Consolidated Review of

A Method for Identifying and Confirming the Use of URL Filtering Products for Censorship

1. Strengths:

The main strong point of this paper is that it proves the use of specific products in some countries for censorship purposes. The study is fairly comprehensive as far as the countries that it covers and the use of these technologies.

Methodology can work without a VP in the network in question. Validation of some results where VPs are available (2nd technique).

2. Weaknesses

As the paper admits, the techniques only apply to products visible on the global Internet, and thus, the methodology may be evaded easily by a more sophisticated installation. In addition, the results discussed in the case studies seem more-or-less obvious.

Coverage of methodology is unclear: how many products aren't visible publicly (either all the time or depending on configuration). Is it possible to at least quantify for places where you do have VPs? It seems like you could submit a unique URL to be blocked to each product, and then test which one, if any, was blocked, or whether blocking seemed to happen via some other mechanism. Then, you could know what appeared to be in place in each place, and you could see whether your (Shodan) crawls detected it.

Easy to evade detection if desired. There are essentially two techniques in the paper: crawling to identify possible deployment of software, then testing using in-country VPs and test sites under the control of the researchers. However, it is unclear what we can conclude from the first technique about possibly deployments without any VPs in the country in question (can't tell if the software is actually used to block anything for anyone), and it's unclear why we need the first technique in cases when the second applies (can just test all known products without having done a crawl first). So the contribution of the first part seems to be "someone in this country tried installing this software," which doesn't seem to be a strong claim.

Much of what is reported in these studies is already known: many countries use censoring Web proxies. In fact, more accurately, many ISPs use censoring Web proxies, something that is already well known.

It is unclear what this study offers as far as more general insights---which ISPs are blocking, what triggers blocking, how URL lists are generated and updated, and so forth.

The tools (WhatWeb, etc.) are not developed by the authors and are well-known. Therefore, both the conclusions and the method are already common knowledge, so it is not clear what this paper offers beyond a "data dump".

As the authors state, the approach proposed doesn't look easily scalable. Doesn’t solve the problem of testing from all (or some specific) countries/AS, which is a typical problem of measuring/detecting censorship. Results from sections 3 and 4 look quite disconnected. Specifically, the results from section 3 don't seem to be used in the rest of the paper. The analysis in section 4 is based on behavior previously detected. Moreover, the results of section 3 don't look much significant: the numbers from 3 are pretty small - I doubt that Web Sense has installations in only 7 ISPs. It seems evident that the methodology of looking for publicly accessible proxies doesn't yield to representative results. The link between results of sections 4 and 5 is small but effective (they show that specific ISPs perform censorship and they can attribute such behavior to the use of specific products). However, the results from section 5 are based on previous methodology and are limited to few cases.

3. Comments

This is a nice paper with a clever methodology to tackle a novel, timely problem. My main concern is that it might be very easy to evade this methodology (especially once the paper is published), as this methodology depends primarily on URL filtering installations being configured to be visible on the external Internet. This is a limitation that the authors acknowledge in the paper (page 2).

It may be also be easy to evade the paper's methodology to confirm the continued use of these products, e.g., if the URL filtering products require that a certain minimum (unknown) number of users submit the URL for filtering/analysis.

The techniques described in this paper are well-known and fairly simple, but, perhaps more worrying, they are also evadable---once the paper is published, the techniques for detecting the presence of a Web proxy described in this paper will be immediately obsolete. This begs the question of whether the paper should be proposing a more robust method

The depth and importance of these findings isn't particularly interesting, beyond simply reporting the evidence that these tests found of different proxies being deployed in different countries. It's possible that all of the filters are being used in all countries, and that the testing method simply didn't uncover the use of all of the proxies.

Detection depends not only on placement, but also on the proxy filters conforming exactly to the signatures outlined in the paper. Isn't it possible that different versions of the software might have different signatures? A proper study would require some validation in the lab of the behavior of different versions of filtering software, and whether the signatures vary across versions and installations.

It's unclear whether use in some large North American ISPs is surprising, since you haven't characterized at all which users are behind the software - for example, AT&T blocking porn for their employees wouldn't be that interesting. Related, when you discover blocking in an ISP in a country that has a repressive regime, how do you differentiate (without a VP) between software that might be configured to block for certain users (say, in gov't offices) vs. something that is being used for widespread censorship? Given that you require confirmation from a VP to gain confidence, it's unclear how much your crawls are buying you. What if you just submitted different test URLs to each product and used that to figure out what was used where, without doing the crawl first?

Table 1, Table 3: Bold the ones that weren't previously known and were discovered only with this paper. Do you really need both these tables? Is Table 1 what was previously known? Perhaps using bold you can combine the two.

Given that [21] found companies evading detection by modifying their product, the signatures in Table 2 don't seem very robust. Do you expect companies or ISPs to react to your paper? Can you characterize Shodan's coverage a bit? I poked around a bit online but it wasn't obvious to me.

3: it's not clear what is the contribution of section 3 to the paper (see comments in paper weaknesses). The "8080/webadmin" criterion seems too generic. Can you provide an estimate of the percentage of the deployments you are able to identify? -

4: In section 4 sometimes is not clear what is the URL category considered/submitted. In some cases it is missing (2nd paragraph of section 4.4). - In section 4 it's not clear what is the meaning/impact of paragraphs starting with "Challenge #" - typo: "further, the these"

4.1: At this point, I was wondering about webpages that would be naturally different for users in different locations. I later realized that you are testing webpages under your control, but you might want to state that clearly here. In general, I'm not sure that you need this section separate from 4.2. Minor comment: in Section 4.1, can you clarify whether you have users in various countries testing the websites for your experiments?

4.2: It took me awhile to realize that the submission of URLs for blocking was centralized for a product, not specific to a given installation. Make this clear up front. Is the idea that a URL is submitted and some verification of category happens at the company? I like the anecdote about the Yemeni ISP with limited licenses.

5: How did you get the set of URLs to test?

4. Summary from PC Discussion

The TPC acknowledges that the methodology has useful advantages (e.g., not needing a vantage point in the network analyzed). However, the analysis could have been deeper, elaborating more about the networks found to filter URLs and their motivations. The paper also presents several important limitations (coverage and easy to evade) that the authors are encouraged discuss further in the camera-ready version.

5. Authors' Response

A key issue raised in the PC comments is a lack of clarity in the purpose of the analysis in Section 3. We have revised this section to make it clear that the goal is to identify candidate networks for the profiling discussed in Section 4. While this identification step is a useful filter, the analysis of Section 4 is not dependent on it.

We have also added a discussion section, elaborating on the potential ways in which our techniques may be evaded by vendors, how this impacts the approaches we propose, and potential mitigation strategies. We have also added more recent confirmation results to Section 5 and clarified various aspects of Sections 4 and 5 mentioned in the above review.