Workshop Session #107 - Background Paper

Internet blocking:

When well-intentioned measures go too far

Introduction

The economic and public policy impacts of Internet access blocking by state actors have been well studied[1],[2],[3],[4]. Receiving less study to date are the economic and public policy impacts of Internet policing by third party non-state actors. The lack of a universal definition of "due process" or a common policy framework[5] has led to occasional collateral damage that undermines the security and stability of the Internet. This is a form of "digital culture clash" between those who value unobstructed access over all else vs. those who wish to control access (by others) to specific information.

This text will explore the state of play in third-party Internet-access blocking such as Internet reputation systems, whether motivated by commerce, a sense of duty, or legal requirements. Examples of collateral damage will be drawn from the public record, including the impact of Spamhaus' block of significant Swedish IP space in early 2014, Microsoft's court ordered takedown of No-ip.com, and the US Congress' COICA/SOPA work in 2010/2011. Although blocking can be essential to maintaining the usefulness of the Internet, such as filtering the abusive email that has accounted for over 90% of email connections at times,[6] blocking can also go too far. The operative question is: "at what point does organized Internet access blocking do more harm than good to either the public interest or the organizer's own agenda, due to foreseeable collateral damage, lack of care, or lack of investigatory or research resources to validate the action?"

The moderators of IGF 2014 WS107 hope to reach a broad understanding and brief set of recommendations for those who might block or restrict Internet reachability between otherwise consenting parties, for those who might participate in such events by subscribing to an Internet reputation system, for those who might be targeted by such intentional blockages, and also for policy makers and shapers who need to know the limits and the risks of collective third party action in Cyberspace.

---

[1] *Information controls during Thailand's 2014 Coup*, https://citizenlab.org/2014/07/information-controls-thailand-2014-coup/

[2] *Monitoring Information Controls in Iraq in Reaction to ISIS Insurgency*, https://citizenlab.org/2014/06/monitoring-information-controls-in-iraq/

[3] *Security and Other Technical Concerns Raised by the DNS Filtering Requirements in the PROTECT IP Bill*, http://www.redbarn.org/files_redbarn/PROTECT-IP-Technical-Whitepaper-Final.pdf

[4] *SAC056 – SAC Advisory on Impacts of Content Blocking via the Domain Name System*, https://www.icann.org/en/system/files/files/sac-056-en.pdf

[5] *Internet and Jurisdiction*, http://www.internetjurisdiction.net/about/mission/

[6] *MAAWG. Email Metrics Program: Report #15*, http://www.maawg.org/sites/maawg/files/news/MAAWG_2011_Q1Q2Q3_Metrics_Report_15.pdf

History

The Internet system and the allocation systems for Internet resources such as IP addresses and DNS names are available for any use case; these systems are as easily accessible for abusive purposes as for constructive purposes. Because these systems have become far more efficient with each passing year, traffic filtering keyed to the abuse of Internet resources is now a necessary and universal commodity. Traffic filtering technology varies in complexity from simple user-configured software or devices, to fully adaptive and autonomous machine learning systems capable of independent detection and policy supervision. One extremely attractive point on the complexity/economics curve is to have a large number of simple filters which are remotely configured by some central apparatus that can, because of its great leverage, exercise high cost human investigation powers in order to isolate patterns among Internet resource identifiers which can be treated as "known bad". This construction is called by various names but we will use the term Network Reputation Service (NRS) herein.

The first NRS was built in 1996 by Paul Vixie, Eric Ziegast, and David Rand, and was called MAPS (Mail Abuse Prevention System, or "spam" spelled backward, depending on the reader's sense of humor). This well intentioned non-profit public resource was commercialized after a large volume of inbound civil lawsuits. However, this effort also inspired first dozens and then hundreds of other NRS services, many non-profit, and many for-profit. The growth of the NRS industry was due to the Internet's attractiveness for commerce and communication, since the Internet was equally attractive for abuse including crime, spam, malware, botnets, denial-of-service attacks, and any other predatory human action that found its way onto the Internet. NRS makes it possible for a small team of investigators to determine with high confidence that some set of Internet resources such as IP addresses or DNS names are so dedicated to abusive behavior that their attendant services have either no value to anyone except their operators, or have a societal cost much higher than any societal benefit. In general, the use of a NRS is compatible with maintaining consensual Internet use, because an NRS only affects the reachability of its consensual subscribers, who both understand and intend that exact result.

Another growing facet of Internet defensive countermeasures is "takedown". It is now widely known that DNS names and IP address blocks are necessary and valuable to Internet abusers and criminals. In fact, these resources are easy for criminals to acquire since other criminals are in the business of providing such resources. To resist this growing phenomenon, there is a segment of the Internet security industry who specializes in "takedown". This can be done via:

- Appeal to a service provider to revoke a service due to breach of terms of service between the abusive user and the service provider. When applied to a DNS registry or registrar, this effectively causes a DNS name to go out of existence.

