O Pakistan, We Stand on Guard for Thee:
An Analysis of Canada-based Netsweeper’s role in Pakistan’s Censorship Regime

KEY FINDINGS

- Netsweeper filtering products have been installed on Pakistan Telecommunication Company Limited (PTCL)'s network. PTCL is Pakistan's largest telecommunications company and also operates the Pakistan Internet Exchange Point.

- Netsweeper technology is being implemented in Pakistan on PTCL for the purposes of political and social filtering, including websites relating to human rights, sensitive religious topics, and independent media.

- In addition to using Netsweeper technology to block websites, ISPs also use other less transparent methods, such as DNS tampering.

INTRODUCTION

In February 2012, Pakistan’s Ministry of Information Technology solicited proposals for a national level URL filtering system. The request, originating from the Ministry’s National ICT R&D Fund, specified that the filtering system should be capable of blocking websites at the domain and subdomain level, and allow for the filtering of specific files, file types, and IP ranges. A petition by Access, an advocacy group, called on technology companies that produce network filtering technology to publicly announce that they would not submit a bid for the project. While the petition won support among several major IT companies, five firms—China-based Huawei and ZTE, US-based Blue Coat Systems and McAfee, and Canada-based Netsweeper—declined to comment.

In the past, research by the OpenNet Initiative (ONI) has uncovered evidence of the filtering of blasphemous and anti-Islamic content, and sites promoting Balochi, Sindhi, and Pashtun human rights and political
Due to blasphemous content being circulated on major sites like Facebook, Twitter, and YouTube, Pakistani authorities have on different occasions blocked the entire domains of these sites as opposed to just the offending content. While the blocking of Twitter\(^5\) and Facebook\(^6\) was rescinded after periods of filtering, YouTube\(^7\) continues to be blocked. Authorities have also shut down communications in the name of national security. For example, on Pakistan Day 2012, authorities shut down cellular communications in Balochistan for 14 hours.\(^9\) A few months later, on Pakistan Independence Day, authorities shut down cellular communications once again in Balochistan.\(^10\)

On May 2, 2013, a number of Twitter users reported that Tumblr was blocked in some regions of Pakistan, including Balochistan and Sindh.\(^11\) Based on previous findings in several Middle Eastern countries that technology provided by Netsweeper categorizes and blocks Tumblr as pornography,\(^12\) Citizen Lab searched the computer search engine Shodan for Netsweeper installations on networks in Pakistan and discovered a Netsweeper installation on Pakistan Telecommunication Company Limited (PTCL). PTCL is Pakistan’s leading telecommunications provider, offering broadband Internet connectivity to over one million subscribers and accounting for 60 percent of broadband market share in 2012.\(^13\) It also serves as a major Internet backbone provider along with Transworld Associates (TWA).\(^14\) This formerly government-owned ISP was privatized in 2006, with the Government of Pakistan still retaining a 62 percent stake in the company.\(^15\) Etisalat International Pakistan, a subsidiary of the UAE-based Emirates Telecommunication Company Corporation, owns an additional 26 percent stake\(^16\) and effectively controls the management of the company.\(^17\)

This report documents our finding of Netsweeper filtering technology in Pakistan, describes how Netsweeper devices censor content, and explains our most recent results from Internet censorship testing in the country. We conclude our report with specific questions directed to Netsweeper in the hope of encouraging greater transparency about its product and services, especially in human rights-concerning jurisdictions like Pakistan.

**METHODOLOGY**

As a founding partner of the ONI, the Citizen Lab has been studying and reporting on Internet censorship and surveillance worldwide since 2003. Since that time, the ONI has tested for Internet filtering in 74 countries and found that 42 of them—including both authoritarian and democratic governments—implement some level of filtering.\(^18\) The ONI combines network measurements with contextual field research to investigate the prevalence, depth, and breadth of state-mandated Internet filtering regimes.\(^19\) Our research has uncovered the use of technologies developed by Western companies to perform Internet censorship and surveillance in a number of countries governed by regimes with poor human rights records.\(^20\) This includes products produced by companies such as Netsweeper, Blue Coat Systems, Websense, and McAfee.

In our earlier research, Internet filtering products were often easily identified, as block pages would explicitly show the company name or logo. (See Figure 1).
Figure 1: Block page from Saudi Arabia’s STC ISP in 2009, clearly displaying McAfee SmartFilter’s logo.

