JR01-2008

Information Warfare Monitor
ONI Asia

Joint Report

BREACHING TRUST:

An analysis of surveillance and security practices on China's TOM-Skype platform

Nart Villeneuve, Psiphon Fellow, the Citizen Lab

TheSecDevGroup

University of Toronto
Munk Centre
For International Studies
At Trinity College

http://www.infowar-monitor.net/breachingtrust/

October 1, 2008

Author Bio

Nart Villeneuve is the CTO of psiphon inc and the psiphon research fellow at the Citizen Lab, Munk Centre for International Studies, University of Toronto. He is a graduate of the University of Toronto and the former Director of Technical Research at the Citizen Lab where he analyzed the Internet filtering policies of over forty countries as part of the OpenNet Initiative (ONI). Nart is also a senior research associate at the Information Warfare Monitor. His research focuses on Internet censorship around the world as well as the evasion tactics used to bypass Internet filtering systems.

Projects

The Information Warfare Monitor is a joint project of The SecDev Group, and the Citizen Lab, at the Munk Centre for International Studies, University of Toronto.

http://www.infowar-monitor.net

ONI Asia is a sub-project of the OpenNet Initiative, focusing on censorship and surveillance in the Asia region, funded by IDRC Canada, and executed by the SecDev Group.

http://www.idrc.ca/panasia/ev-120961-201-1-DO_TOPIC.html

The OpenNet Initiative is a collaborative partnership of four leading academic institutions: the Citizen Lab at the Munk Centre for International Studies, University of Toronto, Berkman Center for Internet & Society at Harvard Law School, the Advanced Network Research Group at the Cambridge Security Programme, University of Cambridge, and the Oxford Internet Institute, Oxford University.

http://opennet.net

The Citizen Lab is an interdisciplinary research and development laboratory based at the Munk Centre for International Studies, University of Toronto, focusing on the Internet, global security, and human rights.

http://citizenlab.org

The SecDev Group is an operational "think tank" based in Ottawa, Canada focusing on emerging security issues including new media and information warfare.

http://secdev.com

Foreword

We are very pleased to introduce the first Information Warfare Monitor/ONI Asia joint report, ***Breaching Trust: An analysis of surveillance and security practices on China's TOM-Skype platform***, written by Nart Villeneuve, Psiphon Fellow, the Citizen Lab, Munk Centre for International Studies, the University of Toronto.

Surveillance is a practice often shrouded in secrecy. Although many people may be vaguely aware that governments and corporations regularly engage in surveillance (indeed, often in collusion) it is a practice that is difficult to identify and document directly. Not surprisingly, therefore, surveillance practices are often the subject of speculation and conspiracy theories. Our lives today are surrounded by mediated communications, serviced by third parties and private entities, sent through channels that pass through multiple political jurisdictions, each step of which offers an opportunity for surveillance. Can we rely on the assurances of the service providers and technology companies who tell us they are secure and private? Should we trust the assurances of a well-known global brand?

The findings unearthed and documented by Nart Villeneuve in *Breaching Trust* suggest that we cannot. Here we have a major software tool used to make telephone calls and send instant messages over the Internet, advertising secure end-to-end encryption, and widely touted by activists and dissidents as a safe way to communicate sensitive information, logging sensitive keywords and uploading entire transcripts of conversations to servers in China, which themselves are insecure. How insecure? Villeneuve was able to view, download, and archive millions of private communications, ranging from business transactions to political correspondence, along with their identifying personal information. Although some have mooted that Skype is equipped with a backdoor for intelligence, and that TOM-Skype in particular contained a Trojan Horse for the Chinese government, the company publicly denied these suspicions. Villeneuve's research definitively shows these denials are untrue. Although Villeneuve's trail runs cold at the doorstep of eight TOM-Skype servers in China, the underlying purpose of such widespread and systematic surveillance seems obvious. Dissidents and ordinary citizens are being systematically monitored and tracked.

While there have been other recent revelations of corporate complicity in China's censorship and surveillance regime – the Yahoo case involving Shi Tao and others comes to mind — the facts laid out in *Breaching Trust* are of such massive proportions that these other cases pale in comparison.

The lessons to be drawn from this case are numerous and issues of corporate social responsibility will be raised. If there was any doubt that your electronic communications – even secure chat – can leave a trace, *Breaching Trust* will put that case to rest. This is a wake up call to everyone who has ever put their (blind) faith in the assurances offered up by network intermediaries like Skype. Declarations and privacy policies are no substitute for the type of due diligence that the research put forth here represents.

