Three Researchers, Five Conjectures: An Empirical Analysis of TOM-Skype Censorship and Surveillance

Jeffrey Knockel, Jedidiah R. Crandall, and Jared Saia
University of New Mexico
Dept. of Computer Science
{jeffk, crandall, saia}@cs.unm.edu

Abstract

We present an empirical analysis of TOM-Skype censorship and surveillance. TOM-Skype is an Internet telephony and chat program that is a joint venture between TOM Online (a mobile Internet company in China) and Skype Limited. TOM-Skype contains both voice-over-IP functionality and a chat client. The censorship and surveillance that we studied for this paper is specific to the chat client and is based on keywords that a user might type into a chat session.

We were able to decrypt keyword lists used for censorship and surveillance. We also tracked the lists for a period of time and witnessed changes. Censored keywords range from obscene references, such as 二女一杯 (two girls one cup, the motivation for our title), to specific passages from 2011 China Jasmine Revolution protest instructions, such as 成都 春熙路麦当劳门前 (McDonald's in front of Chunxi Road in Chengdu). Surveillance keywords are mostly related to demolitions in Beijing, such as 灵境胡同拆迁 (Ling Jing Alley demolition).

Based on this data, we present five conjectures that we believe to be formal enough to be hypotheses that the Internet censorship research community could potentially answer with more data and appropriate computational and analytic techniques.

1 Introduction

How effective is keyword censorship at stifling the spread of ideas? Is constant surveillance necessary for effective Internet censorship? What are the computational, linguistic, political, and social problems faced by both the censors and the people seeking to evade censorship?

A good understanding of how Internet censorship works, how it is applied, and what its impacts are will require both ideas from the social sciences and computational ideas. Consider a relatively simple question such as if keyword-based censorship is effective at stopping protests when censorship keywords target specific advertised protest locations, *e.g.*, 西大直街康宁路路口世纪联华 (Corning West and Da Zhi Street intersection, Century Lianhua gate). Estimating the effectiveness of this entails an understanding of psychology to quantify the effects of perceived surveillance and uncertainty, meme spreading, social networking, content filtering, linguistics to anticipate attempts to evade the censorship, and many other factors.

In this paper, we propose five conjectures about censorship. Our conjectures are based on our recent results in reverse-engineering TOM-Skype censorship and surveillance, combined with past studies of Internet censorship. TOM-Skype is an Internet telephony and chat program that is a joint venture between TOM Online (a mobile Internet company in China) and Skype Limited. TOM-Skype contains both voice-over-IP functionality and a chat client, the former of which implements keyword-based censorship and surveillance that we have reverse-engineered.

We present these conjectures in a formal way, in an attempt to propose them as testable hypotheses on which future research can focus. We do not expect all of our conjectures to be true. However, all of them have the properties that: 1) they can in principle be empirically tested; and 2) determining whether they are true or false will advance our understanding of Internet censorship. We contend that the enumeration of such testable conjectures is critical in order for the study of Internet censorship to continue as a viable area of scientific research.

1.1 TOM-Skype results

In this paper, we give preliminary results from reverse-engineering different versions of TOM-Skype. Our results include the cryptography algorithms used for both censorship and surveillance, differences between TOM-Skype versions, fully decrypted lists of keywords with translations, changes to the lists over time, and a rough

categorization of three of the lists.

Recently, Nart Villeneuve demonstrated that the chat functionality of TOM-Skype triggers on certain keywords, preventing their communication and uploading messages to a server in China [16]. He provided some high-level analysis of what is censored and how this mechanism works. In this paper, we provide a more detailed analysis of TOM-Skype, including the algorithms for protecting the keywords that trigger censorship and surveillance. All versions of TOM-Skype have at least one of two separate lists: one that triggers both censorship and surveillance and one that only triggers surveillance. We have decrypted all lists for all versions of TOM-Skype that we analyzed, and we have translated the most recent version's lists and tracked changes in the lists. The encryption for protecting the keyword lists in earlier versions is based on a simple algorithm involving additions and exclusive-or operations on each byte, whereas the encryption for later versions is AES-based. The encryption for uploading conversations that trigger surveillance is DES-based. By overcoming the anti-debugging functionalities built into both Skype and TOM-Skype, we also have a detailed understanding of how the censorship and surveillance is implemented. Based on our analysis, we propose a set of five conjectures.

