UNIVERSITY OF TORONTO

MUNK SCHOOL OF GLOBAL AFFAIRS

Join the Global Conversation

The Citizen Lab

Research Brief
November 2013

Asia Chats:

Analyzing Information Controls and Privacy in Asian Messaging Applications

INTRODUCTION

Across Asia, a new class of instant messaging (IM) mobile applications are rapidly growing in popularity and amassing enormous user bases. These applications encompass more than text, voice, and video chat as they offer social networking platforms that include expressive emoticons and stickers (known as "emoji"), photo and video sharing, e-commerce, gaming, and other features that provide a more sophisticated user experience than previous generations of IM clients.

Currently, the three most popular chat applications developed by companies based in Asia are WeChat (developed by Tencent Holdings Ltd., based in China), LINE (developed by LINE Corporation, based in Japan[1]), and KakaoTalk (developed by Kakao Corporation, based in South Korea). These applications are not just dominating their respective domestic marketplaces, but are also expanding into markets in countries across Asia and beyond the region.

The growth of these applications and their strategies to attract an international user base raise questions regarding the kind of pressures they face in specific jurisdictions, particularly on implementing censorship or surveillance features and providing user data, and how they will respond to these demands.
This post is an introduction to a series of research reports analyzing information controls and privacy in mobile messaging applications used in Asia. Our series of reports will include analysis based on our technical investigation of censorship and surveillance, assessment on the use and storage of user data, and comparison of the terms of service and privacy policies of the applications. This series will begin with a focus on WeChat, LINE, and KakaoTalk.

THE RISE OF ASIAN CHAT APPS

WeChat, LINE, and KakaoTalk are all under three years old, but have become extremely popular in the short period that has passed since their release. All three have user bases in excess of 100 million registered users, with each application adding millions of users per month.

KakaoTalk

KakaoTalk, founded by South Korea-based Kakao Corp., was released in March 2010 and now has 110 million users, with over 40 percent of those users based outside of South Korea. Kakao Corporation was founded by Kim Beom Soo who was formerly CEO of NHN Corporation. While its user population is impressive it is relatively low compared to the growth of WeChat and LINE, which have been faster to focus efforts on international expansion than KakaoTalk.

WeChat

Following its launch as Weixin (微信) in January 2011, Tencent's WeChat has seen its popularity soar, in part due to its integration with the existing 700 million-strong user base of Tencent's flagship portal QQ. WeChat is estimated to have over 600 million registered users, of which 271.9 million are active users of the application. This number is double the estimated registered user base of 300 million reported in January 2013, a sign of WeChat's staggering growth.

The application has also been highly popular outside its home market of China. In recent reports Tencent claims that WeChat has attracted over 100 million overseas users, drawing from South and Southeast Asian markets including India, Thailand, and Malaysia. WeChat has also launched in Australia, Europe, Latin America, US, and Africa. A Tencent representative noted "We see the mobile Internet industry as the biggest chance to take our business abroad" and "Tencent is ready to enter the world stage", as evident by their recruitment of international soccer superstar Lionel Messi as a spokesperson.

LINE

LINE, produced by LINE Corporation, a subsidiary of South Korea-based Naver Corporation, has similarly seen a rapid increase in its user base. Following its release in June 2011, the application now has 280 million registered users. Expanding outside of its original market of Japan, it has gained 18 million users in Thailand, 17 million in Taiwan, and 14 million in Indonesia. LINE launched a Chinese-branded version of the app, Lianwo (连我), in December 2012 in partnership with Qihoo 360 Technology Co., Ltd, a software company who develop a suite of products, including antivirus software, web browsers and a mobile applications store and management service. Through this arrangement Qihoo 360 has obtained a business license for LINE to operate in China with LINE providing a license to Qihoo 360 to promote and provide LINE services. The head of LINE's Chinese division has stated that the company aims to be the second biggest mobile IM company in China, behind Tencent.

LINE is publicly targeting a user base of 300 million users, and seeking to expand its reach outside of Asia to South America and Europe, including Spain where it already has 15 million users. LINE's CEO Akira Morikawa has stated he hopes LINE will "become a universal language" and "serve as a communication infrastructure for billions of global smartphones".

INFORMATION CONTROLS AND PRIVACY

Citizen Lab has previously examined censorship and surveillance in instant messaging clients used in China. In 2008, Citizen Lab researcher Nart Villeneuve documented surveillance and censorship of text messages in TOM-Skype (the version of Skype available in China) and found that millions of chat records were being collected and stored on a publicly accessible, unsecured server based in China. As a follow-up to this research and in collaboration with Professor Jedidiah Crandall and Jeffrey Knockel at the University of New Mexico, we reverse-engineered and analyzed the censorship and surveillance mechanisms of Sina UC, another Chinese

chat client, and TOM-Skype. We were able to monitor changes to keywords that trigger censorship and surveillance for more than a year and a half. While these results provide a unique window into how censorship and surveillance features are implemented in two chat clients used in China, the popularity of these programs pales in comparison to WeChat, LINE and KakaoTalk, and they are only focused on text chatting and voice / video (in the case of TOM-Skype), rather than providing a social networking platform.

