-down companies do not exchange their lists with each other. We published a paper highlighting the adverse effect of firms' refusal to cooperate and called for greater data sharing [16], but we decided that it was entirely inappropriate for us to even consider being the long-term conduit for exchanging relevant information.

For a long time, researchers conducting clinical trials have balanced individual ethics – the needs of the next eligible patient – and collective ethics – the obligation to develop correct policies for the future [24]. Clinical trials can and should be stopped prematurely once the results become statistically significant and the divergence in treatment outcome is substantial. While security researchers are rarely afforded the luxury of controlling a randomized trial (for more on that see Dilemma 3), we could still learn from the procedures adopted by clinical researchers. In particular, we recommend that researchers avoid direct interference during data collection, but once the conclusions have been drawn, assistance to relevant stakeholders should be encouraged.

Dilemma 2: Should researchers intervene to assist victims? Security researchers often stumble upon information that identifies victims. For example, we gathered 414 user responses with personal information published on phishing websites [14]. We used the information to answer a research question (what proportion of user responses to phishing emails are legitimate?). Having gathered the information, were we obligated to notify the victims of their risk? We notified banks where we had existing contacts and where fast intervention was possible, but we did not notify all banks.

Unfortunately, ours is a common dilemma. Researchers investigating the Torpig botnet observed 180 000 infections and gathered over 70 GB of data collected by the bots, including over 1 million Windows password logins, 100 000 SMTP account logins and 12 000 FTP credentials [29]. In a subsequent presentation one co-author lamented the headaches introduced by collecting such data [11]. The co-author's conclusion is that researchers should go out of their way to avoid collecting such data because of the resulting obligation to notify victims.

If victims should be notified, what form should the notification take? One option might be to attempt to notify victims via pop-up messages on victim computers. In March 2009 the BBC's Click program purchased access to a botnet and demonstrated the evil that could be done with it. They then changed the 'wallpaper' on the individual members of the botnet to warn the owners that their machines were compromised. They have vigorously defended their actions as being in the public interest [22] – but they were heavily criticized for paying money to criminals and for the likelihood that in accessing machines without permission and altering content they had committed offenses under the UK's Computer Misuse Act 1990 [23].

We must beware the unintended consequences of intervention. Rod Rasmussen, CTO of InternetIdentity, a take-down company, relates [26] that they

had problems getting a phishing site disabled on a machine in a small West Virginia county:

> The normal admin for the machine had been deployed to Iraq as part of his National Guard unit, and his backup was busy and hundreds of miles away that weekend because of his father's funeral. There were plenty of people looking at the machine (as in had their physical eyeballs on it) including the local sheriff, but no one was touching it since it ran the 911 Dispatch system and no one had the knowledge (as in passwords and expertise) to fix it.
>
> We've also had take-downs on machines that were in hospitals, railroad stations, airports, and government facilities. While those could be just public access terminals, there's no way we can tell from the outside if that is the case or they are running life-saving equipment, switching operations, air-traffic control systems, or have sensitive data on them respectively. That's why we have a very bright line barring any sort of 'write access', resetting or otherwise monkeying with content on compromised servers. Not only is it usually illegal in the US, someone's life can literally be on the line!

Dilemma 3: Should researchers fabricate plausible content in order to conduct 'pure' experiments of take-down? With few exceptions, empirical research in information security does not employ a similar design to that of randomized experiments. Instead, researchers must rely on observational data, as is commonly done in the social sciences. Incidents regularly occur, which prompts defenders to respond. Researchers observe this process and collect data describing it. Observational studies present many difficulties, notably the potential for sample bias, correlation/causation issues, and the presence of multiple confounding dependent variables.

Some researchers looking at the response to copyright violations have attempted to create 'pure' experiments by introducing content and observing the take-down response. The US and EU have different regimes for dealing with claims of copyright violation online, and some have argued that the EU approach encourages greater ISP compliance with dubious take-down requests. At least two groups have performed experiments to test the claim.

