Jeffrey Knockel
Candidate

Computer Science
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Jedidiah Crandall , Chairperson

Ronald Deibert

Stephanie Forrest

Jared Saia

Measuring Decentralization of Chinese Censorship in Three Industry Segments

by

Jeffrey Knockel

B.A., Philosophy, University of New Mexico, 2008

B.S., Computer Science, University of New Mexico, 2009

M.S., Computer Science, University of New Mexico, 2011

DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Computer Science

The University of New Mexico

Albuquerque, New Mexico

May 2018

Acknowledgments

I would like to thank my advisors, Professor Jedidiah R. Crandall and Professor Jared Saia for supporting my research for so many years and for their invaluable insight. I would also like to thank Professor Ronald Deibert and Professor Stephanie Forrest for serving on my dissertation committee and for providing valuable feedback on this dissertation.

I wish to extend many thanks to all of my friends and colleagues at the Citizen Lab. I would particularly like to thank Professor Ronald Deibert for supporting my research, and (in alphabetical order) collaborators Masashi Crete-Nishihata, Sarah McKune, and Adam Senft for contributing Section 3.1 of my dissertation; Masashi Crete-Nishihata, Adam Senft, and Diana Tseng for assisting in translating and categorizing keywords in Chapter 3 and Greg Wiseman for generating figures; Masashi Crete-Nishihata for contributing Section 4.1; Masashi Crete-Nishihata, Jason Q. Ng, and Lotus Ruan for assisting in translating and categorizing keywords in Chapter 4; Masashi Crete-Nishihata and Lotus Ruan for contributing Section 5.1; and Lotus Ruan for her translation and categorization of keywords in Chapter 5. Without their contributions, this cross-disciplinary work would not have been possible.

I would like to thank Professor David Ackley for teaching me how cool it could be to be a computer scientist and for inspiring me to pursue computer science professionally.

Finally, I would like to thank my parents for instilling in me the value of science and education.

Measuring Decentralization of Chinese Censorship in Three Industry Segments

by

Jeffrey Knockel

B.A., Philosophy, University of New Mexico, 2008

B.S., Computer Science, University of New Mexico, 2009

M.S., Computer Science, University of New Mexico, 2011

PhD, Computer Science, University of New Mexico, 2018

Abstract

What is forbidden to talk about using Chinese apps? Companies operating in China face a complex array of regulations and are liable for content voiced using their platforms. Previous work studying Chinese censorship uses (1) sample testing or (2) measures content deletion; however, these techniques produce an incomplete picture biased toward (1) the tested samples or (2) whichever topics were trending.

In this dissertation, I use reverse engineering to study the code that applications use to determine whether to censor content. In doing so, I can provide a more complete and unbiased view of Chinese Internet censorship. I reverse engineer applications across three Chinese industry segments: instant messaging, live streaming, and gaming. Together this reveals over 100,000 unique blacklisted keywords from blacklists spanning hundreds of different companies.

A common assumption in Chinese censorship research is that observed censorship is the result of a monolithic motive; however, in this dissertation, where I provide a more complete and unbiased view of Chinese Internet censorship, I will test three hypotheses: (1) there is little overlap between the keyword lists used by different companies, (2) there is no China-wide list of banned words or topics largely determining what Chinese companies censor, and (3) provincial-wide lists of banned words or topics do not largely determine what companies censor. These hypotheses suggest that it is largely Chinese companies that are burdened with choosing what topics to censor.

Contents

List of Figures **x**

List of Tables **xii**

1 Introduction **1**

1.1 Dissertation Overview 3

2 Related Work **5**

3 Censorship in Instant Messaging Applications **10**

3.1 Background 11

3.1.1 TOM-Skype and Sina UC 11

3.1.2 Chinese Regulatory Environment of Instant Messaging Apps . 13

3.2 Methodology 16

3.2.1 Sina UC Censorship 17

3.2.2 TOM-Skype Censorship 17

3.2.3 TOM-Skype Surveillance 19

3.3 Technical Analysis 20

3.4 Keyword Analysis 24

3.4.1 Political 25

3.4.2 People 26

3.4.3 Events 27

3.4.4 Social 27

3.4.5 Technology 28

3.4.6 Targeted and Broad Keywords 30

3.4.7 Adaptation to Censorship Evasion 30

3.4.8 Keyword List Changes 31

3.4.9 Changes in Response to Events 36

3.5 Conclusion 50

4 Censorship in Live Streaming Platforms 52

4.1 Background 53

4.1.1 Live Streaming Platforms 53

4.1.2 Legal and Regulatory Environment in China 55

4.1.3 Content Monitoring and Censorship on Live Streaming Platforms 56

4.2 Technical Analysis 57

4.2.1 YY Censorship and Surveillance 58

4.2.2 Sina Show Censorship 60

4.2.3 9158 Censorship 62

4.2.4 GuaGua Censorship 63

4.3 Keyword Analysis 63

4.3.1 Similarity Comparison Across Keyword Lists 64

4.3.2 Keyword Content Analysis 66

4.3.3 Keyword List Changes 68

4.3.4 Content Analysis of Keyword Additions 70

4.4 Conclusion 81

5 Censorship in Mobile Games **83**

5.1 Background 84

5.2 Methodology 86

5.2.1 Analyzing Highly Downloaded Games 86

5.2.2 Analyzing Games from Popular Publishers and Developers . . 90

5.3 Results 92

5.3.1 Results from Analyzing Highly Downloaded Games 92

5.3.2 Results from Analyzing Games from Popular Publishers and Developers 94

5.4 Keyword Analysis 97

5.4.1 Keyword List Provenance 98

5.4.2 Content Analysis 99

5.5 Summary 102

6 Conclusion and Future Work **103**

6.1 Future work 106

References **109**

List of Figures

3.1 Algorithm for decrypting TOM-Skype 3.6–3.8 keyfiles 18

3.2 Themes by client 25

3.3 Political categories by client 26

3.4 People categories by client 27

3.5 Event categories by client 28

3.6 Social categories by client 29

3.7 Technology categories by client 29

3.8 The number of unique keywords in TOM-Skype lists over time . . . 33

3.9 The number of unique keywords in Sina UC lists over time 34

4.1 Keyword lists clustered by Jaccard similarity 64

4.2 Keyword lists clustered by *Intersection over Smallest* 65

4.3 Breakdown of list source by theme 67

4.4 Distribution of keyword updates between May 18 2015 and September 30 2016 (Universal Standard Time) 68

4.5 Distribution of themes across three live streaming platforms 70

4.6 Percentage of Social theme keywords by category 71

4.7 Percentage of Event theme keywords by category 72

4.8 Percentage of Political theme keywords by category 78

4.9 Percentage of Technology theme keywords by category 79

4.10 Percentage of People theme keywords by category 80

5.1 Cosine similarity of blacklists from highly downloaded games hierarchically clustered according to the centroid linkage method 93

5.2 Cosine similarity of blacklists from games from top publishers and developers hierarchically clustered according to the centroid linkage method 96

5.3 For both datasets a histogram showing the number of keyword lists versus the number of keywords in that list 97

5.4 Breakdown by theme of 7,000 randomly sampled keywords (±1.1 error at 95% confidence) 100

List of Tables

3.1 Lists used by TOM-Skype and Sina UC and their HTTP URLs . . . 20

3.2 Lists used by TOM-Skype and Sina UC and their functions 21

3.3 List of cryptographic algorithms and keys used by the TOM-Skype and Sina UC clients 22

3.4 List of cryptographic algorithms and keys used for surveillance by TOM-Skype clients 22

3.5 Surveillance message triggered when JaneDoe receives “fuck you” from JohnDoe 23

3.6 Breakdown of character types in the keyword lists 24

3.7 Jaccard similarity between first and last lists 37

4.1 Live streaming users by platform 54

4.2 Keyword list size (May 17, 2015) 63

4.3 Content themes and related categories 66

4.4 Keywords added by platform 68

4.5 Additions to keyword lists compared by Jaccard similarity and by *Intersection over Smallest* 69

4.6 Keyword additions to Event theme in three live streaming platforms compared by Jaccard similarity and by *Intersection over Smallest* . 73

5.1 The number of games downloaded and lists found for each publisher 95

5.2 The number of games downloaded and lists found for each developer 95

5.3 Content themes and related categories 99

Chapter 1

Introduction

One way to measure which topics are forbidden on a communication platform is for a researcher to use *sample testing*. In this method a researcher attempts to communicate using the platform and then measures which communication is filtered. When communication such as posts are publicly available, another way to measure which topics are forbidden is to measure which posts are deleted that were previously seen available. While these methods provide insight into censorship and its effects on online communication, they are limited in answering the question of which topics are forbidden because neither method provides an unbiased view of the topics the censor wishes to suppress.

If researchers send messages over a platform to measure which are censored, then their results will be biased toward which messages they test, which may reflect the suspicions of the researchers. Moreover, measuring which posts are deleted is biased toward whatever topics users are currently discussing, which may give the illusion that popular topics are more heavily censored and that a forbidden topic that is not discussed during the measurement period is not forbidden at all. If we wish to infer the motives of the censor and understand the censor's intentions, then we require a

complete, unbiased list of any forbidden topics or blacklisted words.

In this work, I directly analyze, through reverse engineering, the rules for how chat software is programmed to perform censorship. This approach provides a more complete view of which topics and keywords the producers of this software intend to suppress.

A dominant academic theory [89, 90] posits that the motive of Chinese censorship is to suppress collective action. The implicit assumption of this theory is that there is a monolithic motive to censorship in China exerted by the Chinese government.

In this dissertation, I will analyze over 100,000 blacklisted keywords from blacklists spanning hundreds of different companies. These lists are used to trigger censorship across three different Chinese industry segments: instant messaging, live streaming, and games. These industry segments were chosen because they have downloadable applications that are amenable to my reverse engineering methodology and because these industry segments are popular, each having hundreds of millions of users in China. Each operates under a unique Chinese regulatory environment, and together they provide a representative view into Chinese censorship in general.

This more complete and unbiased dataset allows me to test the following three hypotheses: (1) **there is little overlap between the keyword lists used by different companies**, (2) **there is no China-wide list of banned words or topics largely determining what Chinese companies censor**, and (3) **provincial-wide lists of banned words or topics do not largely determine what companies censor**. These hypotheses suggest that censorship in China is decentralized and that the responsibility to choose what is censored is largely pushed down to individual companies and employees.

It is important for researchers to have a correct understanding of the nature of Internet censorship in China. In December 2016, China had over 730 million Inter-

net users [57]. With the largest population of Internet users in the world, China in cooperation with private companies operating in the country maintains the world's largest ecosystem of censorship. By understanding the nature of Internet censorship in China, users in the country may make better informed decisions because if, for example, users are aware that individual companies have a large amount of freedom in choosing what to censor, then they may attempt to evade censorship on one platform by switching to another. Moreover, understanding Internet censorship in the world's largest country provides an important data point for understanding the future of Internet censorship in other countries. Many countries besides China are increasingly exerting their "Internet sovereignty," *i.e.*, the right of a country to exert boundaries and information controls on the Internet. Since 2015, Russia has collaborated with Chinese officials to design and implement their own country-wide information controls [122], and governments in India, Turkey, and Pakistan [95] have become increasingly reliant on private Internet companies such as Facebook to police their platforms and remove content in accordance with those countries' local laws.

1.1 Dissertation Overview

In Chapter 2, I review previous work related to Chinese Internet censorship. This includes work testing for censorship on Chinese search engines, blogs, and chat platforms.

Chapters 3 through 5 present my research into Chinese censorship of realtime chat across three different industry segments. Chapter 3 analyzes Chinese instant messaging products. I present our results studying censorship mechanisms built into two Chinese instant messaging products, TOM-Skype and Sina UC. We find that each product uses thousands of keywords to trigger either censorship or surveillance. However, we found that when we compare the keywords used in each product, there

is little overlap [91, 59].

Chapter 4 looks at Chinese live streaming applications. In this chapter, I present our analysis of four of the most popular live streaming apps in China. Between all four of the apps, we found over ten thousand unique keywords that trigger censorship in one or more of the apps. We again compare the lists used between each product, and we find that keyword lists differ in terms of size and keyword content and that over time keywords are added to these lists in response to different events, except for when products have a shared parent company or have shared employees [92, 62].

Chapter 5 examines Chinese mobile games. We analyze hundreds of popular games in China, including Chinese-developed games and international games adapted to the Chinese market. We test a number of different hypotheses attempting to explain the observed overlap between some lists. Among the hypotheses tested, we found that only by looking at whether games shared a common publisher or developer adequately predicted whether they had similar keyword lists [93].

I summarize our findings in Chapter 6. In this chapter, I also give some concluding remarks and set out some potential directions for future research.

Chapter 2

Related Work

Most research on Internet censorship in China focuses on the *Great Firewall of China* (GFW), its national Internet filtering system. Some of this research studies the technical implementation details of the GFW. Zittrain *et al.* [143] documented the types of filtering methods used by the GFW. Clayton *et al.* [56] showed that GFW blocking based on TCP resets can be evaded if each host in the TCP connection ignores the resets. The authors also leveraged stateful blocking behavior of the GFW to implement a denial of service attack. Weaver *et al.* [133] introduced a method for detecting forged TCP packets, including those used by the GFW to terminate TCP connections. In 2009, Park *et al.* [111] found that the GFW had discontinued keyword filtering of HTTP responses, although filtering of HTTP request URLs remained. Winter *et al.* [135] determined how the GFW blocks Tor and how the GFW discovers Tor bridge nodes. Wright [136] detected regional variations in the GFW's DNS poisoning behavior. Ensafi *et al.* [68] used a side-channel technique to test if there exists geographic variation in the implementation of the GFW, finding no obvious geographic patterns.

Other research studies what content is blocked by the GFW. Crandall *et al.* [60]

used latent semantic analysis to uncover keywords censored by the GFW related to sensitive topics. Xia Chu [137] tested to see which Wikipedia article URLs are censored by the GFW, finding over 900 rules specifically targeting Wikipedia URLs and 18 keywords targeting all URLs. In 2014, anonymous authors [22] studied the GFW's implementation of DNS blocking, tracing their locations and generating a DNS blacklist of over ten thousand keywords. One ongoing project, the GreatFire Analyzer [76], routinely probes different URLs to see which are blocked by the GFW across time. In a work related to probing for blocked URLs, Weinberg *et al.* [134] used LDA (Latent Dirichlet Allocation) to automatically categorize into basic themes URLs from URL probe lists. Their study revealed that probe lists were often biased toward different topics.

Compared to research targeting the GFW, the number of studies on surveillance and censorship performed by private Chinese companies is limited. The microblogging service, Sina Weibo, has been the focus of a number of studies that show the dynamic nature of content filtering on the platform. Bamman, *et al.* [28] conducted statistical analysis of deleted Weibo posts and found that posts with sensitive words and from certain geographic locations (*e.g.*, Tibet and Qinghai) have a higher deletion rate. Zhu, *et al.* [142] measured censorship on Weibo and found that retroactive post deletions occur within minutes and the censors use a variety of automated tools. The University of Hong Kong has developed WeiboScope, a data collection and visualizations system for tracking censorship on Weibo [12]. Fu *et al.* [73] use this system to show that real name registration policies on Weibo may have caused some users to self-censor. Ng [106] identified numerous ways that a Weibo message could be censored or held in review both before and after being posted, confirming the usage of both automated and manual review processes.

Tencent's WeChat, a popular chat and blogging platform in China, has also been the focus of many studies. Ng [107] analyzed deleted blog posts in WeChat and found

that a large number of the deleted posts were related to rumors, even when the topic of the rumors seemed apolitical. Ruan *et al.* [115] used sample testing to measure WeChat censorship over chat and found that WeChat only enables chat censorship for accounts registered to phone numbers. They also found that group chat was more heavily censored than one to one chat, hypothesizing that this was because messages in group chat reached larger audiences. A follow-up work [116] studying censorship related to an event in China known as the "709 Crackdown" found that many of the keywords filtered on WeChat's chat platform were similar to those filtered by Sina Weibo's microblog search feature.

Other research on Internet censorship in China has studied search engines and various blog providers. Villeneuve [128] analyzed keyword filtering in non-Chinese search engines localized for the Chinese market and found the implementation of censorship inconsistent. Zhu *et al.* [141] measured keyword and URL blacklisting across four search engines in China and similarly found inconsistent implementations of censorship, although the authors had suspicions that there may exist a shared URL blacklist. Xia Chu [138] analyzed Bing in China and documented multiple different kinds of censorship techniques used by the site. MacKinnon [98] examined 15 different Chinese blog providers and found that tested keywords were inconsistently censored and that a significant amount of sensitive political content survived censorship. King *et al.* [89] collected posts from 1,382 Chinese social media Web sites and, through statistical analysis comparing censored and uncensored posts, contends that censorship in China focuses on content that represented, reinforced, or encouraged collective action.

The previously outlined studies relied on testing samples or observing changes (*e.g.*, deletions) in a subset of content over a fixed period. However, in this work I have focused on client-side implementations of censorship and surveillance. This allows me to extract the entire keyword list used to trigger these functions and

analyze changes to it over time.

In at least two cases, leaks have revealed the complete keyword lists used by some Chinese applications to trigger censorship. In 2004, a keyword list used to trigger censorship in a game bundled with QQ Chat was retrieved through a string dump of one of QQ Chat's dynamically linked libraries, which was made possible due to a lack of encryption [36]. In 2006, a list from an unknown blog provider was published by the Washington Post [132]. In both of these cases, the complete keyword list from the application was leaked. However, the two leaked lists were from two different types of products (instant messaging versus blog provider), and they were leaked years apart. In this dissertation, I compare contemporaneous keyword lists between applications in the same industry segment in three different industry segments. In two of these industry segments, I am also able to compare changes to the lists across time to determine if they are updated in response to the same news events.

The work most resembling that of this dissertation was that of Hardy [78]. Following my work described in Chapter 3, which reverse engineers censorship in Chinese instant messaging apps, he found and reverse engineered another instant messaging app called "LINE." The app was developed in Japan but included a Chinese blacklist to target users with Chinese phone numbers. Hardy observed four updates [61] to the blacklist over its lifetime. I incorporate this data in my analysis in Chapter 4.

Another approach to understanding Chinese censorship is to look at first- or second-hand accounts from company operators in China. King *et al.* [90] purchased Chinese social media software from a popular Chinese software application company in order to understand how it performed censorship. The software included various technical means to censor blog posts, but it did not include any keyword blacklist. Between the software's user guides and software support teams, King *et al.* deduced that the keyword lists different companies used with this software were typically hand-curated.

Many second-hand accounts have revealed that Chinese companies receive "directives" from Chinese officials instructing them to censor certain topics. Under the pseudonym "Mr. Tao," a technician at a Chinese Internet company documented certain directives that have been sent by Chinese authorities [102]. The Chinese Digital Times maintains a directory of leaked directives called "Directives from the Ministry of Truth" [53]. However, in both of these cases, it is unclear the extent to which Chinese companies follow these directives and to what degree they shape their censorship policy. By analyzing the amount of overlap between the keyword lists that different companies use, my work in this dissertation measures the maximum extent to which companies could be receiving and acting upon the same directives.

Chapter 3

Censorship in Instant Messaging Applications

In this chapter we[1] describe the censorship and surveillance mechanisms built into TOM-Skype and Sina UC, two instant messaging (IM) chat applications primarily used in China. We reverse engineered both applications, revealing the exhaustive lists of keywords used to trigger censorship and/or surveillance in these applications. Between April 2011 and January 2013, we discovered a total of 4,256 unique blacklisted keywords.

We wanted to address two questions: what topics do the producers of Chinese IM products censor by filtering keywords? And do censors receive keywords from a common source such as the Chinese government—or are they tasked with choosing which keywords to censor themselves? To answer these questions, we translated all non-English keywords into English and then categorized them according to their topic to determine what topics each app censors. We then compared the January

[1] I use "we" instead of "I" in this and the following chapters since they are published collaborative work.

2013 versions of TOM-Skype and Sina UC's keyword lists to determine how many keywords were blocked by both clients, finding that only 3.2% of the keywords in our dataset were blocked by both.

We summarize our major contributions as follows:

1. We introduce a new reverse engineering technique for sidestepping the "anti-debugging" measures that Skype's software uses to try to prevent reverse engineering.
2. We reveal 4,256 unique keywords used by TOM-Skype and/or Sina UC to trigger censorship or surveillance.
3. We categorize all of these words according to their topic to determine which topics are censored by each application.
4. We analyze our dataset for shared keywords amongst TOM-Skype and Sina UC, finding that only 3.2% of the keywords are featured in both. This strongly suggests that keyword lists are not directed to these companies from a common source.

3.1 Background

In this section we provide an overview of TOM-Skype and Sina UC and summarize the Chinese regulatory environment for instant messaging apps.

3.1.1 TOM-Skype and Sina UC

Skype is a popular chat and VoIP application developed in 2003. TOM-Skype, a version of Skype modified by TOM Online to target the Chinese market, is what

a user in mainland China is offered when attempting to download Skype. Skype and TOM Online established a joint venture, Tel-Online Limited, in 2005 to provide instant messaging and VoIP services in China. TOM Online is the majority partner in this joint venture; according to Skype, "TOM Online provides access to Skype for Chinese customers, using a modified version that follows Chinese regulations, called TOM-Skype" [120].

Sina UC is a chat program developed by Sina Corporation, the Chinese company behind the well known Sina Weibo micro-blogging service used in China.

We chose TOM-Skype[2] and Sina UC for analysis because these two instant messaging programs implement censorship (and surveillance in the case of TOM-Skype) inside the client software. According to 2011 reports, Tencent's QQ Chat is the most popular, and TOM-Skype is ranked the tenth most-used instant messaging program in China with 2.1 million unique daily users, and Sina UC currently holds 1.1% of the market and does not appear in the top ten most used instant messaging programs in China [33, 67]. Although these apps do not command a large amount of the Chinese market by proportion, they do in absolute numbers, each having millions of daily users. Moreover, as TOM-Skype was created specifically to deal with China's unique regulatory environment, comparing it to ordinary Skype will reveal what additional rules China's regulatory environment requires. As TOM-Skype is used to communicate with Skype users outside of China who may be more likely to introduce sensitive topics, it may especially be under scrutiny from the Chinese government. Sina UC is noteworthy for being produced by Sina, the company responsible for operating the immensely popular Sina Weibo micro-blogging platform, and so insights into censorship on Sina UC reveals how one of China's largest technology companies implements censorship in their instant messaging application.

[2]TOM-Skype includes Voice-over-IP (VoIP) features; however, our analysis focuses only on text chat functionality.

3.1.2 Chinese Regulatory Environment of Instant Messaging Apps

All Internet companies in China are held responsible for the content they host [98]. Authorities often provide directives to companies to direct them to censor certain content. News websites, for example, receive directives from "local departments in charge of news propaganda" or "public security departments" to remove articles with objectionable content [30]. Company censors are then required to adjust their list of "filter words" to account for the recently deleted content. Search engines similarly maintain lists of keywords and web addresses that cannot appear in result pages [81]. In 2009, documents leaked by an employee of Baidu, China's leading search engine, provided a glimpse into the company's censorship and monitoring policies. The documents contained guidelines on identifying information for censorship and lists of filtered keywords and URLs [37]. Among those terms and subjects identified for censorship were words concerning collective assembly and social mobilization (*e.g.*, "demonstration"), government repression (e.g, "The use of force to suppress"), specific events and people (*e.g.*, "9.12 events"), and various other terms (*e.g.*, "AIDS," "land").

Both TOM Online and Sina Corporation summarized Chinese government controls relevant to their operations in filings with the U.S. Securities and Exchange Commission, which offer some insight into the labyrinthine information control requirements that Internet companies in China must navigate. For example, Sina's annual report for 2011 illustrates the following circumstances of relevance [119]:

- Companies must obtain appropriate licensing from government authorities to conduct operations. Licenses are required to provide "basic" or "value-added" telecommunications services, the latter of which are defined as "telecommunications and information services provided through public networks." Additional

licensing requirements apply to provision of Internet content services, which vary according to the content at issue. Specific approvals are also required to provide BBS services, namely, "electronic bulletin boards, electronic forums, message boards and chat rooms." Some licenses are subject to annual inspection.

- Companies must comply with extensive laws and regulations governing the provision of various Internet services. The filing details the numerous measures relevant to Sina's services, including thirteen laws and regulations specific to information security and censorship. "According to these laws and regulations, it is mandatory for Internet companies in the PRC [People's Republic of China] to complete security-filing procedures and regularly update information security and censorship systems for their websites with the local public security bureau."

- Companies must actively guard against disclosure of "state secrets." "[T]he newly amended Law on Preservation of State Secrets which became effective on October 1, 2010 provides that whenever an Internet service provider detects any leakage of state secrets in the distribution of online information, it should stop the distribution of such information and report to the authorities of state security and public security. Internet service providers are required to delete any content on its website that may lead to disclosure of state secrets. Failure to do so on a timely and adequate basis may subject us to liabilities and penalties."

- Companies that maintain news websites or Internet portals must rely on continued cooperation with state-owned media for certain types of content. "[T]he PRC government has the ability to restrict or prevent state-owned media from cooperating with us in providing certain content to us, which will result in a significant decrease of the amount of content we can publish on our web-

> sites. We may lose users if the PRC government chooses to restrict or prevent state-owned media from cooperating with us, in which case our revenues will be impacted negatively." Such reliance provides government authorities with additional leverage in controlling industry.

Facilitating company compliance with government mandates are "self-discipline" drives and the presence of Party branches and committees internal to the companies. For example, the Internet Society of China issued the Public Pledge on Self-Discipline for the Chinese Internet Industry in 2002, which obliges "voluntary" signatories to refrain "from producing, posting, or disseminating pernicious information that may jeopardize state security and disrupt social stability, contravene laws and regulations and spread superstition and obscenity…" and to "monitor the information publicized by users on websites according to law and remove the harmful information promptly" [83, 81]. At the same time, the Communist Party of China (CPC) maintains a heavy presence within Internet companies, organizing its members and establishing Party Committees within these entities [87, 88]. Sina, Baidu, social network Kaixin, and at least six other Internet companies are reported to have formed internal Party organizations [88]. And in November 2012, a new CPC "Capital Internet Society Committee" was established to expand Party presence and strengthen the Party's governing capacity and development work in the Internet industry in Beijing, including among smaller Internet companies [140, 87].

With such extensive involvement of industry, however, and the evolving nature of the online environment, application of information controls, *i.e.*, how censorship and surveillance within the industry are actually implemented, appears to vary according to the circumstances of the platform or company at issue, activities of users, and the fluctuating policy priorities of local and central authorities. For example, Sina noted it had encountered certain difficulties in fully complying with microblog real-name registration requirements: "Although we have made significant efforts to

comply with the verification requirements, for reasons including existing user behavior, the nature of the microblogging product and the lack of clarity on specific implementation procedures, we have not been able to verify the identities of all of the users who post content publicly on Weibo" [119]. Additionally, in January 2013, a Sina Weibo manager responded to user criticism regarding Weibo censorship of references to the Southern Weekly incident by publicly expressing frustration over the government's censorship requirements and attempting to explain the company's considered trade-offs in the monitoring process [94]. Continuously changing legal requirements and vaguely defined content categories, necessitating individual company interpretation, may result in additional variations in censorship practices across platforms and providers.

