**The Citizen Lab**

## *Asia Chats: LINE Censored Keywords Update*

This report is the third in a [series](#) which analyzes regionally-based keyword censorship in LINE, a mobile messaging application developed by LINE Corporation (a subsidiary of South Korean Naver Corporation) based in Japan. We document recent changes to the list of keywords used by LINE to trigger regionally-based keyword filtering for users with accounts registered to Chinese phone numbers.

## PREVIOUS KEYWORD LIST CHANGES

In November 2013, [we reported](#) on results from reverse engineering LINE in which we reveal that when the application is registered to a Chinese phone number, censorship functionality is enabled.

The code analysis in our first report was performed on LINE v3.8.5 for Android and the keyword blocking behaviour was confirmed on an Android device running v3.9.3 downloaded directly from the Google Play store. We confirmed the presence of censorship functionality going back to v3.4.2, released on January 18 2013.

The LINE keyword lists were modified in v3.9.4 released on November 18 2013. In our original analysis we found LINE has an [internal keyword list](#) built in to the APK. If the user's registered phone number is set to a Chinese number the application will download an additional keyword list from Naver's server and block transmission of any messages that contain any of those keywords. The downloaded keyword file is stored in the application's cache directory as cbw.dat. If this list is unavailable, LINE will default to using an internal list of 50 keywords. In LINE v3.9.4 the internal list was [removed](#) from the application.

The [content of the list](#) includes keywords that relate 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 Communist Party of China (CPC). A number of these keywords relate to lightly reported incidents that did not go viral, which raises questions as to why they were included. 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 or could potentially have been intended for testing and not production use. Thus, the internal list may have been removed because these keywords no longer merit inclusion.

# KEYWORD LIST V22 ANALYSIS

On April 8 2014, the keyword list that the application retrieves from Naver servers was updated from v21 to v22. This change is current as of LINE v4.3.0 released on April 26 2014. As with the previous version, list v22 is Base64 encoded and encrypted using AES in cipher block chaining mode with PCKS#7 padding. Decryption is done through a static key stored in the binary that remains the same as the previous list version.

We translated each keyword from Chinese to English and assigned them content categories using a set of categories we developed to analyze keywords used to trigger both keyword filtering and surveillance in TOM-Skype and keyword filtering only in Sina UC.

List v22 contains 535 keywords in total. Comparing list v21 and v22 reveals that 312 new keywords have been added and 147 keywords have been deleted. The keywords are almost entirely in Chinese script; only 7 of 535 keywords do not contain Chinese characters. Some are combinations of scripts, such as '天安门1989' ('Tiananmen 1989'), a reference to the 1989 Tiananmen Square massacre.

All of the 147 removed keywords relate to the Bo Xilai scandal, which involved a prominent Chinese politician being jailed for corruption while his wife was convicted of murder. In previous research on the microblogging platform Sina Weibo we found that the keyword '薄熙来' ('Bo Xilai') was blocked and unblocked in patterns that appear to be correlated with authorities filtering his name when online conversations got too unpredictable to control and unblocking it when Bo fell out of favor with the CPC. For example, following the official expulsion of Bo Xilai from the CPC in September 2012, his name was unblocked on Sina Weibo, which possibly reflects authorities easing censorship requirements around the scandal to provide netizens a space to discuss and criticize the disgraced leader. The removal of 147 keywords from LINE related to Bo Xilai may also be the result of directives allowing discussion of Bo. However, 10 new keywords relating to Bo Xilai were added to list v22, and 2 of the Bo Xilai keywords from v21 were not removed. The 12 remaining Bo Xilai keywords on list v22 do not appear to be qualitatively different from the 147 deleted keywords and it is unclear why they are retained while the others are removed.