More recently, some Asian governments have explicitly endorsed measures to gain greater control over these IM applications in an effort to monitor private conversations. For example, in August 2013, the Royal Thai Police's Technology Crime Suppression Division announced its plan to surveil LINE conversations to "ensure the rule of law, order and national security." This proposal came after the police opened an investigation into four people for allegedly posting rumors on Facebook of a possible military coup. Although the police said that they would focus "only on suspicious activity and not pry into private conversations", such moves raise further concerns over the privacy and safety of IM application users. Given that these applications are rapidly expanding their markets and growing their client base, an analysis of the extent of information controls that are present in these applications is needed.

WeChat

All Chinese Internet companies are held responsible for the content they host and are expected to establish entire departments devoted to surveillance and censorship of their platforms' users.
Tencent is one of China's largest companies with a market capitalization of approximately USD 100 billion. The company—like many private sector actors in China—has a very close relationship with the Communist Party of China (CPC). For example, Ma Huateng, chairman and CEO of Tencent, recently became deputy of the National People's Congress, the main legislative body of the state.

Dissidents and at-risk communities in China are uneasy regarding the increasing popularity of WeChat due to censorship and surveillance requirements that Internet companies are beholden to in the country, the dominance of the application in the domestic marketplace, and the close relationship between Tencent and the CPC. Chinese dissident Hu Jia expressed concerns that the domestic division of China's Public Security Bureau is actively surveilling WeChat based on suspicions that security officers were following his movements through WeChat. In addition, there were reports from his colleagues that messages transmitted through WeChat were recited back to them by authorities in full detail almost immediately after they took place.

WeChat has been named by Chinese law enforcement officials during recent crackdowns on "online rumors", who stated that WeChat is "also a public space on the Internet….if you spread rumors or false information that you haven't verified in a public space that is illegal".

The Tibetan community also has concerns regarding surveillance of WeChat communications, especially as Tibetans living in Tibetan areas of China are subject to intensive surveillance and persecution by Chinese authorities. Despite growing evidence of privacy and security issues, Lobsang Gyatso of Tibet Action Institute and a member of our Cyber Stewards Network writes that distressingly Tibetans living both within Tibetan areas inside China and in exile have adopted WeChat as a convenient way to communicate and share information. Recent reports indicate that surveillance of WeChat is likely taking place as part of repressive actions from Chinese authorities against the Tibetan community. For example, a Tibetan woman was arrested by Chinese authorities in October 2013, who alleged that she expressed anti-Chinese sentiments over WeChat and stored prohibited photos of His Holiness the Dalai Lama on her mobile phone.

WeChat's use is expanding in countries with tense diplomatic relations with China, such as Taiwan, India, and

Vietnam. Taiwanese opposition legislators expressed concerns that WeChat might threaten individual's privacy and the island's security if messages were spied upon. An unnamed Indian official said that the government was reviewing whether it should block access to WeChat over national security concerns. "But it is too premature to say whether this application will be banned," the official added. Vietnam has also mulled banning chat apps like WeChat and LINE.

Despite mounting concerns, WeChat surveillance mechanisms have not been verified technically. This lack of technical information is due to the fact that surveillance is typically impossible to measure, since it can be performed by the server and its effects are not visible outside that server. For blogs, microblogs, email, web searches, chat, and most other Internet applications, technically the server is most logical place to implement surveillance functions. The exception is peer-to-peer applications such as TOM-Skype. Based on our previous research on Chinese IM clients, only TOM-Skype and Sina UC were found to have surveillance and censorship features implemented on the client side. We suspect that server-side controls are the industry standard for Chinese IM programs.

Surveillance provides users with no indication of its existence and is difficult to empirically verify, but censorship can be more easily observed and measured. WeChat is required to block sensitive keywords from its service like any other Internet service operating in China. However, unlike the majority of Chinese Internet applications, WeChat is successfully growing its user base outside of the country. Controversy arose when reports in January 2013 indicated that users outside of China were experiencing keyword censorship on WeChat. The keywords 南方周末 (nan fang zhou mo) "Southern Weekend", which refers to an outspoken news outlet based in China, and 法轮功 (Falun Gong), which refers to the religious group, were found to trigger blocking notifications for users both within and outside China. Following media reports of this keyword censorship, Tencent responded that the blocking of users outside of China was a "technical glitch" and immediate actions to rectify it were underway.

LINE

Compared to other Asian IM applications, LINE is noted for its cultural sensitivity. Its emojis and stickers are customizable to cater to local customs and respond to the diverse nature of Asian societies. LINE in Indonesia, for instance, created stickers to celebrate the holy month of Ramadhan.