In 2003 an Oxford research group posted material onto UK and US websites [1]. The material was an extract of John Stuart Mill's 1869 'On Liberty', discussing freedom of speech. The experimenters then wrote anonymously to the two hosting ISPs, falsely claiming that the material was still in copyright. The UK ISP removed the material, whereas the US ISP insisted upon the provision of the legally necessary "on pain of perjury" declaration on the DMCA notice. In 2004 a similar experiment was performed by the Netherlands-based 'Multatuli Project' [20]. They placed some out-of-copyright material from a famous 1871 tract onto webspace provided by ten different Dutch ISPs. Their results were mixed, with some ISPs losing their first complaint and only acting on a follow-

up message. By the end of the experiment, seven of the ten ISPs had removed the material, taking between 3 hours and 3 days to do so.

While both copyright experiments had substantial design issues that undermined the results obtained, the underlying ethical question raised is whether such experimentation is appropriate for researchers to pursue. The main argument in favor is that such methods may be used to improve scientific knowledge. Yet this begs the question whether a randomized experimental design is really necessary to advance knowledge. Might we instead make do with using observational studies instead? Observational studies have the advantage of studying the response to real incidents. Fabricating plausible incidents may be hard, and there is a history of difficulty in other areas of information security that have tried to rely on 'synthetic' data, notably in intrusion detection [12].

Furthermore, individual ethics must also be considered. Wasting the time and energy of frontline responders on fabricated requests suggests real harm is caused by the experiments. In particular, the responders typically have substantial resource constraints and already find it difficult to keep up with the number of legitimate take-down requests. In sum, we believe the fabrication of reports to study take-down is usually unethical.

Dilemma 4: Should researchers collect world-readable data from 'private' locations? Our final 'measurement' ethical dilemma concerns the type of data that is suitable for collection. Researchers must often be creative to identify suitable data sources.

Many websites make use of The Webalizer, a program for summarizing web server log files. It creates reports of how many visitors looked at the website, what times of day they came, the most popular pages on the website, and so forth. It is not uncommon to leave these reports 'world-readable' in a standard location on the server, which means that anyone can inspect their contents. We have repeatedly collected Webalizer reports from websites that have been compromised and loaded with phishing pages.

But is it ethical to collect such 'world-readable' data in order to conduct research? In practice, most website operators who make their Webalizer logs public did not take an explicit decision to do so, and we expect that many would choose to make them private once they became aware they were in fact publicly available. We decided to collect the data for two main reasons. First, it enabled us to answer research questions that otherwise would not have been possible. Second, the data available through logs does not include personally-identifiable information, which lessens the scope for harm by collecting the data.

On balance, we feel the opportunities for scientific advancement outweigh the risks to an individual website operator in collecting the data. However, it is a judgment call, and one that should be weighed on a case-by-case basis.

3 Explaining Take-down

This section considers the two main dilemmas that we have encountered when analyzing criminal activity.

The first dilemma is that our analysis may be superior to that of the criminals, and our insights will assist them in becoming more efficient. This concern about having a superior analysis is subtly different from the ethical question of whether criminality should be studied at all, which is usually seen as being a case of catching up with the bad guys. The usual answer put forward in defense of a 'full disclosure' policy is that the criminals already know how to be bad, and it is beneficial for the good guys to have a fuller understanding. As Hobbs put it back in 1853 in the context of studying the insecurity of locks [9]:

> Rogues are very keen in their profession, and know already much more than we can teach them respecting their several kinds of roguery.

Or as Bishop John Wilkins put it two hundred years earlier [34]:

> If all those useful Inventions that are liable to abuse, should therefore be concealed, there is not any Art or Science which might be lawfully profest.

The second dilemma is that we may be explaining weaknesses in the criminals' systems that can be used by investigators to get an 'edge'. Once those weaknesses are explained, the criminals may be able to fix them. As researchers, we may be completely unaware of what use the weaknesses are being put to, and so any public discussion will carry the risk that we are making the situation worse, rather than better.