The majority of the 312 new additions to the keyword list relate to Chinese government officials or notable political events. These keywords include references to government officials (22.7% of the total new keywords), criticism of the CPC (8.9%), references to the June 4, 1989 Tiananmen Square massacre (7.6%), references to the relatives of political figures (8.3%) and references to political scandals (5.1%). References to Tiananmen Square previously accounted for 15% of the keywords on list v21, second only to the Bo Xilai scandal. After these 5 categories, the next most common categories of keywords added to list v22 were those relating to the CPC generally (4.8%) and content relating to dissidents/activists (4.1%).

See Figure 1 for a breakdown of the content categories of the new keywords added to list v22:
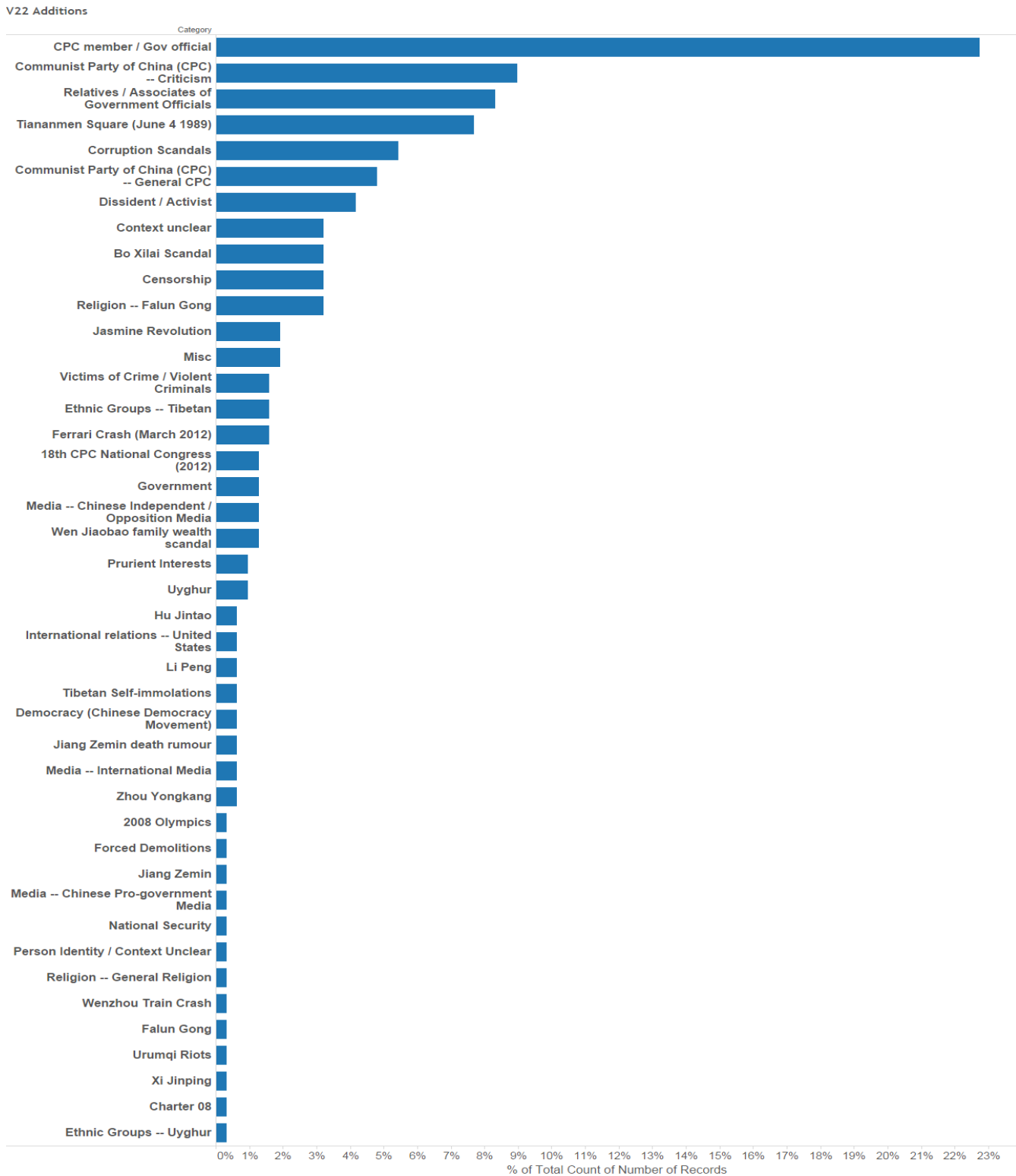
V22 Additions