Concerns have been raised about the privacy and security of user communications mediated by LINE. As previously mentioned, law enforcement in Thailand has expressed their intent to monitor LINE communications to "safeguard order, security and morality of Thailand". But a response from LINE's representatives suggested that such monitoring would only be permitted with a Japanese court order, stating, "We have not received an official request from the Thai government. However, LINE's guidelines focus on user privacy and that's the company's priority".

In light of these security and privacy issues, others have drawn attention to the lack of encryption used in messages sent in LINE over 3G networks, as well as the potential for group chat history to be obtained by a party intercepting these messages. Citizen Lab researchers verified that LINE chat traffic is sent unencrypted over 3G networks on the latest version of the client. This behaviour is unusual given the fact that the client does encrypt chat traffic over wifi connections. Mobile 3G networks are encrypted by default but this encryption is implemented at the carrier level, which means Internet service providers (ISPs) and telecoms have the means to decrypt traffic if desired. Commentators have speculated that LINE may have intentionally left 3G unencrypted to allow authorities to access data at the ISP level and thereby provide easier access to jurisdictions with restrictive communications regimes.

KakaoTalk

KakaoTalk has received less attention than WeChat or LINE, but privacy issues have been raised following reports that the application collected the MAC address of users, presenting a risk that user's location information could be revealed. Kakao denied this claim and said that they use a hardware hash code as a means of preventing multiple downloads of the client. Kakao's privacy practices were called into question following the use of KakaoTalk user data in several high profile police investigations. Following public concerns about the company's retention of user data, Kakao announced that it would decrease the amount of time that messages are stored to as little as one day.

ASIA CHATS PART 1: INVESTIGATING REGIONALLY-BASED KEYWORD CENSORSHIP IN LINE

The first report in the Asia Chats series by Seth Hardy (Senior Security Analyst, Citizen Lab) examines the mechanisms behind keyword censorship in LINE for users based in China. On May 20, 2013, Twitter user @hirakujira (a Taiwanese iOS developer) reported the discovery of a list of 150 blocked words within the iPhone version of Lianwo (连我). This finding was prompted by a string in the program related to blocking capability. We explored if similar mechanisms were present in the latest versions of LINE available for Android. Reverse engineering of the most recent Android version of LINE (v3.9.3 downloaded directly from the Google Play store, current as of November 14, 2013) revealed that when the user's country is set to China during installation of the application it will enable censorship functionality by downloading a list of censored words from Naver's server, and then block the transmission of any messages that contain any of those keywords. The region data is encrypted in newer versions of LINE, likely due to users changing it in order to get free in-app downloadable content. The presence of censorship functionality has been confirmed as far back as v3.4.2, released on January 18 2013, using APK files found at AndroidDrawer. Lianwo was launched in December 2012, so given the data available to us it appears keyword censorship was enabled soon after the launch.

In our analysis, we were able to retrieve two lists from the server: Version 20 (223 keywords) and 21 (370 keywords), which suggests that there have been at least 21 iterations of the keyword block list.

Citizen Lab Research Fellow Jason Q. Ng has translated both the original keyword list discovered by @hirakujira, and the latest versions we extracted from Chinese to English and describes the context behind them. The first keyword list discovered by @hirakujira is described in a series of blog posts (full list available here) and the most recent keyword list uncovered by Citizen Lab and translated by Ng is available here.

The first keyword list from @hirakujira contains content related to domestic Chinese politics, human rights, and sensitive political events--many of which are rather obscure and only mentioned in media known for being critical of the CPC. For example, 浙江签单哥 (Zhejiang's receipt-signing Brother) refers to Zhejiang's Vice Minister of Propaganda Bao Hongjun (鲍洪俊), who was accused by netizens of charging over 54 million yuan in expenses to his public office and illegally embezzling hundreds of millions in other corrupt activities. Netizens posted images of his receipts, which contained his signature, thus meriting him the nickname of "Zhejiang's receipt-signing Brother." However, it is unclear whether or not this is merely a fabricated rumor or contains a kernel of truth since no reliable sources have corroborated the few unofficial user forums that mention this supposed scandal. Thus, it raises questions as to who ordered or decided that such lightly-reported incidents merited positions on LINE's list of keywords to be censored. The fact that some of these censored incidents are not high profile seems to indicate that they have been added by LINE as a pre-emptive,

preventative measure.