Moreover, such amorphous application of information controls may lend itself to unanticipated abuses, as evidenced in recent reporting examining China's "black PR" industry and documenting instances of corruption surrounding practices of keyword blocking and content deletion [65].

3.2 Methodology

We analyzed each of TOM-Skype and Sina UC for censorship and surveillance behavior. Using packet sniffing, we discovered from which URL each client downloads its keyword lists and, for each client that sends surveillance messages, to which URL it uploads those messages. To decrypt the keyword lists and surveillance messages, we used a variety of reverse engineering techniques as appropriate for each client, which we describe below.

3.2.1 Sina UC Censorship

Sina UC had no built-in measures to resist reverse engineering, and so we were able to directly use reverse engineering tools such as IDA Pro [23] and Ollydbg [24] to directly analyze the application. Sina UC like most contemporaneous IM programs contained a large amount of executable code. Sina UC specifically contained over 80 Windows Dynamic Link Library (DLL) files, which are library files containing compiled machine code that can be updated modularly. These files together comprised over 30 megabytes. Since the code responsible for decrypting Sina UC's downloaded keyword lists was expected to be only kilobytes in size, finding it was equivalent to searching for a needle in a haystack.

To facilitate this search, we used tools for finding cryptographic constants used by well-known cryptographic algorithms inside a program's address space as it is running. We found a number of cryptographic constants for different algorithms. Among them were constants for the Blowfish algorithm. By familiarizing ourselves with the Blowfish algorithm, we determined the constants that should be referenced by the algorithm's key scheduler. By looking at which functions referenced those constants, we found the function implementing the key scheduler. We then set a breakpoint on the key scheduler function, ran Sina UC, and witnessed the Blowfish key passed to the scheduler as a function argument.

3.2.2 TOM-Skype Censorship

We reverse engineered the TOM-Skype 3.6–4.2 and TOM-Skype Mobile's keyword list's cryptography by employing a type of attack called a *chosen ciphertext attack*, *i.e.*, an attack that obtains information by observing the decryption of specifically chosen ciphertexts. These versions of the TOM-Skype clients perform a DNS lookup for the address of the server used to download the keyword lists. By modifying the

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**

Figure 3.1: **Algorithm for decrypting TOM-Skype 3.6–3.8 keyfiles**

client's operating system's "hosts" file, we hardcoded the address of our own server to be returned in response to the lookup. This causes the client's download request to instead download from our own web server, we were able to completely control the ciphertext that the client downloads.

The initial ciphertext that we sent to the client was identical to that which TOM-Skype provided. We knew from previous work [129] that the word "fuck" was censored, and so by deleting half of the list at a time, we were able to use binary search to determine which line corresponded to the keyword "fuck." From there, we made perturbations to the ciphertext until we were able to infer the algorithm described in Figure 3.1.

We found that TOM-Skype 5.0–5.1's keyword lists were downloaded from and decrypted in ContentFilter.exe, a separate process from Skype.exe. They are encrypted using a 256-bit key that was originally known to have been used in TOM-Skype 2.5 [70]. The key appears to have been intended to be 32 ASCII-encoded characters, but the 32 characters were UTF16-encoded, and so only the first 16 characters fit into the 256-bit key, where the other 16 of the 32 bytes are null bytes.

Reverse engineering the cryptography for TOM-Skype 5.5–6.1's keyword lists was challenging. These lists were downloaded and decrypted inside Skype.exe itself, not a separate process as with TOM-Skype 5.0–5.1. The ordinary Skype client is known to contain sophisticated anti-debugging measures that resist traditional

reverse engineering techniques [69], and we found that TOM-Skype inherits these measures.

We circumvented these measures by using DLL injection, a technique for running arbitrary code inside of another process's address space. We used this technique to hook API functions, which allows us to substitute an API function's behavior with behavior of our own. We first hooked the API function that the client uses to download the keyword lists, which allowed us to obtain information about the thread used to download and decrypt the lists including the return address for each caller on that thread's call stack. We then hooked into the API function used to create threads and, when it created a thread matching the criteria that we previously discovered, we had it create the thread in a suspended state. From there, we attached with a debugger, suspended all other threads to avoid anti-debugging measures, and resumed our thread of interest. We were then able to analyze the behavior of the resumed thread using standard reverse engineering tools.

3.2.3 TOM-Skype Surveillance

We were able to reverse engineer the cryptography used for surveillance messages in TOM-Skype 5.1 using standard tools, since the surveillance was done in Content-Filter.exe, a separate process from Skype.exe that does not contain the same anti-debugging measures. Although TOM-Skype 4.0–4.2 and 5.5–6.1 perform surveillance inside of Skype.exe, we found that they used the same cryptography for surveillance as 5.1.

Reverse engineering the cryptography used for surveillance messages in TOM-Skype 3.6–3.8 was more challenging, since they perform surveillance inside of the Skype.exe executable itself. However, we used a similar DLL injection strategy as we used to reverse engineer TOM-Skype 5.5–6.1's keyword list cryptography. We

List	HTTP URL
TOM-Skype 3.6–3.8	`skypetools.tom.com/agent/newkeyfile/keyfile`
TOM-Skype 4.0–4.2	`a1.skype.tom.com/installer/agent/keyfile`
TOM-Skype 5.0–5.1	`skypetools.tom.com/agent/keyfile`
TOM-Skype 5.1 (Surveillance-only)	`skypetools.tom.com/agent/keyfile_u`
TOM-Skype 5.5–6.1	`a1.skype.tom.com/installer/agent/keyfile5.5/keyfile`
TOM-Skype 5.5–6.1 (Surveillance-only)	`a1.skype.tom.com/installer/agent/keyfile5.5/keyfile_u`
TOM-Skype Mobile	`skypetools.tom.com/agent/newkeyfile/keyfile_a`
Sina UC Lists 1–5	`im.sina.com.cn/fetch_keyword.php?ver=8.3.4.22616`

Table 3.1: **Lists used by TOM-Skype and Sina UC and their HTTP URLs**

knew from looking at other versions of TOM-Skype that they reseed the random number generator before sending surveillance messages to generate random padding bytes, and so we hooked an API function normally used to generate the seed to instead suspend the thread when it was called with a specific return address that we had determined through trial by error. Then, as before, we were able to attach with a debugger, suspending all other threads except the thread of interest, and reverse engineer that thread avoiding Skype's anti-debugging measures.

3.3 Technical Analysis

In this section we outline how the censorship and surveillance mechanisms of each app operate.

We collected a number of different lists for both the TOM-Skype and Sina UC clients. For TOM-Skype, different versions of the client use different lists downloaded

List	Function
TOM-Skype 3.6–3.8	Chat censorship, chat surveillance
TOM-Skype 4.0–4.2	Chat censorship, chat surveillance
TOM-Skype 5.0–5.1	Chat censorship, chat surveillance
TOM-Skype 5.1 (Surveillance-only)	Chat surveillance
TOM-Skype 5.5–6.1	Chat censorship, chat surveillance
TOM-Skype 5.5–6.1 (Surveillance-only)	Chat surveillance
TOM-Skype Mobile	Chat censorship, chat surveillance
Sina UC List 1	Combined functionality of Sina UC Lists 2–5
Sina UC List 2	Username and mood indication censorship
Sina UC List 3	–
Sina UC List 4	One-to-one chat censorship
Sina UC List 5	Group chat censorship

Table 3.2: **Lists used by TOM-Skype and Sina UC and their functions**

from different URLS (see Table 3.1), and in later versions of the client, use separate lists for censorship and/or surveillance. All versions of the Sina UC client use the same set of lists, with the lists serving different functions (see Table 3.2).

The TOM-Skype client contains built-in lists of keywords and downloads new lists using HTTP requests. The client uses one of these two lists to censor incoming or outgoing text chat; however, the various versions of the clients differ in their built-in lists, the source of the list updates they download, and whether they censor incoming and/or outgoing chat messages or perform surveillance. Most versions of the client, upon censoring an incoming or outgoing chat message, will send a log of the message content and sender information to TOM-Skype servers through HTTP requests.

List(s)	Cryptography	Cryptographic key
TOM-Skype 3.6–3.8 and 4.0–4.2	"Homebrew" XOR	–
TOM-Skype 5.0–5.1 and 5.1 Surveillance-only	AES+ECB	`"0\0s\0r\0␣\0T\0M\0#\0R\0" + "W\0F\0D\0,\0a\04\03\0␣\0"`
TOM-Skype 5.5–6.1 and 5.5–6.1 Surveillance-only	DES+ECB	`"\x7a\xdd\xe7\xdc" + "\x23\x25\x53\x75"`
TOM-Skype Mobile	"Homebrew" XOR	–
Sina UC Lists 1–5	Blowfish+ECB	`"H177UC09VI67KASI"`

Table 3.3: **List of cryptographic algorithms and keys used by the TOM-Skype and Sina UC clients**

The Sina UC client censors incoming and outgoing messages as well as usernames. While the client does not perform surveillance itself, it is possible that server-side surveillance is performed. The client contains five built-in lists that are updated through HTTP. Each list serves a different purpose. The first list censors incoming and outgoing one-to-one text chat, group chat, and usernames (replacing the username with an ID number); the second list censors only usernames; the fourth list

Client(s)	Cryptography	Cryptographic key
TOM-Skype 3.6–4.2	DES+ECB (using only first 6 of 8 bytes of each plaintext block)	`32bnx23l`
TOM-Skype 5.0	No surveillance	–
TOM-Skype 5.1–6.1 and Mobile	DES+ECB (using only first 6 of 8 bytes of each plaintext block)	`X7sRUjL\0`

Table 3.4: **List of cryptographic algorithms and keys used for surveillance by TOM-Skype clients**

Clients	**Example surveillance message**
TOM-Skype 3.6–4.2	`JohnDoe fuck you 12/31/2011 6:00:00 PM 1`
TOM-Skype 5.0	No surveillance
TOM-Skype 5.1	`fuck you 12/31/2011 6:00:00 PM 1`
TOM-Skype 5.5–6.1	`JohnDoe fuck you 12/31/2011 6:00:00 PM 1 JaneDoe`
TOM-Skype Mobile	`JohnDoe fuck you 2011-12-31 18:00:00 1`

Table 3.5: **Surveillance message triggered when JaneDoe receives "fuck you" from JohnDoe**

censors incoming and outgoing one-to-one text chat; and the fifth list censors group chat. (The third list has no functionality in the version of Sina UC we analyzed.)

Clients used a variety of cryptographic algorithms to encrypt keyword lists and surveillance messages, ranging from well-known algorithms to a "homebrew" algorithm (see Figure 3.1) that does not provide much security.

We found that all versions of TOM-Skype analyzed except 5.0 perform surveillance, and that versions differ in what information they send in their surveillance logs. Most clients include the sender of the triggering message, the triggering message in its entirety, the date and time, and a 0 or 1 to indicate if that message was outgoing or incoming, respectively. TOM-Skype 5.1 sends the least comprehensive surveillance logs that do not include the sender, whereas TOM-Skype versions 5.5–6.1 send the most comprehensive surveillance logs, including the recipient of the message in addition to the sender.

Character type(s)	**Number of keywords**	**Example(s)**
CJK including spaces only	(72%) 3069	遛似
ASCII only	(12%) 518	`six 4`
ASCII and CJK	(15%) 645	`six`月`four`日
Cyrillic	5	Восемь-Девять-Шесть-Четыре
Unicode roman numerals	6	Ⅷ Ⅸ Ⅵ Ⅳ
Unicode full-width latin	10	ｑ ｑ ｑ ｑ ｑ
Other	3	`six`—`four` (em dash), ⑥④, ⑥㈣

Table 3.6: **Breakdown of character types in the keyword lists**

3.4 Keyword Analysis

Collection of the keyword lists began on April 24, 2011 (TOM-Skype) / August 8, 2011 (Sina UC) and ended on January 31, 2013, with the latest changes occurring on December 20, 2012 (TOM-Skype) / October 11, 2012 (Sina UC). In total, the dataset consists of 88 lists, which combined contain 4,256 unique keywords. The lists range in size from 1 to 1,421 unique keywords.

Of the 4,256 keywords, 3,070 keywords (72%) contain only Chinese characters (specifically, characters in the CJK range of Unicode), 518 (12%) contain only ASCII characters and 645 (15%) have a combination of both. Among the 518 containing only ASCII characters, 52% of these were URLs or URL-like strings. Five words were in Cyrillic, six keywords contain Unicode Roman numeral characters and 10 keywords comprise Unicode fullwidth Latin characters.

Each keyword was translated from Chinese to English by a fluent Chinese speaker and accompanied with descriptions of the political and social context behind the keyword. Based on these contextual descriptions, we coded the keywords into 61

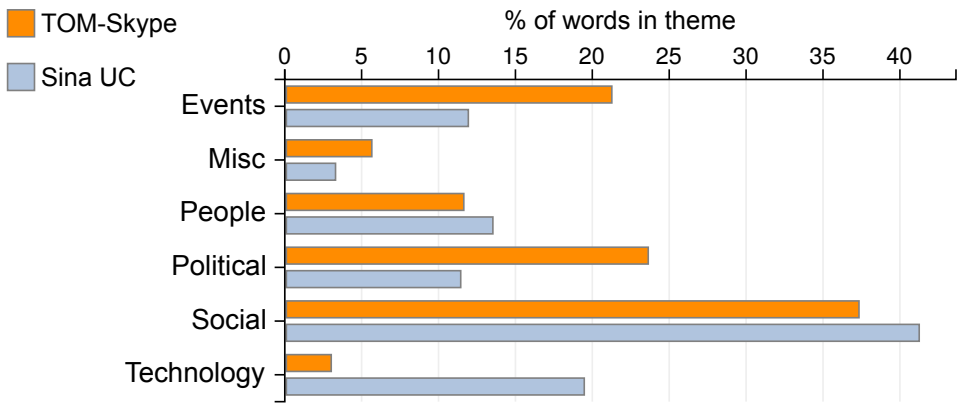


Figure 3.2: **Themes by client**

content categories grouped under six broad themes: Political (content related to the Chinese government or political issues, *e.g.*, human rights, freedom of expression, ethnic groups, religious movements, *etc.*); Social (content perceived as socially sensitive or undesirable, *e.g.*, pornography, gambling, illicit weapons and drugs, *etc.*); People (names of individuals, *e.g.*, government officials, political dissidents); Events (scheduled events, recurring events, current events); Technology (*e.g.*, general technical terms, websites, spyware, URLs, *etc.*) and Miscellaneous (*e.g.*, terms without clear context).

Overall we found very little overlap in keywords between the clients: 138 keywords (3.2%) distributed across 29 categories and all six themes were shared in common between TOM-Skype and Sina UC lists, represented primarily in the categories of CPC member / Government official (21 keywords), Prurient interests (19), Dissident / Activist (18), Religion (15), and Tiananmen Square (13) (see Figure 3.2).

3.4.1 Political

Within the "political" theme a wide range of issues are covered, including CPC politics, Chinese democracy movement, corruption scandals, ethnic groups, and religious

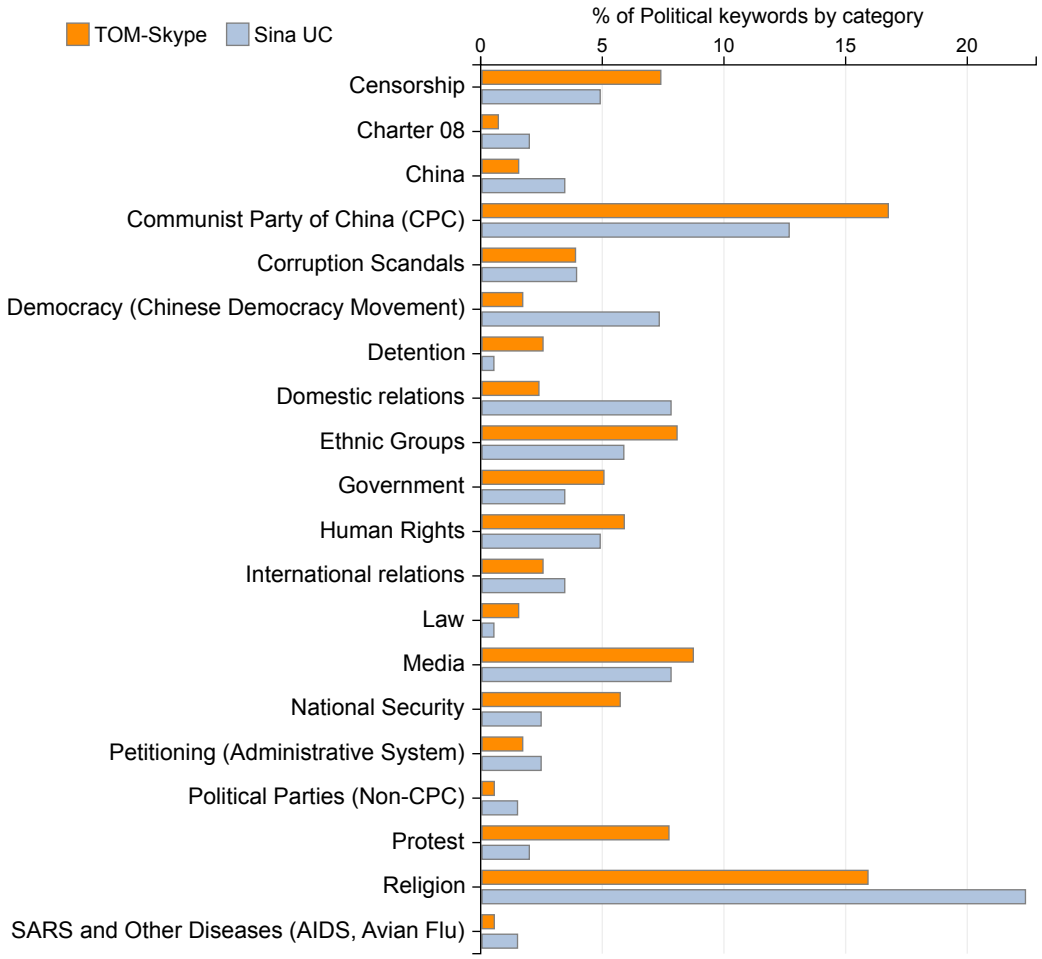


Figure 3.3: **Political categories by client**

movements. While there is little overlap in unique keywords between the clients, the keyword lists show common concern for the category issues. Figure 3.3 shows a breakdown of political categories.

3.4.2 People

Within the "people" theme the most prominent category references members of the CPC. Second to that are of names of individual activists or political dissi-

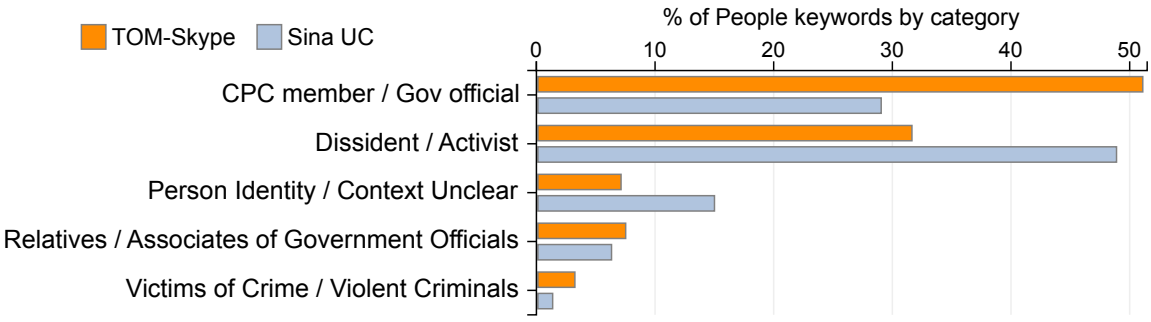


Figure 3.4: **People categories by client**

dents. Between the TOM-Skype and Sina UC lists combined there are 179 keywords with the names of individual activists or political dissidents (*e.g.*, Ai Weiwei, Chen Guangcheng, Wu'erkaixi). This theme also references relatives of CPC members, and perpetrators and victims of violent crimes. (See Figure 3.4 for a breakdown.)

3.4.3 Events

In general, TOM-Skype lists contained more references to specific events, with the exception of the 190 words in Sina UC lists relating to the June 4, 1989 Tiananmen Square massacre. TOM-Skype lists included keywords relating to all 21 events we identified, while Sina UC lists referred to only eight of these events. (See Figure 3.5 for a breakdown.)

3.4.4 Social

Keywords in the "social" theme were primarily in two categories: Illicit goods and services, which included the trafficking of illicit materials like narcotics, weapons and counterfeit goods; and Prurient interest, which generally referred to pornography and prostitution. Illicit goods and services is the largest single category in the dataset with 677 total keywords; Prurient interest is the second largest category with 663

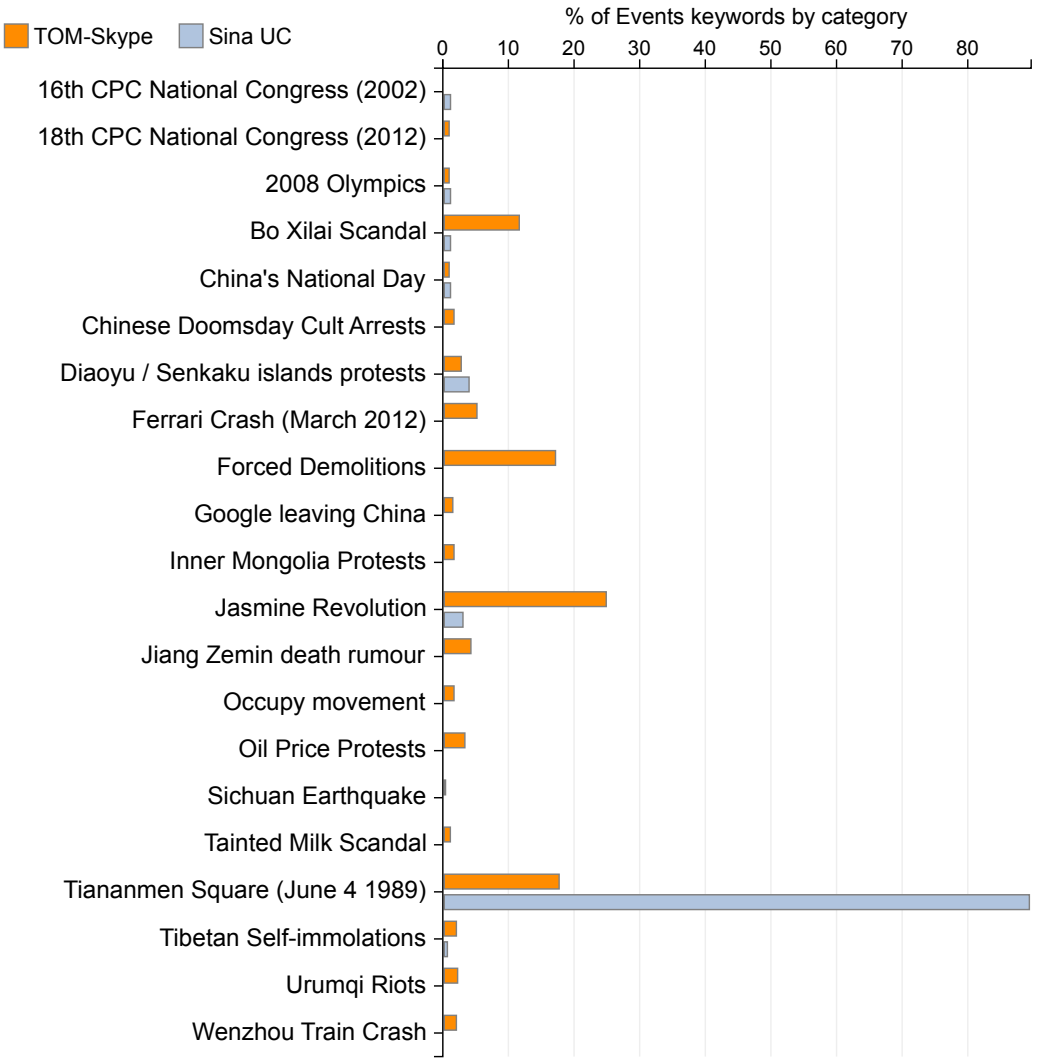


Figure 3.5: **Event categories by client**

total keywords. (See Figure 3.6 for a breakdown.)

3.4.5 Technology

The largest proportion of keywords in the "technology" category were URLs in Sina UC lists. The inclusion of URLs on censorship lists was likely a mechanism of preventing the spread of malicious links or spam. Next most common were generic

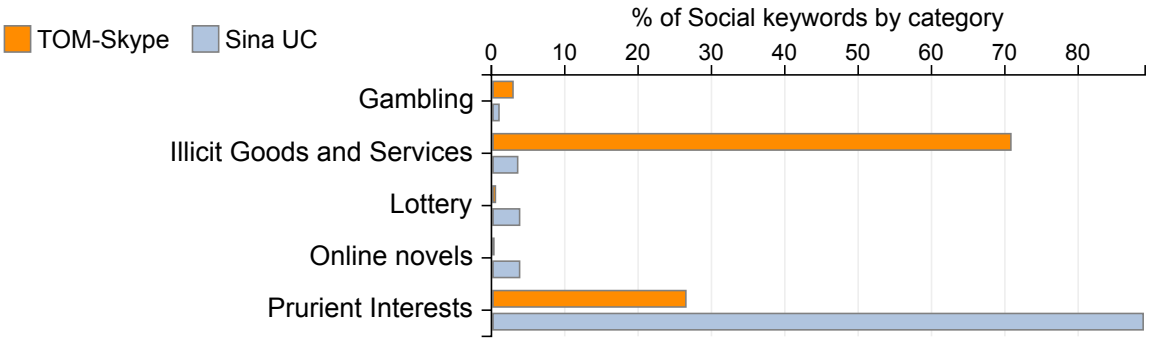


Figure 3.6: **Social categories by client**

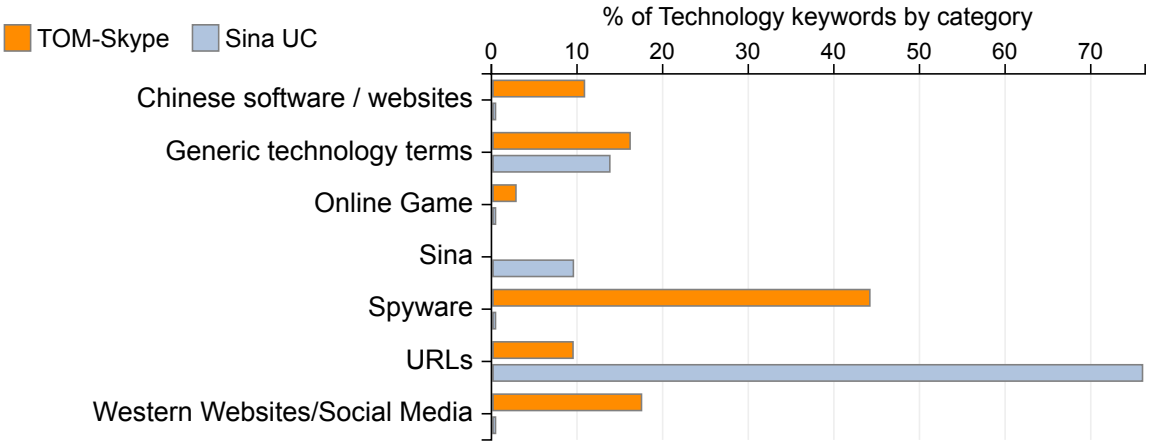


Figure 3.7: **Technology categories by client**

technical terms, which also appeared most frequently in the Sina UC lists. Many of these were very general, including "Administrator" ("管理员") and "System notification" ("系统通知") on Sina list 2, which may be a means of preventing users from creating usernames that allow them to impersonate Sina administrative accounts. Also in this category were the names of prominent websites, including "Chinese language Wikipedia" ("中文维基百科") and "Google Blogger" ("谷歌博客"). (See Figure 3.7 for a breakdown.)