However, such instances have become increasingly rare. As Western companies have come under fire for supplying censorship tools to repressive governments, ISPs have begun to remove traces of their partners from block pages. (See Figure 2).

Figure 2: This updated Saudi standard block page does not display the product used for filtering.
In response, the Citizen Lab has been developing reproducible methodologies for identifying filtering technology and determining whether they are being used for censorship. The methodology used in this report leverages key tools and datasets, such as Shodan, a search engine that lists IP addresses of externally visible devices on the Internet. We used Shodan to scan for keywords and URL fragments that have been associated with censorship devices in the past (e.g., “8080/webadmin” for Netsweeper). Results are limited by top-level domain to return only those pertaining to the country of interest. Several tools and sources are then used for validation, including WhatWeb (signatures), Internet Census 2012 (IP history), and Team Cymru (whois data).

We are also assisted by in-country testers running measurement software that we have developed to document instances of Internet filtering. This software attempts to access lists of predefined URLs simultaneously in both Pakistan and a control test from the University of Toronto. Two URL lists are tested in each country: a ‘global list’ (constant for each country) and a ‘local list’ (different for each country). The global list is comprised of a wide range of internationally relevant and popular websites including sites with content that is perceived to be provocative or objectionable. The local list is specific to each country, and in this case contains content specific to the Pakistani linguistic, political, social, and cultural context. Once completed, the results of these tests are sent to servers in the University of Toronto, where they are analyzed for evidence of filtering.

**TECHNICAL ANALYSIS**

Using the Shodan search engine, a Netsweeper installation was found in May 2013 on the IP address 202.125.134.154, which belongs to the PTCL autonomous system AS17557. It was identified as Netsweeper based on a URL pattern which is consistent in all Netsweeper installations: the return of a block page when visiting http(s)://IP_IN_QUESTION/webadmin/deny/index.php or a netsweeper login page when visiting the IP directly. Upon visiting this IP with the block page URL pattern, a clear block page is seen referencing restricted content in Pakistan (See Figure 3), as well as a Netsweeper login page when visiting the IP directly (See Figure 4).
Figure 3: A block page is seen when visiting http://202.125.134.154/webadmin/deny/index.php

Figure 4: A Netsweeper login page is seen when visiting http://202.125.134.154/webadmin/start
These URL patterns and results are consistent with previously observed installations of Netsweeper\textsuperscript{24} and are highly consistent with Netsweeper’s own documentation on configuring block pages.\textsuperscript{25} PTCL customers have also posted on Pakistani Internet forums in a number of different cases, each time reporting the occurrence of this block page upon trying to visit content such as file sharing and proxy sites.\textsuperscript{26} The screenshots and IP addresses reported in these cases are identical to the identified installation of Netsweeper. In June 2013, we scanned Pakistan’s network again using Shodan, and another IP address hosting Netsweeper emerged (http://119.159.224.68:8080/webadmin/start). This IP address also belongs to the PTCL autonomous system AS17557. Both IP addresses are hosted on PTCL, which also operates the Pakistan Internet Exchange (PIE). Pakistan's PIE was created by the PTCL, managing a significant portion of Pakistan’s overall Internet traffic.\textsuperscript{27}

As can be seen in Figure 5, AS17557 (PTCL) provides the bulk of Pakistan’s connectivity to ASes outside of the country:

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{A map of the Pakistani Internet infrastructure. PTCL’s AS17557 (highlighted along with connected autonomous systems) has most of the links to ASes outside of the country.}
\end{figure}
Additionally, the text that is present on the block page is consistent with the messaging on other observed block pages in the region on different Pakistani ISPs (See Figure 6). The installations on these particular IPs are not consistent with the use of Netsweeper, but the language used does seem to indicate a uniform requirement for all block page text.
We also examined another public source of information about Internet-connected devices, the Internet Census 2012, for the presence of Netsweeper devices. The Internet Census 2012 data is controversial, as it used a vast network of unsecured, publicly-accessible devices without consent to collect data about Internet-connected devices.28 We have used the dataset for this research as it is in the public domain and contains an unreplicated view of connected devices; however, Citizen Lab does not condone research methods that may be unethical or illegal.29

We searched the Internet Census 2012 dataset for information related to the IP addresses of the PTCL Netsweeper devices. The first IP address, 202.125.134.154, returned no data for a GET request on port 80. This may indicate that the device was not present at the time of the Internet Census 2012 data collection, which took place from June 2012 to December 2012. This absence may indicate that the IP address represents a new Netsweeper installation installed sometime between December 2012 and May 2013, but further research of the Internet Census dataset is required.