**Figure 1: Categories of new keywords added to LINE list v22.**

The complete list v22 shows that keywords relating to Tiananmen Square (15%) make up the largest single category.

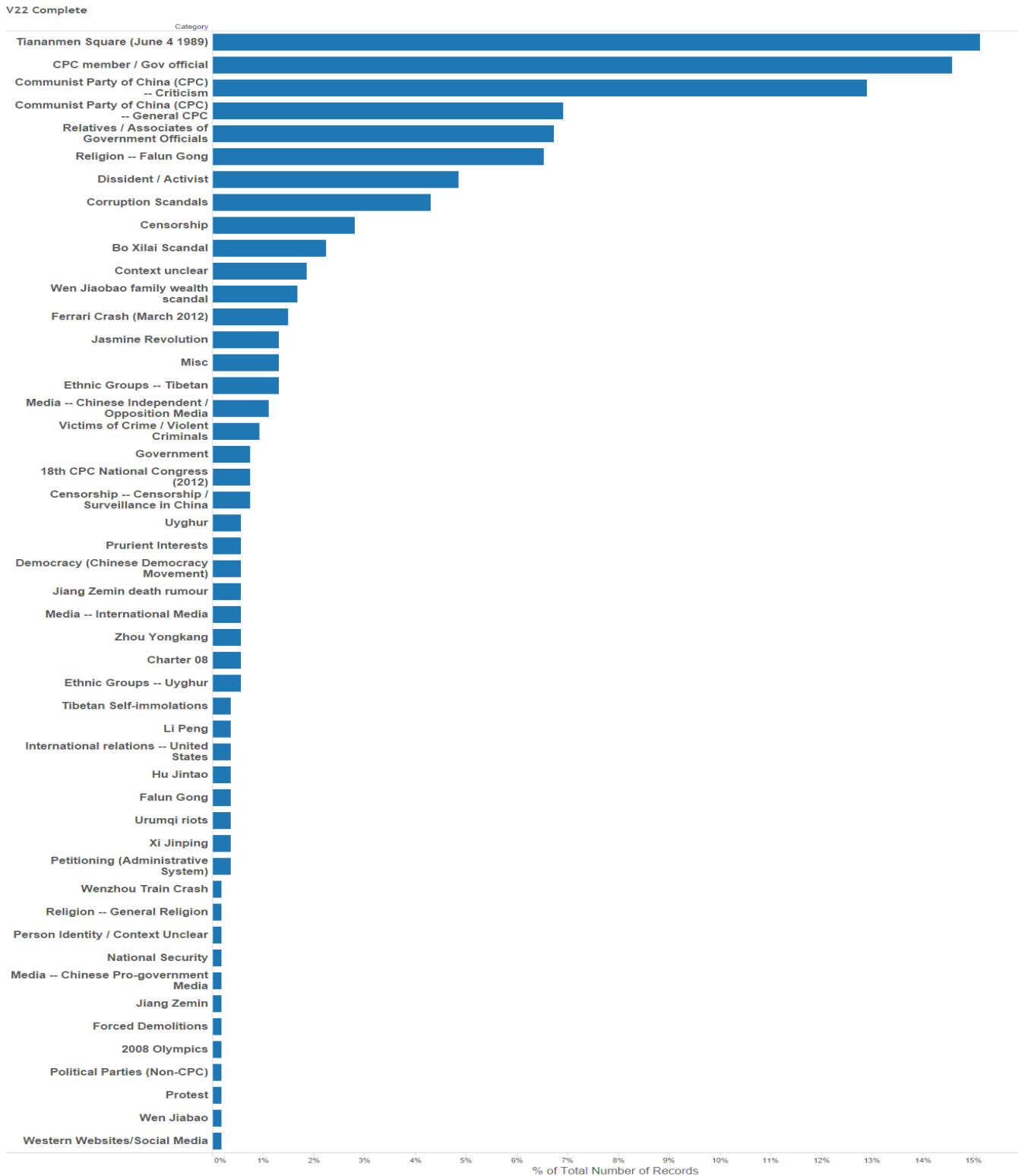See Figure 2 for a breakdown of the all the categories in list v22.