- Causing the name to be reallocated to a "sinkhole,"[7] which supplies a controlled DNS response for malicious domains, in order to prevent access or to operate a victim notification service.
- Causing an IP route advertisement to be dropped or blocked, or causing a virtual or physical hosting customer to have their contract and service cancelled.

"Takedown" means mitigating abuse close to its source rather than filtering it at the thresholds of every victim, and as such, gives even better leverage than NRS. However, Internet criminals always adapt. Today, Internet abusers and criminals deploy lightweight services that they can afford to lose. The current battle being fought is to try see whether criminal infrastructure can be destroyed as soon as and as fast as it is created. This battle is likely a never-ending struggle as defenders and attackers continuously adapt, but continuing to fight it is a necessary aspect of keeping the Internet open and usable.

At the time of this writing, no mature Internet user or operator goes without digital defense, of which NRS and "takedown" are extremely attractive and therefore common forms. And this new equilibrium provides a foundation for an entirely new set of challenges for the next generation of Internet users and operators to cope with.

Problem Statement

The Internet is a richly interdependent system in which a failure in one service can have cascading effects on other services. The downstream dependencies of an Internet service might not be known or knowable, even to its operator. An IP address that distributes malware might also be a benign DNS server that serves other domain names, on which many web sites might depend. Blocking access to that IP address might block all those web sites, affecting all of their operators and users. A domain name might also be used as an non-criminal Internet "mail exchange" in which case a loss of that domain name will have an impact on all e-mail using that exchange even if the e-mail server itself remains completely reachable. The permutations here are endless. Some researchers in this area call this "collateral damage" or "unintended consequences".[8]

Because NRS and "takedown" are deliberate unilateral impositions of service loss, these necessary and legitimate forms of digital defense burden every active defender with the responsibility to investigate and avoid collateral damage. The bright center of the problem statement is: *this responsibility is not practical.* In many cases it is simply not possible to know what other users or other services depend upon some digital asset whose existence or reachability a defender plans to interrupt. This observation yields an unpleasant moral dilemma for which there is no obvious solution.

---

[7] *InfoSec Reading Room: DNS Sinkhole,* http://www.sans.org/reading-room/whitepapers/dns/dns-sinkhole-33523

[8] *SAC050 – DNS Blocking: Benefits Versus Harms,* https://www.icann.org/en/system/files/files/sac-050-en.pdf

If a defender can't coherently categorize a given digital asset as "solely malicious" then she has the unenviable choice between leaving the asset in place (thus enabling further abuse and crime) or risking an unknowable interruption of benign services (and thus harm to innocent third parties). Abusers and criminals know this, and so they will often place their assets in mixed-use digital neighborhoods, where the risk of NRS or "takedown" is shared with the defenders. Overreach and collateral damage are real risks for defenders, because attackers arrange their assets deliberately in order to heighten those risks.

So an active defender seeking to identify and qualify targets for NRS or "takedown" must either err on the side of intervention – in order to tip the scales against complacent service providers and their abusive criminal customers – or on the side of caution, in order to earn a "net good" and also to avoid de-legitimizing NRS and "takedown" as accepted tools for digital defense.

Actions by an NRS consist of a number of steps, as listed below. If any of these steps are omitted, the NRS will be considered to have too high a "false positive" incidence.

1. Identification of issues – Often by having automatic or manual reporting mechanisms.
2. Evaluation – A decision making process that gives as a result that action is to be taken.
3. Action – The action that results in the filtering being done by the subscribers of the NRS.
4. Reaction – Removal of the Internet resource from the NRS filter as soon as the original cause has been addressed *or sooner* if severe collateral damage is discovered.

In many cases where objections to NRS systems have happened, it is usually not so much steps 1-3 that are the problems, but step 4. The original Issue is remediated by a system operator, or a third party who is otherwise not involved is injured by the Action, either of which lead to re-Evaluation and might result in the filtering be removed. If this is too slow or too cumbersome a process, the filter will be in place much longer that what is necessary to deal with the malicious activity, and the ratio of unintended to intended impact has time to rise.

We explicitly note that "severity" of both intended and unintended consequences of blocking must be evaluated in a larger context than the networks and systems whose abuse triggered the Action, but also to the suppliers, customers, and partners of resources. This is captured in SAC050 [*ibid*] as four recommendations for good organizational choices:

1. The organization imposes a policy on a network and its users over which it exercises administrative control (i.e., it is the administrator of a policy domain).
2. The organization determines that the policy is beneficial to its objectives and/or the interests of its users.
3. The organization implements the policy using a technique that is least disruptive to its network operations and users, unless laws or regulations specify certain techniques.
4. The organization makes a concerted effort to do no harm to networks or users outside its policy domain as a consequence of implementing the policy.