Ron Deibert, Director, the Citizen Lab, Munk Centre for International Studies, University of Toronto.
Rafal Rohozinski, Principal, The SecDev Group, Ottawa, Canada.

Major Findings

- **The full text chat messages of TOM-Skype users, along with Skype users who have communicated with TOM-Skype users, are regularly scanned for sensitive keywords, and if present, the resulting data are uploaded and stored on servers in China.**
- **These text messages, along with millions of records containing personal information, are stored on insecure publicly-accessible web servers together with the encryption key required to decrypt the data.**
- **The captured messages contain specific keywords relating to sensitive political topics such as Taiwan independence, the Falun Gong, and political opposition to the Communist Party of China.**
- **Our analysis suggests that the surveillance is not solely keyword-driven. Many of the captured messages contain words that are too common for extensive logging, suggesting that there may be criteria, such as specific usernames, that determine whether messages are captured by the system.**

Summary

Our investigation reveals troubling security and privacy breaches affecting TOM-Skype—the Chinese version of the popular voice and text chat software Skype, marketed by the domestic Chinese company TOM Online. TOM-Skype routinely collects, logs and captures millions of records that include personal information and contact details for any text chat and/or voice calls placed to TOM-Skype users, including those from the Skype platform. These records are kept on publicly-accessible servers, along with the information required to decrypt these log files. These files contain the full text of chat messages sent and/or received by TOM-Skype users that contain particular *keywords* that trigger TOM-Skype's content-filtering capability.

Our investigation revealed eight servers that are part of the TOM-Skype surveillance network. In addition, we found one server hosting a special version of TOM-Skype designed for use in "net bars" or cybercafés. This server contained log files and information that revealed the list of the words that the system censored. Another server captured data from TOM Online's wireless services, and contained logs of SMS messages and other sensitive information.

The log files obtained during the course of the investigation reveal information such as the IP addresses, usernames (and land line phone numbers) used to place or receive TOM-Skype calls, as well as the full content of filtered messages and the time and date of each message. The collected data affects all TOM-Skype users and also captures the personal information of any Skype users that interacted with registered TOM-Skype users. This represents a severe security and privacy breach. It also raises troubling questions regarding how these practices are related to the Government of China's censorship and surveillance policies. The captured messages contain keywords relating to *sensitive* topics such as Taiwan independence, the Falun Gong, and political opposition to the Communist Party of China.

Security problems appear to be endemic at TOM Online. The publicly-accessible servers accessed by our investigation are insecure and contain information that can be used to exploit the TOM-Skype server network. It is possible that a malicious attacker could exploit vulnerabilities in the system and access the millions of logged communications and, possibly, detailed user profiles. In fact, evidence suggests that the servers used to store captured data have been compromised in the past and used to host pirated movies and *torrents* (for peer-to-peer file sharing).[1]

These findings raise key questions. To what extent do TOM Online and Skype cooperate with the Chinese government in monitoring the communications of activists and dissidents as well as ordinary citizens? On what legal basis is TOM-Skype capturing and logging this volume and detail of personal user data and communication, and who has access to it?

1 See screen captures in Appendix.

Background

Skype is a popular voice-over-IP software program that lets users make free peer-to-peer phone calls over the Internet. In 2004 Skype developed a relationship with TOM Online, a leading wireless provider in China, and announced a joint venture in 2005.[2] Skype and TOM Online produced a special version of the Skype software, known as TOM-Skype, for use in China. By 2006 it became clear that TOM-Skype was censoring the text chat feature of its software. It was reported that sensitive words, such as those referring to the banned Falun Gong movement, the Dalai Lama, and the Tiananmen Square incident triggered a filtering mechanism that prevented such messages from being displayed. Human rights groups criticized Skype, suggesting that the company was "legitimizing China's system of censorship",[3] while others suggested that TOM-Skype contained Trojan horse capabilities that could be used for surveillance by the Chinese Government.[4] Skype responded to the criticisms emphatically stating:[5]

- The text filter does not affect in any way the security and encryption mechanisms of Skype.
- Full end-to-end security is preserved and there is no compromise of people's privacy.
- Calls, chats and all other forms of communication on Skype continue to be encrypted and secure.
- There is absolutely no filtering on voice communications.

While Skype specifically stated that censored messages are "simply discarded and not displayed or transmitted anywhere" [6], this report demonstrates that not only are filtered messages transmitted to and stored on TOM-Skype servers located in China, but also that the servers themselves are configured with such poor security that it is possible to retrieve and decrypt these logs.