3.4.6 Targeted and Broad Keywords

Many of the keywords are highly targeted. For example, the TOM-Skype lists contain the keyword "Corning West and Da Zhi Street intersection, Century Lianhua gate" ("西大直街康宁路路口世纪联华"), the address of a planned Jasmine rally.

Other keywords are extremely generic. For example, the TOM-Skype 5.1 surveillance lists include "Han People" ("汉人") (the majority ethnic group in China), "Chinese person" ("华人"), "world wide web" ("万维"), and "Internet" ("互联网"). The inclusion of these common terms likely causes the messages of a large number of users to be surveilled. In addition, these keywords appear too broad to be useful for routine surveillance, and suggest that either TOM-Skype is overzealous in its efforts to enforce government-mandated surveillance or is using additional criteria to identify messages or users to be surveilled [129].

Sina UC List 2 contains a number of generic keywords such as "system" ("系统") and "chat" ("聊天"). This list is used for censoring usernames, and it is possible these keywords exist to prevent users from impersonating system administrators or functions of the client.

3.4.7 Adaptation to Censorship Evasion

The keywords also indicate that censors adapt to the censorship evasion techniques employed by Chinese Internet users who frequently use creative language and homophones to evade censorship [35]. As Chinese dialects are tonal languages, Internet users often use similar words with variations in tone or different characters represented by the same tone to impart the meaning of a character that is otherwise censored. Users will also take advantage of visual similarities between different characters to imply the meaning of a banned character or word. The variation in keywords

seen in the lists demonstrates that censors are highly aware of and adaptive to the techniques that users employ to attempt to evade censorship and surveillance.

For example, keyword lists included the name of disgraced politician "Bo Xilai" ("薄熙来"), as well as "博西莱" ("Bo Xilai" with the same pronunciation and tones, but different characters), and "B○稀莱" ("BO Xilai" with character variation of two kinds). Similarly, numerous homonyms for "Jasmine" (茉莉花) appeared on the various TOM-Skype lists. In some cases the censored words included combinations of Chinese characters with English words, numbers and symbols. The breadth of these combinations is seen in keywords related to the June 4, 1989 Tiananmen Square massacre, a highly sensitive topic that is the subject of intense censorship. The censored lists contain many terms referring to June Fourth, such as variations of six-four expressed in Chinese (六四, six四); Roman numbers (Ⅵ Ⅳ), equations (6.2+2, 32x2), symbols (⑥④), and dates (May 35th (五月三十五), March 96th (三月九十六号)).

3.4.8 Keyword List Changes

We began monitoring for daily changes to the TOM-Skype lists beginning April 24, 2011, and the Sina UC lists beginning August 8, 2011, with the collection period ending January 31, 2013. Lists from both clients underwent significant fluctuations over the course of the study, in some cases increasing rapidly in size and shrinking to a single keyword within a short time frame. The result of these changes is that the censorship and surveillance mechanisms in the clients are significantly altered. In addition, there were a number of anomalies observed in the results that are difficult to explain but that may reflect technical misconfiguration on the part of TOM-Skype or Sina UC administrators.

TOM-Skype Keyword List Changes

In May 2011, the TOM-Skype 5.1 Surveillance-only list rapidly increased in size to 1,421 unique keywords, only to decrease the next day back to 399 keywords. The censorship lists for versions 5.0–5.1 decreased to a single keyword (at one point containing a string of seemingly random characters) in April 2011, while the censorship lists for versions 5.5–6.1 increased from one keyword "Rong Shoujing" ("荣守京") to that same keyword 1,134 times and back to that single keyword a day later.

On September 20, 2012, an update to the TOM-Skype 5.5–6.1 Surveillance-only list was sent to clients with a number of added keywords relating to the island dispute with Japan. However, these keywords used an encryption scheme from TOM-Skype 3.6 that was no longer in use, which rendered these keywords unusable for triggering surveillance. These keywords were added again the next day with the same incorrect encryption scheme. It is possible that TOM-Skype administrators noticed that surveillance for these keywords was not functioning and attempted to add them again, but without success. As of January 31, 2013, these keywords are still incorrectly encrypted, despite newer keywords having been added to the list with the correct encryption.

The censorship keyword lists for the recent versions of the client have also been reduced to a single keyword, meaning that these clients are now effectively only performing surveillance. TOM-Skype versions 3.6–4.2, which, unlike the recent versions, do not have a separate surveillance-only list, still have censorship keyword lists containing hundreds of keywords that were last updated in March 2012.

Sina UC Keyword List Changes

The Sina UC lists similarly underwent a number of changes for unknown purposes that represent a shift in the functionality of censorship in the client. Over the course

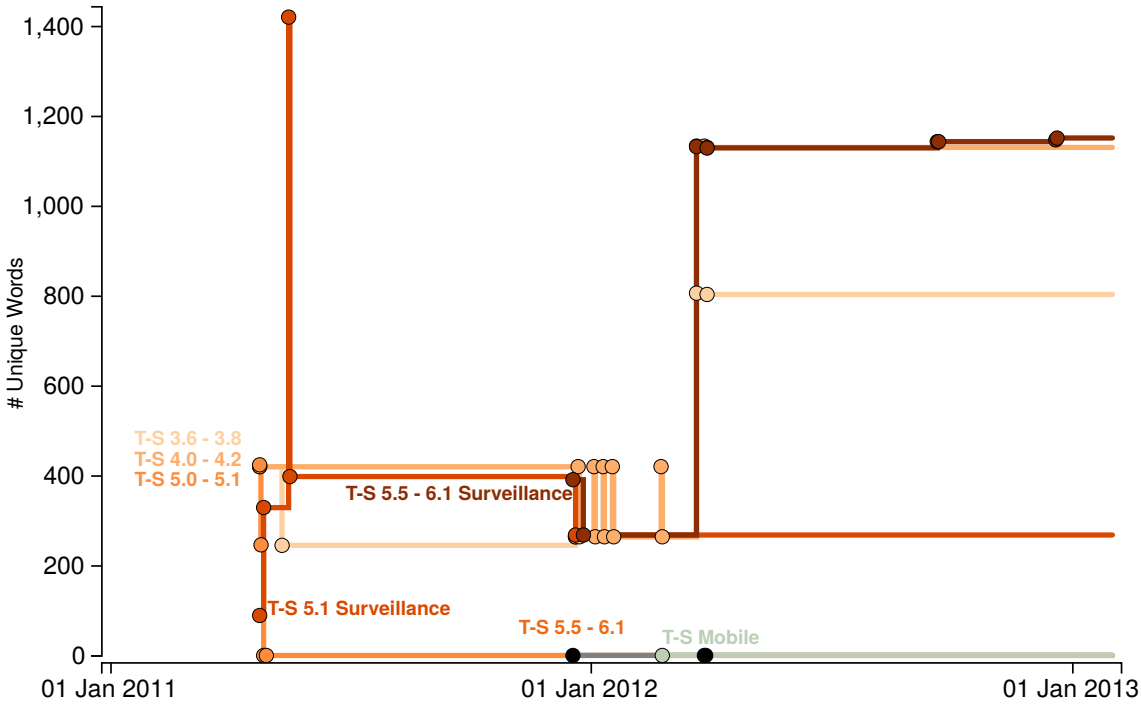


Figure 3.8: **The number of unique keywords in TOM-Skype lists over time**

of seven months in 2012, Sina UC List 3 shifted from 50 keywords to 880 before eventually being reduced to a single keyword. List 1 shrank to a single keyword in September 2012. List 4 went from 380 to 21 to 855 keywords over the span of three months, before also shrinking to a single keyword. List 5, like lists 1, 3 and 4, was reduced to a single keyword in September 2012. The single keyword was different on each of these four lists, ranging from a reference to Tiananmen Square to keywords in the prurient interest category. It is conceivable that a single keyword was used because the censorship functions of the client require a non-zero list to function. List 2, which censors usernames, is the only list with multiple keywords remaining (476 keywords as of October 11, 2012). As a result, most client-side messaging censorship has been eliminated in the client. We discuss the possible implications of this change in the following section.

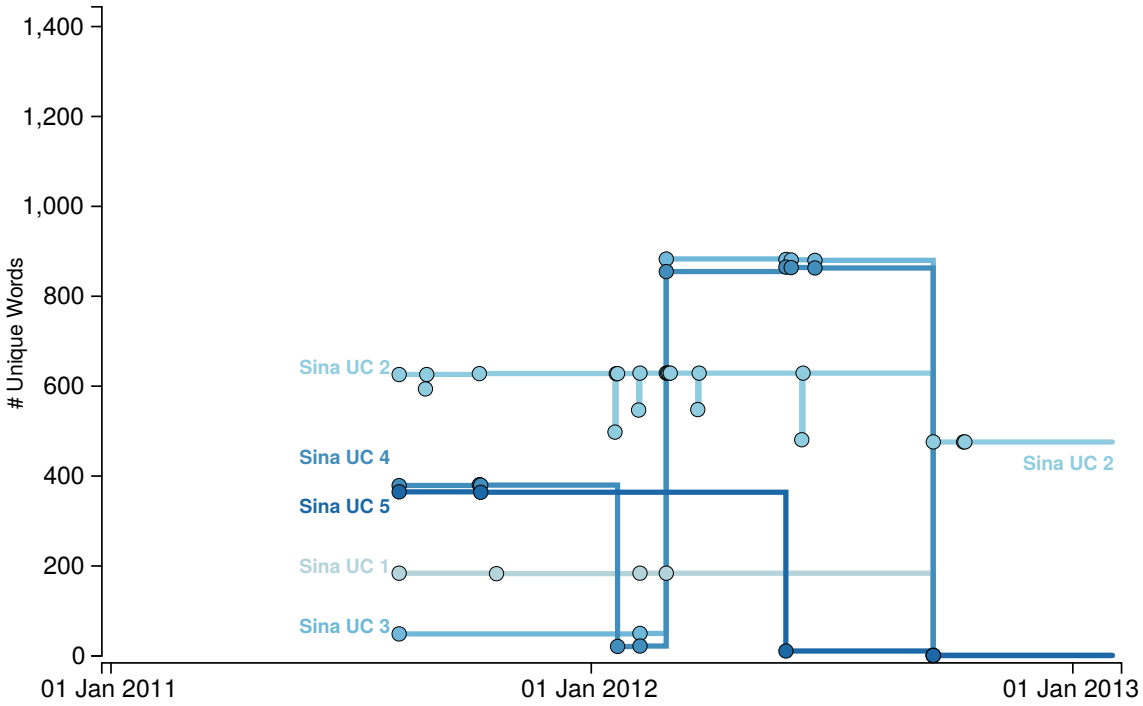


Figure 3.9: **The number of unique keywords in Sina UC lists over time**

Changes in Censorship and Surveillance Focus

One of the unexpected changes we observed was fluctuation in the keyword lists that effectively rendered the latest versions of TOM-Skype to be focused on surveillance only (rather than also focusing on censorship), and seemingly caused Sina UC to only focus on username censorship.

In the case of TOM-Skype the shift to surveillance-only keyword lists was correlated with the Jasmine Rallies, which could potentially signify pressure from authorities or an independent decision made on the part of the company to monitor discussions of sensitive events, particularly those that may lead to social mobilizations.

For Sina UC the changes are difficult to explain. Given China's legal and reg-

ulatory restrictions it is unlikely that the company would discontinue censorship features. However, it is possible that, like TOM-Skype, Sina UC has also switched to a surveillance focus but is implementing these features on the server side, which would not be detected by our reverse engineering methods. Due to the peer-to-peer architecture of (TOM-)Skype, surveillance and censorship must be implemented on the client side. Of all the Chinese IM programs on the market, TOM-Skype and Sina UC are the only ones we are aware of that implement censorship or surveillance features on the client side.

Additional exploratory testing we conducted further supports the hypothesis that there may be a move away from a censorship focus in IM clients. In April 2012, we attempted to send messages containing the keywords from the TOM-Skype 5.5–6.1 Surveillance-only list through QQ Chat. In total, nine keywords from that list were found filtered, mostly relating to the Falun Gong. In a similar experiment, 15 words from the Sina UC lists were found censored on QQ Chat, also relating to the Falun Gong. Repeating this experiment in February 2013, zero words from either list were found filtered on QQ Chat, even when performing the experiment tunneled through a VPN in China. Both experiments demonstrate the lack of overlap in censored content between the different clients, and the most recent results suggest that the focus may have shifted to server-side surveillance.

China's most popular IM program, QQ Chat, and new applications quickly rising in popularity such as WeChat, are suspected to have surveillance features [86, 75], but no technical analysis has yet been able to confirm their existence or operation as server-side surveillance by its nature is difficult to measure. Unlike server-side censorship, which can be at least partially measured with sample testing, server-side surveillance is difficult to measure because it has no observable effect on messages under surveillance. Unlike client-side surveillance, which can be reverse engineered, the code implementing server-side surveillance is running on a remote machine, making

it inaccessible to reverse engineering methods.

If the majority of Chinese companies providing IM programs are engaging in surveillance, the potential for massive violations of privacy is acute. It is clear that Chinese companies are obliged to cooperate with government investigations, maintain and disclose records and reports to security authorities, and terminate transmission of state secrets [80]. Yet it is unclear how these regulations affect how private companies decide what specific content to target for surveillance, and what level of government oversight into company practices exists. These obligations to the government and the risk of penalties for non-compliance could be an incentive for overly broad keyword triggers to ensure persistent capture of user data. At the same time, however, our analysis observed inconsistent patterns in how sensitive topics and events were targeted for surveillance.

Jaccard similarity between initial and current lists

Calculating the Jaccard similarity coefficient (the size of the intersection of two sets divided by the size of their union) between the set of words in the most recent versions of lists and the first versions reveals that in most cases, the current lists are significantly different from what they started as. The exceptions to this are Sina UC List 2, which has a similarity coefficient of 0.76, and TOM-Skype 5.5–6.1, whose first and last lists contain the same single word (discussed above). All other lists have a coefficient of 0.09 or lower. Coefficients (excluding sources with only one version of a list) are shown in Table 3.7.

3.4.9 Changes in Response to Events

One of the unique aspects of this dataset is that it provides visibility into how the censorship and surveillance keyword lists change over time. Analyzing these changes

Source	**Similarity**
TOM-Skype 3.6–3.8	0.03
TOM-Skype 4.0–4.2	0.09
TOM-Skype 5.0–5.1	0
TOM-Skype 5.1 (Surveillance-only)	0
TOM-Skype 5.5–6.1	**1.0**
TOM-Skype 5.5–6.1 (Surveillance-only)	0.09
Sina UC List 1	0.005
Sina UC List 2	**0.76**
Sina UC List 3	0
Sina UC List 4	0
Sina UC List 5	0

Table 3.7: **Jaccard similarity between first and last lists**

provides insight into how these two companies respond to dynamic political and social events through updates to the keyword lists.

Events referenced in the keyword lists include the following types: past events (*e.g.*, 16th National Congress of the Communist Party of China), national holidays and anniversaries (*e.g.*, June 4, 1989 Tiananmen Square massacre, National Day of the People's Republic of China), scheduled events (*e.g.*, 18th National Congress of the Communist Party of China) and current events (defined as events that occurred within our data collection period).

TOM-Skype lists include substantially more event-related keywords (including current and recurrent events). Sina UC lists have very little focus on current events but a greater number of keywords related to recurrent events. For example, the Sina UC lists include 190 keywords (10.5%) related to the June 4, 1989 Tiananmen Square massacre, whereas TOM-Skype included 95 such keywords (3.7%).

In order to track how the two companies implemented keyword changes in response to current events, we identified six events referenced in the dataset that occurred within our data collection timeframe. Keywords related to only two of these events appeared on both Sina UC and TOM-Skype lists.

Across the selected cases we observed inconsistent patterns in how changes were made around event timelines. In some cases keyword updates were implemented within a single day of a sensitive event. In others, keywords were added weeks or months after the event took place, potentially indicating the censors only responded after an issue developed sufficient political salience. In some cases, seemingly important and sensitive political events that were clearly of concern to the Chinese government either did not appear in any testing lists or were only represented with a small number of terms.

The following sections provide analysis of the context behind each event and correlate the event timelines with keyword updates or lack thereof.

Jasmine Rallies

Following the uprisings in the Middle East and North Africa in 2010 and early 2011, calls to gather for "Jasmine Rallies" circulated online beginning on February 20, 2011, with locations of planned rallies in a number of major cities throughout China [84]. While none of the planned events developed into protests, the event scheduled in Beijing gathered widespread attention after video of U.S. Ambassador to China, Jon Huntsman, apparently in the area of the designated rendezvous point, was circulated. Later calls would suggest participants "stroll" near designated locations in a number of cities so as not to attract police attention. These gatherings, while reportedly sparsely attended, were met with a significant police presence and saw numerous reports of arrests and police violence against journalists. Notably, prominent artist

Ai Weiwei was arrested on April 3 following several Tweets he made discussing the "Jasmine Revolution" [114].

As our data collection period began in April 2011, two months after the first rallies, we cannot identify when related keywords were first added to the lists. A large number of keywords relating to the Jasmine Rallies were already present on the first lists gathered for both of the clients. In total, 132 keywords were on the TOM-Skype lists, four were on the Sina UC list and two were present on both lists. Keywords on the lists include "Next Sunday Jasmine" ("下周日茉莉"), "Western Thai Square on the 20th" ("西临天泰广场 20日") and "Jasmine revolution written backwards" ("命革花莉茉"). Lists from both clients contained "Hold a microphone to indicate liberty" ("拿着麦克风表示自由"), an instruction for rally participants.

Between April and May 2011, there is a notable change in the presence of Jasmine-related words on the TOM-Skype lists. On April 25, 69 keywords were removed from the censorship list for TOM-Skype 5.0–5.1, and on May 16, 75 keywords were added to the TOM-Skype 5.1 Surveillance-only list. This update followed a general pattern of the keyword lists for the recent versions of the TOM-Skype client transitioning to surveillance only. It is possible that this change happened in response to the Jasmine Rallies as a strategy for monitoring mobilization and discussion of sensitive events.

Bo Xilai Scandal

On November 14, 2011, Neil Heywood, a British businessperson based in China, was found dead in his hotel room in Chongqing province [29]. Heywood had long ties with the family of Bo Xilai, then the high-profile leader of the Chongqing branch of the CPC and at one time touted to be in line to join the Politburo Standing Committee, the CPC's top leadership committee. Heywood's death would later be deemed a homicide related to a soured business deal involving Gu Kailai, Bo's wife, who would

eventually be convicted of murder. In February 2012, Chongqing police chief Wang Lijun met with U.S. consular officials in Chengdu, reportedly to provide information on Heywood's murder and potentially to seek asylum and protection from Bo. Wang would later be sentenced to 15 years in prison for corruption. Additionally, leaked reports from the CPC indicated that Bo had been conducting surveillance of high-level party officials, including tapping the phone of president Hu Jintao. On March 18, 2012, Bo was dismissed from his position as Chongqing party chief, and in September 2012 he was expelled from the CPC. The ouster of such a high-ranking politician marked one of China's biggest political crises in decades and threatened to disrupt the carefully-planned leadership transition taking place in November 2012. Leaked instructions from government authorities on the topic called for media to refer only to state-sanctioned sources when covering the story [40].

A total of 62 keywords relating to Bo Xilai and the Heywood murder scandal appear on the lists, predominantly on TOM-Skype. Some of these terms were already present on the first TOM-Skype and Sina UC lists collected during our data collection period, which predated the Heywood murder scandal. This is not unexpected, as Bo was already a prominent and often controversial figure seen as a rising star within the CPC. On March 21, 2013, 50 keywords were added, including numerous homophones and variations on the name "Bo Xilai," such as B○稀莱 (Bo xī lái), 泊稀莱 (Po xī lái), 己厚天下 ("not thick, the word below," a reference to "Bo" literally meaning "thin"), and "bullshitliar." On March 29, 2012, nine additional keywords were added to the same list, including terms calling for collective action in support of Bo, such as "To support Bo [Xilai] go to Chongqing People's Square" ("挺薄去重庆人民广场") and "March 17 Chongqing People's Grand Hall" ("3月17日重庆人民大礼堂").

Although a number of events related to the Bo scandal occurred in late 2011 and early 2012, the additions to the keyword lists followed shortly after Bo's March 16, 2012 dismissal as Chongqing party chief.

Ferrari Crash

On March 18, 2012, Ling Gu, the son of high-ranking CPC official Ling Jihua, was killed in a car crash outside of Beijing [26]. Ling, as well as two women in the car who were injured, were reported to be naked, and photographs of the crumpled Ferrari began circulating online. The incident was politically sensitive for a number of reasons: Ling Jihua, a close political ally of leader Hu Jintao, was expected to be promoted during the November 2012 leadership transition. Further, the son of two government officials driving a luxury car touched on widespread public criticism over government corruption and inappropriate behavior of the family members of government officials. Ling Jihua would later be demoted from his position at the General Office of the CPC Central Committee.

Within a day of the crash, reports emerged that searches for "Ferrari" ("法拉利"), "Master Ling" ("令公子"), and other related terms had been blocked on Sina Weibo and search engines Baidu and Soso [72]. While Chinese state media initially published stories covering the crash, by March 20 a Global Times article about the incident had been removed.

TOM-Skype lists were updated within a few days of the event. On March 21, 2012, 24 keywords related to the incident were added to the TOM-Skype 5.5–6.1 Surveillance-only list, including several variations on "Beijing Ferrari car accident" ("北京法拉利车祸"). Eight days later, an additional three keywords were added to this list, while three of the previous keywords were removed. Notably, none of the terms added referenced the names of Ling Jihua or Ling Gu specifically.

Church of Almighty God Arrests

On December 19, 2012, Chinese state media reported on the arrest of 500 individuals associated with the religious group Church of Almighty God, on allegations they had

spread rumors that the world would end on December 21 in accordance with the last day of the Mayan calendar [85].

Reports from China Digital Times on December 10, 2012 indicate that official instructions were issued to media outlets to guard against the creation and spread of rumors relating to the December 21 prediction, and requesting media to "discontinue reporting on recent public conversion assemblies and other illegal activities orchestrated by the Almighty God cult" [42]. On December 19, the day media reports of the arrests emerged, four keywords were added to the TOM-Skype 5.5–6.1 Surveillance-only list which related to the religious group, followed the next day by the addition of four more keywords. The keywords included "red dragon gospel" ("大红龙福音"), which refers to the group's term for the CPC, and "God in Henan" ("真神在河南"), referring to the province where the group was founded. These December 19 and 20 keyword additions were the last instance of TOM-Skype list updates we observed, and as of January 31, 2013, remain on the list. Other keywords relating to religious organizations or practices are also present on both lists, most notably 99 keywords relating to Falun Gong.

Wenzhou Train Crash

On July 23, 2011, two high-speed trains travelling near the city of Wenzhou collided, killing 40 people. The government response to the accident was met with widespread criticism, including allegations that sections of the damaged trains were ordered to be buried as a means of hiding evidence [110]. Zhang Dejiang, vice premier in charge of transportation (and later Bo Xilai's replacement as party chief of Chongqing) received criticism for his handling of the rescue operations.

Zhang Dejiang's name (张德江) was a consistent presence on many of the TOM-Skype lists that predated the train crash and subsequent controversy. There were

no additions of new keywords following the Wenzhou train crash until March 21, 2012, eight months after the incident. A number of terms, all linking Zhang and the train crash, were added eight days after Zhang replaced Bo Xilai as party chief in Chongqing. It is notable that the additions on March 21 included nine terms referring to Zhang's role in the Wenzhou train crash, including "Vice Premier Zhang, train" ("张副总动车") and "Dejiang buried train crash" ("德江动车埋").

It is unclear why it took eight months for these terms to be added to the TOM-Skype lists. It is possible that the controversy achieved additional political salience after Zhang's promotion to Chongqing party chief. A number of other terms relating to Zhang were also added on March 21, 2012, including "Dejiang SARS" ("德江 SARS"), referring to Zhang's position as party secretary of Guangdong province where the SARS crisis broke out in 2003, and "Dejiang Nanducase" ("德江南都案"), referring to a prior controversy involving Zhang and a corruption case at a Chinese newspaper. Thus, we see that upon being promoted, a number of terms relating to Zhang were added that reference prior controversies. However, the train crash itself was a highly sensitive event, coverage of which government officials tried to limit. Reports leaked online days after the crash indicate that government authorities issued instructions to print and online media to limit publication of stories about the incident [38].

Tibetan Self-immolations

Over the course of 2011–2012 an unprecedented wave of self-immolation protests took place in Tibetan areas of China. The self-immolation of 20-year old monk Phuntsog on March 16, 2011 marked the beginning of this wave and the first instance of this controversial form of protest in the Tibetan community since February 2009. Since March 2011, 119 Tibetans have self-immolated as a form of protest against CPC policies around Tibet and Tibetan culture, undermining CPC assertions that

Tibetans favor and benefit from Chinese government policies. Of the 119 Tibetans who have self-immolated, 100 were confirmed dead following their protest [82]. This series of self-immolations has been met with aggressive crackdowns by the Chinese government.

In Sina UC lists, "self-immolation" ("自焚") is the only keyword related to the issue, and has been present in Sina UC List 2 since our data collection began. Two months after the self-immolation of Phuntsog, the keyword "self-immolation" ("自焚") was added to the TOM-Skype 5.1 Surveillance-only list; it was subsequently removed May 17, 2011.