Dilemma 5: What if our analysis will assist criminals? In our first paper about phishing [14] we analyzed the relative take-down performance of traditional phishing websites and botnet-hosted fast-flux systems. We observed that although the lifetime of individual servers was less on the fast-flux systems it was still substantial. The criminals clearly expected to have much lower lifetimes because they used five or more servers in parallel to compensate for failures. We didn't spell out that our figures showed this was unnecessarily cautious, and the criminals have continued to use multiple servers in parallel, which has considerably simplified detecting the use of fast-flux systems.

In that first paper we also observed that since the server lifetimes were so high, setting appropriate time-to-live values would be likely to keep the criminal sites available in the DNS caches of large ISPs, even when domain names were taken down. Once again we didn't especially stress this point, and we are not aware of criminals taking this approach, but once again our measurements and analysis had shown the use of sub-optimal trade-offs in criminal system design, and our dilemma had been the extent to which we should improve the intellectual value of our paper versus the help we might give the criminals. In the event, we chose to sacrifice some clarity in our exposition, and the criminals missed our point, and have yet to do anything imaginative with time-to-live values.

Dilemma 6: Should investigatory techniques be revealed? It is not appropriate to write computer security papers which keep some parts of the methodology

secret. Experiments should be reproducible by others to confirm the accuracy of results (never mind that in computer security it is almost impossible to find a venue that will publish papers that merely confirm what others have found, triggering doubt as to how often such reproduction is attempted).

The main effect has been that we have failed to conduct some research or publish some results because we would need to reveal how we knew what we did. For example, the existence of 'back doors' in phishing kits was widely known about before Netcraft decided to write about it on their blog [19]. The freely available phishing kits sent details of compromised victims not only to the criminal who deployed them, but also, in various obfuscated ways, to webmail accounts operated by the 'Mr Brain' gang. We speculate that tipping off criminals to the existence of such back-doors will have impeded law-enforcement investigations by eliminating central repositories of victim details.

As a further example, the location of the rock-phish gang's 'motherships' could only be determined by inspecting traffic that traversed one of the relay machines. Consequently, as part of our research we spent some effort in providing live feeds of their location to police forces so they could visit active machines and monitor the traffic. We then discovered a technique for remotely identifying the mothership location.[2] Even now this technique may be of use to future investigators, so we still have chosen not to reveal it here.

Many researchers do not see these types of issues as a dilemma, falling back on 'full disclosure' arguments. Recently Billy Rios, a "security engineer for a major software firm" took a look at a kit for Zeus, a major component of criminal attacks on banking customers that are netting many millions of dollars [30] (and pounds [8] and rubles [32]) a month. His blog post [27] explained how a file injection vulnerability in the code could be leveraged to inspect the internals of this crimeware. There are said to be several hundred Zeus-related botnets and many are under active investigation by law enforcement (only a few days later a major series of arrests were made). Disclosing this vulnerability could well have jeopardized some of these investigations. Rios, however, did not seem to mind:

> There are some fascinating things to consider when finding bugs in software that is used primarily by criminals, but I won't bore you with that now. Instead I'd like to share with you some of the more interesting parts of my research.

Rios argues that the information he has disclosed will assist in defending against the Zeus threat, although in what he has published thus far he hasn't explained what this assistance might be.

4 Fixing Take-down

During the course of our research, we have gained a better understanding of how to make take-down a more effective tool for improving information security.

[2] For some time, one of the motherships was located at a hosting provider in suburban Seattle, a most convenient place for investigators to visit.

This has led to several dilemmas over the choices we face in 'fixing' take-down. We discuss three of them here. First, those who identify content that should be removed must decide whether to publish lists of this content, since this will help researchers and defenders but can also aid criminals. Second, when recommending a change in policy, we must balance between what we believe to be right and what we know to be achievable. Third, while take-down may improve security, it can sometimes conflict with the principles of a free and open society. Consequently, the benefits of any mechanism that expedites take-down in the short term must be weighed against the potential for its abuse by others.

Dilemma 7: When should datasets be made public or kept secret? Given the many types of online material targeted for take-down, a natural question arises: should a public record be kept, or would society's interests be better served by keeping the information secret? In a few contexts, public lists are kept. For example, PhishTank (`phishtank.com`) reports known phishing websites, Chilling Effects (`chillingeffects.org`) documents DMCA take-down requests, and Artists Against 419 (`aa419.org`) publishes records of websites that support advanced-fee fraud.