**V22 Complete**

**Figure 2: Categories of all keywords on LINE list v22.**


## COMPARING LINE KEYWORD LISTS TO TOM-SKYPE AND SINA UC

Our dataset on TOM-Skype and Sina UC comprises 88 separate keyword lists, which combined contain 4,256 unique keywords. Comparison of TOM-Skype and Sina UC keyword lists revealed that of the 4,256 unique keywords, only 138 terms (3.2%) were shared in common between two clients.

Of the 535 total keywords on LINE list v22, 45 are an identical match on TOM-Skype lists, 19 on Sina UC lists and 34 on the lists of both clients for a total of 98 (18%) keywords on list v22 matching the China Chats dataset.

Compared to the 370 total keywords of LINE list v21, 8 are an identical match on TOM-Skype lists, 10 on Sina UC lists and 9 on lists of both clients for a total of 27 (0.7%) keywords on list v21 matching the China Chats dataset.

The top categories for keywords which appear on both the v22 list and either of the TOM-Skype/Sina UC lists are content relating to CPC members/government officials (22% of these 98 keywords), content relating to the Tiananmen Square massacre (12%), content relating to dissidents/activists (12%) and keywords related to the Falun Gong (10%).

We observe a similar lack of overlap between the LINE, TOM-Skype, and Sina UC lists. The inconsistencies between the lists used for the three clients suggests that no common keyword list is provided to companies operating chat programs in the Chinese market.

## CONCLUSION

It is unclear how the content of LINE keyword lists are determined. LINE previously had a partnership with Chinese software company Qihoo 360 Technology Co., Ltd to distribute a Chinese branded version of the application called Lianwo (连我). This week LINE announced that it changed domestic Chinese partners from Qihoo 360 to Beijing Zhuoyi Xunchang Technology Co. Ltd also know as "wandojia" (豌豆荚). In the official announcement of this new partnership its noted that amongst its responsbilities Wandojia will provide technical support to LINE operations in China. How this new relationship affects the implementation and content of keyword filtering for Chinese users is uncertain.

Following our first report we sent LINE Corporation a letter asking a number of questions including a request for clarification of the relationship between the LINE and Qihoo 360 and information on the process for determining the content of keyword lists. We received a terse reply:

"LINE had to conform to local regulations during its expansion into mainland China, and as a result the Chinese version of LINE, 'LIANWO,' was developed. The details of the system are kept private, and there are no plans to release them to the public".

Despite the lack of information provided by LINE Corporation around its operations in China, it is clearly maintaining keyword filtering features for users in the country. Previous work on the censorship practices of chat clients, blog services, and search engines in China reveal inconsistencies in the specific keywords and content that are targeted for blocking, but general similarities in content categories. These differences suggest

that companies may be given general guidelines from government authorities on what types of content to target but have some degrees of flexibility on how to implement these directives. The LINE keyword lists appears to fit these findings, but the process of developing and implementing content filtering policies and the interactions between LINE Corporation, its domestic partners, and Chinese authorities remain unknown.

## RESOURCES

- [Raw and translated LINE keyword list data on Github](#)

- [LINE Region Code Encrypter Tool](#)  for changing regions in the LINE client to disable regionally-based keyword censorship in the application