Censorship Arms Race: Research vs. Practice

Sadia Afroz[1], David Fifield[1], Michael C. Tschantz[2], Vern Paxson[1,2], J. D. Tygar[1]

[1]University of California, Berkeley
[2]International Computer Science Institute

ABSTRACT

Government censorship of the Internet has been prevalent in many countries in recent years. Researchers have deployed numerous censorship evasion tools to provide open access to information. Censors, in response, have blocked access to evasion tools. In our talk, we will survey the evaluation techniques and censor models that have been used by researchers, and compare them against the blocking techniques used by real censors, as informed by an in-depth timeline of the blocking of Tor. The goal of the discussion is to illuminate the lack of principled evaluation criteria and the mismatch between research and practice, and to propose a set of criteria that we consider are important.

1. INTRODUCTION

Recent years have seen significant efforts on the part of both practitioners and researchers in countering large-scale Internet censorship imposed by nation-states. The censors seek to block open access to certain types of information, while citizenry in censored countries in turn attempt to *evade* (circumvent) the censor's restrictions. The arms race between censors and evaders often has observable real-world effects.

We seek to build better models of censors: of their incentives, their capabilities, and the costs they incur in carrying out censorship. Improved models will help us design practical and principled criteria for the evaluation of censorship circumvention systems. We are working toward criteria that are grounded in the censorship realities that users face today, while anticipating the future capabilities of sophisticated censors. A common evaluation will facilitate the comparison of different systems and tools. Currently such comparison is hard because there is no common evaluation.

We argue that the censor models under which circumvention tools are evaluated should be informed by the real world as much as possible. We observe that system designers tend to place undue emphasis on censor capabilities that may become important in the future, but are not often seen in practice today, and put too little focus on actual blocking techniques used by current censors. In particular, we identify four **disconnects** between practice and research:

1. Research has emphasized forms of steganography irrelevant for today's arms race.
2. Real censors employ very simple features; research often wrestles with complex ones.
3. The practical interplay between censorship and evasion does not depend upon deep properties of background traffic.
4. Research emphasizes development costs, whereas censors likely emphasize operational costs.

Our talk will focus on resolving these disconnects. To build realistic censor models, we have made a survey of real blocking attacks against Tor and categorized each attack. We have built up a picture of the landscape of current evaluation practices through a survey of more than 50 proposed and deployed circumvention systems. For each system, we cataloged and categorized its evaluation techniques, finding around 60 different evaluations that have been variously applied. Based on our survey, we will discuss properties of good evaluation criteria. The creation of a specific set of criteria is a work in progress.

2. IN RESEARCH

To assess the landscape of research on approaches to censorship evasion, we conducted an extensive literature survey. Table 1 summarizes the space of such efforts, arranged along two dimensions, and with systems that have seen actual deployment highlighted in **bold**. From each work, we extracted the evaluation techniques and the attacks they are meant to defend against, and categorized each along the two dimensions: *Who*, *What*, or *How*; and *polymorphism* or *steganography*. *Who* refers to blocking based on the communicating parties (*e.g.,* a Tor user and the bridge they connect to). *What* refers to the "user payload", that is, the actual substantive content that the user wishes to transmit or receive, contrary to the censor's policy. *How* refers to the protocol used to convey that content (*e.g.,* HTTP-over-TLS). The second dimension is based on whether the systems works by taking on many different forms (*polymorphism*) or by hiding within an existing allowed form (*steganography*).

In principle, censors can target any of *Who*/*What*/*How* to block disallowed communication. In abstract terms, censors do so by selecting and examining or controlling features of Internet traffic. As concrete examples, blocking traffic sent to known Tor bridge nodes (based on the IP address in a packet's header) targets *Who*, as does altering DNS replies to prevent lookups of prohibited destinations. Determining that a Skype flow has additional information beyond pure audio encoded within it targets *What*. Flagging TLS connections whose certificates use particular parameters targets *How*.

We found that techniques developed in the research literature have heavily emphasized steganography over polymorphism, and particularly to mask *What* features (FTE, Infranet, SkyF2F, Collage, CensorSpoofer, DEFIANCE, SkypeMorph, StegoTorus, Freewave, Identity-based Steganographic Tagging, Message In A Bottle, SWEET, Facade, Trist, Facet). Only one of these (FTE) has seen practical deployment. Techniques addressing *How* features have received attention only from practitioners, such as when Tor changed details of its protocol in order to better evade censorship in January and September 2011, on both occasions changing features of the TLS handshake that had been targeted by censors.