1.2 Related work

The Open Net Initiative is an excellent source of information about censorship in a variety of countries [6]. However, the descriptions of what is filtered are relatively high-level. The report by Zittrain and Edelman [17] is a good overview of some of China's censorship implementations.

The methods of China's HTTP keyword filtering were first published by the Global Internet Freedom Consortium [10]. Clayton *et al.* [3] published a more detailed study of this mechanism. The ConceptDoppler project [4] studied multiple routes and also used Latent Semantic Analysis [12] to reverse-engineer 122 blacklisted keywords by clustering around sensitive concepts and then probing. ConceptDoppler has generated two more recent lists as well.

Human Rights Watch [11], Reporters Without Borders [14], and others [13, 1] have released reports describing China's censorship regime. These reports often include insider information about what is censored [14] or perhaps full leaked blacklists, such as the list of keywords blocked in the QQChat chat program [11, Appendix I] or by a particular blog site [11, Appendix II]. In the case of the QQChat list, hackers found the list in the QQ software by doing a simple string dump of QQChat's dynamically linked libraries, *i.e.*, there was no encryption.

There are several factors that make the list we have obtained from TOM-Skype unique among the lists that have been made public thus far. The first is that, not only is the TOM-Skype list more up-to-date, but in the three weeks that we have been monitoring it we have recorded many updates to it. We plan to record daily updates for a long period of time. Also, we believe our list is the first to provide more than anecdotal evidence of some of the shorter-term applications of censorship, such as the censorship of specific intersections where protesters planned to meet or events in the news. Our list is not only complete, but there is a clear separation between censorship keywords and surveillance keywords.

The prior work that is closest to ours is the aforementioned study of TOM-Skype by Villeneuve [16]. That analysis was based on obtaining the uploaded conversations, which were available for download on the server that TOM-Skype uploaded them to at the time, and performing clustering and other aggregate analyses of this data. In contrast, the analysis we present in this paper is based on reverse-engineering of the TOM-Skype implementation of censorship and of the cryptography for protecting the keyword blacklist. Thus, we are able to present exactly what the keywords are and make a clear distinction between keywords that evoke censorship and surveillance and those that only evoke surveillance.

There has been some amount of historical, political, economic, and legal discussion of the potential effectiveness and applications of Internet censorship in China and elsewhere [1, 13, 7, 2, 5]. Our goal in this paper is to present data that can help the research community move toward formalizing conjectures about Internet censorship that can be tested computationally. We propose five such conjectures after presenting our data.

This paper is structured as follows. In Section 2 we present our findings from reverse-engineering TOM-Skype. Section 3 discusses the two keyword blacklists for the most recent version of TOM-Skype, which we have translated and analyzed in detail. Then we propose five conjectures and some recommendations for future work in Section 4.

2 Empirical analysis of TOM-Skype

In this section we describe the basic empirical results from our efforts to reverse-engineer TOM-Skype.

2.1 Censorship mechanisms

Each version of TOM-Skype that we analyzed features shared censorship mechanisms. Each binary contains a built-in, encrypted list of keywords to censor, and, via HTTP, each client downloads at least one additional encrypted list of censored keywords called a "keyfile" from TOM's servers. Each client uses at least one of these lists to censor incoming and/or outgoing chat messages at any

time.

However, we also found stark differences in their censorship implementations. Different versions use different encryption algorithms, different built-in keyword lists, and download keyfiles from different locations. Moreover, implementations vary in whether they censor incoming and/or outgoing chat messages.