For TOM-Skype, no further self-immolation related keywords were added until March 21, 2012. Between March 16, 2011, and March 21, 2012, 29 Tibetans self-immolated. However, the keywords added to TOM-Skype on March 21 focus on only one incident, the self-immolation of a 30-year old monk named Jamyang Palden. On March 14, 2012, Jamyang Palden self-immolated, marking the 27th immolation in Tibetan areas of China since February 2009 and the first in Rebkong (in Tibetan) / Tongren (in Chinese) county. The incident was followed by demonstrations of Tibetan monks and lay persons against Chinese rule. Later on the same day, approximately 4,000 students engaged in protests over Tibetan language rights across three counties in the Qinghai province: Rebkong (in Tibetan) / Tongren (in Chinese), Tsekhog (in Tibetan) / Zeku (in Chinese), and Kangtsa (in Tibetan) / Gangcha (in Chinese) [112]. In the Tsekhog protests, students demanded equality for all nationalities and freedom of language, and called for the end of Chinese military barracks in the area. The protest's specific reference to the Chinese military presence was reported as the first known reference to be made in a protest since the post 2009 self-immolations and subsequent government response began [117]. Additionally, on March 16, 1,000 Tibetans demonstrated in Gepasumdo county (Tongre in Chinese), Qinghai province demanding the release of 50 monks who had been detained the

previous day for raising the Tibetan flag and engaging in peaceful protest [113].

On March 21, 2012, nine keywords were added to TOM-Skype lists (3.6–3.8, 4.0–4.2, 5.5–6.1 Surveillance-only) referencing Jamyang Palden's self-immolation and the subsequent protests, including: "Jamyang Palden, Monk" ("加央班旦僧人"), "students demonstrations" ("学生示威游行"), "Amdo - pay respects - Longwu monastery" ("安多日贡隆务寺"), "Qinghai, student" ("青海学生"), and "Zeku county, students" ("泽库县学生"). These keywords remain on the TOM-Skype lists. However, since March 21, 2012, no further self-immolation related keywords have been added.

A possible explanation for the focus on the self-immolation issue at this specific time is that Jamyang Palden's self-immolation and the following protests over March came at a particularly sensitive period. March 10 marks the anniversary of the 1959 Tibetan uprising and 2008 unrest in Lhasa, which began as observance of the March 10 anniversary, but turned to riots on March 14. The reportedly large protests that followed Jamyang Palden's self-immolation may have prompted Chinese authorities to relay specific instructions to private companies around censoring and / or surveillance of content related to the events. Therefore, while the issue did not gain attention previously, the protests in March 2012 could have brought attention from authorities to enact pressure on companies like TOM-Skype.

However, if Jamyang Palden's immolation and surrounding protests did force pressure to react, it is surprising that no further keywords related to the issue are added to the client lists, as from the last keyword update (March 21) to January 31, 2013, 70 more Tibetans self-immolated, similar demonstrations occurred in Tibetan areas, and aggressive government responses continued.

Diaoyu / Senkaku Island Protests

China and Japan have been involved in a territorial dispute over a group of islands in the East China Sea for decades. The uninhabited islands, referred to as "Diaoyu" in Chinese, or "Senkaku" in Japanese, have been a source of considerable political tension and public protest. Diaoyu / Senkaku Islands-related content has been targeted for censorship and surveillance in Chinese IM programs in the past. Keywords related to the issue were present on the 2004 QQ keywords list [36] and in the TOM-Skype logs collected in 2008 [129].

The island dispute is one of the few current events reflected in the lists of both TOM-Skype (14 keywords) and Sina UC (8 keywords). Three of these keywords are common between the clients: "Protect Diaoyu" ("保钓"), "Anti-Japan" ("反日"), and "Diaoyu Islands" ("钓鱼岛"). All of the Sina UC keywords appear on the earliest collected list (Sina UC List 2 from August 8 2011, used for censoring usernames), and were therefore likely present before our data collection began. Keywords specific to the island dispute on the TOM-Skype lists include "Protect Diaoyu" ("保钓") and "Anti-Japan" ("反日"), which were added to the TOM-Skype 5.1 Surveillance-only list on May 16, 2011 and removed the following day. Outside of May 16, 2011, there were no keyword updates related to this topic until September 2012 when tensions around the dispute began to escalate.

Relations between China and Japan over the islands deteriorated following a public campaign launched by Tokyo governor Shintaro Ishihara in April 2012, which sought to raise donations to purchase the islands and place them under control of the Tokyo municipal government. On September 11, 2012, the Japanese government bought and nationalized three of the islands. The purchase was claimed to be conducted from the "viewpoint of peaceful and stable management of the Senkaku Islands" and was also perceived as an attempt to block a purchase by Ishihara [27].

This move led to the Chinese government denouncing the purchase as an infringement on Chinese territorial sovereignty [139]. Subsequently, Chinese surveillance ships entered Japanese territorial waters around the islands, and large-scale anti-Japanese protests broke out in more than 80 cities across China.

The scale and often violent character of the protests led some commentators to question how the protests were allowed to take place, and speculate on the possibility of some level of direct or tacit government approval. Suspicions were furthered by anecdotal reports that in early September previously blocked keywords including "Anti-Japan Protest" ("反日示威") and "Boycott Japanese Goods" ("抵制日货") were accessible on Weibo search [74, 71].

On September 15, however, amidst growing and aggressive protests, a leaked directive from the State Council Information Office requested all websites "to inspect and clear every forum, blog, Weibo post, and other form of interactive content of material concerning mobilizing anti-Japan demonstrations, stirring up excitement, rioting and looting…" [39]. Anecdotal reports indicate that on September 18, the following keywords were blocked on Weibo Search: "beating, smashing and looting" ("打砸抢"), "Liangmaqiao" ("亮马桥"), the location of the Japanese embassy in Beijing, "thug" ("暴徒"), and "school closure" ("封校"), apparently related due to a number of schools being as a result of the escalation of protests [45]. On September 19, a further number of keywords were reported to be blocked on Weibo search: "anti-Japan" ("反日"), "anti-Japan" ("抗日"), "smash + car" ("砸+车"), and "smash" ("打砸") [44].

These events correlate with changes in the Sina UC and TOM-Skype keyword lists. On September 17, 10 related keywords were removed from the Sina UC list: "protect Diaoyu Islands" ("保钓"), "anti-Japan" ("反日"), "vandalism" ("打砸抢"), "boycott Japanese products" ("抵制日货"), "Japanese embassy" ("日本大使馆"), "Japanese embassy" ("日本使馆"), "Japanese consulate" ("日本领事馆"), "demon-

strate" ("游行"), and "Diaoyu islands" ("钓鱼岛").

On September 20, 11 related keywords were added to the TOM Skype 5.5–6.1 Surveillance-only list: "protesting at embassy" ("使馆游行"), "protect Diaoyu islands" ("保钓"), "sailing out and landing on the island" ("出海登岛"), "anti-Japan" ("反日"), "throwing eggs" ("扔鸡蛋"), "protest" ("抗议"), "slogan" ("标语"), "banner" ("横幅"), "demonstrate" ("游行"), "molotov cocktail" ("燃烧瓶"), "demonstration" ("示威"), "joint" ("联署").

The TOM-Skype keyword updates follow the pattern of increased restrictions around the issue following the September directive. However, the removal of keywords used to trigger username censorship on Sina UC do not appear to have any sensical purpose and could be the product of a technical or human operator error.

Sensitive Events Without References on Keyword Lists

In contrast to the above cases, notable political developments occurred during the collection period that were either not represented or seemingly underrepresented in the keyword lists. Given the importance of these events, which included organized protests and China's once-a-decade leadership transition, relative to other events that appeared in the keyword lists, their exclusion is unexpected.

In January 2013, controversy emerged after a New Year's editorial from prominent Guangdong-based newspaper Southern Weekly calling for strengthened constitutional rights was censored. Protests outside the offices of the newspaper led to arrests, as well as a notice from central government authorities instructing media and websites to publish a government-sanctioned editorial on the story [46]. Reports indicated that many terms relating to the controversy were blocked on Sina Weibo. However, no keywords relating to the controversy were found on any of the keyword lists.

A number of sensitive events relating to Hong Kong occurred during the data collection period but did not appear in any keyword lists. Legislative elections occurred in Hong Kong in September 2012, which led Chinese government authorities to issue instructions restricting media coverage of the elections [41]. That same month, widespread protests occurred following a plan by Hong Kong authorities to introduce changes to the educational curriculum, which were criticized as a means of indoctrinating students into CPC doctrine [32]. Neither of these events appeared in the keyword lists, and no keywords relating to Hong Kong were added to the lists after May 2011.

In November 2012, the 18th National Congress of the Communist Party of China was held, hosting the once-a-decade leadership transition that saw Xi Jinping become paramount leader of the CPC. This meeting is one of the most important political events in China, made even more sensitive following the Bo Xilai scandal earlier in the year. In total, only four keywords relating to the event were added to the TOM-Skype lists, including "18 great" ("十八大") and "name successor" ("立接班人"). These words were added in May 2011, a full year and a half before the event took place. That these words were added so far in advance of the event is not necessarily surprising, as China's leadership transition process is scheduled long in advance. However, given the heightened sensitivity and significance of the event and TOM-Skype's response to other important political developments, the addition of so few keywords related to the event, as well as the lack of words added in the period leading up to the event, is unexpected. Reports have indicated that during the run-up to the Congress in November 2012, Sina Weibo blocked a number of terms relating to the event [43] and manipulated the results of dozens of CPC officials' names [108], most of which do not appear in our dataset.

The absence and limited representation of these events on the keyword lists illustrates the challenge in identifying which political events have sufficient importance to

be added to keyword lists and calls into question how TOM-Online and Sina determine which events to target in their chat clients, as well as how official instructions are given around them.

3.5 Conclusion

In this work we reverse engineered the censorship mechanisms of Sina UC, an instant messaging app developed by Chinese technology giant Sina. We also developed a novel application of DLL injection to reverse engineer TOM-Skype, a product of a joint partnership between Microsoft and TOM to modify Skype to comply with Chinese regulations. This method revealed the exact nature of TOM-Skyp's censorship and surveillance mechanisms, whose existence had been first speculated about years prior.

Unlike other studies of censorship and surveillance in China, our work draws on complete lists of keywords used to trigger censorship and surveillance in two instant messaging applications, offering a less biased and more complete picture of censored and surveiled topics in those programs. Our dataset also enables a comprehensive view of how information controls were modified in Sina UC and TOM-Skype over a period of 21 months, providing insight into how and when content was targeted for censorship or surveillance. We observed some keyword updates that appeared to be in reaction to politically sensitive events, which in some cases were in line with official directives given to media and Internet companies. However, other events that were clearly issues of concern for the CPC and targeted by other Chinese Internet services were not present in the dataset. This inconsistency raises questions regarding if, how, and when official directives may be communicated to TOM-Online and Sina and the level of discretion with which the companies operate.

As an exploratory exercise we compared our dataset to two datasets of words

found blocked on Weibo collected by Jason Q. Ng [105] and China Digital Times (CDT) [47]. Only 330 unique keywords from our dataset were found in either of those datasets. Of the 282 keywords found in both our dataset and [47], 132 were in TOM-Skype lists, 84 in Sina UC lists and 66 were on the lists of both clients. Of the 100 keywords found in both our dataset and [105], 29 were in TOM-Skype lists, 47 in Sina UC lists and 24 were on the lists of both clients. Fifty-two keywords were shared in common between all three lists. Including the preliminary analysis of QQ Chat described in Section 3.4.8, these initial comparisons (albeit exploratory and incomplete) raise questions regarding how the operation and targeting of information controls may vary between different applications and companies.

Overall our findings suggest that the implementation of censorship and surveillance features in Chinese services can be impacted by the actions of the private companies and operators who manage them. These decisions may affect what keywords are targeted, from highly specific content that could be used to monitor discussion of social mobilizations (*e.g.*, Jasmine Rally locations and instructions) to overly broad keywords that could result in inadvertent blocking and mass surveillance.

Chapter 4

Censorship in Live Streaming Platforms

In this chapter we describe the censorship and surveillance mechanisms built into Chinese chat software by analyzing four of the most popular live streaming platforms in China: YY, Sina Show, 9158, and GuaGua. We reverse engineered these applications, revealing the exhaustive lists of keywords used to trigger censorship and/or surveillance in them. We discovered 17,547 unique keywords overall.

We wanted to revisit two questions: what topics do the producers of chat products censor by filtering keywords? And do censors receive keywords from a common source such as the Chinese government—or are they tasked with coming up with the keyword lists themselves? In our previous chapter, although we compared the exhaustive keyword lists of two contemporary chat apps, these apps were not among the most popular in their industry segment. In this chapter, we examine the most popular live streaming apps in China which have hundreds of millions of users. The 17,547 unique sensitive keywords we discover is an order of magnitude larger than the dataset in the previous chapter.

We summarize our major contributions as follows:

1. We reveal 17,457 unique keywords used to trigger censorship and/or surveillance in the most popular apps in the live streaming industry segment.
2. We cross-compare these keyword lists in addition to TOM-Skype, Sina UC, and another available dataset and find mutually little overlap except for products from the same companies. This demonstrates that these keyword lists cannot be largely directed to these companies from a common source.
3. We track additions to these lists for 16 months and categorize new keywords according to their topic.
4. We find that additions to these lists have little overlap in terms of specific keywords or in terms of the high level topics or events that these keywords are added in response to.

4.1 Background

In this section, we give a overview of live streaming platforms and summarize the regulatory framework governing live streaming platforms in China.

4.1.1 Live Streaming Platforms

This study provides a broad look into keyword censorship and surveillance across live streaming platforms, a popular class of applications in China. Live streaming platforms combine real-time video streaming and social networking features that enable users to broadcast content and create interactive groups. One of the most popular uses is broadcasting karaoke performances. Live streaming platforms are

Company	**Product(s)**	**Registered Users**	**MAUs**
YY Inc.	YY	861.4 mn.	117.4 mn.
Tian Ge	9158, Sina Show	245.0 mn.	14.4 mn.
Jinhua Changfeng	GuaGua	70 mn.	–

Table 4.1: **Live streaming users by platform**

primarily monetized through the sale of virtual goods (such as virtual roses) that users give to performers during broadcasts. While musical performances account for the majority of revenues, live streaming platforms are expanding to gaming, education, financial analysis, and online dating applications.

The most popular live streaming platforms in China include YY, Sina Show, 9158, and GuaGua. YY is developed by YY Inc. based in Guangzhou, China, and is the largest platform in terms of user population. As of December 2014, YY had 861.4 million registered users and 117.4 million average monthly active users (MAUs) [13]. In November 2012, YY Inc. announced an initial public offering on the Nasdaq stock market. It is currently the only Chinese live streaming company to be traded on the US stock market.

Tian Ge Interactive Holdings Limited based in Hangzhou, China owns and operates two live streaming platforms: 9158 and Sina Show. In 2010, Sina Corporation invested 10 million dollars (representing a 25% stake) in Tian Ge and provided the company a sole license for the operation of of Sina Show. Tian Ge reports user numbers as aggregates across its platforms and in 2014 had 245 million registered uses and 14.4 million MAUs [7, 8]. In July 2014, Tian Ge went public on the Hong Kong Stock Exchange.

Jinhua Changfeng Information Technology Co., Ltd., is a privately held company based in Zhejiang Province, China that provides the GuaGua platform, which as of

2013 had 70 million registered users [20]. See Table 4.1 for a breakdown of user bases across live streaming platforms.

4.1.2 Legal and Regulatory Environment in China

In public filings for YY and Tian Ge both companies underline the risk that their businesses face from potential legal sanctions being brought against them for hosting prohibited content [6, 13]. The companies also highlight the risk of being affected by government campaigns such as "Clean the Web 2014," which was an government effort to crack down on the creation and dissemination of pornographic content online. During this campaign, Sina Corporation received notices regarding prohibited content on its platforms and was subsequently fined 5.1 million RMB[1], had licenses temporarily revoked, and saw its stock price drop as a result. This campaign demonstrates the dynamic nature of Internet regulations in China, and shows companies are subject to unpredictable enforcement, which can impact their bottom line.

Unlike most other social media platforms in China, live streaming platforms have an added dimension of sharing revenues with performers. This business model and the general live streaming user experience encourages performers to keep audiences engaged and spending on virtual goods. The popularity of live streaming applications, the diversity of real-time media content, and the virtual goods business model puts these platforms under particular pressure to monitor and manage user activity.

[1]RMB, or *renminbi*, is the Chinese system of currency.

4.1.3 Content Monitoring and Censorship on Live Streaming Platforms

To comply with China's laws and regulations live streaming platforms manage content through a combination of terms of service (TOS), automated content monitoring and filtering systems, and dedicated review teams.

YY has an extensive TOS that includes descriptions of prohibited content and a five-level system for penalties that range from freezing the account from 7 days (level 1), 30 days (level 2), 120 days (level 3), 360 days (level 4), or permanently (level 5). Serious violations that warrant a level 4 or 5 response include publishing pornography; publishing content that endangers national security or undermines national unity, social stability, or national religious policy. Offenses that carry lower level punishments include vulgar jokes, verbally abusing other users, and copyright infringement [13]. Performers are expected to obey an additional list of regulations that include the same high level prohibitions and other specific guidelines on inappropriate attire and performance material. Failure to comply can result in fines and account suspensions [14].

To enforce these TOS, YY has a team within the data security department that maintains "24-hour surveillance" on content and is supported by a system that periodically "sweeps" the platform for offensive content and "automatically" filters keywords. The company also describes a voice monitor system that provides "various alerts on sensitive words or abnormal activities of users, channels, or groups" [13].

Tian Ge has a similar combination of controls. It employs a team of 74 content monitors who identify TOS violations and enforce internal policies. In public filings the company describes an image processing system used to detect skin tone and facial features to flag nudity or "sexually suggestive partial nudity." Screenshots of video chat rooms are randomly captured every one to three minutes and processed through

the detection system. Flagged content is then sent to the content monitoring team for further review. The company also generally describes audio monitoring and keyword filtering systems. In addition, it provides the same level of access that its content monitoring team has to the Jinhua City Municipal Public Security Bureau to allow the authorities means to "monitor and supervise the activities" on the platform [6].

As GuaGua is a private company less information is available on its internal operations. In a 2013 interview, co-founder Dong Guanjie claims the platform has a content management team of over 100 staff [20]. GuaGua's TOS explains the company performs automated and manual inspection of content and deletes any infringements [1]. Similar to YY and Tian Ge emphasis is placed on incentives for user self-regulation and financial penalties for violations.

4.2 Technical Analysis

In this section, we describe the technical implementations of keyword censorship and, if present, keyword surveillance in each of the live streaming applications we analyzed. We reverse engineered these apps using the same general techniques that we described in the previous chapter to reverse engineer instant messaging apps.

We found that three Chinese live streaming platforms not described in this paper, VV (*51vv.com*), Sixroom (*6.cn*), and BoBo (*bobo.com*), perform keyword censorship on the server-side. We determine that an application censors on the server-side by first ensuring that there are no obvious signs of a client-side censorship implementation such as one of its censored keywords appearing in plain text in any of the application's files. Then we compare the packet trace of sending a censored keyword (*e.g.*, "falun") with sending a nearly identical uncensored keyword ("galun") and, if the application censors by asterisking out sensitive keywords, with sending that keyword asterisked out ("*****"). If the first comparison is repeatedly similar and,

if performed, the second comparison is less so, we conclude that the censorship is server-side. If the application displays a warning message upon entering a censored message, we also ensure that the warning message does not display when the application has no access to the Internet. We leave it as future work to analyze the server-side censorship implementations used by these platforms.

4.2.1 YY Censorship and Surveillance

YY 7.1 downloads three different keywords lists with the following names: *Finance*, *Normal*, and *High*.

The *Finance* keyword list is downloaded from the following URL:

`http://do.yy.duowan.com/financekwordlist`

These keywords are downloaded in plain text in UTF8-encoded XML. Keywords in this list are related to phishing scams. When a user receives a keyword from the list the following warning message is displayed in the chat window: "YY安全提示: 聊天中若有涉及财产的操作，请一定要先核实好友身份，谨防受骗!" (YY Security Tip: This chat seems to involve managing assets; please be sure to verify the identity of a friend to avoid being cheated!).

The *Normal* keyword list is downloaded from the following URL:

`http://do.yy.duowan.com/NormalKWordlist.txt`

It is as a base64-encoded list of UTF16-encoded keywords each separated by a carriage return followed by a line feed. If an outgoing or incoming message contains a keyword from this list, those keywords are asterisked out in the chat window. In the

case of an outgoing message, those keywords are also asterisked out in the message sent over the network.

The *High* keyword list is downloaded from the following URL:

`http://do.yy.duowan.com/HighKWordlist.txt`

Like the *Normal* list, it is a base64-encoded list of UTF16-encoded keywords each separated by a carriage return followed by a line feed. If an outgoing message contains a keyword from this list, that message is silently filtered. If an incoming message contains a keyword from this list, it appears in the chat window as a blank message.

YY Surveillance

The keywords from both the *Normal* and *High* lists are also used to trigger surveillance. When attempting to send a message containing keywords from either of these lists, a surveillance message is sent in an HTTP GET request to a URL of the form:

`http://sere.hiido.com/do.action?id=`**<id>**`&content=`**<content>**

<id> is a hex encoding of a hash computed as

md5(⌊**<seconds since unix epoch>**/1000⌋+
`";username=report"`+
`";password=pswd@1234"`).

Note that the username and password appearing in the hashed string are hardcoded; these are not the username and password of the sender or receiver of the triggering message. **<content>** is a base64 encoding of the following string:

type=2;uid=**<sending user id #>**;touid=**<receiving user id #>**;keyword=**<triggering keyword>**;txt=**<entire triggering message>**

Type is hardcoded to 2.

4.2.2 Sina Show Censorship

Sina Show 3.4 comes installed with keyword lists for censoring messages and also downloads additional keywords remotely from its servers. When an outgoing message is censored, the message is not sent, and the following warning message is displayed in the chat window: "系统过滤，你发送的信息含有非法字符，请重新输入！" (System filter: the message you sent contains illegal words; please re-enter!) When an incoming message is censored, the contents of the incoming message are replaced by the following message in the chat window: "发送的消息有非法词汇，已经被自动屏蔽" (The message sent contains illegal vocabulary; it has been automatically blocked.)

Sina Show comes installed with a binary database of keywords in a file named `Word_410.ucw` and downloads updates for it from the following URL:

`http://www.51uc.com/uc_interface/down_policy/Word_410.ucw`

This file is a custom binary container storing sensitive GBK-encoded keywords that have been encrypted using Blowfish-like algorithm in ECB mode with the 8-byte key `Dey,1blE`. A standard library implementation of Blowfish cannot be used to decrypt these keywords, however, as the Blowfish implementation used by Sina Show is atypical, containing byte endianness inconsistencies and in multiple places shifts bits by different amounts.

While these inconsistencies may have been unintentional bugs, they significantly increased the difficulty of reverse engineering the decryption algorithm. In order to implement the algorithm we needed to implement Sina Show's Blowfish algorithm bug for bug. We first used the Hex-Rays decompiler [2] to decompile the decryption code to a C-like pseudocode. We then carefully modified the code into compilable C code while being cautious to not change the program's semantics. While at this point we had code that could successfully decrypt Sina Show's modified Blowfish, the code was humanly incomprehensible and revealed little insight into what made this implementation of Blowfish different from others. To flesh out the differences, we iteratively modified this C code to resemble the source code of a standard Blowfish implementation, after each modification verifying that the compiled C code could still decrypt Sina Show's modified Blowfish. After we could no longer modify the C code to resemble the standard implementation without changing its behavior, only the relevant semantic differences remained, revealing the byte endianness inconsistencies and differences in bit shifts that made Sina Show's Blowfish implementation different.

After decrypting the keyword database file, we found that each keyword in the file is associated with a category number from 1 to 8, inclusive, or 12. For this reason, keywords often appear more than once in this file if they belong to multiple categories. However, Sina Show at present only utilizes category 5, which it uses to censor chat messages. If the original purpose of the additional categories was like that found in Sina UC, it may be the case that the other categories were originally used to censor usernames or other strings in an older version of the program. As of May 11, 2015, 888 of the 2709 keywords in this file are in category 5.

Sina Show also has GBK-encoded lists of keywords included in plain text built into many of its binaries. `SinaShow.exe` contains a list of 1224 keywords, which are also included in `ChatRoom.dll` and `Props.dll`. `UCClient.dll` also includes another list of 910 keywords. However, among all of these built-in keywords, only

108 keywords, the 1114th through the 1221st keywords in `SinaShow.exe` are actually referenced by any binary code, and these are also used to censor chat messages. The other unused keywords may correspond to presently unused categories as we saw with the downloaded keywords.

4.2.3 9158 Censorship

9158 6.9 is installed with two lists of keywords, `filnick.xml` and `filter.xml`. Although these XML files self-identify as being GB2312-encoded, they are really GB18030-encoded. The former list is used to replace sensitive keywords with asterisks in user names, whereas the latter list is used to replace sensitive keywords with asterisks in both outgoing and incoming chat messages. Updates to the latter list are also downloaded from the following URL:

`http://mimtenroom.9158.com/web9158/filter.zip`

The `filter.xml` file also includes a version number of the list, which is an integer in the hundreds that we have found to increase by a few every time the list is updated. The version number of the list installed with the program, however, is greater than the version number in any of the updates we have seen offered for download, suggesting that the sequence may have reset or forked at some point.

In addition to keyword censorship, we found that if a chat message contains six or more English alphabet letters, then all of its English alphabet letters are asterisked out. The intent of this filtering is not clear. Aside from stifling English conversation, this may be intended to filter out URLs. Given that their keyword lists filter keywords like *http*, *www*, and *com*, it would seem they intend to filter all URLs.

Keyword List	**Keywords**	**Unique**
YY *Finance*	48	18
YY *Normal*	20	20
YY *High*	13,482	13,242
9158 *Nick*	65	59
9158 *Chat*	318	318
Sina Show `SinaShow.exe`	1,224	910
Sina Show `UCClient.dll`	910	910
Sina Show *Downloaded*	3,711	3,206
GuaGua	58	58

Table 4.2: **Keyword list size (May 17, 2015)**

4.2.4 GuaGua Censorship

GuaGua 6.2.38 has keywords built into `RuleCenterPlug.dll`. These keywords appear in plain text, GBK-encoded. Any outgoing message containing one of these keywords is never sent and the following warning message is displayed in the chat window: "消息发送失败,含有违法或不文明字符！" (Failed to send message; it may contain illegal or uncivilized words!) GuaGua does not filter incoming messages.

4.3 Keyword Analysis

In this section, we perform a similarity comparison between the live streaming keyword lists and blacklists from other products. We then categorize the keywords on each live streaming list according to high level themes. Finally, we track changes to the live streaming lists over a period of 16 months and analyze which events or topics these changes are in response to.