A few things can help to make the NRS-based actions less failure-prone, including:

- A known set of evaluation criteria.

- Well-formed and highly available contact mechanisms (also for parties not having any relationship with the NRS).
- Information mechanisms where current information such as NRS policies and current NRS Actions can be found.

This would not only make it easier for parties to evaluate whether a specific NRS service is valuable, but also for the Internet community as a whole to have visibility into an NRS whose Actions might have widely felt effects.

Case Studies

The following three case studies provide insights into actual behavior and responses in the Internet ecosystem that can provide us with more detailed lessons learned derived from the core recommendations about suitable DNS blocking put forward in SAC 050.

Spamhaus

The Spamhaus project is "an international nonprofit organization whose mission is to track the Internet's spam operations and sources" and their most visible act is to disseminate lists of Internet identifiers considered to be a source of reliably abusive traffic.[9] Although Spamhaus is not a network operator itself, many service providers utilize Spamhaus's lists of abusive IP addresses directly in their decision to deliver or block data. Thus, being listed by Spamhaus as abusive has cascading effects that significantly degrade the Internet connectivity of the targeting IP addresses. When the listings are precise and accurate, this is exactly what is desired. However, the costs of any mistake is high and is felt widely.

A mistake is exactly what happened on February 27, 2014 when 221,184 IP addresses allocated by the Swedish organization Resilians were listed by Spamhaus and suffered an intentional and massive loss of Internet connectivity.[10] The Swedish government and several Swedish universities and international companies have networks which use IP address blocks allocated by Resilians. The disruption lasted for "less than 12 hours,"[11] which imposed a significant cost in lost connectivity to the disrupted Swedish users. The number of addresses actually sending abusive traffic was 17,664 or about 8% of what was listed.

Although reducing abusive traffic supports the security and stability of the Internet ecosystem, disruption of large, well-behaving address spaces clearly degrades the security and stability of the Internet ecosystem. This Spamhaus-Resilians case study elucidates desirable features and common pitfalls to avoid for any NRS in order to support the security and stability of the Internet ecosystem:

[9] *About Spamhaus.* http://www.spamhaus.org/organization/

[10] *Resilians. Report of Spamhaus incident.* http://webb.resilans.se/documents/spamhaus-incident-20140227-en.pdf

[11] *Resilans Incident Report*. http://www.spamhaus.org/news/article/710/resilans-incident-report

A. Accuracy – The blocking behavior targets the resources that are causing the abusive behavior (note, the Spamhaus actions did do this).
B. Precision – The blocking behavior targets exactly the resources that are causing the abusive behavior, includes all relevant resources (Spamhaus likely did) and only those relevant resources (Spamhaus did not).
C. Monitoring – The results of each precise targeted blocking action must be watched to ensure precision and accuracy, with a very rapid "undo" function whenever overreach is indicated.

Microsoft

Personal Computers (PCs) running Microsoft's Windows software have been the primary target of malicious software for some time, with adversarial actors creating botnets of tens of thousands Windows machines in 2004[12] and 2014[13], and sometimes of millions[14],[15]. Microsoft has a legitimate interest in reducing the exploitation of its software, and so has enacted several programs designed to both discourage infection of Windows machines and disrupt botnets.

No-IP.com is a service that provides free domain names and DNS resolution services under a variety of domains it controls. The company describes itself as helping "home users, small and large businesses and even fortune 500 companies take control over all aspects of their DNS and domain services."[16] Reliable, free, rapidly-configurable DNS is a requirement for many Internet services, including botnets and other malicious aspects of the Internet.

On June 19, 2014 Microsoft filed a civil suit for an *ex parte* temporary restraining order that would permit it to take control of 18,000 allegedly maliciously domains administered by Vitalworks, better known as No-IP.com.[17] Two feature of this case are remarkable. One, an *ex parte* motion means that Microsoft argued it was necessary to grant an injunction against No-IP without the defendant, No-IP, being able to first defend itself in court. Thus, when Microsoft took control of the No-IP namespace on June 30 in order to "classify the identified threats,"[18] No-IP had no forewarning of the seizure. The second remarkable feature is that due to what Microsoft