The IP address of the second identified Netsweeper device, 119.159.224.68, returned the following to a GET request on port 80 in the Internet Census data:

HTTP/1.1 302 Found
Date: Sat, 15 Dec 2012 16:06:58 GMT
Server: Apache
Location: http://127.0.0.1/webadmin/
Content-Length: 269
Connection: close
Content-Type: text/html; charset=iso-8859-1

<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN"
The two responses are identical with the exception of the IP address that it redirects to, which changed from 127.0.0.1 (localhost) to 119.159.224.68. This similar response suggests that the same device was present in December 2012 and June 2013. We hypothesize that the December 2012 redirection to 127.0.0.1/webadmin represents a device which has yet to be properly configured. We are conducting a more detailed scan of the Internet Census dataset and will write a follow-up with our findings.

Altogether, the evidence suggests that Netsweeper filtering devices are being actively used to censor content on an ISP-wide level in Pakistan.

IN-COUNTRY TEST RESULTS

In June 2013, we conducted in-country tests on PTCL. These tests were run from a number of different locations throughout the country. Testers ran a software tool which queried a list of 1,465 URLs simultaneously on both PTCL and a control test at the University of Toronto. The results of these tests were analyzed at the University of Toronto to determine if any URLs were blocked. Lists of the URLs tested as well as blocked URLs can be found in the Data section at the end of this report.
Test results were consistent with prior test results documented by the ONI. June 2013 testing found evidence of 123 URLs blocked through a variety of methods. In some cases, URLs were blocked by different methods on different tests. 90 URLs were found to be blocked by one of two explicit block pages on at least one occasion.

39 unique URLs were found blocked with the block page seen in Figure 7.

**Figure 7: Block page found on PTCL during June 2013 testing**

31 URLs were found to be blocked by consistently failing to resolve or by resolving to localhost (127.0.0.1) during the DNS resolution process. Two URLs either did not see a response to the TCP handshake or were forcibly closed through RST packets.

**NETSWEeper-Enabled Blocking**

In addition to the above mentioned methods of blocking, 51 URLs received the block page seen in Figure 8 on at least one test:

**Figure 8: Block page found on PTCL during June 2013 testing**

The HTML source code of the block page shown in Figure 8 above is as follows:

```html
<iframe src="http://202.125.134.154/webadmin/deny/?dpid=1&dpruleid=78&cat=104&ttl=0&groupname=PTCL2&policyname=PTCL2-policy&username=MMBB-9-WLL&userip=X.X.X.X&connectionip=127.0.0.1&nsphostname=X&protocol=policyprocessor&dplanguage=-&url=X" width="100%" height="100%" frameborder=0></iframe>
```
The format of this block page is highly consistent with Netsweeper’s documentation for configuring block pages. 32

URLs from a variety of content categories were found blocked using this method. 15 YouTube videos and four websites dealing with the Muhammad cartoon controversy were blocked, as were a number of websites containing content critical of Islam. Several websites relating to Baloch issues, including news sites http://balochwarna.com and http://thebalochhal.com, were also found blocked. Also found blocked were proxy and circumvention tools, bittorrent file-sharing sites, and pornography.

These results show that URLs are blocked on PTCL through a variety of methods, and with varying degrees of transparency. Some URLs were accessible on some tests while blocked on others. While all testing took place on PTCL, the tests were run from several different regions of the country, which could potentially explain some inconsistency in results. However it is clear from these results, particularly given the block page seen in Figure 8, that Netsweeper technology is being used to filter web content on PTCL.

Other ad hoc tests have documented other types of content blocked via Netsweeper. For example, the website Answering Islam, which is critical of the Islamic faith, was found to be blocked with this same block page, as was the website of British evolutionary biologist Richard Dawkins, likely because of objections to the theory of evolution and Darwinism among Pakistani authorities.