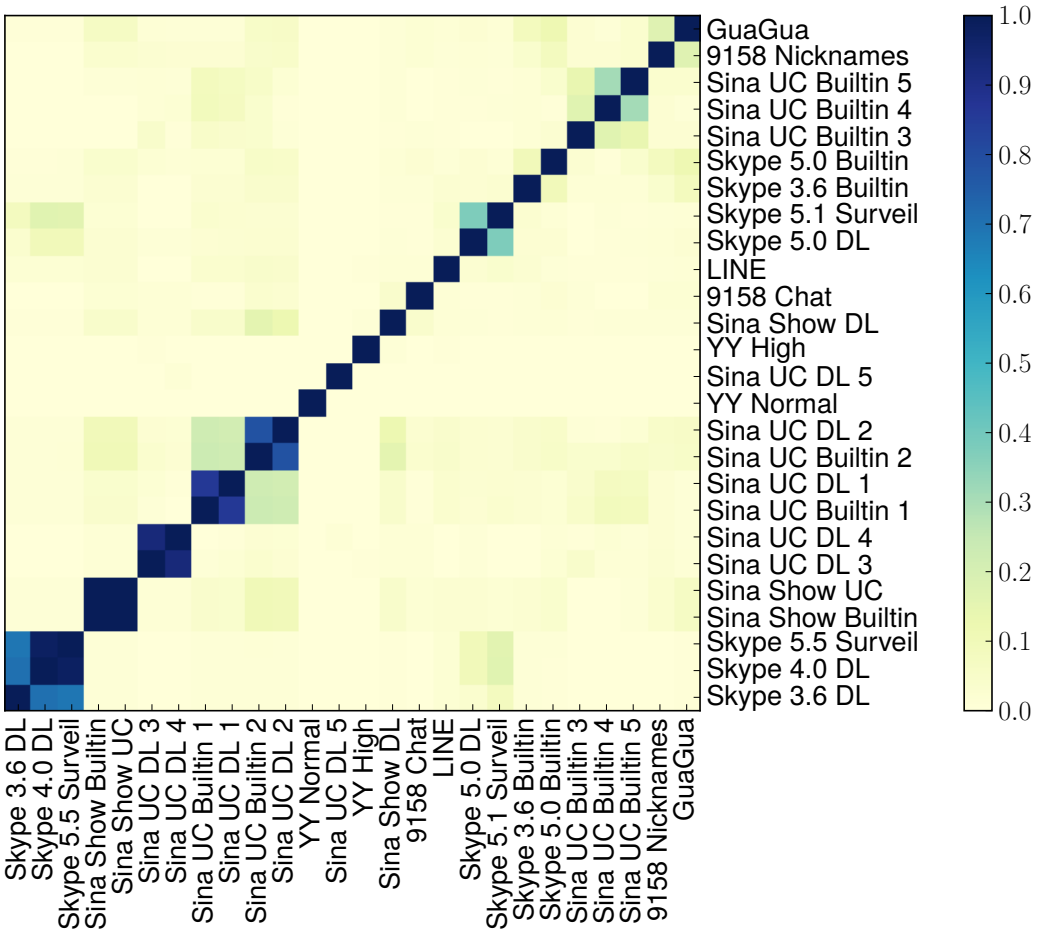


Figure 4.1: **Keyword lists clustered by Jaccard similarity**

4.3.1 Similarity Comparison Across Keyword Lists

The live streaming keyword lists vary significantly with respect to size. The YY *High* list is the largest live streaming list, whereas, if we exclude YY's smaller lists, GuaGua's list is the smallest. See Table 4.2 for a summary of the size of each live streaming list. When including TOM-Skype and Sina UC lists and the latest list from LINE [61], we have a dataset consisting of 42 lists, which together contain 17,547 unique keywords. The lists range in size from 20 to 13,244 unique keywords.

In Figure 4.1, using the centroid linkage method [103], we hierarchically cluster

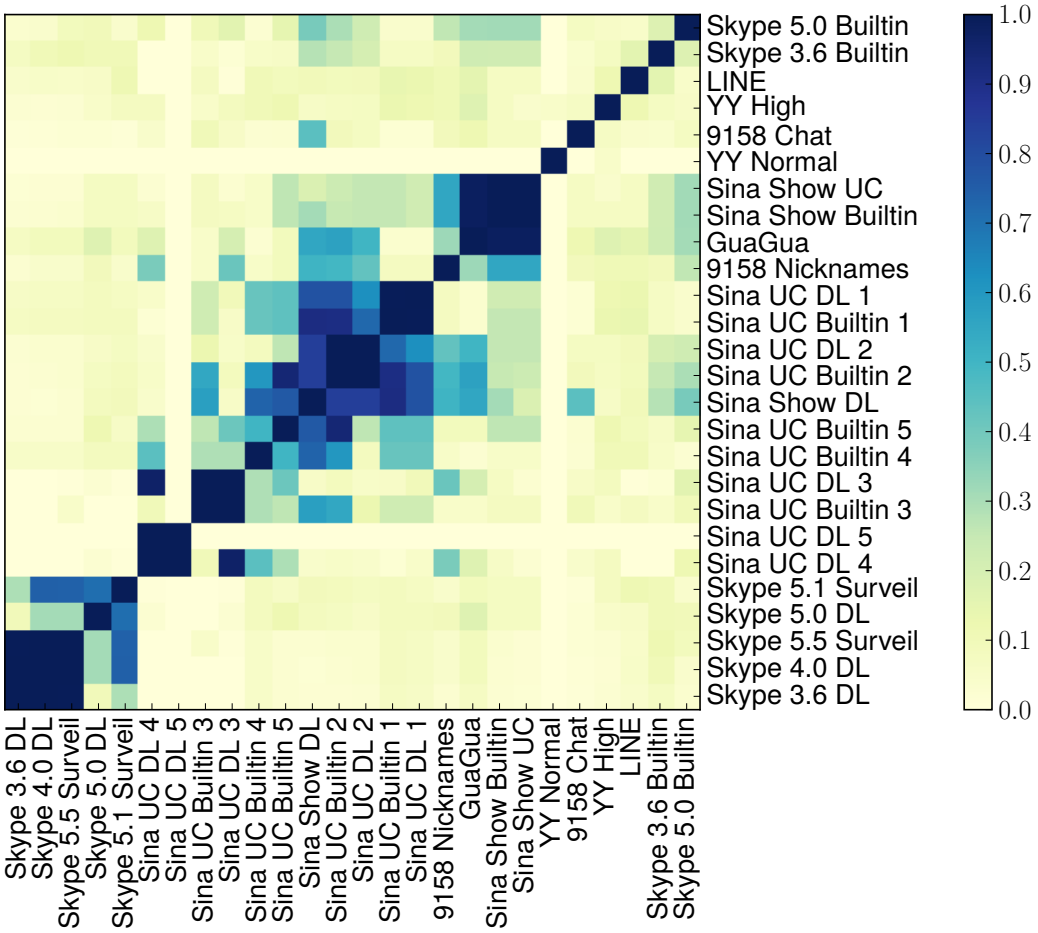


Figure 4.2: **Keyword lists clustered by *Intersection over Smallest***

the live streaming keyword lists along with TOM-Skype, Sina UC, and LINE by the Jaccard similarity coefficient, also known as the *Intersection over Union* metric, by computing

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

Using this method, we find very little similarity between lists, and when lists are similar they are lists within the same company.

In Figure 4.2, we cluster the same keyword lists using a different similarity metric we call the *Intersection over Smallest* metric. We compute list A's similarity to B

Theme	Example Categories
Event	Scheduled events, current events
Political	Communist Party of China, ethnic groups
People	Government officials, dissidents
Social	Gambling, prurient interests
Technology	URLs, apps
Misc	No clear context

Table 4.3: **Content themes and related categories**

as max(% of A in B, % of B in A). The intuition behind this metric is that it would tease out lists that inherit from other lists. We call this metric *Intersection over Smallest* because it is equivalent to computing

$$IoS(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)}.$$

Although using this method we see more lists similar to each other within companies, lists from different companies remain mostly dissimilar with one exception: GuaGua is similar to many Sina Show lists. Closer inspection reveals that the GuaGua list is a near exact duplicate of a 2004-era list built into Sina UC that Sina Show's built-in lists build upon. The only difference in the GuaGua list is the addition of a single keyword. Both of the founders of GuaGua formerly worked on audio chat software at Langma UC (acquired by Sina Corporation in 2004 to become Sina UC) and Sina [20]. This employment history may explain why the GuaGua and Sina lists are so similar.

4.3.2 Keyword Content Analysis

We used a combination of machine and human translation to translate any Chinese keywords in the live streaming dataset to English, and then we analyzed the context

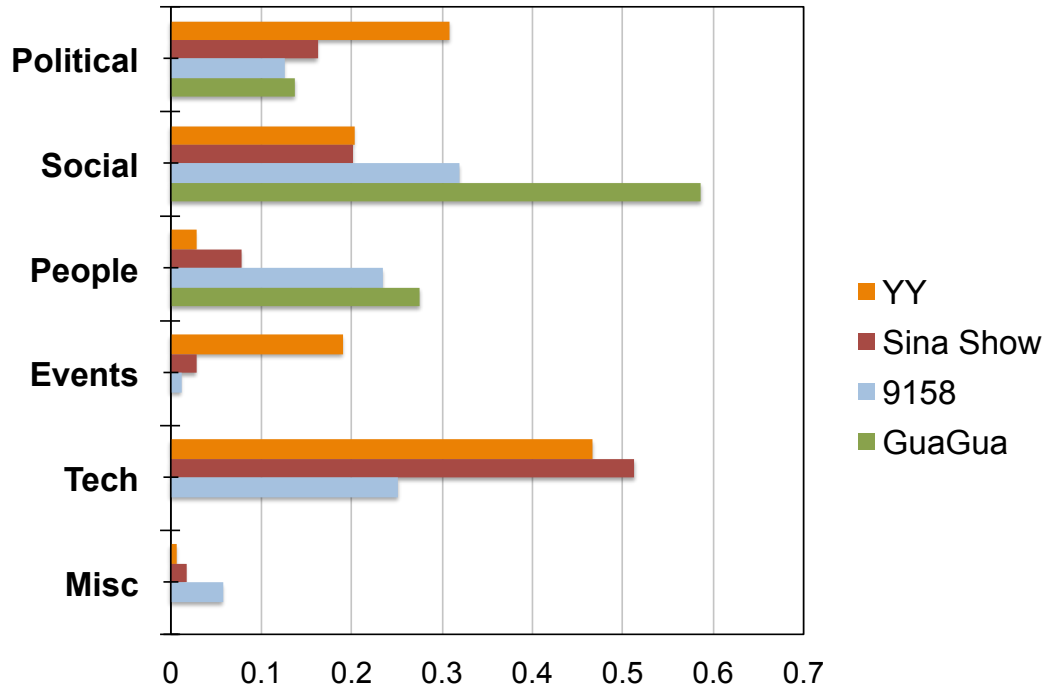


Figure 4.3: **Breakdown of list source by theme**

behind each one. Based on these translations and contextual information three researchers coded each keyword into one of six general themes based on the code book developed in the previous chapter (see Table 4.3). We performed interrater reliability checks throughout the categorization process.

Figure 4.3 shows for each of the four platforms the percentage of its blacklisted keywords that fall into each theme. Keywords related to the social, political, and people themes are the most common across all four platforms.

Keywords were typically either Chinese, English, or both; however, we found that 0.5% of the YY *High* list are Uyghur keywords in Arabic script, which we had not seen in our earlier work. They include references to terrorism and Islam. For example:

پارتلىقۇچ ياساش دەرسلىكى

which is a Uyghur phrase that translates to "instructions to create explosives." The

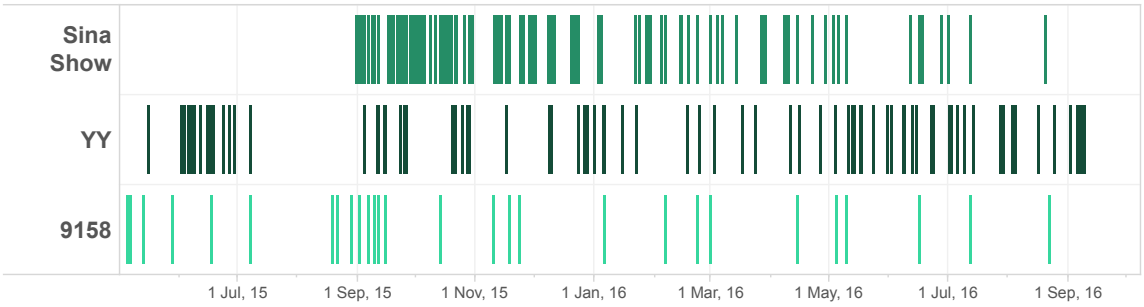


Figure 4.4: **Distribution of keyword updates between May 18 2015 and September 30 2016 (Universal Standard Time)**

Platform	New keywords
YY	1,468
Sina Show	*266
9158	310

Table 4.4: **Keywords added by platform.** *The total for Sina Show is 1,239 when including strings of numbers that we suspect are primarily phone numbers, which we excluded from our analysis.

focus on Uyghur related content relative to previously available keyword lists may have been motivated by a June 2014 government campaign to censor terrorist content following attacks in the Xinjiang region. Thirty Chinese Internet companies signed a "letter of commitment" to block such content [21].

4.3.3 Keyword List Changes

In this section, we analyze changes occurring to keyword lists between May 18, 2015 to September 30, 2016. Over this collection period, we downloaded YY, Sina Show, and 9158 lists hourly. (GuaGua does not download updates to its keyword list, and so it is not analyzed in this section.)

Metric	**9158 *vs.* Sina Show**	**YY *vs.* 9158**	**YY *vs.* Sina Show**
Jaccard similarity	17.28%	1.21%	0%
Intersection over Smallest	41.56%	4.02%	0%

Table 4.5: **Additions to keyword lists compared by Jaccard similarity and by *Intersection over Smallest***

We collected a total of 2,044 additional keywords. Table 4.4 provides a breakdown of unique keywords added by each application since May 18, 2015, excluding any keywords we had already seen on that platform up until that date.

While YY added the most new keywords, Sina Show had more frequent updates to keyword lists (158 updates) compared to YY (138 updates) and 9158 (33 updates). Figure 4.4 shows the distribution of updates over the collection period by each application. Sina Show changed the URL of their keyword list download in an updated version the client, and as a consequence we have no data for the first three months of Sina Show updates until we discovered the new URL.

Analyzing similarity in unique keywords between the lists extracted from YY, Sina Show, and 9158 between May 2015 and September 2016 reveals limited overlap between the companies (YY and Tian Ge), but does show commonalities between Sina Show and 9158.

In Table 4.5, we compare the additions made to the keyword lists according to their Jaccard similarity coefficient. The results are very little to no overlap in keyword list between YY and 9158, and YY and Sina Show, with some commonalities between 9158 and Sina Show. We also compare additions to the keyword lists using the *Intersection over Smallest* metric. This metric further confirms our previous result showing greater overlap between 9158 and Sina Show, and limited overlap to

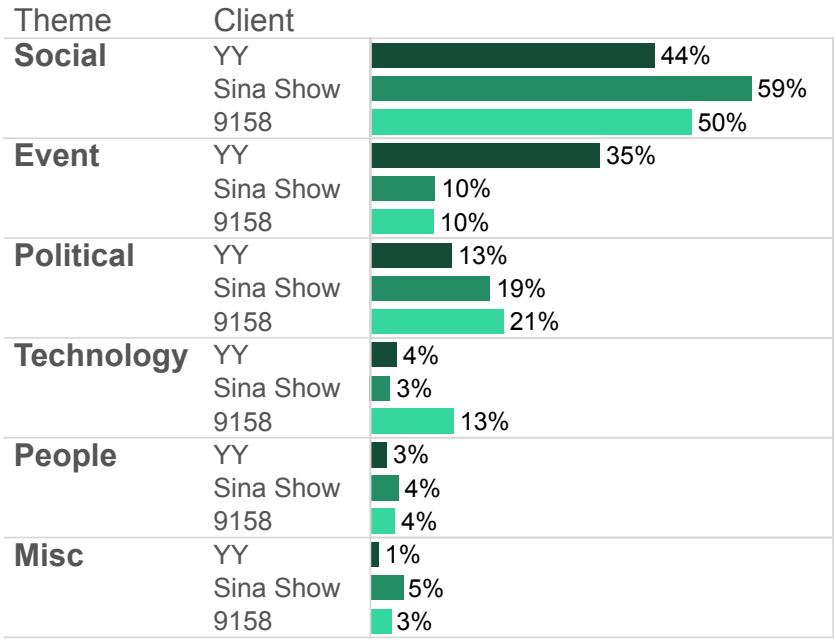


Figure 4.5: **Distribution of themes across three live streaming platforms**

YY, over time.

The overlap between Sina Show and 9158 is not surprising since the platforms are owned by the same company. Despite this common ownership and degree of similarity the keyword lists and timing of list updates are not identical, which suggests Tian Ge does not manage content on the platforms in completely the same way.

4.3.4 Content Analysis of Keyword Additions

We analyzed each of the keyword additions to the live streaming keyword lists, categorizing them into the same high level themes as the other keywords; however, this time we also broke each theme down into smaller sub-categories. Figure 4.5 shows for each of the three applications what percentage of the additional blacklisted keywords belong to each theme. In the following sections, we provide descriptive statistics describing the keyword content and examine each theme in detail.

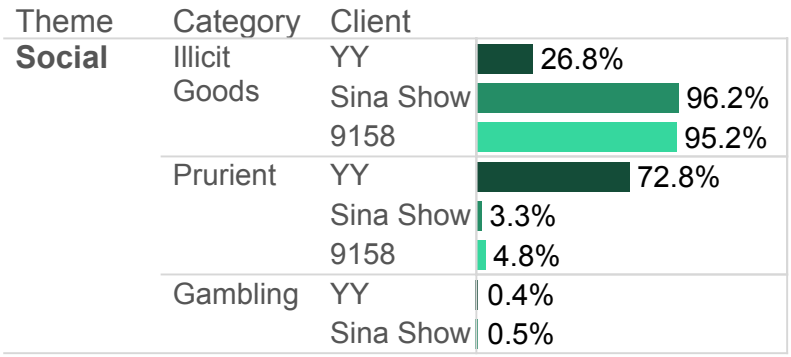


Figure 4.6: **Percentage of Social theme keywords by category**

Social Theme

The Social theme is divided into three categories: gambling (*e.g.*, online casinos), illicit goods and services (*e.g.*, narcotics, weapons, counterfeit products), and prurient interests (*e.g.*, sexuality, pornography, prostitution). Figure 4.6 shows the percentage of Social theme keywords by category.

The Social theme accounts for the highest percentage of keywords added to each application relative to other themes (Sina Show: 59%, 9158: 50%, YY: 44%). The focus on this theme may reflect a reaction of companies to the new regulatory campaigns that specifically target pornography, drugs, and weapons.

Event Theme

The Event theme includes reference to 20 distinct events in keywords added to the applications' keyword lists. We correlate the timing of keyword list updates to events that happened within our collecton period and find reactive censorship driven by current events.

Reporting on current events in China is tightly controlled by government authorities. Media organizations are routinely provided directives on how to report

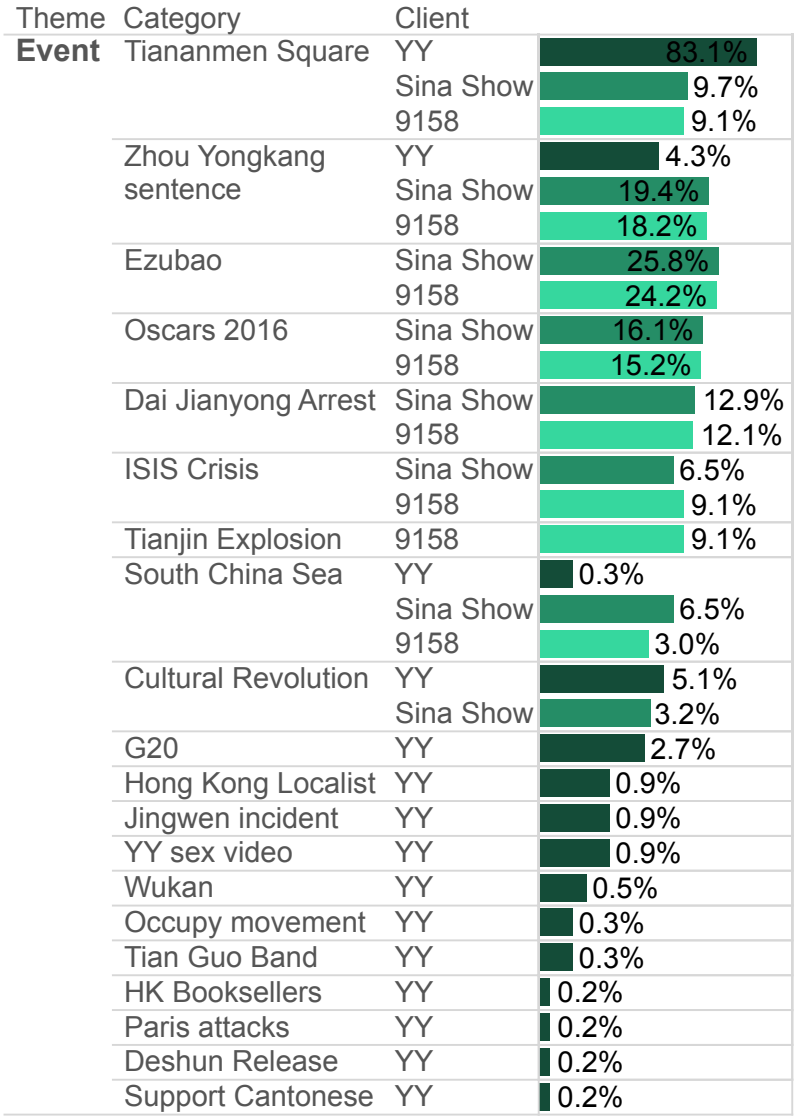


Figure 4.7: **Percentage of Event theme keywords by category**

the news. China Digital Times, an independent media group, occasionally publishes leaked directives sent to Chinese news organizations, which provide a glimpse into how this system works [54]. There have also been leaks from social media companies, such as Sina Weibo, which describe censorship instructions from company managers that purportedly correspond to state directives [58]. However, it is unclear in what form or at what frequency directives are provided and if companies receive the same

Metric	**9158 vs. Sina Show**	**YY vs. 9158**	**YY vs. Sina Show**
Jaccard similarity	73.52%	0%	0%
Intersection over Smallest	92.59%	0%	0%

Table 4.6: **Keyword additions to Event theme in three live streaming platforms compared by Jaccard similarity and by *Intersection over Smallest***

ones.

YY added the largest number of event related keywords (632 keywords) compared to 9158 (33 keywords) and Sina Show (31 keywords). YY also referenced in its additions more unique events (15) than 9158 (8) or Sina Show (8).

In Table 4.6, we compare the additions of event-related keywords across the platforms by Jaccard similarity and our similarity metric. Our results show no overlap in additional event keywords referenced between YY and Tian Ge operated applications, which suggests there are either no common directives provided to these companies or there is varying compliance with directives. However, we do see close similarity between Sina Show and 9158 Event keywords, most likely due to their shared parent company.

Only three events are referenced by all applications in keyword additions (June 4 1989 Tiananmen Square Massacre, the sentencing of Zhou Yongkang, and the Hague Verdict on the South China Sea arbitration).

In keyword additions, Sina Show and 9158 reference the same seven events. The only difference between them is 9158 references the Tianjin Explosion and Sina Show references the Cultural Revolution. Event updates on the two applications are often made within the same period and sometimes on the same day. The close similarities between these applications can explained by common ownership. However, the lack

of overlap in event-related keywords between the platforms shows they still do not share an identical list. Below we examine the three events that all three applications did reference.

The June 4, 1989 Tiananmen Square Massacre remains one of the most taboo events in China. Reactive censorship on social media in China often accompanies the anniversary [50], and the Chinese government continues to push revisionist narratives of what happened.

Between late May and the first week of June 2015, leading up to the 27th anniversary of the Tiananmen Square Massacre, YY added 525 keywords related to the event. Comparatively, 9158 and Sina Show each added three keywords on dates that did not fall close to the anniversary.

At the start of our data collection period, YY keyword lists already had a heavy focus on June 4, accounting for over 90% of YY's event keywords and 32% of YY's lists overall. June 4 related keywords on YY's lists include a number of ways to refer to the event including numerals ("89VIIV"); homonyms (陆4, "Land 4," the character (陆 Lù) sounds similar to six (六 Liù) in Chinese); locations of annual memorial events (维园烛光, "Victoria Park Candle"); and references to recent discussion of the event such as "Trump June 4" (川普六四), which is likely related to Donald Trump referring to Tiananmen Square as a "riot" in an election debate.

On June 11 2015, Zhou Yongkang, who was once one of China's most powerful political figures, was sentenced to life in prison on corruption charges [131]. On June 11, YY added 23 keywords related to the sentencing (*e.g.*, 無期徒刑 "life imprisonment"). Prior to the date of the sentencing, Sina Show and 9158 added references to associates of Zhou who were also implicated in his corruption case including former People's Liberation Army general Xu Caihou (徐才厚) and former Party official Ling Jihua (令计划).

In a case known as the South China Sea Arbitration, the Philippines under provisions of the United Nations Conventions on the Law of the Sea brought complaints against China over territorial claims in the South China Sea [127].

On July 12, 2016, an international tribunal in the Hague ruled in favor of the Philippines and concluded that China has no legal basis to claim historical rights in the South China Sea [104]. China rejected the ruling. On the same day, Sina Show added two keywords (南海仲裁 "South China Sea Arbitration", 海牙 "Hague") and 9158 added one (南海仲裁 "South China Sea Arbitration"). On July 13, YY added two keywords, one referencing China's rejection of the ruling (习总的拒绝 "President Xi's rejection") and another related to a fake news story that went viral on Chinese social media following the verdict, which claimed China and the Philippines had declared war on each other and the Chinese army successfully wiped out a unit of the Philippine Air Force (全歼菲方空军 "Wipe out the Philippine Air Force").

YY keyword lists include reference to 11 events that do not appear on the other applications. Some of these events are clearly sensitive topics.

Wukan is a fishing village in southern Guangdong that has earned renown for activism. In 2016, villagers took to the streets calling for the release of detained democratically-elected local leader Lin Zulian and the resolution of a long-simmering dispute over land sales. China Digital Times published [51] a leaked directive that was issued to news organizations on June 21, 2016, (China Digital Times does not disclose the issuing bodies to protect the sources of the leaks):

> "Regarding former village committee chief of Wukan, Guangdong, Lin Zuluan being investigated and admitting his guilt, websites are strictly prohibited from releasing or re-publishing any news, photos, video, or information related to the mass incident in the village."

On June 22, YY added one keyword (林祖銮 "Lin Zulian") followed by the

addition of two keywords on June 23 (林祖戀 "Lin Zulian," 還我書記 "Return our secretary"). It is unclear if YY received similar directives for handling the Wukan protests. While it is plausible, the lack of any Wukan-related keywords on Sina Show or 9158 suggests distribution of these directives or compliance to them varies.

We observe a similar pattern in the censorship of President Xi's gaffe during his opening speech at the 2016 G20 summit in Hangzhou [79]. During the September 4, 2016 speech Xi mistakenly said "reduce taxes and make roads easy [to travel on], facilitate commerce and loosen clothing" (轻关易道通商宽衣), when he should have read "reduce taxes and make roads easy [to travel on], facilitate commerce and be lenient to farmers" (轻关易道通商宽农).