---

2 http://about.skype.com/2005/09/tom_online_skype_announce_join.html

3 http://www.hrw.org/reports/2006/china0806/5.htm#_Toc142395828

4 http://en.epochtimes.com/news/7-9-29/60228.html

5 http://share.skype.com/sites/en/2006/04/comments_about_skype_chat_text.html

6 http://share.skype.com/sites/en/2006/04/comments_about_skype_chat_text.html

Censorship: How Does it Work?

The TOM-Skype software from skype.tom.com contains a *keyfile* that appears to be an encrypted list of banned keywords as well as a filtering component created by TOM Online. When a TOM-Skype user sends or receives a text chat message that contains a keyword, the message is not displayed. However, when such a message is sent or received by a TOM-Skype user an HTTP connection is made to a TOM-Skype server and encrypted data is uploaded. These messages are stored in log files on the TOM-Skype servers.

Security and Privacy

TOM-Skype maintains eight servers to which data is uploaded from the TOM-Skype client software. Much of the interaction between the client and the servers focuses on product features, such as letting the user's contacts know when he or she is online, storing text messages that users send to contacts that are not currently online, and displaying advertisements to the user. However, there are a variety of different log files that contain stored personal information:

- **contentfilter*.log** - ip, username, message, date, time (+ unknown parameters)
- **skypecallinfo*.log** - ip, username, version, username/phone number, date, time (+ unknown parameters)
- **skypelogininfo*.log** – ip, version, username, date, time
- **skypenewuser*.log** – ip, version, username, date, time
- **skypenewusersendmoneytest*.log** - unable to decrypt
- **skypeonlineinfo*.log** – ip, username, version date, time (+ unknown parameters)
- **skypeversion.log** – version, ip, date, time (not encrypted)

These log files contain information such as the IP addresses and usernames as well as the date and time when the entry in the log was recorded. The most damaging information concerns the log files that record call information and the content filter logs that contain full text chat messages. The call information logs date from August 2007 and contain a record of the IP addresses and usernames of all those that participated in voice calls as well as the username and/or phone number of the recipient of the call. There are additional parameters that are logged whose function is unknown at this time. The content filter logs dating from August 2008 contain similar identifying information as well as the full content of the logged text messages. These messages contain sensitive information including email addresses, passwords, phone numbers, package tracking numbers and bank card numbers.

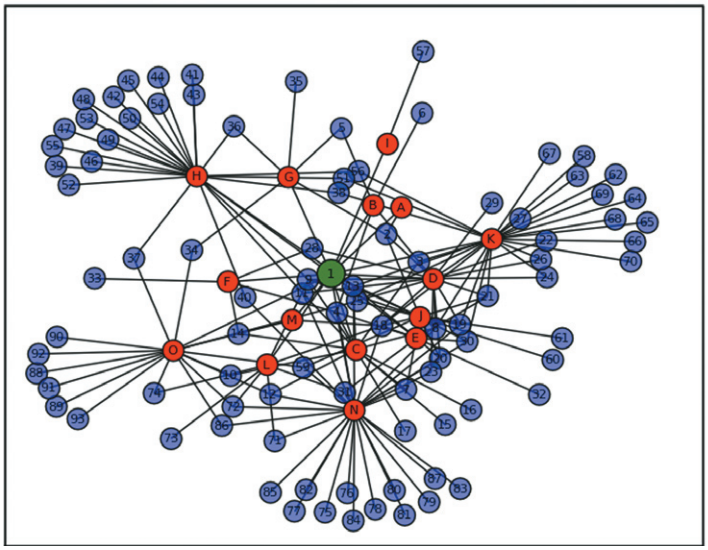
Censorship and Surveillance

We analyzed the content filter log files containing censored messages from August and September 2008. These log files, spread across eight servers, contained 1,045,800 messages. However, there is some duplication across the eight servers and users often send the same message several times after it is censored because it fails to be displayed.

The log files for the content filter contain the full content of censored messages along with the IP address of the TOM-Skype user, the username that sent the message and the date and time the message was sent. If a TOM-Skype user is sent a message by a "normal" Skype user that contains a banned keyword, that message is also logged and the Skype user's username is logged along with the IP address of the intended recipient.

It is possible to map the social network of each user that appears in the log files. We cannot prove that this data is actively being used for politically motivated surveillance but with simple social networking tools it is possible to identify the relationships between users. With just one username it is possible to identify all the users that have sent messages to or received messages from the original user. More complex mapping could track the relationships between all those users as well.