2.1.1 TOM-Skype 3.6 and 3.8

We first analyzed TOM-Skype 3.6.4.316 and 3.8.4.44. We found that these clients censor both incoming and outgoing chat messages by failing to render them in one's chat window and by failing to record them in one's chat history. Censored outgoing messages are additionally never sent.

When the client starts up, words are initially censored according to the keyword list built into the binary. After a keyfile is downloaded from TOM's servers, the downloaded keyfile *substitutes* the built-in keyword list. By using a packet sniffer, we found that these clients download keyfiles from the following URL:

skypetools.tom.com/agent/newkeyfile/keyfile

To decrypt this file, we redirected `skypetools.tom.com` DNS queries to our own HTTP server, allowing us to force TOM-Skype to load keyfiles of our choosing. Then, through binary search, we were able to locate the ciphertext entry for the keyword "fuck" in the keyfile, which is the keyword that Villeneuve [16] used. We started with a known-plaintext analysis of this keyword. We then employed a chosen ciphertext attack by adding initially single-character words to the list to see which words were filtered by TOM-Skype. As we recognized patterns and became more familiar with the decryption algorithm, we were soon able to censor entire words. The decryption algorithm that we discovered follows:

Algorithm 1 Decrypting TOM-Skype 3.6 keyfiles

1: **procedure** DECRYPT($C_{0..n}, P_{1..n}$)
2: **for** $i \leftarrow 1, n$ **do**
3: $P_i = (C_i \oplus 0\text{x}68) - C_{i-1} \pmod{0\text{xff}}$
4: **end for**
5: **end procedure**

The ciphertext always has one more byte than the plaintext, since the ciphertext's first byte serves as an initialization vector. We found that this algorithm also decrypts the keyword lists built into the binaries.

2.1.2 TOM-Skype 4.0 and 4.2

Next we analyzed TOM-Skype 4.0.4.226 and 4.2.4.104. These versions implemented censorship similarly as our tested 3.6 and 3.8 versions, except they download a keyfile from the following URL:

`a`*n*`.`skype.tom.com/installer/agent/keyfile

where n is a pseudorandom, uniformly-distributed integer between 1 and 8, inclusive. This file can be decrypted using the previous algorithm, as can these clients' built-in keyword lists.

2.1.3 TOM-Skype 5.0 and 5.1

Finally, we analyzed TOM-Skype 5.0.4.14 and 5.1.4.10. These versions delegate censorship to a separate out-of-process binary `ContentFilter.exe`. We found these versions to only perform censorship on incoming messages.

We found that `ContentFilter.exe` downloads keyfiles from the following URL:

skypetools.tom.com/agent/keyfile

which, as before, substitutes the list built into `ContentFilter.exe`'s binary. Moreover, we found that only TOM-Skype 5.1.4.10's `ContentFilter.exe` additionally downloads a keyfile from the following URL:

skypetools.tom.com/agent/keyfile_u

The words in the latter keyfile are not used for censorship but only for surveillance, detailed in the next section.

We found that both of these keyfiles and the keyword lists built into `ContentFilter.exe` are encrypted with a 256-bit AES key in ECB mode. This UTF-16LE-encoded key is originally known to have been used to encrypt the downloaded keyfile in TOM-Skype 2.5 [9]:

$\texttt{CENSOR\_KEY}_{5.0}$ = "0sr TM#RWFD,a43 "

In UTF-16LE encoding, this key is 256 bits, although half of the bytes are null.

2.2 Surveillance mechanisms

We found all versions of TOM-Skype analyzed in the previous section to perform surveillance except 5.0.4.14. Each of the other versions, whenever it performs censorship, reports back encrypted text in a query string to the following URL:

a*n*.skype.tom.com/installer/tomad/ContentFilterMsg.php

where again n is a pseudorandom, uniformly-distributed integer between 1 and 8, inclusive.