This slip of the tongue was clearly embarrassing for Xi. China Digital Times published a September 4 leaked directive that instructed online media to "filter and intercept content" related to "tongshang kuannong [通商宽农]," and strictly delete comments, photos, videos, and "related information" [52]. On September 5, YY added 17 keywords including "Xi undress" (習寬衣), "loosen the clothing and undo the belt" (寬衣解帶), and other references to the speech. However, Sina Show and 9158 did not add keywords related to this event.

Events like the Wukan protest and G20 speech gaffe are clearly sensitive to Chinese authorities, and it is surprising to see them only referenced on one application. Other keywords added by YY are related to sensitive events specific to the application. In September 2015, YY added 6 keywords referencing an August 2015 incident during which a YY user apparently forgot to turn off her webcam and had sex with her partner while live streaming (yy出事视频 "yy accident video," 忘关视频被啪 "forgot to turn off the video while having sex"). Videos of the incident circulated on Chinese social media causing a scandal. In this case, it is obvious that YY would be motivated to attempt damage control over the incident as it brings unwanted attention from authorities.

Overall, we find that censorship of events is dynamic and reactive. However, we observe a lack of overlap in the unique events censored by different companies suggesting that there are no centralized directives given to companies or differing levels of compliance. YY is registered in Guangzhou whereas Tian Ge is registered in Hangzhou. Each company has to follow respective municipal and provincial regulations. The companies may therefore be given different directives based on the location of their registration, which other studies have suggested may account for variance in how censorship is implemented. These results demonstrate that events are catalysts for censorship but the ways in which they are managed is not uniform.

Political Theme

The Political theme includes 18 categories related to issues including the Communist Party of China (CPC), ethnic minority groups in China, religious movements, and terrorism.

All three applications added keywords related to the CPC. This includes general references to the structure of the party and its various departments (*e.g.*, 中央政治局 "politburo," 中共中央 "CPC Central Committee"); allusions to factional struggles within the party (e.g, 习近平阵营和江派 "Xi and Jiang faction camp"); and pejoratives (*e.g.*, 共匪 "Communist bandits").

Keywords related to the Uyghur ethnic minority were also added by all of the applications. These keywords appear in Chinese and in the Uyghur language in both Arabic and Latin script. The content of other Uyghur-related keywords range from religion (东突穆斯林 "East Turkestan Muslim"), violence (partila "explode"), to separatism ("Turkestan Islamic Party," an Islamic separatist organization founded by Uyghur militants). Other keywords are more cryptic without clear context such as Uyghur words for "cloudy weather" and "sweet potato."

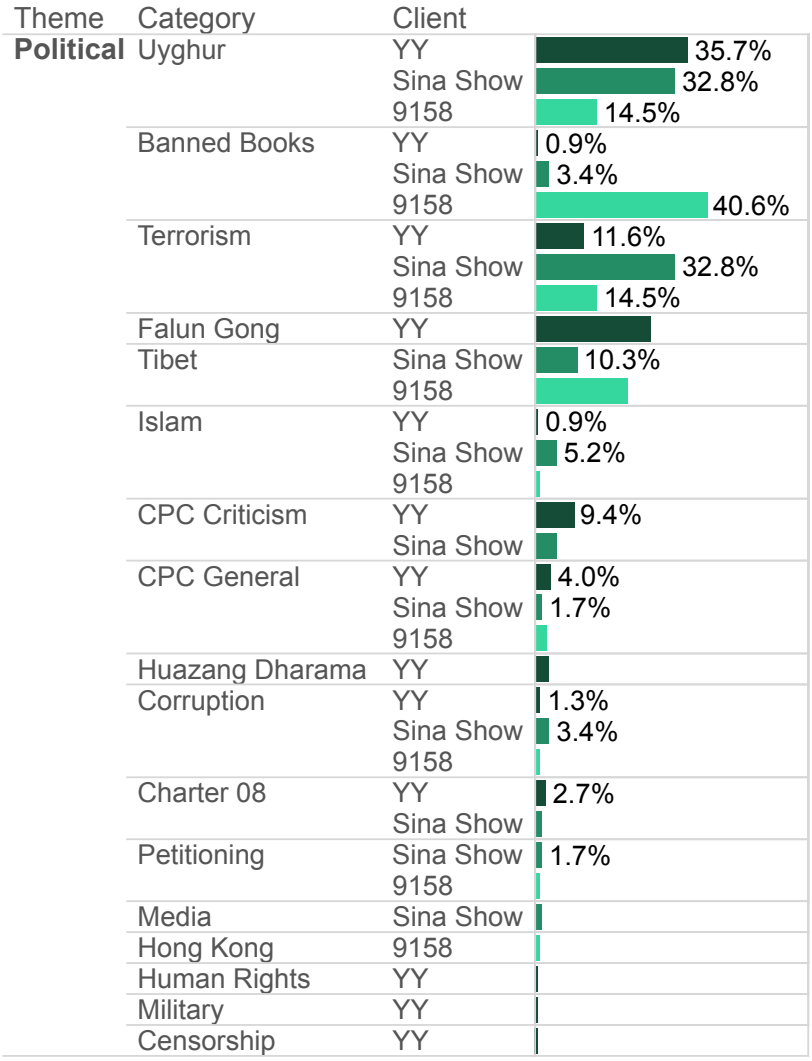


Figure 4.8: **Percentage of Political theme keywords by category**

Titles of books dealing with sensitive topics that have been banned in China also appear on each application. These books, predominantly published in Hong Kong and Taiwan include discussions of power struggles within the CPC (*e.g.*, 老江气杀习大, "Old Jiang Enrages Uncle Xi"), and fiction critical of communist rule (*e.g.*, 黃禍, "Yellow Peril" written by Wang Lixiong). China has strict regulations on the publishing industry [126], pushing dissident and tabloid authors to Hong Kong and Taiwan to publish on sensitive topics. The sale of banned books was

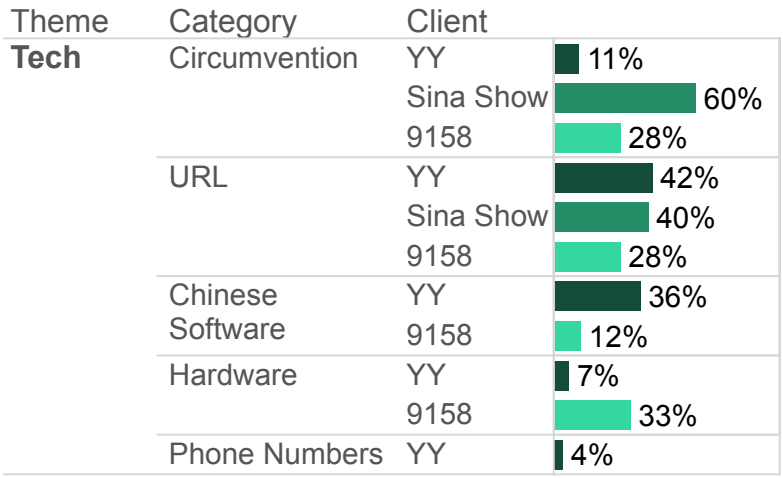


Figure 4.9: **Percentage of Technology theme keywords by category**

highlighted in 2015 after five employees of a book shop and publishing firm in Hong Kong specializing in taboo titles went missing, only to later emerge in custody in mainland China. Their disappearances had a chilling effect on publishers in Hong Kong who pulled sensitive titles from their shelves [125]. One of the booksellers, Lam Wing-kee, revealed details of his detention at a press conference in Hong Kong on June 16, 2016. His name (林荣基) is included in the keywords lists on YY under the event theme.

Technology Theme

The technology theme has five categories including censorship circumvention tools, URLs, hardware devices, Chinese software and websites, and phone numbers.

The hardware category includes 25 references to drones and other unmanned aircraft (e.g, 四旋翼无人机 "quadcopter";). While it is unclear why these keywords are censored, there is rising concern in China regarding safety, privacy, national security issues and increasing regulations on drone technology [66].

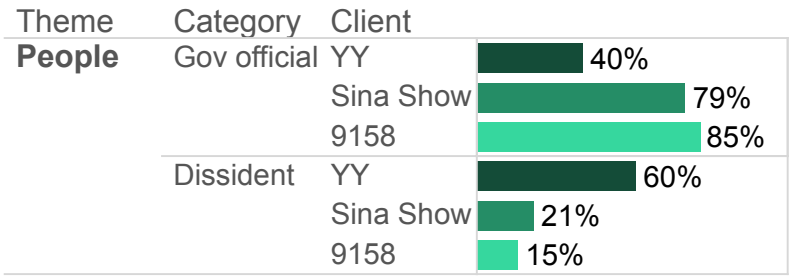


Figure 4.10: **Percentage of People theme keywords by category**

In the "Chinese websites and software" category we see instances of what may be companies using censorship to gain competitive advantage. YY lists include 25 keywords that reference competing live streaming services in China (*e.g.*, 美拍直播, "Mei Pai Live," 熊猫TV, "Panda TV"), and 9158 includes two keywords (*e.g.*, 六间房 "Six Room"). The addition of these keywords may be attempts to prevent users from being lured away from the provider's platform.

People Theme

The People theme includes two categories: names of CPC officials and names of dissidents.References to dissidents include the renowned artist Ai WeiWei (艾未未), Chinese human rights lawyer Guo Feixiong (郭飛雄), and gender activist Ye Haiyan (referred to by her nickname "rogue yan" 流氓燕).

References to officials includes current and former leaders (*e.g.*, 李克强 "Li Keqiang" current Premier of the State Council of China, 胡锦涛 "Hu Jintao" former Chinese President).

There are also numerous examples of playful, derogatory, and creative ways to refer to party leaders in the keyword lists. Keywords related to President Xi Jinping include an endearing nickname, "Daddy Xi" or "Uncle Xi" (习大大), which has been used in state propaganda [34], but recently has been reportedly banned from official

use to tone down Xi's populist image [31]. Other nicknames are more derogatory such as "Bun Ruthless" (包子心狠手辣). The word steamed bun (包子) is used to refer to Xi following the circulation of a photo showing him ordering lunch at a steamed bun shop that was subsequently criticized as a political show [48]. Whereas "ruthless" (心狠手辣) criticizes Xi's hardline rule over China. Chinese netizens often make creative use of the Chinese language in efforts to evade censorship [49]. We see examples of this practice in reference to Xi by reversing the order of the characters in his name (平近习), and using homoglyphs (刁近乎, diāo jín hū) that appear similar to his characters (习近平, xí jìn píng).

Researchers have argued that automated keyword censorship is ineffective, because through creative use of language users find means to circumvent the filters [89]. The keyword lists we collected show that censors are clearly picking up on these practices, engaged in a cat and mouse game between users. The censors will never be able to comprehensively censor speech through keyword filtering, nor will users always be able to evade these controls.

4.4 Conclusion

Our earlier work studying instant messaging apps found inconsistencies in client-side keyword filtering. However, the work was limited to a small number of applications that were not very representative in terms of usage numbers. In this chapter, we confirmed these earlier findings by analyzing the keyword lists of much more popular applications, the top applications of an industry segment.

We reverse engineered the top four live streaming applications, finding a total of 17,547 unique keywords that trigger these applications to censor or surveil. We combined this dataset with our earlier data obtained by analyzing instant messaging apps in Chapter 3. After cross-comparing these lists, we found very little consistency

in what each app censored with the exception of apps that shared parent companies or the case of GuaGua being founded by former Sina employees who seemingly took with them the list Sina was known to be using at the time they left the company.

The inconsistencies we found between keyword lists extend beyond the exact keywords each app censors. They also exist when we examined over time which topics each list was updated to censor. Over a period of 16 months, we found that apps without shared parent companies rarely updated their lists in response to the same topics or events. The lack of any large amount of consistency between these keyword lists demonstrates that the content of these lists cannot be largely the result of common directives received from a shared source such as the central Chinese government.

Chapter 5

Censorship in Mobile Games

In this chapter we describe the censorship and surveillance mechanisms built into Chinese mobile games. We analyzed over 200 games, including both games developed in China and international games adapted for the Chinese market. We discovered over 250 keyword lists and 183,111 unique keywords overall.

The results from previous chapters suggest that keyword lists that different companies use to perform censorship have little overlap. This suggests that these lists are not from some central source such as the Chinese government. But other work, such as by Mackinnon [98] and "Mr. Tao" [102], has hypothesized that city or provincial authorities may play a larger role in determining what to censor than those from the central Chinese government.

To further investigate these hypotheses we looked at the Chinese mobile gaming industry, which has recently come under increased government pressure [55, 144]. We found a large number of games implement keyword censorship client-side, which provided the opportunity to collect hundreds of keyword blacklists. Facilitated by the large number of keyword lists, we analyzed the similarity between them and assessed four new hypotheses for how the lists are created: (1) Content directives

are determined at the city or provincial level and may vary depending on where companies are based; (2) Content directives are determined for specific genres of games; (3) Content directives are related to the date that games are approved by regulators; (4) Companies are under general regulatory pressures, but have a degree of flexibility in determining which specific content to block.

We summarize our major contributions as follows:

1. We analyze over 250 keyword lists collected from over 200 games together comprising 183,111 unique keywords.
2. We present a novel application of a statistical technique and use it to test for correlation between keyword list similarity and other features in games such as whether they share the same publisher or developer.
3. We show that Chinese companies generating keyword lists themselves is the only plausible explanation for similarity between certain keyword lists among all of the hypotheses we tested.

5.1 Background

With an estimated value of over 27.5 billion US$ in 2017 [101], China represents the largest gaming market in the world. Lucrative as it is, China's market presents unique challenges to companies due to its strict regulatory environment. In 2010 China's Ministry of Culture (MOC) published a regulation [55] listing a number of prohibited topics for online games:

> "violating basic principles set by the Constitution; jeopardizing national unity, state sovereignty and territorial integrity; leaking state secrets,

> endangering state security or damaging state honor and interests; instigating ethnic hatred or discrimination, jeopardizing ethnic unity, and infringing ethnic rituals or customs; promoting heretical or superstitious ideas; spreading rumors, disrupting social order and stability; disseminating obscenity, pornography, gambling, violence or abetting crime; humiliating or slandering others, infringing the lawful rights of others; transgressing social morality; and other contents forbidden by laws and administrative regulations."

The vague definitions of prohibited topics make it unclear how to stay within the line and have been called "pocket crimes" (口袋罪), because anything can fit into them [144].

On June 2, 2016, China Audio-video and Digital Publishing Association, a non-governmental organization established in 1994, suggested a list of topics to be filtered on mobile games, including "attacking leaders of the Community Party of China (CPC)", "opposing Maoism", "referring to Taiwan, Hong Kong and Macau as an independent country". No specific keywords are listed in the document [18].

Games released in China require registration approval from MOC and a publication license from China's State Administration of Press, Publication, Radio, Film and Television (SAPPRFT). In May 2016, SAPPRFT extended the requirement to China's fast-growing mobile gaming industry. Games without a license will be removed from app stores. Previously, authorization from SAPPRFT was not mandated for individual games as long as the operators registered the games with the MOC within 30 days of publication [124]. In line with the new regulations, Apple Inc. notified all developers in mainland China that they would need to submit the approval number issued by SAPPRFT to list a game in the iTunes App Store.

Both high-level game content such as the story line and low-level content including

all scripts and conversations are subject to content controls [77, 25]. To ensure compliance with China's laws and regulations, gaming companies manage content by enabling keyword filtering systems on their products. Keyword filtering may be enabled to block content in chat features or to prevent players from creating user names and scoreboards containing sensitive keywords. To obtain a publication number for a game, the applicant must submit a list of blocked keywords enabled in the game and administrative accounts for regulators to test and review all scripts and features in the game [123]. In addition to content censorship, since August 1, 2016, Chinese mobile application developers are required to give the Cyberspace Administration of China (CAC) access to basic user information, which must be verified with a user's mobile phone number or real identification.

5.2 Methodology

Our analysis of keyword censorship in Chinese mobile games consists of two experiments. First, we collected the top games from a popular Chinese app store to evaluate which variables best predict the similarity between any two games' censored keyword lists. In the second experiment, we analyzed games from top Chinese game publishers and developers to further explore how well the "same publisher" and "same developer" variables predict keyword list similarity.

5.2.1 Analyzing Highly Downloaded Games

To understand which commonalities between games best predict the similarity of those games' keyword lists, we collected a variety of popular Chinese games. First, in December 2016, we acquired games from a Top 20 list of games for November 2016 [10] compiled by Newzoo, a gaming market analytics company. We also acquired

games from the earliest month Newzoo had a Top 20 list for, October 2014 [9]. In March 2017, to expand the collection of games, we collected highly downloaded games from Hi Market (安卓市场) [17], a popular [11] app store owned by Baidu [64]. We downloaded games from the site by scraping the first 500 search results from a search query we designed to restrict the results to "Chinese" games with over two million downloads and a rating of 4 to 5 stars. We repeated the process again for what the site termed "English" games, which consisted more broadly of all international games that generally had been adapted to meet Chinese regulations. (Our search queries were not designed to include games with exactly five stars as those tended to be games with very few reviews.) Together these methods produced a total of 836 unique games.

We then analyzed the games. Since there were too many games to reverse engineer each by hand, we first automatically searched for strings in the games to narrow down which warranted closer analysis and reverse engineering effort. We searched for sensitive words including "falun", "法轮" (Falun), "fuck", and "肏" (fuck), and general words or partial word stems that based on experience in previous work we expected to find in program code implementing censorship specifically, `blacklist`, `censor`, `dirty`, `filter`, `forbid`, `illegal`, `keyword`, `profan`, and `sensitiv`. We then manually reviewed the search results, collecting obvious blacklists, and manually reverse engineering games that appeared to implement censorship but where the blacklists were not immediately apparent, such as if the blacklists were obfuscated or encrypted. Using this method, we found keyword lists in a variety of formats, including text formats such as plain text, XML, JSON; binary formats such as compiled Lua or C++ code; and in files encrypted with a variety of different algorithms that required reverse engineering to decrypt.

We next tested which commonalities between two games best explained the similarity of those games' keyword lists. The commonalities we tested were whether two

games had the same publisher city, same publisher province, same developer city, same developer province, same genre, same publisher, same developer, or similar approval date. For the publisher and developer city, if a publisher or developer had offices in multiple cities, we used the city that they were primarily headquartered in. For genre and approval date, we used the genre and approval date under which the game was approved by the MOC [15].

To test for correlation between each of these commonalities and keyword list similarity, we used a statistical test called a partial Mantel test [121, 96], an extension of the Mantel Test [99]. A Mantel test is a statistical test for the presence of Pearson correlation between two similarity or distance matrices, a response variable Y and a potential explanatory variable X. The result of this test is the Mantel r statistic. The r statistic is related to the Mantel z statistic,

$$z = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij} Y_{ij}$$

except standardized by the variances of both matrices to be between -1 and 1, where the closer r is to -1, the more negative correlation exists, and the closer to 1, the more positive correlation exists (see [96]). The Mantel test also provides a method for computing the test statistic's p value, the probability that at least as extreme of correlation could have occurred by chance, using a Monte Carlo test method. A partial Mantel test extends the Mantel test by being able to control for additional matrices $Z_1, Z_2, \ldots$.

To create our response variable Y representing keyword list similarity, we first vectorize each list by creating a *count vector* v with dimension equal to the total # of unique keywords we have seen across all lists and where v_k is the # of times the kth word in our dataset appears in that list, which in our dataset could be greater than one. Then we define Y_{ij} as the cosine similarity between the ith list's count

vector u and the jth list's count vector v, *i.e.*,

$$C(u,v) = \frac{u \cdot v}{\|u\|_2 \|v\|_2} = \frac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\sum_{i=1}^{n} u_i^2}\sqrt{\sum_{i=1}^{n} v_i^2}}.$$

Geometrically, the cosine similarity is the cosine of the angle between two vectors.

In the previous chapter, we compared lists by using two different metrics. By considering each list as a set of keywords, we compared two lists by computing the Jaccard similarity (also known as the *Intersection over Union* metric), *i.e.*,

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|},$$

and computing the *Intersection over Smallest* metric, *i.e.*,

$$IoS(A,B) = \frac{|A \cap B|}{\min(|A|,|B|)}.$$

The advantage of the *Intersection over Smallest* metric was that it teases out relationships between lists, such as when one large list entirely contains a small list, that might otherwise be obscured by Jaccard similarity. However, the disadvantage of *Intersection over Smallest* is that very small lists comprised of a lot of very common words are very similar to most large lists, even though there is no interesting relationship between these lists.

The cosine similarity metric can be seen as striking a compromise between Jaccard similarity and *Intersection over Smallest*. If A and B are the corresponding keyword sets for count vectors u and v, respectively, then, if every keyword appears in each list at most once (keywords in our dataset rarely appear more than once in the same list), then the cosine similarity between u and v is identical to computing

$$C(u,v) = \frac{|A \cap B|}{\sqrt{|A|}\sqrt{|B|}}.$$

When $|A| < |B|$, if B acquires n new words, each of them not in A, then, with Jaccard similarity, the metric decreases by a factor of $1 + n/|B|$, in *Intersection*

over Smallest, the metric does not decrease (*i.e.*, decreases by a factor of 1), but with cosine similarity, the denominator decreases by a factor of $\sqrt{1 + n/|B|}$, striking middle ground between $1 + n/|B|$ and 1. Thus, the choice of cosine similarity to compare keyword lists can be seen as partially having both the advantages of Jaccard similarity and *Intersection over Smallest*.

With the response variable Y defined as a cosine similarity matrix between each of the lists, we must still define X, the matrix we will test for correlation with Y. Depending on the variable that we wish to test for correlation with Y, we define X differently. For all but the test of nearby approval dates, X is a binary matrix where X_{ij} is one if and only if the ith list has a property in common with (*e.g.*, if common publishers is being tested, then has the same publisher as) the jth. For testing nearby approval dates, we first construct X as the matrix such that X_{ij} is the distance between the ith and jth lists' approval dates as measured in days. We then normalize the matrix by dividing by X's largest value, and then add -1 to each value to turn it into a similarity matrix.

We often observed multiple lists included in each game. Some games used separate lists for censoring different features of the game. Other games seemed to accidentally include an older, outdated version of a list. As the partial Mantel test allows us to control for variables, in each of our tests we control for a binary matrix Z where Z_{ij} is one if and only if the ith list is from the same game than the jth.

5.2.2 Analyzing Games from Popular Publishers and Developers

During April 2017, after observing the results from the last experiment (see Section 5.3.1), we noticed that many games did not share a publisher or developer with any other game. We decided to perform another experiment focusing on further

testing the correlation between games' keyword list similarity and games with the same developers and publishers. In this experiment, we specifically look at the games from five popular publishers and seven popular developers to increase the number of games that share the same publisher or developer.

To determine the list of publishers, we returned to the November 2016 Top 20 list [10] and looked at all publishers from this list who both appeared in the Top 20 list and in whose games we had previously found at least one blacklist. This search resulted in us looking at Giant Interactive Group Inc. (巨人网络科技有限公司), Happy Elements (乐元素游戏), iDreamSky Technology Ltd. (乐逗游戏), NetEase (网易), and Tencent (腾讯).

To find a similar list of developers, we could not immediately use the same Top 20 list as it only listed publishers and not developers. Instead, we used the data from our previous analysis of highly downloaded games to collect the five developers who had the highest number of games containing keyword blacklists in our analysis. Since there was a four-way tie for fourth place, this yielded seven developers: CatCap Studio (达唯科技股份有限公司), Chukong Technologies (触控科技), Joymeng (乐堂动漫), Ourpalm Co. Ltd. (掌趣科技), Smile Games (乐人游戏), Ultralisk (雷兽互动), and Xiao Ao (小奥游戏).

We compiled a list of games for each publisher from the comprehensive list of mobile games for which each publisher had ever successfully obtained approval from the MOC, information available on their website [15]. We compiled a preliminary list of games for each developer from each developer's website and again from the MOC website [15]. However, often it was unclear whether a game on a developer's website or attributed to them by the MOC was published or developed by that company, as the MOC only provides data concerning the publisher and/or operator of a game. To exclude such games, we obtained copyright and ownership information for games using a tool called Tianyancha, a privately-owned platform that consolidates all

public information of companies registered with China's State Administration for Industry and Commerce and relevant regulators [16].

After downloading these games for each publisher and developer, we found 341 and 240 games, respectively, and 574 unique games combined. We used the same techniques as before to find lists in these games, and we again performed partial Mantel tests to assess correlation between keyword list similarity and a variety of different commonalities between games (again controlling for lists from the same game). However, in this case, we only tested keyword list similarity with correlation between same publisher, same developer, and nearby approval dates.

5.3 Results

In this section we describe our results from both of our experiments.

5.3.1 Results from Analyzing Highly Downloaded Games

From the 836 games that we analyzed in this experiment, we found 132 lists in 113 different games together containing 152,114 unique keywords. (This should not be interpreted to mean that other games had no censorship, as their client-side censorship may not have been discovered by our methods or as they may have performed censorship server-side.) These lists came from a number of popular Chinese games such as 天天酷跑3D (Tiantian Dash), and popular international games adapted to comply with Chinese regulations such as Ski Safari 2 and Candy Crush Saga. Figure 5.1 shows a heatmap of the pairwise cosine similarity of each blacklist hierarchically clustered according to the centroid linkage method.