Publishing reports offers several advantages. It increases transparency, important given that so much take-down is coordinated by the private action of volunteers. It enables research to be reproducible, while creating the potential for richer investigation by people who otherwise would not have access to the information. Finally, it can help defenders expedite the take-down process. For instance, we found that phishing websites reported to PhishTank were less likely to be recompromised than websites which only appeared in secret lists [17].

Of course, there are downsides to publishing take-down lists. The list could help the attacker by revealing what the defender knows, as well as providing a source of future targets. It could even help criminals copy caches of credentials from each other, much as we ourselves did and discussed in Dilemma 2. Sometimes knowledge of the content being taken down is problematic: publishing locations of child sexual abuse images would certainly be harmful.

Finally, publishing a record of offending content will 'name and shame' responsible parties and victims, which they may not appreciate. In [17], we explained how `chat2me247.com` had been repeatedly compromised through targeted web searches and loaded with phishing pages. In October 2010, the webmaster of `chat2me247.com` wrote us to complain that we had publicly discussed the website's security problems. We chose not to ask for the website's permission before writing about it because the information we gathered was public: some of the phishing pages on `chat2me247.com` had appeared on PhishTank, and the Webalizer logs we collected were also made publicly available.

Given these downsides, some have devised alternative arrangements to public disclosure. One popular compromise is to publish a cryptographic hash of the records, so that anyone can still verify whether a record is present in the list without making the entire list public. For example, the Google Safe Browsing API[3]

[3] `http://code.google.com/apis/safebrowsing/`

only allows users to verify whether suspected URLs are malicious. This strikes a balance that protects users without letting the world know which websites have been infected. There are significant downsides to this approach, however, which should be considered. Secondary research is severely limited. Outside researchers cannot answer even basic questions, such as whether the number of take-down requests grows or shrinks over time. Defenders can also be hampered by the hashing arrangement. For example, a hosting provider who manages many websites cannot easily determine whether they are present in the blacklist, and so cannot proactively respond.

Some researchers have opted for a completely private exchange of information. For instance, groups such as Team Cymru (`team-cymru.org`) and Shadowserver (`shadowserver.org`) directly pass along lists of machines suspected of participation in botnets to the relevant ISPs. These informal arrangements have low overhead and ensure a timely response and shield the ISP from any fear of public humiliation. The downsides to this arrangement are similar to those of hash-based arrangements, with the added cost of requiring explicit cooperation between all partners for positive gains to be realized.

Why is selecting a method of publishing an ethical issue? We argue that those involved in take-down should consider how to protect individuals from harm while creating an opportunity for research to advance the understanding of how to better perform take-down. Opting to keep information private can be even more dangerous than the reckless publication of information that aids attackers. The harm may be more difficult to directly observe (slowed take-down speed, lack of pressure to improve practices, etc.) but equally destructive.

Dilemma 8: Is the fix realistic, and does it consider the incentives of all the participants? As we explained earlier, we believe that security stems in a large part from getting the economics of the situation correct, and that often means aligning incentives and making an entity that is able to fix the security issue responsible for doing so – even if the original problem was not of their making.

This analysis has led us to make controversial recommendations for information sharing in the anti-phishing community. Our paper conclusively demonstrated that damage was being done by the compartmentalizing of data about where the websites were located – our 'sound-bite' is that "bank phishing websites are taken down in four hours when the banks know about them, and four days if they don't". We suggested that the take-down companies should be forced to share information by their customers (the banks) renegotiating their contracts. This suggestion did not impress the take-down companies, and Eric Olsen wrote a comprehensive rebuttal [21], suggesting that we would destroy the incentives to improve the quality of feeds and permit 'free riding'. We responded to this [13] with a suggestion as to how feeds could be shared for payment in such a way as to keep the incentives in place, and the following year one of us co-authored a paper on a data sharing protocol, with strong information-hiding guarantees, that would support the sharing-for-payment concept [18].