As reported in the previous section, TOM-Skype 5.1.4.10 has an additional downloaded keyfile containing words only used for surveillance of but not the censorship of incoming messages.

By reverse engineering TOM-Skype 3.8.4.44's `Skype.exe` and TOM-Skype 5.1.4.10's `ContentFilter.exe` binaries, we discovered

two DES keys used to encrypt surveillance text, where which is used depends on the version. Being outside of the main Skype binary, we first targeted `ContentFilter.exe`, as `Skype.exe` is known to contain anti-debugging measures that cause the program to crash when attached with a debugger [8]. In `ContentFilter.exe`, we discovered that before surveillance text is encrypted, each sequence of six bytes of the text is used as the first six bytes of each eight-byte DES block to be encrypted. The remaining two bytes are pseudorandom, uniformly-distributed between 0x27 and 0x73, inclusive. DES encryption is performed on all blocks in ECB mode using the following 64-bit key ASCII-encoded:

$$\texttt{SURVEIL\_KEY}_{4.0} = \texttt{"X7sRUjL\0"}$$

which we also found to be used by TOM-Skype 4.0.4.226 and 4.2.4.106. Note that the 8th byte of $\texttt{SURVEIL\_KEY}_{4.0}$ is a null byte. Although we express this null byte for clarity, in TOM-Skype's implementation, this byte is the null-terminating byte of the string. This string, ASCII-encoded, also appears as a literal in the binary.

To discover the other DES key and circumvent the anti-debugging measures in TOM-Skype 3.8.4.44's `Skype.exe` executable, we used DLL injection, a technique where we cause TOM-Skype's calls to library functions to instead call code that we have written. We previously observed that, when stuffing each eight-byte DES block with the two random bytes, `ContentFilter.exe` reseeds the random number generator with a hardware time that it retrieves via a library call. We similarly found and then exploited this behavior in `Skype.exe` by causing each of these library calls for the time to call our code and sleep for ten seconds, allowing us to attach with a debugger while TOM-Skype slept. After attaching, we suspended all other threads not sleeping in our code. Then we observed the encryption process in the debugger before TOM-Skype's anti-debugging measures activated. We found the eight bytes of the DES key embedded in instructions in eight cases of a compiled switch statement. When ASCII-encoded, the following 64-bit DES key is used in ECB mode to encrypt surveillance text in TOM-Skype 3.6.4.316 and 3.8.4.44:

$$\texttt{SURVEIL\_KEY}_{3.6} = \texttt{"32bnx23l"}$$

After decrypting the surveillance text, we found that less information was reported in Skype 5.1.4.10 versus older versions. Here is example surveillance plaintext for a censored outgoing message for tested versions before 5.*x*:

```
jdoe falungong 4/24/2011 2:25:53 AM 0
```

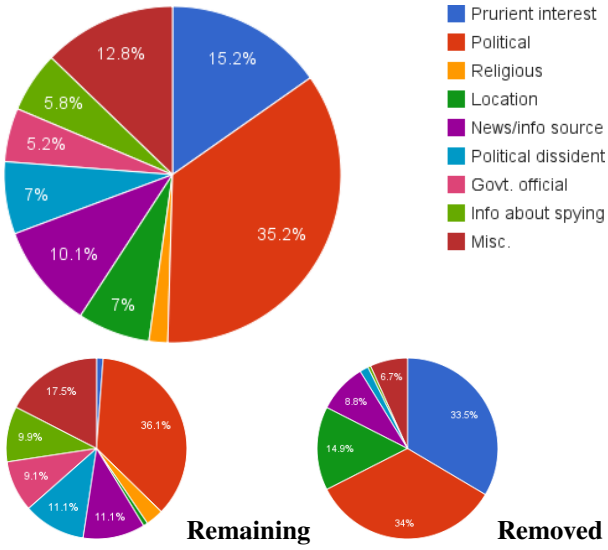

Figure 1: **Distribution of keywords on the keyword list that evokes both censorship and surveillance.**

Here "jdoe" is the sender of the outgoing message, "falungong" is the offending message in its entirety, followed by the date, time, and a "0" to indicate that the censored message was outgoing. When an incoming message is censored, then that message's sender is reported instead and the trailing "0" becomes a "1."