[12] *Botnet detection and response*, http://www.caida.org/workshops/dns-oarc/200507/slides/oarc0507-Dagon.pdf
[13] Kerkers, M., J.J. Santanna, & A. Sperotto. *Characterisation of the Kelihos. B Botnet,* Monitoring and Securing Virtualized Networks and Services, pp. 79-91. Springer. 2014.
[14] *Analysis of a "/0" Stealth Scan from a Botnet*, http://www.caida.org/publications/papers/2014/analysis_slash_zero/
[15] *Conficker Working Group*, http://confickerworkinggroup.com/wiki/
[16] *About No-IP*, http://www.noip.com/about
[17] *Microsoft v. Mutairi. US District of Nevada. Case # 2:14-cv-00987-GMN-GWF,* http://www.noticeoflawsuit.com/
[18] See original post: http://blogs.microsoft.com/blog/2014/06/30/microsoft-takes-on-global-cybercrime-epidemic-in-tenth-malware-disruption/

later called a "technical error"[19] Microsoft failed to maintain connectivity for the other 5 million domains in the zone that were not part of the indictment, causing an outage that lasted through July 3rd.[20]

Fighting to reduce botnets is certainly a good goal, and when done safely and responsibly would contribute to the security and stability of the Internet. In light of the general wisdom that reactively taking down individual domains is an unsustainable defensive strategy,[21] the approach to go after name servers like No-ip.com's seems tempting. In addition to the three features derived from the Spamhaus-Resilians case study (Precision, Accuracy, Collateral damage avoidance), this Microsoft--No-IP case study elucidates further desirable features and common pitfalls to avoid for any traffic filtering behavior:

D. Transparent accountability – This massive outage of well-behaving address space likely would have been avoided had the injunctive relief not been sought on an *ex parte* basis. Very few judges have the technical capability to measure or predict "irreparable harm" when Internet "takedown" is proposed, and the adversarial model of justice gives opposing technical experts a chance to argue their positions.

E. Technological capacity – Good intentions are not a substitute for proper capacity planning or good system administration. Some DNS experts have privately reported that Microsoft's self-declared "technical error" was no such thing – rather, that the intended result of Microsoft's proxy attempt is technically bankrupt and could not have worked.

F. Anti-abuse – If you create a service that is widely abused, you should expect to be targeted for disconnection. Every service provider will inevitably pay some cost for anti-abuse, either via providing sufficient and responsive anti-abuse measures or as the Internet community decides not to associate with a place hosting abuse. Poor execution of dissociation measures is not an endorsement of abusive behavior.

COICA/ProtectIP/SOPA

Intellectual Property protections have been a thorny issue with digital technologies for some time. The Combating Online Infringement and Counterfeits Act (COICA),[22] the Preventing Real Online Threats to Economic Creativity and Theft of Intellectual Property Act (PROTECT IP Act),[23] and the Stop Online Piracy Act (SOPA)[24] are three related pieces of proposed intellectual property legislation in the United States that did not pass into law. They are remarkable as they

---

[19] *Joint Statement*. https://www.noip.com/blog/2014/07/09/vitalwerks-microsoft-reach-settlement/

[20] *Update: Details on Microsoft Takeover*, http://www.noip.com/blog/2014/07/10/microsoft-takedown-details-updates/

[21] Spring, JM, *Modeling malicious domain name take-down dynamics: Why eCrime pays*, eCrime Researchers Summit (eCRS), 2013.

[22] *S. 3804 (111th): Combating Online Infringement and Counterfeits Act*.

[23] *S. 968 (112th): Preventing Real Online Threats to Economic Creativity and Theft of Intellectual Property Act of 2011.*

[24] *H.R. 3261 (112th): Stop Online Piracy Act*

proposed several forms of mandated Internet blocking to attempt to achieve their stated goals of reducing abuse of intellectual property.

The abusive behavior targeted by COICA/ProtectIP/SOPA was of a different sort than abusive email or malicious botnets. In many cases of intellectual property abuse, there has been no computer security violation on either the user's or server's machine and both parties to the communication are intent on seeing the communication through. This has different technical requirements for blocking than the other two issues, as there is no software to remediate and the user does not need to be notified of an infection or abusive behavior. Indeed, the proposed measures in COICA/ProtectIP/SOPA were technically bankrupt and would not have effectively remediated problem they were seeking to solve.

The legislation, as proposed, did not contain many of the desirable features or avoid common pitfalls already laid out in the previous two case studies for wrong-headed Internet blocking. However the legislation also highlights further wrong-headedness, which perhaps explain why the proposals did not become law.

G. Effectiveness – the proposed measures should be demonstrably effective against the behavior they are designed to prevent.
H. Jurisdiction – The authority proposing the measures should either have the authority to impose the measures on all parties to the communication, or should incentivize voluntary cooperation for those parties not within the authority's jurisdiction.

Thus we arrive at eight desirable features and common pitfalls, summarized as:

A. Accuracy
B. Precision
C. Collateral damage avoidance
D. Monitoring
E. Technological capacity
F. Anti-abuse
G. Effectiveness
H. Jurisdiction