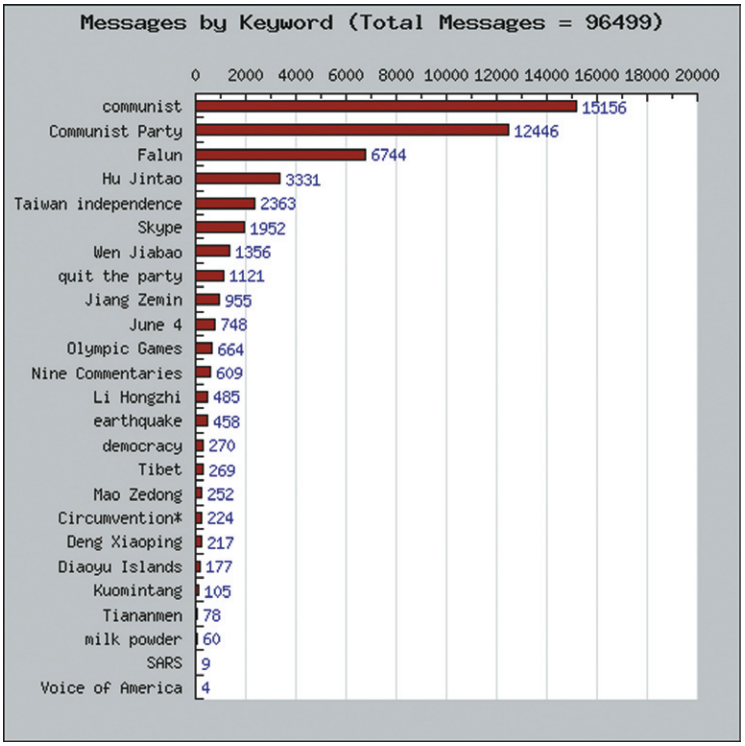
The example below shows the social network map of user "1" (the green circle) and is based on messages from the content filter logs. The red circles represent unique IP addresses that are associated with user "1" and the blue circles represent other user names that are associated with these IP addresses. All these "blue" users have had some form of connection with user "1" and all have had their messages logged by TOM-Skype. The form of the connection between each user is unclear. In some cases the "blue" users may represent user "1" employing a different username. In other cases it may represent users that received messages from or sent messages to user "1" in either a one-to-one conversation or in a public chat.



Messages were logged from IP addresses allocated to 59 different countries. However, 95% of all the logged messages were from IP addresses allocated to China; the U.S. followed with 0.78%. The reason that this percentage is so high is because when a user—a *normal* Skype user for example—sends a message to a TOM-Skype user that contains a keyword, the message is logged with the username of the sender but the IP address of the receiver. The non-Chinese IP addresses appear to represent TOM-Skype users in other countries.

In total, these log files contain 166,766 unique messages from 71,237 unique IP addresses and 44,254 unique usernames. The messages were processed to remove messages containing English language obscenities leaving 102,196 (61%) messages, mostly in Chinese. While some pornographic and obscenity-laden messages remain, the majority appear to be a combination of innocuous messages captured by overzealous filtering, and messages containing keywords relating to *sensitive* topics such as Taiwan independence, the Falun Gong, and political opposition to the Communist Party of China.

Of the 96,499 messages that were successfully translated with machine translation, 15,156 messages (15.71%) contain the word "communist", 6,744 contained "Falun" (6.99%) and 2,363 (2.45%) contained "Taiwan Independence."[7]



7 The keyword matching was done in English against machine-translated messages. These numbers serve as an indicator of the type of content. They do not reflect the content of the conversation (for or against the topic). Processing the messages using Chinese characters would provide a much more accurate data set.

However, when we pasted politically sensitive keywords from the log files into TOM-Skype text messages, we were unable to trigger the filtering or the uploading of the data to TOM-Skype servers. This is consistent with the findings of Human Rights Watch from 2006.[8] Despite Skype's claim that keywords concerning Taiwan, the Dalai Lama and Falung Gong were censored, Human Rights Watch was able to send and receive messages containing such keywords.

It is possible that the filtering of politically sensitive keywords is restricted in some way, perhaps geographically. However, it may also indicate that TOM-Skype is performing surveillance: logging messages containing keywords but continuing to display the messages to users. Based on an analysis of the messages in the content filter logs it is clear that messages containing such keywords are being delivered to users, as many of the messages appear to be questions and responses. A conversation is taking place and it is being logged.