It is unclear which version number is the list that @hirakujira uncovered, but since it was collected in May 2013, it is obvious the list is an older version than v20 and v21. Lists v20 and v21 extracted by the Citizen Lab are identical to each other except for the addition of 147 keywords to v21 that are all related to disgraced former CPC politician Bo Xilai.. The previously revealed keyword list and v21 have 40 keywords in common. Though the v21 list also contains its fair share of topical keywords related to specific current events, for the most part the keywords are much more general and refer to more well-known incidents than those found on the list that @hirakujira uncovered. Besides keywords related to the Bo Xilai scandal, other major categories of keywords include those related to the June 4 crackdown on Tiananmen Square, infighting or factions within the Communist Party, Falun Gong, and various controversies like the fatal Ferrari car crash that involved the son of a party official close to President Hu Jintao and Wen Jiabao's secret wealth.

Our research results introduce a number of open questions. Content filtering and surveillance is mandated by the CPC for Internet companies active in the Chinese market. The presence of regionally-based keyword censorship and its compliance with government regulations indicate that LINE is serious about expanding its user base in China. Indeed, it aims to be the second most popular IM application in China after WeChat. The relationship between LINE Corporation's parent company Naver and Qihoo 360 reflects a regular practice of foreign companies partnering with domestic entities to provide services in the Chinese marketplace (as Skype did with TOM-Online, for example). Qihoo 360 appears serious about this partnership evidenced in part by mandating all employees to use LINE for work related discussions. It is likely that management of keyword censorship for users based in China is managed by Qihoo 360, but how this relationship functions in practice is unclear. Comparing the mechanism and targeted keywords for censorship in LINE with those practiced by Qihoo 360 and WeChat would be an interesting area for further research.

In our previous analysis of surveillance and censorship keyword lists in Sina UC and TOM-Skype, we found that in our total dataset of 4,256 unique keywords, there were only 138 terms (3.2 percent) shared between the two programs. Similarly, between the LINE keyword lists for v20 (223 keywords) and v21 (370 keywords), there are only 27 exact matches with keywords in the TOM-Skype and Sina UC dataset. However, many of the topics are similar (e.g. CPC officials, June 4, Falun Gong). The list extracted by @hirakujira has only three identical keywords in common with the dataset. They are the following: 北京频传江病危 (Běijīng pín chuán jiāng bìngwēi) "Frequent reports from Beijing that Jiang critically ill", a reference to rumors of Jiang Zemin being severely sick; 网络封锁 (wǎngluò fēngsuǒ) "Internet blocked"; and 八九民主 (bājiǔ mínzhǔ) "Eighty-nine democracy", a reference to June 4, 1989. The lack of matches between this list and the TOM-Skype and Sina UC lists make sense given the obscurity of the majority of the content. The keywords that do match are more general and reflect typical topics targeted for censorship by domestic Chinese products. This lack of overlap suggests that no common keyword list was provided to these companies by government authorities. Previous research on censorship in Chinese blog services and search engines localized for the Chinese market have found similar inconsistency between products. Overall, these findings further reinforce the validity of the speculation that Chinese companies may be given general guidelines from authorities on what types of content to target, but have some degree of flexibility on how to implement these directives.

In the case of LINE, we see a company actively trying to attract a diverse user base not only through culturally-relevant customizations (e.g. country-specific emojis and stickers), but also complying with government mandated information control regulations (e.g. in China). In addition, it is encountering pressures to provide user data to law enforcement (e.g. in Thailand) or risking being banned entirely (e.g. Vietnam). Conversely with WeChat, Tencent is encountering criticism for information controls within its domestic market and facing distrust in countries with poor relations with China. These challenges demonstrate the difficulty of operating in multiple jurisdictions and the unique characteristics of Asian markets with restrictive

communications environments.

As Asian IM applications continue to grow in popularity at staggering rates, so too will governmental pressures to enact information controls and provide user data. The Citizen Lab will continue to analyze the rise and implications of these applications in the Asia Chats series.

Our research outputs on regionally-based censorship in LINE include:

- Detailed technical report by Seth Hardy
- Keyword list translated from Chinese to English with contextual descriptions by Jason Q. Ng
- Blog series by Jason Q. Ng on context behind the blocked keywords
- LINE Region Code Encrypter Tool developed by Seth Hardy and Greg Wiseman (Senior Data Visualization Developer, Citizen Lab) for changing regions in the LINE client to disable regionally-based keyword censorship in the application.

ASIA CHATS RESEARCH TEAM

Contextual, Legal and Policy Research: Masashi Crete-Nishihata, Andrew Hilts, Irene Poetranto, Jason Q. Ng, Adam Senft, Aim Sinpeng.

Technical Research: Jakub Dalek, Seth Hardy, Katie Kleemola, Byron Sonne, Greg Wiseman.

This project is financially supported by the John D. and Catherine T. MacArthur Foundation

FOOTNOTES

1. Originally the application was developed by NHN Japan the Japanese arm of Naver Corporation (formerly NHN) based in South Korea. Following the success of the application Line Corporation was formed as a subsidiary of Naver.

MEDIA COVERAGE

Media coverage includes TIME Magazine, TechPresident.