In contrast to versions before 5.*x*, here is example surveillance for version 5.1.4.10:

```
falungong 4/24/2011 2:29:57 AM 1
```

No username appears to be reported. Since all surveillance-related text in version 5.1.4.10 is incoming, a "1" will always trail the message.

3 Keyword analysis

We found that the built-in keyword lists in Tom-Skype 3.6.4.316, 3.8.4.44, 4.0.4.226, and 4.2.4.104 were identical and contained 108 censored words in either English or Chinese. Moreover, although they retrieve keyfiles from different URL's, as of 4/29/2011, each retrieved keyfile was identical and contained 442 words in either English or Chinese. These keyfiles have not been modified since 4/22/2011, and, if we believe their HTTP last-modified headers, they were last modified on 3/11/2011.

Since we began downloading Tom-Skype 5.0.4.14 and 5.1.4.10 keyfiles on 4/22/2011, we have noticed substantial changes to both the censorship keyfile and the 5.1-specific surveillance-only keyfile. We do not know the reason for the changes, but one possible reason is the human rights talks between China and the United States that were scheduled for 4/27/2011 and 4/28/2011 [15]. We focus on these lists and their changes in this section.

Before 4/22/2011, the keyword list that evokes both censorship and surviellance contained **Prurient interests**, *e.g.*, 两女一杯 (Two girls one cup), 二男一马 (Two men one horse), and 操烂 (Fuck rotten); **Political** terms, *e.g.*, 六四 (Liu Si, in reference to the Tiananmen Square protests that occurred on June 4th, 1989—this is literally the numbers "64"), 陆肆 (Lu Si, a homophonic way of writing 六四), and 河蟹社会 (River Crab Society, a corruption of "和谐社会", which means "harmonious society"); **Religious** terms, *e.g.*, 法轮 (Falun) and 观音法门 (Quan Yin Method, a Buddhist meditation method); **Locations** of planned events such as protests, *e.g.*, 广州 天河体育中心正门 (The main entrance of the Guangzhou Tianhe Sports Center in Guangzhou) and 杭州 湖滨路 凯悦酒店前至音乐喷泉旁一带 (Hyatt Regency Hubin Road, next to the area in front of the musical fountain in Hangzhou); **News/information sources**, *e.g.*, 维基百科 (Wikipedia) and 加拿大广播公司 (Canadian Broadcasting Corporation); **Political dissidents**, *e.g.*, 刘晓波 (Liu Xiaobo) and 吾尔开希 (Wu'er Kaixi, a student leader from the Tiananmen Square protests of 1989); **Government officials**, *e.g.*, 刘延东 (Liu Yandong, she is the highest ranking female in the communist party and a member of the politburo—she is caught up in a scandal involving her son-in-law) and 影帝温 (Oscar best actor winner, a nickname for Wen Jiabao after he appeared to cry insincerely on television); **Information about spying**, *e.g.*, 手机窃听软件免费下载 (Phone tapping software free download) and 三利普 (Three gain universal, part of a product name at sunlips.com, 三利普加强版二代橡皮, that appears to be a remote microphone for spying), and other **Miscellaneous** keywords. The contents of the original keyword list (before 4/22/2011) is shown in Figure 1. Figure 1 also shows the distribution of the words that were taken away (right) and that remained (left) on 4/22/2011.