In this case: at best our initial paper exemplified classic academic naïveté in proposing a system that would not work in the real world, at worst we were

wrong to put the idea forward without further explanation, and through the unnecessary controversy we reduced the impact of some important measurements whose implications needed to be understood by the whole take-down industry.

Measurements we made (in [15]) of data from the Internet Watch Foundation mean that we can add to our sound-bite, "and child sexual abuse image websites are removed in four weeks". This quite scandalous statistic arises, we believe, because the 'hot-lines' who share information about this type of content fail to tell the hosting companies that they are providing services to criminals. The hot-lines do tell the police, but their lines of communication can be slow, and policing incentives are more to do with catching the criminals than in getting the website taken down.

Once again our recommendations for improvement have been seen as unrealistic – in particular we understand that INHOPE,[4] the international association of hot-lines, forbids members from contacting hosting providers located in a country where another INHOPE member operates. This is clearly obstructive and so we continue to believe that our recommendations about sharing are the ethical ones, even though they do not currently appear to be practical without the disbandment or restructuring of INHOPE.

Dilemma 9: What if the fix is worse than the problem? We believe that when we put forward solutions to problems we have an ethical responsibility to ensure that we will not be making things worse.

We could, for example, make a clear case for the benefits of restricting the registering of misleading domain names, and of being able to precisely identify who was making the registrations, but we believe that such measures would be incompatible with a free and open society. For every fake `barklays.com` that was blocked[5] there would also be the restriction to free speech of blocking, say, `barclayssucks.com`.[6] The Peoples' Republic of China has chosen to make it compulsory to provide photographic identification of `.cn` domain registrants [3] with the stated intention of tackling pornography, but many commentators have suggested that the real intent is to suppress dissident use of the Internet.

In this context we note with some alarm the recent RPZ proposal by Paul Vixie which codifies a method for suppressing DNS results [31]. Vixie envisages that the new system will be used to disrupt malware rendezvous and command-and-control mechanisms. However we believe that politicians worldwide could immediately understand the message to be "there is now a standard and easy-to-deploy method of insisting that particular domains must be censored". They might bring forward a whole raft of censorship initiatives, to block access to child sexual abuse sites, offshore gambling, adult pornography, political dissidents, and even, in Turkey, `richarddawkins.com`.

[4] `http://www.inhope.org`

[5] In fact `barklays.com` is owned by a small business in Oshawa, Canada and currently redirects to the website for a Canadian fishing (`f` not `ph`) TV show. It has nothing to do with the global banking group Barclays plc

[6] `barclayssucks.com` currently appears to be registered by Barclays themselves, but at present it is 'parked' with no Barclays-related content.

This type of blocking, 'DNS poisoning', has been implemented in ad hoc mechanisms for years despite clear evidence of unintended collateral damage [6] and relatively simple evasion [4]. The difference Vixie has made is to make the mechanism standard, to provide the functionality in the normal codebase, and to provide authorities with the ability to identify conformance by the inspection of configuration files. The RPZ will probably have an impact on botnet design, and will be outflanked as other rendezvous systems come to the fore. Time will tell if the short-term disruption to the criminals is outweighed by the harm that we predict will stem from this well-meaning, but ethically dubious, proposal.

5 Related Work

Information security researchers have encountered numerous ethical dilemmas. In one early work, Spafford argued that, under most circumstances, unauthorized access to a computer is unethical [28]. Indeed, in most countries such unauthorized access is also illegal. More recently, following the rise of botnets, researchers and practitioners have argued over whether defenders could or should intervene to remediate botnet-infected computers. Dittrich et al. [5] discuss this and several related open ethical questions regarding how best to fight botnets.

Perhaps most closely related to the dilemmas we discuss in this paper is the quandary facing researchers who infiltrated a portion of the Storm botnet in order to measure its activity [10]. To obtain a more accurate measurement of Storm activity, the authors took control of a portion of the botnet and allowed it to continue operation rather than shut down the machines under its control. The authors followed a self-declared ethical principle of "strictly reducing harm": no additional spam was sent out than otherwise would have been, and they blocked purchase of goods advertised by the spam. Nevertheless, consumer machines under their control sent slightly modified spam as directed by the botnet's controller. This led to unprecedented measurements of botnet activity, but at the cost of permitting some harm that they were in a position to reduce.