**HOW NETSWEEPER’S CONTENT FILTERING SOLUTION WORKS**

Netsweeper Inc. is a Canada-based provider of web content filtering and web threat management products. 33 Its filtering service has been used for the purposes of state-sanctioned censorship in several countries. For example, national ISPs in Qatar, United Arab Emirates, Kuwait, and Yemen employ Netsweeper to implement political and religious censorship. 34 As we cited in a previous report, Netsweeper has stated that its product can be used to block inappropriate content to meet government rules and regulations “based on social, religious, or political ideals.” 35

The deployment of Netsweeper at the ISP level can potentially result in pervasive filtering across various categories as a result of the way Netsweeper technology works. Its web filtering solution provides clients with an automated mechanism to bulk-filter entire content categories. Netsweeper’s content categories include: 36

- **Adult** (adult image, alcohol, alternative, criminal skills, extreme, gambling, hate speech, lifestyles, matchmaking, occult, pornography, profanity, substance abuse, weapons)
- **Entertainment** (arts and culture, educational games, entertainment, games, humour, sports)
- **Information** (general news, journals and blogs, political, portals, religion, self help, sex education, social networking, technology, travel)
- **Security** (adware, directory, host is an IP, intranet servers, malformed URL, phishing, anonymizer, viruses and malware)
- **Miscellaneous** (investing, job search, sales, search engine, web chat, web e-mail)
- **Advanced** (general, images, network timeout, network unavailable, new URL, no text redirector page, safe search, search keywords).

- **Custom Categories** (e.g. extreme sites for racism)

ISPs and telecom operators can choose which of these categories they want to block but can also manually add their own categories and URLs. Websites are categorized at a rate of approximately 10 million new URLs every day. \(^{37}\) Netsweeper uses an artificial intelligence engine to categorize new and uncategorized URLs in real time.

When a user attempts to access a website, the “Netsweeper Enterprise Filter” intercepts the request and sends it to the “Policy Server” to determine whether access to the site should be allowed or denied. If the website is in this local database, the Policy Server categorizes the website and gives direction to the Netsweeper Enterprise Filter based on the client’s specific filtering policy. If the client’s filtering policy does not allow the content category that the website request belongs to, the request is denied and the Enterprise Filter returns a block page to the user.

If the Policy Server is unable to categorize the website, it sends the URL to the Category Name Server (CNS) for categorization. If unsuccessful, it continues upstream, requesting categorization from the Master Category Name Server, and finally, for categorization by the Categorization Engine. In real time, the Categorization Engine uses artificial intelligence to scan the words and images, assigns the website a content category and updates the Categorization Database with the URL and its category. The Policy Server then carries out the request based on the categorization and the client’s filtering policy. Netsweeper states that the process of categorization—from user request to categorization engine categorization—only takes a maximum of five seconds (See Figure 9: Netsweeper Filtering Process). \(^{38}\) Netsweeper has categorized over 5 billion URLs in total. \(^{39}\)

![Figure 1: Netsweeper Filtering Process](image)

**Figure 9: Netsweeper Filtering Process. Source:** “Netsweeper Internet Content Filtering”\(^{40}\)
The comprehensiveness of the content categories shows how pervasive censorship can be both in scope and depth. It also shows how a commercial company can aid national censors not only by providing technology but also by defining the parameters of permissible content through categorization. Yet, Netsweeper’s artificial intelligence categorization engine has proven to be unreliable.41 As previously mentioned, research by the OpenNet Initiative has shown how Tumblr.com and all blogs hosted by Tumblr were blocked in Qatar, UAE, Yemen, and Kuwait due to Netsweeper’s miscategorization of the site as “pornography.”42

CONCLUSIONS

Advocacy groups have expressed concern around the lack of accountability and transparency surrounding Pakistan’s censorship and surveillance practices. Pakistan’s 1996 Telecommunications Act has been criticized by rights groups for containing too many broad provisions that justify restrictions on expression and privacy in order to protect undefined national security interests.43 In the past, the OpenNet Initiative has found that Pakistani authorities block content that they consider blasphemous, anti-Islamic, or threatening to internal security—particularly content that relates to Balochi, Sindhi, and Pashtun human rights and political autonomy.44 In addition, the Pakistani advocacy group Bytes for All has expressed concern over the country’s lack of privacy laws and data protection legislation in light of the Citizen Lab’s discovery of a FinFisher command and control server in Pakistan.45 FinFisher is a surveillance technology sold by the UK-based Gamma Group that has been used to surveil journalists, dissidents, and activists in a number of countries.46 The fact that Pakistan has deployed the Netsweeper filtering technology at the national Internet Exchange level is a significant development given the potential of extending Internet censorship to lower-level ISPs in the country. Our report sheds light on the commercial technology used by Pakistani authorities to conduct national-level content filtering, but many ethical and legal questions remain.