Figure 2 shows the distribution of the 158 words for the list that evokes only surveillance. Most of this list is specific demolition sites or other references to the demolitions in Beijing, where people have reportedly been forced from their homes and their houses demolished to make room for future construction. The only words on this list that are not related to these demolitions are five keywords related to the Shouwang church, a Christian church in Beijing that illegally holds congregations outdoors in public places, and two references that appear to be names of companies or parts of a company name: 西屋国际 (Westinghouse International) and 大恒 (Daheng).

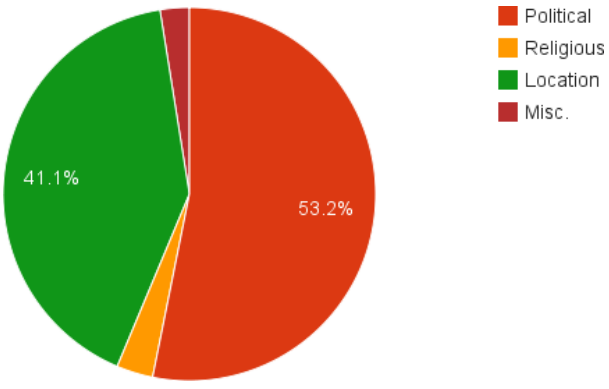

Figure 2: **Distribution of keywords on the keyword list that evokes only surveillance.**

Another interesting aspect of the Skype lists we analyzed, specifically the one that evokes both censorship and surveillance, is that there are many phrases that appear to be exact phrases taken from online instructions for protesters or calls for sit-ins. For example, one document suggests that protesters should take symbolic actions that are ambiguous so that they will not be arrested by the police, *e.g.*, 拿着麦克风表示自由 (Hold a microphone to indicate liberty—a passage where the document suggests that if people want to signify that they need liberty, they should put a picture of a microphone on their clothes or bag and gesture as if speaking into a microphone when they speak). Another document calls for a sit-in in response to the demolitions in Beijing, and another instructs people on how to make an origami jasmine flower and pleads protesters to not get arrested because this is an early phase of the protests.

4 Conjectures

In this section we present five conjectures that are based on the data we presented in this paper as well as previously available data on Internet censorship. Our aim is for these conjectures to be testable hypotheses so that the research community can confirm or refute them given the right data and appropriate computational and analytical techniques. All of the following conjectures are limited to content and traffic within a country where the censoring occurs. It is unlikely that all of these conjectures are true. However, we believe that each of these conjectures have the properties that: 1) they can in principle be empirically tested; 2) determining whether they are true or false will advance our understanding of Internet censorship. The five conjectures are:

1. **Effectiveness Conjecture:** "Censorship is effective, despite attempts to evade it." More formally, censoring a keyword reduces the number of accesses to content that either contains that keyword or contains related keywords. This may simply be because the quality-of-service for accessing content that is the target of censorship goes down whenever viewers or publishers must change their behavior in

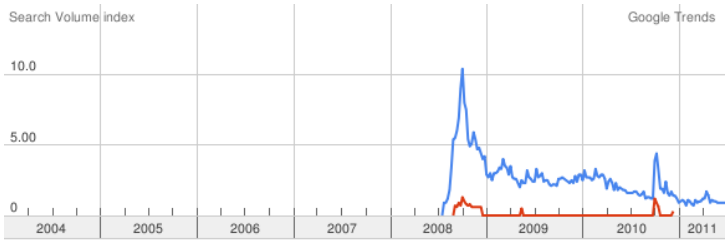

Figure 3: **Google Trends data for the Chinese-language searches for the "2 girls 1 cup" meme. The higher-volume data line is for 两女一杯 and the lower-volume data line is for 二女一杯.**

Figure 4: **Google Trends data for the English-language searches for the "2 girls 1 cup" meme.**

some specific way to access or disseminate the content.