For example, questions appear, such as:

你喜欢共产党?
(Do you like the Communist Party? [machine translation])

And responses that appear to be answers:

我不恨共产党也不喜欢国民党
(I do not hate the Communist Party do not like the KMT [machine translation])

Many of the politically sensitive messages logged by TOM-Skype make reference to the *The Epoch Times* and Falun Gong linked campaign that encourages Chinese citizens to quit the Chinese Communist Party (CCP). In 2005 the "Global Service Center for Quitting the CCP" was set up to encourage Chinese citizens to quit the CCP.[9] This campaign uses "[t]elephone, mobile phone text messages, chatting online" to contact Chinese citizens and provide them with information that is critical of the CCP and highlights the persecution of the Falun Gong.[10] This campaign has made extensive use of Skype including Skype "public chats" and SkypeCast.[11]

> Skype has become a popular communication tool among democracy activists in mainland China in recent years. Due to its excellent vocal clarity, fewer imposed restrictions, and an end-to-end encryption feature making it difficult to monitor, many Chinese democracy activists have favored Skype over traditional telephones and other similar communication tools.[12]

---

8 http://www.hrw.org/reports/2006/china0806/5.htm#_Toc142395828

9 http://en.epochtimes.com/news/8-1-4/63609.html

10 http://en.epochtimes.com/news/7-12-29/63377.html and http://en.epochtimes.com/news/8-7-3/72847.html

11 http://en.epochtimes.com/n2/china/skype-shutdown-chat-room-dissappointing-chinese-3788.html

12 http://en.epochtimes.com/news/7-9-29/60228.html, see also http://en.epochtimes.com/news/8-1-19/64311.html

While the campaign uses Skype extensively, they also warn users of the filtering capabilities of the Chinese version distributed by TOM-Skype.[13] However, many users contacted by members of the campaign are using TOM-Skype, resulting in numerous logged text chat messages on TOM-Skype servers in China. In fact, they are quite open about exactly who they are and what they are doing. Thus it is trivial to identify users associated with the campaign.

我是全球退党服务中心的义工
(I am a global service centre for volunteers to quit the party [machine translation])

Without further testing we are unable to conclusively determine that the objective is surveillance rather than filtering. However, we do know that regardless of the process the full messages are being logged and could be used for surveillance. Moreover, many of these messages contain words that are too common for extensive logging, suggesting that there may be criteria, such as usernames or who one has chatted with, or particular public chats, that determine how fine-grain the logging should be. For example, messages were logged that consisted entirely of "123", "谢谢" (Thank You [machine translation]), and just a text smiley face ":)." The fact that messages such as these are being logged suggests that at least some of the logging is focused on criteria other than just general keywords.

13 http://tuidang.net/news/contents.php?aID=2270

Conclusion

The questionable security practices of TOM-Online led to the disclosure of millions of records containing personal information regarding mobile phone accounts, SMS messages, and the usage of TOM-Skype. However, this disclosure also confirms that TOM-Skype is censoring and logging text chat messages that contain specific, sensitive *keywords* and may be engaged in more targeted surveillance. These logged messages contain keywords relating to sensitive topics such as Taiwan independence, the Falun Gong, and political opposition to the Communist Party of China. Many of these messages contain words that are too common for extensive logging, suggesting that there may be targeted surveillance taking place.

This case provides a unique perspective on the battle over information taking place between authoritarian governments and political activists through the medium of new technologies. While new technologies provide an innovative platform for political activists to communicate globally, they also provide governments with the ability to monitor and track political opponents and human rights advocates. The logging of these messages by TOM-Skype reveals the efforts to combat the use of Skype by political activists through censoring and logging messages that contain politically sensitive keywords. While the extent of targeted surveillance by TOM-Skype is not known for certain, many of the logged messages are specific to the "Quit the CCP" campaign. These messages reveal many of the methods used by the campaign including the promotion of circumvention technology and pseudonymous email addresses.

These findings should serve as a warning for groups engaging in political activism or promoting the use of censorship circumvention technology accessed through services provided by companies that have compromised on human rights. Private and politically sensitive messages sent through new communications technologies are only as secure as the robustness of the security of the technology companies themselves. In this case we were able to access volumes of sensitive data without the cooperation of the company involved due to lax security. There is no reason why an inquisitive government could not do the same.