Pakistan is not the only country in which Netsweeper has been used for filtering purposes. Prior Citizen Lab research has identified Netsweeper operating on ISPs in several other countries with questionable human rights records, including Qatar, UAE, Yemen, and Kuwait.47 Netsweeper’s presence in these countries indicates that the company is aggressively moving into the mass political censorship market at the very time that human rights concerns are being raised both about these practices and the companies that serve them. At one time, the company even published on its website a brochure that advertised that its product provides blocking of “inappropriate content using pre-established list of 90+ categories to meet government rules and regulations—based on social, religious, or political ideals.”48 At the time of writing, that brochure is no longer available on the website, possibly indicating the company is becoming less transparent about the services it offers in this particular market.

Companies that provide equipment and services that can be used or otherwise modified to infringe on human rights are increasingly being pressed to provide greater transparency around the decisions they take and the services they provide in specific jurisdictions. To further that end, and consistent with prior Citizen Lab research reports on the issue-area involving other companies,49 we pose the following questions to Netsweeper:

- Does Netsweeper have a human rights policy, and does it implement this policy when developing its technologies and sales strategy?
• Does the company assess the human rights impact of its products during the design phase and has it ever discarded or altered designs given their inherent capability to undermine rights of freedom of expression and access to information?

• What resources does Netsweeper devote to human rights program at the operational level? Does Netsweeper ask staff in relevant departments (e.g., legal, sales, engineering) to undergo human rights training?

• Is Netsweeper aware of the “know your customer” standard, where companies actively investigate whether potential clients may use technology to undermine human rights standards? If so, how does it implement this standard (for example, through active investigation of a government’s human rights record)?

• Has Netsweeper implemented the United Nations Guiding Principles on Business and Human Rights (the so-called “Ruggie Principles”) in building a business strategy that safeguards human rights standards?

• Has Netsweeper explored joining the Global Network Initiative (GNI), a network of business, civil society, and academic stakeholders, in finding solutions for technology companies to uphold standards of privacy and free expression, as the ICT company Websense did in 2011?

We commit to publishing in full Netsweeper’s reply.

DATA

The complete list of URLs found blocked, as well as the list of URLs tested, can be found in the following:

Complete list of URLs found blocked on PTCL:

• [CSV][Google Drive]

• Testing conducted from June 2nd to 13th, 2013

List of URLs tested on PTCL:

• [CSV][Google Drive]

• Testing conducted from June 2nd to 13th, 2013

ACKNOWLEDGEMENTS

Citizen Lab would like to thank Bytes for All (B4A) in Pakistan for their assistance in gathering technical data and assisting in the research for this report.
FOOTNOTES

1 The title is a reference to the Canadian national anthem (“O Canada, We Stand on Guard for Thee”)
2 DNS tampering is the practice of preventing nameservers from returning the correct IP address to a DNS resolution request.
11 Bytes for All (Bytes for All), "Apparently @tumblr is blocked in #Quetta #Pakistan. @satanubis shared this screenshot @citizenlab @Liberationtech pic.twitter.com/ws7CvaRttof," May 2, 2013, 10:16 a.m., Tweet, https://twitter.com/bytesforall/status/330007804901482496.
24 Noman and York, “West Censoring East.”
31 Some information has been redacted from this HTML source code for privacy reasons.
34 Noman and York, “West Censoring East.”
35 Ibid.
39 “About Us,” Netsweeper.
41 Noman and York, “West Censoring East.”
42 Noman, "When a Canadian Company Decides."

"Pakistan," OpenNet Initiative.

“Bytes for All Vs. Federation of Pakistan - Censorship is Not a Solution, it is a Problem!” Bytes for All, April 30, 2013, http://content.bytesforall.pk/node/98; and “Notorious Spy Technology Found in Pakistan,” Bytes for All, May 1, 2013, http://content.bytesforall.pk/node/99. Bytes for All has filed a petition in the Lahore High Court concerning the use of FinFisher in Pakistan, see “Privacy Rights Violations Challenged in Lahore High Court," Bytes for All, May 8, 2013, http://content.bytesforall.pk/node/100.


Noman, “When a Canadian Company Decides.”