We performed the partial Mantel test as described in Section 5.2 testing for cor-

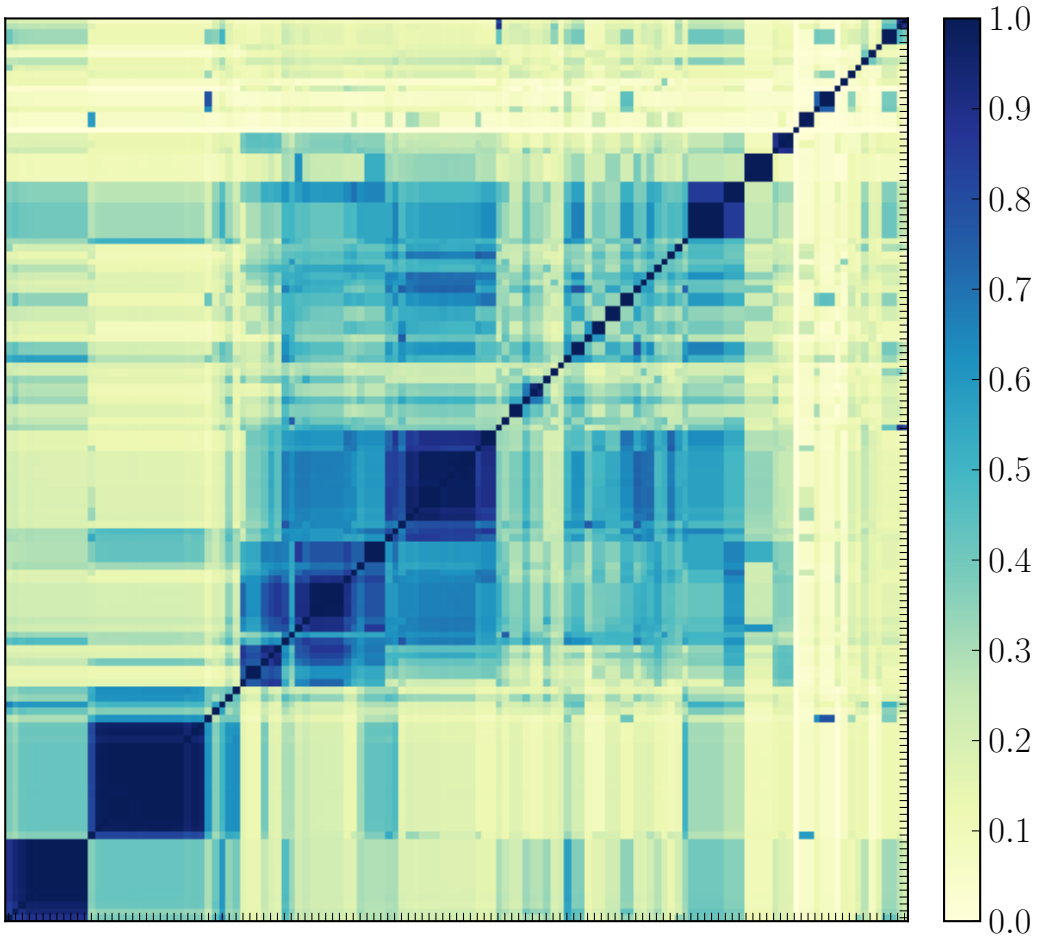


Figure 5.1: **Cosine similarity of blacklists from highly downloaded games hierarchically clustered according to the centroid linkage method**

relation between keyword list similarity and a number of different commonalities between games. The results of this test are the Mantel r statistic, a measurement of correlation between -1 and 1, and its corresponding p value, the probability that at least as extreme of correlation could have occurred by chance. Five game commonalities tested did not show significant correlation: **publisher city**, -0.0097 ($p = 0.60$); **publisher province**, -0.014 ($p = 0.68$); **developer city**, -0.0074 ($p = 0.59$); **developer province**, -0.0037 ($p = 0.55$); and **genre**, -0.013 ($p = 0.65$). Three game commonalities tested did show significant correlation: **approval date**, 0.16

($p = 0.0067$); **publisher**, 0.15 ($p < 0.001$); and **developer**, 0.17 ($p < 0.001$).

Since the game commonalities tested may be correlated among themselves (*e.g.*, games with the same developer tend to have the same publisher), we repeated each of the three tests two times for approval date, developer, and publisher, controlling for first the one and then the other of whichever of the two we were not presently testing. The greatest changes in results happened when testing same developers controlling for same publishers, where the r statistic reduced to 0.095 ($p < 0.001$), and same publishers controlling for same developers, where the r statistic reduced to only 0.047 ($p = 0.0015$). This suggests that, after controlling for each other as confounding variables, similar developers better predicts keyword similarity than similar publishers.

Because many of the games did not share the same publisher (50%) or developer (62%) with any of the other games we found lists from, and many others shared very few of each, we decided to perform a second experiment looking at games from popular publishers and developers in order to increase the number of games that share the same publisher or developer and in order to generally confirm our results.

5.3.2 Results from Analyzing Games from Popular Publishers and Developers

From the 574 unique games that we analyzed in this experiment, we found 167 lists in 129 different games together containing 171,150 unique keywords. This list again includes popular Chinese (*e.g.*, 开心消消乐 or Anipop) and international (*e.g.*, Fruit Ninja and Temple Run 2) games. See Tables 5.1 and 5.2 for a breakdown of lists found from each publisher and developer. Figure 5.2 shows a heatmap of the pairwise cosine similarity of each blacklist hierarchically clustered according to the centroid linkage method.

Publisher	**Downloaded**	**Lists Found**
Giant	26	11
Happy Elements	11	3
iDreamSky	29	8
Netease	87	22
Tencent	188	34

Table 5.1: **The number of games downloaded and lists found for each publisher**

Developer	**Downloaded**	**Lists Found**
CatCap	23	7
Chukong	38	10
Joymeng	63	9
Ourpalm	38	11
Smile	19	4
Ultralisk	21	16
Xiao Ao	38	32

Table 5.2: **The number of games downloaded and lists found for each developer**

Our partial Mantel test results for the tested game commonalities are as follows: **approval date**, -0.056 ($p = 0.83$); **publisher**, 0.21 ($p < 0.001$); **developer**, 0.23 ($p < 0.001$).

Compared to the previous experiment, we saw even more correlation when testing same publishers and same developers; however, we were surprised to see the correlation that we had originally seen in the first experiment when testing similar approval dates disappear in the second, as we expected to see some correlation due to (*e.g.*) developers adding to lists over time. The second experiment's result may be because

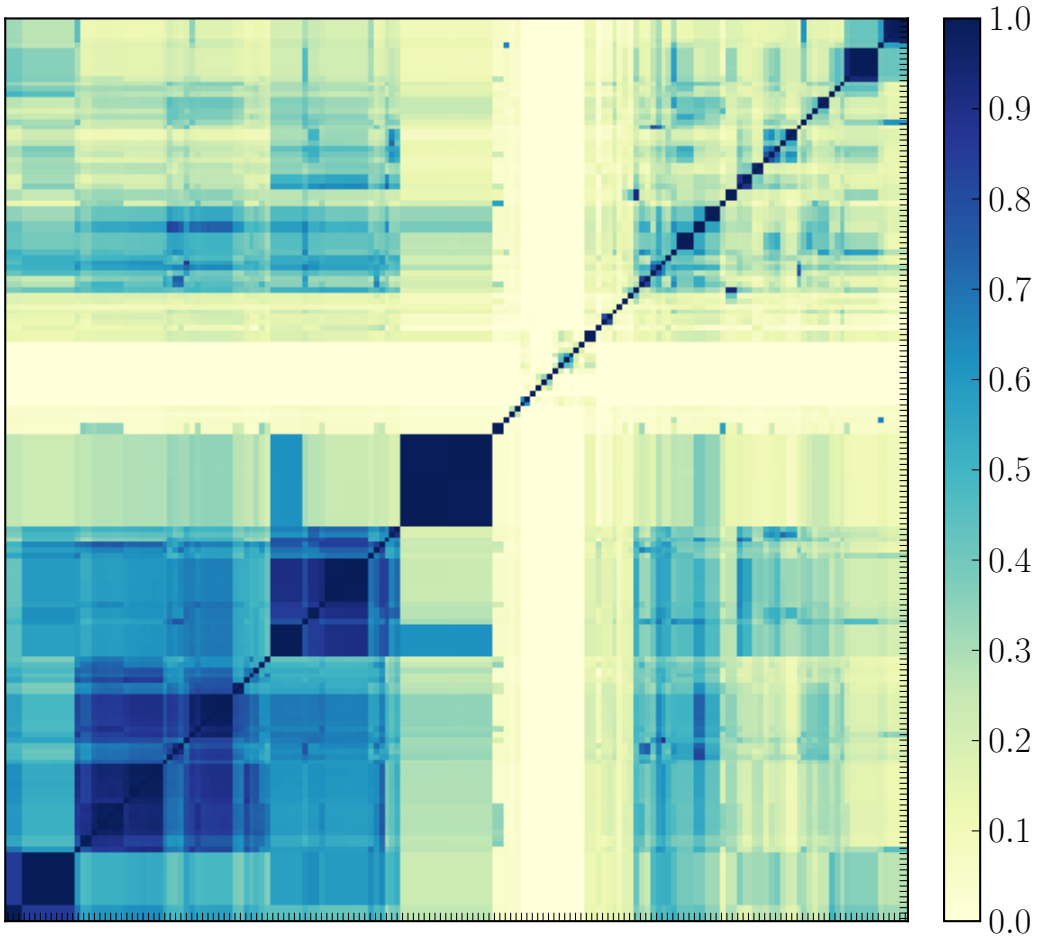
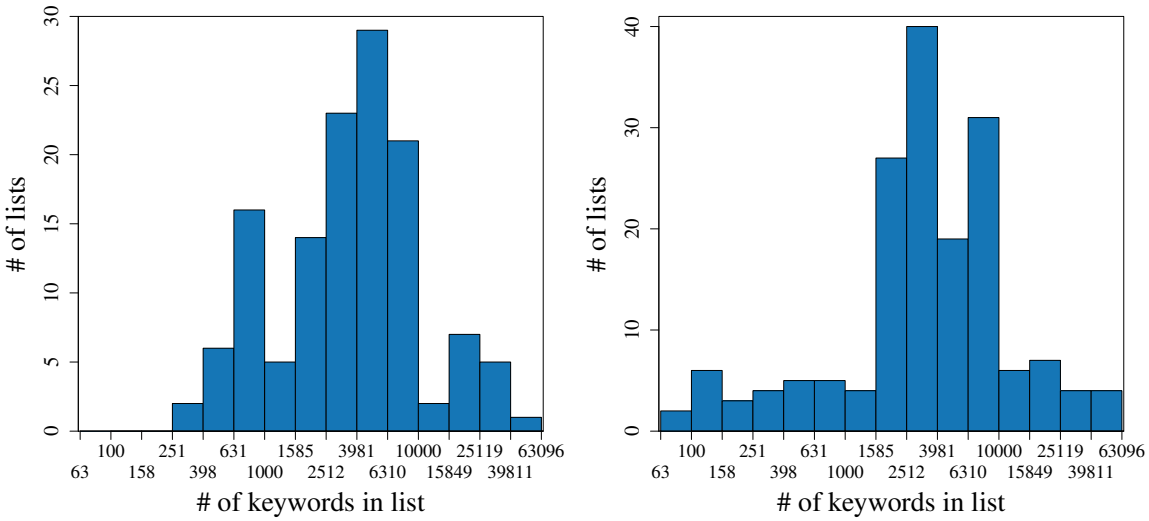


Figure 5.2: **Cosine similarity of blacklists from games from top publishers and developers hierarchically clustered according to the centroid linkage method**

the experiment was specifically designed to further test correlation in games with the same publishers and developers by decreasing publisher and developer variety. However, by focusing on these variables and decreasing variety, we may have limited our ability to observe the effect of other explanatory variables.

We also again tested same publisher controlling for same developer and same developer controlling for same publisher. The same publisher controlling for same developer reduced the correlation to 0.064 ($p = 0.015$). The same developer con-



(a) **Highly downloaded games**

(b) **Games from popular publishers and developers**

Figure 5.3: **For each experiment a histogram showing the number of keyword lists versus the number of keywords in that list. The x axis is log-scaled.**

trolling for same publisher only reduced the correlation to 0.13 ($p < 0.001$). This indicates that, after controlling for each other as confounding variables, games having the same developers better predicts keyword list similarity than similar publishers.

There was considerable diversity in the number of keywords found in each list. In both experiments, most lists have fewer than 10,000 keywords (see Figure 5.3 for the observed distributions). The shortest list seen contained 65 keywords, whereas the longest contained 53,874.

5.4 Keyword Analysis

In this section we discuss the provenance of keyword lists and provide a preliminary content analysis.

5.4.1 Keyword List Provenance

Our results suggest that developers, and, to a lesser extent, publishers are largely responsible for providing the keyword lists that come with their games; however, while our results do not speak directly to how they accumulate these lists, there are clues in our keyword data set that point to possible directions. Text-based data formats typically have escape sequences used to escape otherwise special characters that are used to structure the data. While we took care to unescape keywords from whatever file format it was stored in, some developers did not always do so, leaving the escape sequences in and providing clues as to where the lists were taken from. For instance, we saw C-style escapes in other file formats, *e.g.*, 私\\服 which we found in an XML document with a C-style escape of the \ by preceding it with a second \ (the keyword appears as 私\服 in many lists). Conversely, we saw XML escapes in keywords in non-XML files such as the ampersand in 冠西&艳照 having been replaced with & resulting in 冠西&艳照.

Some keywords may even trace back to originally being shared on old web apps. For instance, old PHP web apps escaped database input by manually calling the *addslashes* [3] function or by enabling "magic quotes" [4] which called it automatically. However, this functionality has long been deprecated due to it being insecure and not multibyte character encoding aware [118]. In 2004, when a leaked keyword list had been published on a bulletin board [19], the published list contained many erroneous \ escapes due to the bulletin board's improper escaping. For instance, the list features 胡錦\濤, since the second character (錦) in the Chinese GBK encoding encodes to the bytes `0xe5` and `0x5c`, the latter byte being \ in ASCII, which the web app erroneously interpreted as a special byte needing to be escaped and preceded it with another \. As keywords such as this one are featured in many of our lists, this suggests that they and potentially many others had been at one time been shared on old web apps that had improperly escaped database input.

Theme	**Examples**
Event	Anniversaries, current events
Political	Communist Party of China, religious groups
People	Government officials, dissidents
Social	Gambling, prurient interests
Technology	Online games, URLs
Misc	No clear context

Table 5.3: **Content themes and related categories**

5.4.2 Content Analysis

Our work in previous chapters performed content analysis of keyword lists by manually grouping keywords into content categories based on contextual information. Using similar methods, we conducted a preliminary content analysis to provide a high level description of the data set. We analyzed a random sample of 7,000 keywords from our data set of 183,111 keywords. A native Chinese speaker reviewed the random sample and based on the context of the keywords assigned high level themes according to the code book developed in the previous two chapters. See Table 5.3 for a description of each theme and Figure 5.4 for the distribution of keywords by theme. We describe and provide examples of each theme in the remainder of this section.

Social: This theme accounted for the largest percentage of the keywords we analyzed. Examples include references to illicit goods such as "出售业主信息数据" (selling data information of property owners) and gambling (*e.g.*, "六合彩" Mark Six, a lottery betting system organized by Hong Kong Jockey Club). We found a similar focus on Social theme keywords in live streaming apps in the previous chapter.

Political: This theme includes general references to the CPC and government

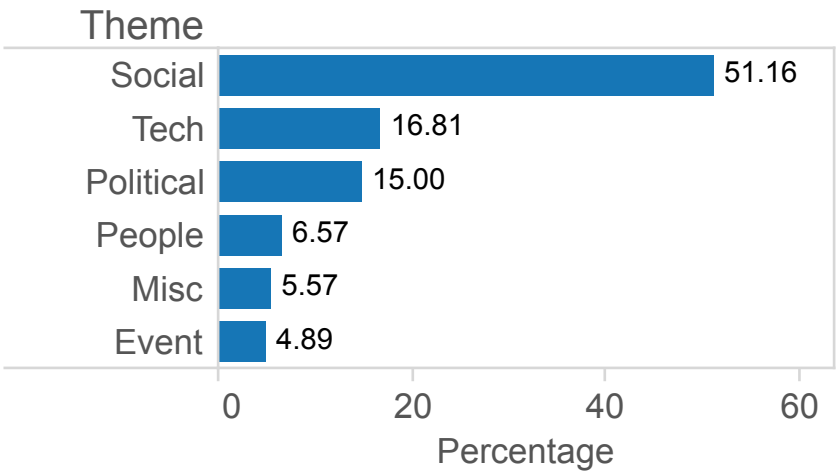


Figure 5.4: **Breakdown by theme of 7,000 randomly sampled keywords (±1.1 error at 95% confidence)**

bodies (*e.g.*, "人民检察院" People's Procuratorate, the agency responsible for both prosecution and investigation under China's legal system), criticism of state policy (*e.g.*, "敏感词屏蔽的社会", meaning "a society where sensitive keywords are blocked"), religion (*e.g.*, Falun Gong and Christianity), Sino-Japanese relations, and ethnic groups in China. Homoglyphs or homonyms were often used when referring to the CPC (共产党) (*e.g.*, "哄铲挡", hǒng chǎn dǎng) and Falun Gong (*e.g.*, "發圇", pronounced fā lún, as opposed to 法轮功, fǎ lún gōng).

We also found keywords in Korean and Japanese related to international relations issues in Asia. Examples include "일진회", which refers to Iljinhoe, a nationwide pro-Japan organization that operated in Korea in the 1900s. The presence of these keywords may be because some games we analyzed were imported from foreign gaming companies or were aimed at foreign markets. It may also have to do with the unique characteristics of gaming platforms where in-group identity and nationalism is often shared and championed by players [5, 109].

People: Similar to our work in the previous two chapters, we found references to people including government officials, relatives of officials, dissidents, and names

without clear context. Keywords included use of homonyms, homoglyphs, coded messages, and pinyin to refer to party leaders and dissidents. For example, "刁净瓶" contains a homoglyph (刁) and a homonym (净瓶, jìngpíng) for President Xi Jinping (习近平, xí jìnpíng). In another example a coded reference is made to China's first Nobel Peace Prize winner Liu Xiaobo who received the award in absentia due to imprisonment ("无法领奖的人", a person who is unable to receive the award).

Event: Our work in the previous two chapters showed evidence that sensitive and critical events can act as catalysts for censorship of social media in China. We tested the prevalence of event-related keywords by searching for terms related to current events that occurred within 2016 to 2017 and that had been found censored on Chinese chat applications in recent studies [115, 63, 116] or that were known to have leaked government censorship directives [54]. We did not find references to these up-to-date news events in the gaming keyword data set. However, we did find keywords referring to other politically sensitive events (*e.g.*, "1989年民运", or 1989 Year Democracy Movement, a reference to the 1989 Tiananmen Square movement; "茉莉花活动", and the Jasmine Revolution, a reference to a series of pro-democracy protests in China inspired by the Jasmine Revolution in Tunisia in 2011).

Chat apps and microblogs provide means to communicate and spread content, whereas games are primarily for entertainment. Therefore, there may be increased concern and scrutiny of how apps used for news consumption manage content related to current events. It is also possible there are references to events we did not find due to the limitation of the scope of our current project. In our analysis we did not analyze historical versions of each mobile game, and we did not track any downloaded updates, if there were any, to its keyword lists.

Technology: This theme includes references to identifiers such as phone numbers and emails. This theme also includes names of websites and various technology services including censorship evasion tools.

Keywords also included references to the names of competing game companies. For example, the keyword lists for Fishing Joy 4 (捕鱼达人4), a casual game developed and published by Chukong Technologies, includes references to games developed by competitors (*e.g.*, "侠客天下", an online game developed by Beijing Xiakehang Network Technology Limited Company; "仙境传说", a reference to Ragnarok Online, a Korean MMORPG). In the previous chapter, analysis of censored keyword lists from Chinese live streaming apps also found references to the names of competing products, possibly to prevent users from being lured away to other platforms. The names of competitors present on the game keyword lists may be motivated by similar reasons.

5.5 Summary

In earlier chapters we analyzed companies' instant messaging and live streaming apps to determine whether these companies were receiving common directives from the central Chinese government that were largely influencing what they censor. Our findings were that this was not the case.

In this chapter, we analyzed censorship in mobile games, finding over 250 keyword lists in over 200 games. This large number of lists allowed us to examine other hypotheses related to Chinese censorship, namely that city or provincial authorities, not ones from the central government, play a large role in determining what companies censor. However, we found no significant correlation between the similarity of two games' lists and whether those games had publishers or developers from the same city or province. We found that only whether two games shared the same developer or publisher correlated with the similarity of those games' keyword lists. This suggests that the responsibility to choose what to censor is pushed down to individual companies or employees.

Chapter 6

Conclusion and Future Work

The work in this dissertation takes a departure from other work in the field by not using sample testing to measure censorship. Many studies employing sample testing choose words from a list of potential words composed by a researcher. The censorship results are biased toward whatever the researcher presupposed could be censored. Others measure content deletion on platforms. These studies are biased toward whatever was trending or popular on the platform during the study and do not reveal forbidden topics that were never discussed or that were never discussed publicly.

Instead of using sample testing, the work in this dissertation uses reverse engineering to determine the entire list of forbidden keywords used to filter realtime chat. This work analyzes over 100,000 keywords from hundreds of different companies across three different industry segments: instant messaging, live streaming, and gaming.

At the beginning of this dissertation, I introduced three hypotheses to be evaluated by the more complete and unbiased dataset provided by this work. The first hypothesis is that **there is little overlap between the keyword lists used by**

different companies. Our analysis of censorship in Chinese instant messaging apps and live streaming apps revealed that there was little overlap between apps' lists from different companies. The only seeming exception was the commonalities between GuaGua's and Sina's lists. This exception disappeared when we discovered that GuaGua was founded by former Sina employees. When over time we looked at what high level topics or events these lists were updated in response to, we again found little overlap. Our analysis of gaming apps also showed little overlap between lists except for what could only be explained by companies sharing the same lists. Therefore, the data we collected in all three of these industry segments strongly supports Hypothesis 1.

The second hypothesis is that **there is no China-wide list of banned words or topics largely determining what Chinese companies censor**. The data in this dissertation strongly supports this hypothesis for largely similar reasons. If there is little overlap between lists, then then there cannot be a China-wide list determining most of what companies censor. If there were such a list largely determining what companies censor, then there would necessarily be a large amount of overlap between lists. Therefore, the data collected in this dissertation also strongly supports Hypothesis 2.

The third hypothesis is that **provincial-wide lists of banned words or topics do not largely determine what companies censor**. Our analysis of censorship in mobile games supports this hypothesis. We saw little overlap between blacklists in this dataset. Because of the large number of blacklists from different companies in this dataset, we were able to test a number of other hypotheses to explain what overlap we did observe. Whether the blacklists' game publishers or developers were from the same province did not explain lists' overlap. We could not transfer this analysis to the instant messaging and live streaming datasets because they do not contain as many blacklists from as many companies. This makes the support for

Hypothesis 3 not as strong as the others. However, if the lesson from all three datasets is to not overestimate the role of the central government in choosing what topics are forbidden, then we should be careful not to overestimate the role of provincial governments as well.

A dominant academic theory developed by King *et al.* [89, 90] from studying blog providers asserts that the motive of Chinese censorship is to suppress collective action. This theory implicitly assumes a monolithic motive. One way of reconciling this theory with the results in this dissertation is to argue that the theory was developed by studying blog providers, a different industry segment than the three analyzed in this dissertation and which might be censored according to substantially different regulations. However, since the King *et al.* work used sample testing, a more biased method than the one used in this dissertation, this more likely explains the discrepant findings. This is especially true since another, earlier work [98] also studying blog providers but also using sample testing found little consistency in the topics forbidden by different blog providers. The two different results from sample testing underscores the need for a complete and unbiased view into which topics are forbidden.

The findings presented in this dissertation parallel Link's notion of the "anaconda in the chandelier" [97], which he uses as a metaphor to describe how the Chinese censorship apparatus puts pressure on individuals to self-censor in order to avoid being punished for violating vague laws. He says that the anaconda's "constant silent message is 'You yourself decide,' after which, more often than not, everyone in its shadow makes his or her large and small adjustments." Although this metaphor was originally used to describe how individuals self-censor themselves, the work in this dissertation shows that this metaphor is equally appropriate in explaining how private companies censor their own platforms.

China could conceivably provide an exhaustive list of blacklisted topics to ev-

ery private company leaving no question what to censor; however, Link's anaconda metaphor provides insight into why they do not. First, by keeping censorship requirements intentionally vague, the Chinese government maintains plausible deniability so that if something a company censors generates backlash, the blame ultimately falls on the company for choosing to censor it and not the Chinese government. Second, in order to avoid penalties, private companies may censor more content than what would appear on a government-provided blacklist as companies are left to try and guess everything that they must censor, inevitably in the process censoring extra things that would not have been on a government-provided blacklist.

Future Chinese censorship researchers should be cautious before presuming a monolithic motive behind Chinese censorship. The Chinese censorship apparatus has a significant effect on controlling what information hundreds of millions of people in China have access to; however, not all significant effects have significant causes, and the work in this dissertation suggests that the reason for why something is censored in a Chinese application may be an arbitrary decision by a private company or one of its employees to protect the company from vague laws. Additionally, as our research finds, companies also use censorship for their own motivations such as to censor competitors. Future researchers would better understand censorship in China by studying the motivations and incentives of private companies who are given little direction in deciding what to censor and are largely deciding what to censor themselves.

6.1 Future work

The work in this dissertation paves the way for future research into Chinese censorship in multiple directions. Our method of extracting complete keyword blacklists has revealed little overlap between lists from different companies, suggesting that

companies are given a large amount of freedom in choosing what to censor. However, there is a small fraction of keywords that appear on most lists. Are these keywords on so many lists because they are well known to be sensitive? Has developers sharing lists, especially early when the first blacklists were being developed, contributed to the number of keywords appearing on a large number of lists? Was there ever a blacklist, no matter how small, provided by the Chinese government to a large amount of Internet companies? These questions are difficult to answer given our current dataset, but one possible future reasearch direction of research is to investigate words that frequently appear together on lists. If words typically appear together with other words across many different lists, then this suggests that they may have originated from an earlier, common list that was shared.

We manually categorized keywords from instant messaging and live streaming apps according to their high-level theme, but we did not do this with the Chinese games dataset due to its large size, instead only sampling and categorizing a fraction of the keywords. Another avenue for future research is to complete keyword categorization of the Chinese games keyword dataset through manual or machine learning assisted methods. This would permit the cross-comparison of lists from this dataset in a high level thematic sense. Although doing this with the instant messaging and live streaming datasets revealed little thematic overlap, it would permit us to verify our previous findings with the new data in the Chinese gaming dataset.

The Chinese gaming dataset also presents an opportunity to measure a "ground truth" regarding how private companies in China choose what to censor. The gaming dataset differs from the previous instant messaging and live streaming datasets we analyzed in that it contains hundreds of different publishers and developers and in that a large number of them are not based in China. Many of these companies may be willing to be interviewed regarding the requirements of developing and publishing a game for the Chinese market. These second-hand accounts could provide further

insight into how private Chinese companies choose to implement censorship.

It remains an open question as to what extent censorship in China mirrors how other countries' censorship ecosystems will develop. As the largest country in the world, China's implementation represents a large data point. However, China is currently unique among other countries in how it has fostered its own ecosystem of domestic Internet apps, often by using Internet-level censorship to block foreign Internet companies competing against domestic alternatives. Nevertheless, other countries such as Russia are also increasingly exerting their "Internet sovereignty" [100], blocking foreign companies for failing to adhere to local laws [130], and becoming increasingly reliant on private companies to police their users' communication in accordance with those countries' local regulations [95]. As countries other than China continue to increasingly exert control over their users' Internet and foster their own local ecosystems of applications, will we see them adopting a decentralized approach leaving private companies to determine exactly what to censor? While decentralizing censorship may at first seem suboptimal, Link's anaconda in the chandelier analogy provides insight into why this may be the most advantageous approach and why other countries may be clever to embrace it.

References

[1] GuaGua, 法律声明. Available at http://www.guagua.cn/other/1907.html.

[2] Hex-Rays Decompiler: Overview. Available at https://www.hex-rays.com/products/decompiler/.

[3] PHP: addslashes. Available at https://secure.php.net/manual/en/function.addslashes.php.

[4] PHP: What are Magic Quotes. Available at https://secure.php.net/manual/en/security.magicquotes.what.php.