The motivation for including this conjecture is that the censorship keywords we found in TOM-Skype that were phrases from specific documents were from documents that are not prevalent on the Web. These documents were presumably important enough that they were targeted by censorship, but there are very few instances of these documents online which suggests that they were not as widely disseminated as the authors of the documents had intended. For example, 拿着麦克风表示自由 (Hold a microphone to indicate liberty) is a phrase from instructions for Jasmine revolution protests in China in 2011. As of July 2011, searching for this exact phrase in quotation marks in the United States version of Google at www.google.com, which is known not to exclude results in response to China's Internet censorship, returns only nine results. The document appears in other places in addition to these nine results online, but with blacklisted phrases divided with dashes or paraphrased.

2. **Spread Skew Conjecture:** "Censored memes spread differently than uncensored memes." More formally, censoring a keyword qualitatively changes the time *vs.* number of accesses plot of the keyword. In particular, the distribution is not simply scaled downwards by a fixed amount, but may also be more or less spread out over time and have a different distribution.

   Because most of the censored keywords in our data do not show significant traffic volume in Google Trends, we cannot support this conjecture with our data. However, the conjecture was inspired by the Internet meme "2 girls 1 cup", which in Chinese is targeted by four keywords in our data (两女一杯, 二女一杯, 俩女一杯, and 两女吃一杯, only the first two of which have enough search volume to appear in Google Trends results). Figures 3 and 4 show the Google Trends results for Chinese and English for this meme, respectively. While the search volume of the Chinese-language versions of the meme is too small to make direct comparisons or extrapolate too much information about the distribution of the meme, the fact that the Chinese-language version of the meme has a lower peak and a taller tail is what inspired the spread skew conjecture. For each distribution, if we consider the 150 weeks after we first have data for that distribution, the English distribution has skewness 2.9511 but the Chinese distribution has skewness 2.0506. This may be a result of censorship effectively removing a portion of the right tail of the distribution over time. Note that Google Trends data is normalized in these graphs and that the English version of the meme has 32.0 times the total traffic volume as the most popular Chinese version.

3. **Interactions of Secrecy and Surveillance Conjecture:** "Keyword based censorship is more effective when the censored keywords are unknown and online activity is, or is believed to be, under constant

surveillance." More formally, for a given word that is perceived to be sensitive, accesses to the content related to that word will be fewer if it is unknown whether the word is censored or not. Further, accesses to related content will be fewer if Internet users believe that their online activities are being recorded and monitored.

This conjecture is inspired by the fact that the entity censoring Tom-Skype has made efforts to keep the list of censored and surveilled keywords and the surveillance traffic private.

4. **Peer-to-peer *vs.* Client-Server Conjecture:** "The types of keywords censored in peer-to-peer communications are fundamentally different than the types of keywords censored in client-server communications."

   For example, the censored keyword list for TOM-Skype, a peer-to-peer application, contains a higher fraction of proper nouns than censored keyword lists for client-server applications (see [4, 11] in addition to our list for examples of both types of lists). In particular, we noticed a high number of names of people and places on the censorship blacklist for TOM-Skype.

5. **Neologism Conjecture:** "Neologisms are an effective technique in evading keyword based censorship, but censors frequently learn of their existence." More formally, if a neologism is used in place of a censored keyword, the content will spread relatively freely until the neologism itself is censored. Phenomena such as "reblogging" and "retweeting" are impacted by this.

   We included this conjecture because of the large number of neologisms present in our data. Examples include 陆肆 (Lu Si, which sounds like 六四, or 64, in reference to the June 4th Tienanmen Square incident) and 影帝温 (Oscar best actor winner, a nickname for Wen Jiabao). Note that some keywords have a large number of possible neologisms, so that this conjecture may not be true for a large number of keywords. For example, we have seen instances in online Web forums of 六四, or 64, being referred to as "32 + 32" or "8 squared". Uncertainty about what keywords are being blacklisted and the possibility of surveillance are also factors in the effectiveness of neologisms, however.