	Who	What	How
Polymorphism	**Tor bridges**, **Flash Proxy**, **VPN Gate**	**Obfs2/3/4**, **ScrambleSuit**, Dust	**Tor Jun, 2012**
Steganography	Cirripede, Decoy routing, **GoAgent**, **Meek**, OSS, TapDance, Telex, CloudTransport	**FTE**, Infranet, SkyF2F, Collage, CensorSpoofer, DEFIANCE, SkypeMorph, StegoTorus, Freewave, Identity-based Steganographic Tagging, Message In A Bottle, SWEET, Facade, Trist, Facet	**Tor Jan, 2011**, **Tor Sep, 2011**

Table 1: Prior research on evading network-based censorship using obfuscation, organized by primary obfuscation method. Columns show the primary type of feature obfuscated. Bold denotes deployed tools.

Attacks	**Target**	**Seen: Description**
Website blocking	Who	Thailand 2006: DNS filtering Tor website; Iran & Saudi Arabia 2007: Block GET request pattern with `/tor/`; China 2008, Iran 2012: Block Tor website.
Block by default	Who	Tunisia 2009: Only allow ports 80/443; Iran 2013: TCP reset all non-HTTP.
SSL throttling/blocking	Who	Iran 2009, 2011 SSL throttled to 2 Kb/s; Iran 2012: Block port 443.
IP address blocking	Who	China 2009: Block public relays and directory authorities; China 2010: Block bridges; Iran 2014: Block directory authorities.
Deep packet inspection (DPI)	How	Iran 2011: On Diffie–Hellman parameter in SSL handshake; Iran 2011, Iran 2013: On SSL certificate lifetime; Syria 2011 and 2012: On TLS renegotiation; China 2011: On TLS cipher list in "Client Hello"; Iran 2012, UAE 2012: On TLS handshake; Iran 2012: On TLS client key exchange; Ethiopia 2012, Kazakhstan 2012: On TLS "Server Hello"; Philippines 2012: On TLS cipher suite.
Active probing	How	Probing is used to populate a blacklist. China 2011, 2013.
Unplug Internet	N/A	Egypt 2011, Libya 2011, Syria 2012.

Table 2: Survey of Known Tor Censorship Incidents

3. IN PRACTICE

To ground our model of censors in reality, we made a list of how historical blocking was achieved. For this part, we focused on the blocking of Tor, because it is familiar and stands as a system that is relevant and used enough to be worth attacking. We undertook a survey aiming to comprehensively enumerate every known censorship attack against Tor. Table 2 summarizes our findings, which are derived from analysis of Tor's issue tracker, blog, presentations, and the OONI censorship wiki. We place each attack in a *Who*/*What*/*How* category according to the traffic features it targets.

The results of our survey highlight an imbalance between research focus and the circumvention challenges users are most likely to face in practice. Real censors tend to target *Who* features, which work despite protocol obfuscation. They almost entirely employ blacklists over whitelists. We speculate the heavy emphasis on blacklisting reflects either sensitivities to over-blocking or practical technical limitations. In addition, the blacklisting that real censors employ focuses on simple *per-flow signatures* that identify traffic as disallowed. Censors have not in practice undertaken any substantive in-depth flow analysis that requires maintaining significant state, much less looking for indications that a flow's *content* (as opposed to protocol framing) diverges from expected patterns.

4. EVALUATION CRITERIA

Our findings question the current need for steganographic approaches that operate over *What* features, despite their popularity in research. They also highlight the need for approaches that handle simple features, such as protocol parameters. We believe this disconnect exists because handling simple features is a tedious process for researchers whereas censors, who must use features at Internet scale, are willing to hunt for the simplest usable features. In short, we must match the evaluation criteria used by circumvention tool designers to the abilities of the censors and censorship evaders.

Our talk will discuss what meaningful and practical evaluation criteria could look like. To date, a number of different criteria have been employed to assess the efficacy of censorship evasion technologies. In our survey we found 60 different evaluation criteria used by ~50 censorship evasion proposals. The criteria fall under the following broad categories: resistance to known attacks (*e.g.,* address blocking, active probing), cost, availability of proxies, obfuscation layer security, packet inspection, performance, traffic analysis, unobservability and usage. But different proposals use different criteria to evaluate their system which makes it hard to assess: 1) how well a system will work in practice? and 2) how does it compare to other tools?

We are working toward finding a principled set of criteria that everyone can follow. The most meaningful criteria will be holistic and capture the tradeoffs made by both the censor and evader. Using existing evaluations as guidance, we will isolate and prioritize a set of recommended criteria. We have so far identified two, goodput and cost, that do not lend themselves to immediate assessment. However, we argue that they can provide significant benefit in helping to identify otherwise potentially overlooked considerations when assessing the efficacy of a circumvention technology. Using criteria such as these—as opposed to more narrow ones based on particular threat models such as in prior work—will help open up the general thinking that one brings to bear when considering the merits of a particular approach to censorship evasion. These criteria will serve as a guide and benchmark for implementers and designers and will provide for comparison between designs, relative to different censor models. Better evaluation will lead to better tools.