[5] Racial and ethnic hate speech thrives in online games. Available at http://www.japantimes.co.jp/news/2017/05/14/business/tech/racial-ethnic-hate-speech-thrives-online-games/.

[6] Tian Ge Interactive Holdings, Global Offering. Available at http://www.tiange.com/Upload/Pigeon-Cover-IPO-ENG-2d.pdf.

[7] Tian Ge Interactive Holdings, Interim Report 2014. Available at http://www.tiange.com/Upload/e101.pdf.

[8] Tian Ge Interactive Holdings, Tian Ge Announces 2014 Third Quarter and Interim Results. Available at http://www.tiange.com/Upload/TIAN%20GE%20ANNOUNCES%202014%20THIRD%20QUARTER%20RESULTS.pdf.

[9] Top 20 Android Games in China: Time & Money spend in a Stable October. Available at https://newzoo.com/insights/articles/top-20-android-games-in-china-time-is-money-stable-october/.

[10] Top Android Games China. Available at https://newzoo.com/insights/rankings/top-20-android-games-china/.

[11] Top Chinese Android App Stores. Available at https://newzoo.com/insights/rankings/top-10-android-app-stores-china/.

[12] University of Hong Kong, WeiboScope. Available at http://weiboscope.jmsc.hku.hk/.

[13] YY Inc, 2014 Annual Report. Available at http://investors.yy.com/annuals.cfm.

[14] YY Inc, YY主播违规管理方法. Available at http://www.yy.com/1309/242131808402.html.

[15] 中国文化市场网. Available at http://ccm.gov.cn/.

[16] 关于我们. Available at http://www.tianyancha.com/property/1/.

[17] 安卓网-中国最大android手机垂直门户. Available at http://www.hiapk.com/.

[18] 移动游戏内容规范(2016年版). Available at http://www.ce.cn/culture/gd/201606/02/t20160602_12463013.shtml.

[19] QQ过滤词列表 zt. Available at https://web.archive.org/web/20040908030852/http://bbs.omnitalk.org/arts/messages/3824.html, 2004.

[20] Guagua: Exploring China's Online Video Community 'Goldmine'. *Knowledge@Wharton*, October 2013. Available at http://knowledge.wharton.upenn.edu/article/guagua-exploring-chinas-online-video-community-goldmine/.

[21] China Launches Campaign to Cleanse Web of Terror Content. Available at http://www.reuters.com/article/2014/06/20/us-china-internet-xinjiang-idUSKBN0EV0TP20140620, June 2014.

[22] Towards a Comprehensive Picture of the Great Firewall's DNS Censorship. In *the 4th USENIX Workshop on Free and Open Communications on the Internet*, San Diego, CA, 2014. USENIX.

[23] IDA: About. Available at https://www.hex-rays.com/products/ida/index.shtml, 2017.

[24] OllyDbg v1.10. Available at http://www.ollydbg.de/, 2017.

[25] D. Ahmad. A look at China's stringent mobile game regulations half a year on. Available at http://www.pocketgamer.biz/asia/comment-and-opinion/65378/chinas-mobile-game-regulations-half-a-year-on/, March 2017.

[26] J. Ansfield. How Crash Cover-Up Altered China's Succession. Available at http://www.nytimes.com/2012/12/05/world/asia/how-crash-cover-up-altered-chinas-succession.html, December 2012.

[27] Asahi Shimbun. UNTIL NOW: Tensions start to rise when China enacts law claiming islands - AJW by The Asahi Shimbun. Available at http://ajw.asahi.com/article/special/senkaku_history/AJ201212260105, 2012.

[28] D. Bamman, B. O'Connor, and N. A. Smith. Censorship and deletion practices in Chinese social media. *First Monday*, 17(3), 2012.

[29] BBC News. Bo Xilai scandal: Timeline. Available at http://www.bbc.co.uk/news/world-asia-china-17673505, 2012.

[30] F. Bei. China's Internet Censorship System. Available at http://hrichina.org/crf/article/3244, April 2010.

[31] Bloomberg. Xi's Nickname Becomes Out of Bounds for China's Media. Available at https://www.bloomberg.com/news/articles/2016-04-28/no-more-big-daddy-xi-as-china-s-spin-doctors-revise-playbook, April 2016.

[32] K. Bradsher. Amid Protest, Hong Kong Retreats on 'Moral Education' Plan. Available at http://www.nytimes.com/2012/09/09/world/asia/amid-protest-hong-kong-backs-down-on-moral-education-plan.html, Sept. 2012.

[33] China Daily. MSN China, Sina link up. Available at http://www.chinadaily.com.cn/bizchina/2010-11/12/content_11539903.htm, November 2010.

[34] China Digital Times. Daddy Xi. Available at https://chinadigitaltimes.net/space/Daddy_Xi.

[35] China Digital Times. Grass-Mud Horse Lexicon. Available at http://chinadigitaltimes.net/space/Grass-Mud_Horse_Lexicon.

[36] China Digital Times. A List of Censored Words in Chinese Cyberspace. Available at http://chinadigitaltimes.net/2004/08/the-words-you-never-see-in-chinese-cyberspace/, August 2004.

[37] China Digital Times. Baidu's Internal Monitoring and Censorship Document Leaked (1) (Updated). Available at http://chinadigitaltimes.net/2009/04/baidus-internal-monitoring-and-censorship-document-leaked/http://chinadigitaltimes.net/2009/04/baidus-internal-monitoring-and-censorship-document-leaked-2/, April 2009.

[38] China Digital Times. Directives from the Ministry of Truth: July 5-September 28. Available at http://chinadigitaltimes.net/2011/10/directives-from-the-ministry-of-truth-july-5-september-28-2011/, October 2011.

[39] China Digital Times. Ministry of Truth: Anti-Japan Protests. Available at http://chinadigitaltimes.net/2012/09/ministry-of-truth-anti-japan-protests/, September 2012.

[40] China Digital Times. Ministry of Truth: Bo Xilai - China Digital Times (CDT). Available at http://chinadigitaltimes.net/2012/11/ministry-of-truth-bo-xilai/, November 2012.

[41] China Digital Times. Ministry of Truth: Hong Kong Elections, Teacher's Day. Available at http://chinadigitaltimes.net/2012/09/143459/, September 2012.

[42] China Digital Times. Ministry of Truth: The "Almighty God Cult". Available at http://chinadigitaltimes.net/2012/12/ministry-of-truth-tackling-almighty-god-cult/, December 2012.

[43] China Digital Times. Sensitive Words: 18th Party Congress. Available at http://chinadigitaltimes.net/2012/11/sensitive-words-18th-party-congress/, November 2012.

[44] China Digital Times. Sensitive Words: Anti-Japan Protests (2). Available at http://chinadigitaltimes.net/2012/09/sensitive-words-anti-japan-protests-2/, September 2012.

[45] China Digital Times. Sensitive Words: Trials, Looting and Liver Cancer. Available at http://chinadigitaltimes.net/2012/09/sensitive-words-trials-looting-and-liver-cancer/, September 2012.

[46] China Digital Times. Ministry of Truth: Urgent Notice on Southern Weekly. Available at http://chinadigitaltimes.net/2013/01/ministry-of-truth-urgent-notice-on-southern-weekly/, January 2013.

[47] China Digital Times. Sensitive Sina Weibo Search Terms. Available at https://docs.google.com/spreadsheet/ccc?key=0Aqe87wrWj9w_dFpJWjZoM19BNkFfV2JrWS1pMEtYcEE, 2013.

[48] China Digital Times. Steamed Bun Xi. Available at http://chinadigitaltimes.net/space/Steamed_Bun_Xi, 2014.

[49] China Digital Times. Decoding the Chinese Internet eBook. Available at https://chinadigitaltimes.net/2015/07/decoding-the-chinese-internet-ebook-2015-edition/, 2015.

[50] China Digital Times. Sensitive Words: June 4th, 2015. Available at http://chinadigitaltimes.net/2015/06/sensitive-words-june-4th-2015/, June 2015.

[51] China Digital Times. Minitrue: Do Not Report on Wukan Mass Incident. Available at http://chinadigitaltimes.net/2016/06/minitrue-former-wukan-chief-admits-guilt/, June 2016.

[52] China Digital Times. Minitrue: Tight Control on Xi's "Loose Clothing" Slip at G20. Available at http://chinadigitaltimes.net/2016/09/minitrue-tight-control-xis-loose-clothing-slip-g20/, September 2016.

[53] China Digital Times. Directives from the Ministry of Truth. Available at http://chinadigitaltimes.net/china/directives-from-the-ministry-of-truth/, October 2017.

[54] China Digital Times. Directives from the Ministry of Truth. Available at http://chinadigitaltimes.net/china/directives-from-the-ministry-of-truth/, 2017.

[55] China's Ministry of Culture. Interim Measures for the Administration of Online Games. Available at http://www.gov.cn/flfg/2010-06/22/content_1633935.htm, June 2010.

[56] R. Clayton, S. J. Murdoch, and R. N. M. Watson. Ignoring the Great Firewall of China. In *6th Workshop on Privacy Enhancing Technologies*, 2006.

[57] CNNIC. Statistical Report on Internet Development in China. Available at https://cnnic.com.cn/IDR/ReportDownloads/201706/P020170608523740585924.pdf, 2017.

[58] Committee to Protect Journalists. The business of censorship: Documents show how Weibo filters sensitive news in China. Available at `https://cpj.org/blog/2016/03/the-business-of-censorship-documents-show-how-weib.php`, March 2016.

[59] J. R. Crandall, M. Crete-Nishihata, J. Knockel, S. McKune, A. Senft, D. Tseng, and G. Wiseman. Chat program censorship and surveillance in China: Tracking TOM-Skype and Sina UC. *First Monday*, 18(7), 2013.

[60] J. R. Crandall, D. Zinn, M. Byrd, E. Barr, and R. East. ConceptDoppler: A weather tracker for Internet censorship. In *14th ACM Conference on Computer and Communications Security, Oct.29-Nov2, 2007*, pages 1–18, 2007.

[61] M. Crete-Nishihata, J. Dalek, S. Hardy, J. Q. Ng, and A. Senft. Asia Chats: LINE Censored Keywords Update. Available at `https://citizenlab.org/2014/04/line-censored-keywords-update/`, 2014.

[62] M. Crete-Nishihata, A. Hilts, J. Knockel, J. Q. Ng, L. Ruan, and G. Wiseman. Harmonized Histories? A year of fragmented censorship across Chinese live streaming applications. November 2016. Available at `https://netalert.me/assets/harmonized-histories/harmonized-histories.pdf`.

[63] M. Crete-Nishihata, J. Knockel, and L. Ruan. Tibetans blocked from Kalachakra at borders and on WeChat. Available at `https://citizenlab.org/2017/01/tibetans-blocked-from-kalachakra-at-borders-and-on-wechat/`, January 2017.

[64] C. Custer. 5 Reasons Why Baidu Spent Nearly $2 Billion to Acquire 91 Wireless. Available at `https://www.techinasia.com/5-reasons-baidu-spent-2-billion-acquire-91-wireless`.

[65] C. Custer. A Shocking Expose of China's Black PR Industry Implicates Government Officials, is Quickly Deleted from the Web, February 2013.

[66] Drone Life. China's New Drone Regulations. Available at `http://dronelife.com/2016/01/19/chinas-new-drone-regulations/`, January 2016.

[67] EMarketer. QQ Continues to Dominate Instant Messaging in China. Available at `http://www.emarketer.com/Article/QQ-Continues-Dominate-Instant-Messaging-China/1009005`.

[68] R. Ensafi, P. Winter, A. Mueen, and J. R. Crandall. Analyzing the Great Firewall of China over space and time. In *Privacy Enhancing Technologies Symposium*. De Gruyter Open, 2015.

[69] D. Fabrice and K. Kortchinsky. Vanilla Skype part 1. Available at `http://recon.cx/en/f/vskype-part1.pdf`.

[70] D. Fabrice and K. Kortchinsky. Vanilla Skype part 2. Available at `http://recon.cx/en/f/vskype-part2.pdf`.

[71] Fei Chang Dao. 2012 in Review: 10 Examples of Free Speech With Mainland Chinese Characteristics (Part 2). Available at `http://blog.feichangdao.com/2013/01/2012-in-review-10-examples-of-free_2.html`, January 2013.

[72] Fei Cheng Dao. The March 2012 Ferrari Crash: Chronicling the Censorship. Available at `http://blog.feichangdao.com/2012/09/the-march-2012-ferrari-crash.html`, September 2012.

[73] K. Fu, C. Chan, and M. Chau. Assessing Censorship on Microblogs in China: Discriminatory Keyword Analysis and the Real-Name Registration Policy. *IEEE Internet Computing*, 17(3):42–50, 2013.

[74] Global Voices. China: Censor Machine Suspended for Anti-Japan Mobilization? Available at `http://globalvoicesonline.org/2012/09/16/china-censor-machine-suspended-for-anti-japan-mobilization/`, September 2012.

[75] Global Voices Advocacy. China: Be aware of QQ! Available at `http://advocacy.globalvoicesonline.org/2009/09/24/china-be-aware-of-qq/`, September 2009.

[76] GreatFire. GreatFire Analyzer. Available at `https://en.greatfire.org/analyzer`.

[77] T. Hancock. China game developers battle crackdown on content. Available at `https://www.ft.com/content/d1452338-ff19-11e6-96f8-3700c5664d30`, March 2017.

[78] S. Hardy. Asia Chats: Investigating Regionally-based Keyword Censorship in LINE. Available at `https://citizenlab.org/2013/11/asia-chats-investigating-regionally-based-keyword-censorship-line/`, November 2013.

[79] Hong Kong Free Press. President Xi Jinping's 'take off clothes' G20 gaffe censored in China. Available at `https://www.hongkongfp.com/2016/09/06/president-xi-jinpings-take-off-clothes-g20-gaffe-censored-in-china/`, September 2016.

[80] Human Rights in China. Nationwide State Secrets Education Campaign Launched as New Law Goes into Effect. Available at `http://www.hrichina.org/content/842`, October 2010.

[81] Human Rights Watch. "Race to the Bottom:" Corporate Complicity in Chinese Internet Censorship. Technical report, 2006. Available at `http://www.hrw.org/reports/2006/08/09/race-bottom`.

[82] International Campaign for Tibet. Tibetan Self-Immolation Fact Sheet. Available at `http://www.savetibet.org/resource-center/maps-data-fact-sheets/self-immolation-fact-sheet`, August 2013.

[83] Internet Society of China. Public Pledge on Self-Discipline for the Chinese Internet Industry. Available at `http://www.isc.org.cn/english/Specails/Self-regulation/listinfo-15321.html`, March 2002.

[84] A. Jacobs. Chinese Security Officials Respond to Call for Protests. Available at `http://www.nytimes.com/2011/02/21/world/asia/21china.html`, February 2011.

[85] A. Jacobs. China Arrests Christian Sect Members Over Doomsday Chat. Available at `http://www.nytimes.com/2012/12/20/world/asia/doomsday-chatter-makes-chinese-government-nervous.html`, December 2012.

[86] J. Kennedy. Hu Jia explains why mobile apps make activism spooky. Available at `http://www.scmp.com/comment/blogs/article/1083025/hu-jia-explains-why-mobile-apps-make-activism-spooky`, November 2012.

[87] J. Kennedy. Communist Party is giving more power to members working in Beijing internet companies. Available at `http://www.scmp.com/comment/blogs/article/1125390/communist-party-giving-more-power-members-working-beijing-internet`, January 2013.

[88] B. Kessler. Baidu, other top mainland internet companies, employ thousands of Party members. Available at `http://www.fcpablog.com/blog/2013/1/30/baidu-other-top-mainland-internet-companies-employ-thousands.html`, January 2013.

[89] G. King, J. Pan, and M. E. Roberts. How Censorship in China Allows Government Criticism but Silences Collective Expression. *American Political Science Review*, 107:1–18, 2013.

[90] G. King, J. Pan, and M. E. Roberts. Reverse-engineering censorship in China: Randomized experimentation and participant observation. *Science*, 345:1–10, 2014.

[91] J. Knockel, J. R. Crandall, and J. Saia. Three Researchers, Five Conjectures: An Empirical Analysis of TOM-Skype Censorship and Surveillance. In *the USENIX Workshop on Free and Open Communications on the Internet*, San Francisco, CA, 2011. USENIX.

[92] J. Knockel, M. Crete-Nishihata, J. Q. Ng, A. Senft, and J. R. Crandall. Every Rose Has Its Thorn: Censorship and Surveillance on Social Video Platforms in China. In *the 5th USENIX Workshop on Free and Open Communications on the Internet*, Washington, DC, USA, 2015. USENIX.

[93] J. Knockel, L. Ruan, and M. Crete-Nishihata. Measuring Decentralization of Chinese Keyword Censorship via Mobile Games . In *the 7th USENIX Workshop on Free and Open Communications on the Internet*, Vancouver, BC, Canada, 2017. USENIX.

[94] O. Lam. China: Sina Weibo Manager Discloses Internet Censorship Practices. Available at `http://advocacy.globalvoicesonline.org/2013/01/07/china-sina-weibo-manager-discloses-internal-censorship-practices/`, January 2013.

[95] J. Lee. India tops countries censoring Facebook content. Available at `https://www.usatoday.com/story/news/nation-now/2014/04/14/facebook-censor-india-turkey-pakistan/7694381/`, April 2014.

[96] P. Legendre and L. F. Legendre. *Numerical Ecology*, volume 24. Elsevier, 2012.

[97] P. Link. China: The Anaconda in the Chandelier. *The New York Review of Books*, 2002.

[98] R. MacKinnon. China's Censorship 2.0: How companies censor bloggers. *First Monday; Volume 14, Number 2 - 2 February 2009*, 2009.

[99] N. Mantel. The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2 Part 1):209–220, 1967.

[100] J. Margolin. Russia, China, and the Push for "Digital Sovereignty". Available at `https://theglobalobservatory.org/2016/12/russia-china-digital-sovereignty-shanghai-cooperation-organization/`, December 2016.

[101] E. McDonald. The Global Games Market Will Reach 108.9 Billion USD in 2017. Available at `https://newzoo.com/insights/articles/the-global-games-market-will-reach-108-9-billion-in-2017-with-mobile-taking-42/`, April 2017.

[102] Mr. Tao. China: Journey to the heart of Internet censorship. Available at `http://www.rsf.org/IMG/pdf/Voyage_au_coeur_de_la_censure_GB.pdf`, October 2007.

[103] D. Müllner. Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*, 2011.

[104] New York Times. Tribunal Rejects Beijing's Claims in South China Sea. Available at `https://www.nytimes.com/2016/07/13/world/asia/south-china-sea-hague-ruling-philippines.html?_r=0`, 2016.

[105] J. Q. Ng. Blocked on Weibo - Search result logs and full list of banned words. Available at `http://blockedonweibo.tumblr.com/post/12729333782/search-result-logs-and-full-list-of-banned-words`, 2012.

[106] J. Q. Ng. Tracing the Path of a Censored Weibo Post and Compiling Keywords that Trigger Automatic Review. Available at `https://citizenlab.org/2014/11/tracing-path-censored-weibo-post-compiling-keywords-trigger-automatic-review/`, 2014.

[107] J. Q. Ng. Politics, Rumors, and Ambiguity: Tracking Censorship on WeChat's Public Accounts Platform. Available at `https://citizenlab.org/2015/07/tracking-censorship-on-wechat-public-accounts-platform/`, July 2015.

[108] J. Q. Ng and P. F. Landry. The Political Hierarchy of Censorship: An Analysis of Keyword Blocking of CCP Officials' Names on Sina Weibo Before and After the 2012 National Congress Selection. In *Eleventh Chinese Internet Research Conference, 2013*, 2013.

[109] H. A. Nie. Gaming, Nationalism, and Ideological, Work in Contemporary China: online games based on the War of Resistance against Japan. *Journal of Contemporary China*, 2013.

[110] E. Osnos. How a High-Speed Rail Disaster Exposed China's Corruption. *New Yorker*, October 2012.

[111] J. C. Park and J. R. Crandall. Empirical study of a national-scale distributed intrusion detection system: Backbone-level filtering of HTML responses in China. In *30th International Conference on Distributed Computing Systems (ICDCS) 2010*, pages 1–12, 2010.

[112] Radio Free Asia. Language Policy Comes Under Scrutiny. Available at `http://www.rfa.org/english/news/tibet/students-03142012213524.html`, March 2012.

[113] Radio Free Asia. Monk Burns Himself Amid Mass Protests. Available at http://www.rfa.org/english/news/tibet/burn-03162012143125.html, March 2012.

[114] K. B. Richburg. Chinese artist Ai Weiwei arrested in latest government crackdown - Washington Post. Available at http://articles.washingtonpost.com/2011-04-03/world/35229738_1_chinese-artist-china-researcher-chinese-human-rights-defenders, April 2011.

[115] L. Ruan, J. Knockel, and M. Crete-Nishihata. One App, Two Systems: How WeChat uses one censorship policy in China and another internationally. Available at https://citizenlab.org/2016/11/wechat-china-censorship-one-app-two-systems/, November 2016.

[116] L. Ruan, J. Knockel, and M. Crete-Nishihata. We (can't) Chat: "709 Crackdown" Discussions Blocked on Weibo and WeChat. Available at https://citizenlab.org/2017/04/we-cant-chat-709-crackdown-discussions-blocked-on-weibo-and-wechat/, April 2017.

[117] Save Tibet. Tensions escalate in Qinghai: Rebkong self-immolation, student protest, monks commemorate March 10. Available at https://www.savetibet.org/tensions-escalate-in-qinghai-rebkong-self-immolation-student-protest-monks-commemorate-march-10/, March 2012.

[118] C. Shiflett. addslashes() versus mysql_real_escape_string(). Available at http://shiflett.org/blog/2006/jan/addslashes-versus-mysql-real-escape-string, January 2006.

[119] SINA Corporation. Form 20-F: Annual Report for the Fiscal Year ended December 31, 2011. Available at https://www.sec.gov/Archives/edgar/data/1094005/000110465912030028/a12-7070_120f.htm, 2011.

[120] Skype. What is TOM Online? Available at https://support.skype.com/en/faq/FA10910/what-is-tom-online.

[121] P. E. Smouse, J. C. Long, and R. R. Sokal. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic zoology*, 35(4):627–632, 1986.

[122] A. Soldatov and I. Borogan. Putin brings China's Great Firewall to Russia in cybersecurity pact. Available at https://www.theguardian.com/world/2016/nov/29/putin-china-internet-great-firewall-russia-cybersecurity-pact, November 2016.

[123] State Administration of Press, Publication, Radio, Film and Television of The People's Republic of China. Review and approval of online game publication. Available at http://www.gapp.gov.cn/govservice/1966/271413.shtml.

[124] T. Tang. China SAPPRFT Issued Implementation Rules to Streamline the Approval Process of Mobile Games. Available at http://www.lexology.com/library/detail.aspx?g=4c3afd11-820f-4b4b-9854-519495ec9d73, June 2016.

[125] The Guardian. Hong Kong bookshops pull politically sensitive titles after publishers vanish. Available at https://www.theguardian.com/world/2016/jan/07/hong-kong-bookshops-pull-politically-sensitive-titles-after-publishers-vanish, January 2016.

[126] The Ministry of Culture of the People's Republic of China. Regulation on the administration of publishing. Available at https://www.cecc.gov/resources/legal-provisions/regulation-on-the-administration-of-publishing-chinese-and-english-text, December 2001.

[127] UNITED NATIONS CONVENTION ON THE LAW OF THE SEA. The South China Sea Arbitration. Available at https://pca-cpa.org/wp-content/uploads/sites/175/2016/07/PH-CN-20160712-Award.pdf, July 2016.

[128] N. Villeneuve. Search monitor project: Toward a measure of transparency. Available at http://citizenlab.org/wp-content/uploads/2011/08/nartv-searchmonitor.pdf, 2008.

[129] N. Villeneuve. Breaching Trust: An analysis of surveillance and security practices on China's TOM-Skype platform. Available at http://www.infowar-monitor.net/breachingtrust/, 2009.

[130] S. Walker. Russia blocks access to LinkedIn over foreign-held data. Available at https://www.theguardian.com/world/2016/nov/17/russia-blocks-access-to-linkedin-over-foreign-held-data, November 2016.

[131] Wall Street Journal. China's Former Security Chief Zhou Yongkang Sentenced to Life in Prison. Available at https://www.wsj.com/articles/chinas-former-security-chief-zhou-yongkang-sentenced-to-life-in-prison-1434018450, June 2015.

[132] Washington Post. Keywords Used to Filter Web Content. Available at `http://www.washingtonpost.com/wp-dyn/content/article/2006/02/18/AR2006021800554.html`, February 2006.

[133] N. Weaver, R. Sommer, and V. Paxson. Detecting Forged TCP Reset Packets. In *NDSS*. The Internet Society, 2009.

[134] Z. Weinberg, M. Sharif, J. Szurdi, and N. Christin. Topics of controversy: An empirical analysis of web censorship lists. *Proceedings on Privacy Enhancing Technologies*, 2017(1):42–61, 2017.

[135] P. Winter and S. Lindskog. How the Great Firewall of China is Blocking Tor. In *the 2nd USENIX Workshop on Free and Open Communications on the Internet*, Bellevue, WA, 2012. USENIX.

[136] J. Wright. Regional Variation in Chinese Internet Filtering. Technical report, University of Oxford, 2012. Available at `http://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2265775_code1448244.pdf?abstractid=2265775&mirid=3`.

[137] Xia Chu. Complete GFW Rulebook for Wikipedia. Available at `https://github.com/cirosantilli/china-dictatorship/raw/master/complete-gfw-rulebook-for-wikipedia-v3.0.pdf`, December 2013.

[138] Xia Chu. An Audit on Bing's China Censorship. Available at `https://docs.google.com/file/d/0B8ztBERe_FUwM2JEN2tZUUtBMEU/edit`, February 2014.

[139] Xinhua. Statement of the Ministry of Foreign Affairs of the People's Republic of China. Available at `http://news.xinhuanet.com/english/china/2012-09/10/c_123697340.htm`, September 2012.

[140] Xinhua. 中共首都互联网协会委员会成立_对话首都_新华网 (Communist Party of China establishes the Capital Internet Society Committee). Available at `http://news.xinhuanet.com/2012-11/06/c\_113615083.htm`, November 2012.

[141] T. Zhu, C. Bronk, and D. S. Wallach. An Analysis of Chinese Search Engine Filtering. *CoRR*, abs/1107.3794, 2011.

[142] T. Zhu, D. Phipps, A. Pridgen, J. R. Crandall, and D. S. Wallach. The velocity of censorship: High-fidelity detection of microblog post deletions. In *the 22nd USENIX Security Symposium (USENIX Security 13)*, pages 227–240, Washington, D.C., 2013. USENIX.

[143] J. Zittrain and B. Edelman. Internet filtering in China. *Internet Computing*, 7(2):70–77, 2003.

[144] 张训. 口袋罪视域下的寻衅滋事罪研究. 政治与法律, (3), 2013. Available at `http://article.chinalawinfo.com/ArticleHtml/Article_78400.shtml`.