5 Conclusion

In conclusion, we have presented new data about Internet censorship in China based on our efforts to reverse-engineer TOM-Skype and proposed five conjectures based on this data. For future work, our hope is that the research community will test these and other conjectures with more data and appropriate computational and analytic techniques.

Note: Complete lists with translations of the censorship and surveillance keywords for TOM-Skype are available at http://cs.unm.edu/~jeffk/tom-skype/.

Acknowledgments

We would like to thank the anonymous FOCI reviewers for their insightful comments. We would also like to thank the many people who helped us improve our translations and gave feedback on other aspects of the paper. This material is based upon work supported by the National Science Foundation under Grant Nos. CCR #0313160, CAREER #0644058, CAREER #0844880, and TC-M #090517.

References

[1] CHASE, M. S., AND MULVENON, J. C. *You've Got Dissent! Chinese Dissident Use of the Internet and Beijing's Counter-Strategies*. RAND Corporation, 2002.

[2] CLAYTON, R. Failures in a hybrid content blocking system. In *Privacy Enhancing Technologies* (2005), pp. 78–92.

[3] CLAYTON, R., MURDOCH, S. J., AND WATSON, R. N. M. Ignoring the great firewall of china. *I/S: A Journal of Law and Policy for the Information Society 3*, 2 (2007), 70–77.

[4] CRANDALL, J. R., ZINN, D., BYRD, M., BARR, E., AND EAST, R. ConceptDoppler: a weather tracker for Internet censorship. In *Proc. of 14th ACM Conference on Computer and Communications Security (CCS)* (2007).

[5] DANEZIS, G., AND ANDERSON, R. The economics of resisting censorship. *IEEE Security and Privacy 3*, 1 (2005), 45–50.

[6] DEIBERT, R. J., PALFREY, J. G., ROHOZINSKI, R., AND ZITTRAIN, J. Access denied: The practice and policy of global internet filtering. *The MIT Press* (2007).

[7] DORNSEIF, M. Government mandated blocking of foreign web content. In *Security, E-Learning, E-Services: Proceedings of the 17. DFN-Arbeitstagung über Kommunikationsnetze* (2003), J. von Knop, W. Haverkamp, and E. Jessen, Eds., Lecture Notes in Informatics, pp. 617–648.

[8] FABRICE, D., AND KORTCHINSKY, K. Vanilla skype part 1. Available at http://recon.cx/en/f/vskype-part1.pdf.

[9] FABRICE, D., AND KORTCHINSKY, K. Vanilla skype part 2. Available at http://recon.cx/en/f/vskype-part2.pdf.

[10] The Great Firewall Revealed. Whitepaper released by the Global Internet Freedom Consortium in December of 2002.

[11] "Race to the Bottom": Corporate Complicity in Chinese Internet Censorship. In *Human Rights Watch* (August 2006). http://www.hrw.org/reports/2006/china0806.

[12] LANDAUER, T. K., FOLTZ, P. W., AND LAHAM, D. Introduction to latent semantic analysis. *Discourse Processes 25* (1998), 259–284.

[13] LIANG, C. Red light, green light: has China achieved its goals through the 2000 Internet regulations? *Vanderbilt Journal of Transnational Law 345* (2001).

[14] MR. TAO. China: Journey to the heart of Internet censorship. Investigative report sponsored by Reporters Without Borders and Chinese Human Rights Defenders, Oct 2007.

[15] PERALTA, E. China, United States to Begin Human Rights Talks. Blog post 26 April 2011, URL: http://www.npr.org/blogs/thetwo-way/2011/04/26/135745130/china-united-states-to-begin-human-rights-talks, accessed 8 May 2011.

[16] VILLENEUVE, N. Breaching trust: An analysis of surveillance and security practices on China's TOM-Skype platform. Available at http://www.infowar-monitor.net/breachingtrust/.

[17] ZITTRAIN, J., AND EDELMAN, B. Internet filtering in China. *IEEE Internet Computing 7*, 2 (2003), 70–77.