Trust in a well-known brand such as Skype is an insufficient guarantee when it comes to censorship and surveillance. This case demonstrates the critical importance of the issues of transparency and accountability by providers of communications technologies. It highlights the risks of storing personally identifying and sensitive private information in jurisdictions where human rights and privacy are under threat. It also illustrates the need to assess the security, privacy and human rights impact of such a decision.

Consistent with most major companies that have acquiesced to China's censorship and surveillance policies, Skype is not transparent or forthcoming about the exact nature of their compliance with Chinese authorities. To what extent is Skype actively complicit in the censorship and surveillance of political discourse in China? What policies, if any, are in place to protect the privacy and human rights of Chinese Skype users? What is clear is that TOM-Skype is engaging in extensive surveillance with seemingly little regard for the security and privacy of Skype users. This is in direct contradiction of Skype's public statements regarding their policies in China.

The fight for the future of freedom of expression is underway and new technologies are the battleground upon which this information war will be waged. It is becoming increasingly clear that companies must be direct and transparent regarding their policies of censorship and surveillance. Failure in this regard leads to little if any internal oversight within the company and, possibly, to severe breaches of privacy and security that affect the global reputation of the company itself.[14] More importantly, it puts real people at risk and exacerbates the potential for human rights abuses.

14 For a detailed analysis of trust and reputation and the case of Yahoo!, see http://rconversation.blogs.com/YahooShiTaoLessons.pdf

Appendix

Tom-Skype v. 3.8.4.44
Download: http://download.skype.tom.com/Tom-SkypeSetup.exe

After downloading TOM-Skype, the Wireshark packet sniffer was used to monitor all the traffic generated by TOM-Skype during the installation and user registration process as well as while text chatting with both Skype and TOM-Skype users. An analysis of the packet dumps revealed that every time the keyword "fuck" was entered into the TOM-Skype Chat (or sent from a Skype user to a TOM-Skype user) a connection was made to a URL with encrypted data being uploaded.

Visiting the directory structure of the URL directly revealed the contents of each of the parent directories. We were able to find eight such servers. On each server, the log files were listed in the directories and were viewable. Browsing through the directory structure revealed the encryption key used to encrypt most of the log files, including the content filter logs. There were a variety of log files:

- **contentfilter*.log** - ip, username, message, date, time (+ unknown parameters)
- **skypecallinfo*.log** - ip, username, version, username/phone number, date, time (+ unknown parameters)
- **skypelogininfo*.log** – ip, version, username, date, time
- **skypenewuser*.log** – ip, version, username, date, time
- **skypenewusersendmoneytest*.log** - unable to decrypt
- **skypeonlineinfo*.log** – ip, username, version date, time (+ unknown parameters)
- **skypeversion.log** – version, ip, date, time (not encrypted)

These log files contain information such as the IP addresses and usernames as well as the date and time when the entry in the log was recorded. The most damaging information concerns the log files that record call information and the content filter logs that contain full text messages. The call information logs date from August 2007 and contain a record of the IP addresses and usernames of all those that initiated voice calls as well as the username and/or phone number of the recipient of the call. There are additional parameters that are logged whose function is unknown at this time.

The content filter logs dating from August 2008 contain similar identifying information as well as the full content of the logged text messages. The messages in the content filter log files from all eight servers were decrypted using the key available on the servers. The messages were then processed to remove duplicates and all unique usernames and IP addresses were extracted. We also mapped all the usernames associated with unique IP addresses as this indicates a network of users that have had some form of contact with one another. We were then able to extract all the messages of the user within this network.

We further processed the file to remove messages containing the string "fuck". The remaining messages, mostly in Chinese, were machine-translated. Queries were run against this set of messages to look for politically sensitive words such as "Communist Party" and "Taiwan Independence" as well as security-sensitive words such as "password", "bank", and "credit card." The logs were parsed and messages containing particular keywords were counted and flagged for analysis. The qualitative analysis of what is contained within these logs was conducted based on these queries:

communist
Communist Party
Falun
Hu Jintao
Taiwan independence
Skype
Wen Jiabao
quit the party
Jiang Zemin
June 4
Olympic Games
Nine Commentaries
Li Hongzhi
earthquake
democracy
Tibet
Mao Zedong
Circumvention*
Deng Xiaoping
Diaoyu Islands
Kuomintang
Tiananmen
milk powder
SARS
Voice of America

* (a group of the following keywords: free door, filter, anti-blockade browser, Garden Network)

Screen shots of Google search results showing that TOM-Skype servers had been previously used for file sharing and torrent hosting.