6 Conclusion

This case study has set out the nature of nine ethical dilemmas we have faced over the past few years as we conducted research into the take-down of criminal websites. Our aim has not been to set out ethical principles that could guide others, or to promote our ethical choices as if we were paragons to be emulated, but rather to provide a rich set of 'war stories' which illuminate the issues we have faced and document the choices we have made. Undoubtedly, others might have done things differently, and with the benefit of hindsight we might have been more cautious in our handling of credentials and more circumspect in our recommendations. We hope that, at the least, our experiences will make others pause before rushing into research activity, and at best we have offered deeper thinkers than ourselves access to practical material for testing their ethical principles.

References

1. Ahlert, C., Marsden, C., and Yung, C.: How 'Liberty' disappeared from cyberspace: the mystery shopper tests Internet content self-regulation, 2004. http://pcmlp.socleg.ox.ac.uk/text/liberty.pdf
2. Anderson, R. and Moore, T.: The economics of information security. *Science*, 314(5799):610–613, 2006.
3. Chao, L.: China Porn Measures Raise Fear Of Censors. Wall Street Journal, page A10, 17 Dec 2009. http://online.wsj.com/article/SB126098577403994051.html
4. Clayton, R.: Anonymity and Traceability in Cyberspace. Technical Report UCAM-CL-TR-653, University of Cambridge Computer Laboratory, 2005.
5. Dittrich, D., Leder, F. and Werner, T.: Ethical decision making regarding remote mitigation of botnets. In: Sion, R. et al. (eds.) Workshop on the Ethics of Computer Security Research (WECSR), Lecture Notes in Computer Science (LNCS), vol. 6054, pp. 216–230 (2010)
6. Dornseif, M.: Government mandated blocking of foreign Web content. In: von Knop, J., Haverkamp, W., Jessen, E. (eds.): Security, E-Learning, E-Services: Proceedings of the 17. DFN-Arbeitstagung über Kommunikationsnetze, Düsseldorf, Lecture Notes in Informatics, pp. 617–648, 2003.
7. Franklin, J., Paxson, V., Perrig, A. and Savage, S.: An inquiry into the nature and causes of the wealth of Internet miscreants. In *Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS)*, pp. 375–388, 2007.
8. Gill, C.: Hi-tech crime police quiz 19 people over internet bank scam that netted hackers up to £20m from British accounts. Daily Mail, 29 Sep 2010. http://www.dailymail.co.uk/news/article-1316022/Nineteen-arrested-online-bank-raid-netted-20m.html
9. Hobbs, A.C. (ed: Tomlinson, C.): Locks and Safes: The Construction of Locks. Virtue and Co., London. 1853.
10. Kanich, C., Kreibich, C., Levchenko, K., Enright, B., Voelker, G. M., Paxson, V. and Savage, S.: Spamalytics: an empirical analysis of spam marketing conversion. In: Proceedings of the 15th ACM Conference on Computer and Communications Security (CCS), pp. 3–14, 2008.
11. Kemmerer, R.: How to steal a botnet and what can happen when you do. Google Tech Talk, 2009. http://www.youtube.com/watch?v=2GdqoQJa6r4
12. McHugh, J.: Testing Intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory. ACM Transactions on Information and System Security 3(4), 262–294, 2000.
13. Moore, T.: How can we co-operate to tackle phishing? Light Blue Touchpaper, 27 Oct 2008. http://www.lightbluetouchpaper.org/2008/10/27/how-can-we-co-operate-to-tackle-phishing/
14. Moore, T., Clayton, R.: Examining the impact of website take-down on phishing. In *Anti-Phishing Working Group eCrime Researcher's Summit (APWG eCrime)*, pp. 1–13, ACM Press, New York, 2007.
15. Moore, T., Clayton, R.: The Impact of Incentives on Notice and Take-down. In: M. Eric Johnson, editor: Managing Information Risk and the Economics of Security, pages 199–223, Springer, New York, 2008.
16. Moore, T., Clayton, R.: The consequence of non-cooperation in the fight against phishing. In: Anti-Phishing Working Group eCrime Researchers Summit (APWG eCrime), pp. 1–14, 2008.

17. Moore, T., Clayton, R.: Evil searching: compromise and recompromise of Internet hosts for phishing. In: Dingledine, R., Golle, P. (eds.) Financial Cryptography and Data Security, Lecture Notes in Computer Science (LNCS), vol. 5628, pp. 256–272, 2009.
18. Moran, T. and Moore, T.: The Phish Market Protocol: Securely Sharing Attack Data Between Competitors. 14th International Conference on Financial Cryptography and Data Security. 25–28 Jan 2010, Tenerife, Spain.
19. Mutton, P.: Mr-Brain: Stealing Phish from Fraudsters. Netcraft Blog, 22 Jan 2008. http://news.netcraft.com/archives/2008/01/22/mrbrain_stealing_phish_from_fraudsters.html
20. Nas, S.: The Multatuli project: ISP notice & take down. In: SANE, 2004. http://www.bof.nl/docs/researchpaperSANE.pdf
21. Olsen, E.: A Contrary Perspective – Forced Data Sharing Will Decrease Performance and Reduce Protection. Cyveillance Blog, 28 Oct 2008. http://www.cyveillanceblog.com/phishing/a-contrary-perspective-%E2%80%93-forced-data-sharing-will-decrease-performance-and-reduce-protection
22. Perrow, M.: Click's botnet experiment. BBC Editors blog, 13 Mar 2009. http://www.bbc.co.uk/blogs/theeditors/2009/03/click_botnet_experiment.html
23. Pinsent Masons: BBC programme broke law with botnets, says lawyer. Out-law news, 12 Mar 2009. http://www.out-law.com/page-9863
24. Pocock, S.J.: When to stop a clinical trial. British Medical Journal 305(6847) 235–240, 1992.
25. Provos, N., Mavrommatis, P., Rajab, M. and Monrose, F.: All your iFrames point to us. In *17th USENIX Security Symposium*, pp. 1–15, 2008.
26. Rasmussen, R.: Personal communication. 13 Aug 2010.
27. Rios, B.: Turning the Tables – Part I. 27 Sep 2010. http://xs-sniper.com/blog/2010/09/27/turning-the-tables/
28. Spafford, E.H.: Are computer hacker break-ins ethical? Journal of Systems and Software 17(1):41–48, 1992.
29. Stone-Gross, B., Cova, M., Cavallaro, L., Gilbert, B., Szydlowski, M., Kemmerer, R., Kruegel, C. and Vigna, G.: Your botnet is my botnet: analysis of a botnet takeover. In: Proceedings of 16th ACM CCS, pp. 635–647, 2009.
30. US Department of Justice: Manhattan U.S. Attorney Charges 37 Defendants Involved in Global Bank Fraud Schemes that Used 'Zeus Trojan' and Other Malware to Steal Millions of Dollars from U.S. Bank Accounts. Press Release, 30 Sep 2010. http://newyork.fbi.gov/dojpressrel/pressrel10/nyfo093010.htm
31. Vixie, P.: Taking Back the DNS. CircleID, 30 Jul 2010. http://www.circleid.com/posts/20100728_taking_back_the_dns/
32. Warner, G.: Is Russia joining the Zeus hunt? Cybercrime & Doing Time, 4 Oct 2010. http://garwarner.blogspot.com/2010/10/is-russia-joining-zeus-hunt.html
33. Weaver, R. and Collins, M.P.: Fishing for phishes: applying capture-recapture methods to estimate phishing populations. In *Anti-Phishing Working Group eCrime Researcher's Summit (APWG eCrime)*, pp. 14–25. ACM Press, New York, 2007.
34. Wilkins, J.: Mercury: Or the Secret and Swift Messenger. Maynard and Wilkins, London. 1641.