

Whiskey, Weed, and Wukan on the World Wide Web: On Measuring Censors' Resources and Motivations

Nicholas Aase
Univ. of New Mexico

Jedidiah R. Crandall¹
Univ. of New Mexico

Álvaro Díaz
Universidad de Grenada, Spain

Jeffrey Knockel
Univ. of New Mexico

Jorge Ocaña Molinero
Universidad de Grenada, Spain

Jared Saia
Univ. of New Mexico

Dan Wallach
Rice Univ.

Tao Zhu
Independent Researcher

Abstract

The ability to compare two instances of Internet censorship is important because it is the basis for stating what is or is not justified in terms of, for example, international law or human rights. However these comparisons are challenging, even when comparing two instances of the same kind of censorship within the same country.

In this position paper, we use examples of Internet censorship in three different contexts to illustrate the importance of the elements of motivation, resources, and time in Internet censorship. We argue that, while all three of these elements are challenging to measure and analyze, Internet censorship measurement and analysis is incomplete without all three.

The contexts we draw examples from are: public wireless networks in Albuquerque, New Mexico, USA; microblogging in China; and, chat programs in China.

1 Introduction

The “tit-for-tat” aspects of Internet censorship are often discussed. For example, if the censor blocks keywords, the censored can use different strings of bytes with the same meaning, and the censors can respond by censoring those, then the censored can employ even more such strings, this causes the censor to use context analysis or human moderators, and so on. This “tit-for-tat” view of Internet censorship ignores that fact that the censor and censored have some level of motivation to accomplish various goals, some limited amount of resources to expend, and real-time deadlines that are due to the timeliness of the information that is being spread.

There has been a considerable amount of work both on measuring censorship for what is censored and how [8, 29, 22, 12, 5, 6, 20, 13, 19, 15, 2, 23, 14, 28, 10, 1, 27, 21, 25] and on the social, historical, political, economic, and legal aspects of Internet censorship [2, 15, 9, 4, 7, 11, 16, 17, 18]. Our position in this paper is that motivation,

resources, and time are three important elements in Internet censorship that technical measurement studies must find a way to incorporate into their measurements in order to bridge the gap between the measurements and the social sciences aspects of Internet censorship.

We explore these three elements and give examples of the importance of each in different contexts in which we have measured Internet censorship. The first context is censorship on public wireless networks in Albuquerque, New Mexico, USA. We performed some very simple measurements to learn what common categories of content were blocked. This is described in Section 2. Section 3 discusses a different context: microblogging in China. The third and final context is chat programs in China, which is discussed in Section 4. We discuss results in Section 5 and conclude in Section 6.

2 Public wireless in Albuquerque, New Mexico, USA

As part of a semester-long independent study project, we developed an application for Android devices which was used to test filtering on publicly available wireless networks in Albuquerque, New Mexico, USA. The application attempts to connect to multiple websites whose contents are related to pornography, gambling, social networking, drugs, shopping, and hacking. The websites for testing these categories were chosen based on Google searches and websites known to us. The application allows the user to check either all, or a subset of, these subjects, and the user can add and remove specific sites to the application’s list of queries as well.

The user is notified by the application if certain classes of sites are blocked, and the user can also specify what percentage of queried sites can be blocked before they are notified of the censorship. Table 1 shows the publicly open wireless networks that we tested, and the results of which categories showed evidence of censorship, if any.

A full list of the websites tested by the application for each category is available upon request. To con-

Location	Filtered Subject Matter
DSL	-
Lobo-Guest	-
Lobo-WiFi	-
Lobo-Sec	-
Albuquerque Public Library	Pornography
Sunport International Airport	Pornography
Albuquerque Police Department	-
Albuquerque Convention Center	-
Albuquerque Townhall Wi-Fi	-
Wendy's	Gambling, Pornography, Alcohol and Tobacco
Satellite Coffee	Peer-to-Peer
Hookah Star Restaurant/Cafe	-
Papa John's	-
Brickyard Bar	-
McDonald's	-
Starbuck's Coffee	-
Bank of America	-

Table 1: Censorship on public wireless networks in Albuquerque, New Mexico, USA.

firm the censorship of gambling, pornography, and alcohol and tobacco at Wendy's we used a web browser to access <http://bwin.com>, <http://xvideos.com>, and <http://marlboro.com>, respectively. In all three cases our browser displayed the message (using Alcohol/Tobacco as an example):

```
This site is blocked by the
SonicWall Content Filter Service.
URL: http://marlboro.com. Reason
for restriction: "Forbidden
Category Alcohol/Tobacco."
```

3 Microblogging in China

Weibo is a microblogging website in China that is similar to Twitter. In this section we draw upon the China Digital Times' tracking of what is censored on Weibo [3], some recent work on the censorship and deletion practices of Weibo [1], and our own tracking of Weibo posts and experiments on Weibo. In our own efforts to track Weibo, we use a public interface and are able to see between 1% and 10% of all Weibo posts over the course of a day depending on the traffic workload. We estimated this coverage when the post IDs were still predictable, making it possible to know what percentage of the posts we had obtained.

One interesting observation from our data for Weibo is that often the sensitive keywords that were reported by the China Digital Times [3] to be censored by Weibo never appeared in our data, or appeared very sparsely. This suggests that the censors are very proactive, and get

ahead of news stories before the interplay between traditional media and social media allows for a sensitive topic to appear on Weibo.

In addition to allowing users to post microblog posts, Weibo has a search feature for searching all of Weibo for posts containing keywords. An interesting observation, which we made from experiments to compare the censorship in the search feature with the censorship of microblog posts, is that the search censorship is more aggressive than the microblog posting censorship. We determined this by testing many examples where searching for a keyword returned no results but posting the same keyword was not blocked.

Another interesting observation was that the censorship of posts on Weibo does not check for wildcards unless the user making the post has recently triggered a keyword. For example, posting “法ccc轮” (Fa-ccc-lun as a neologism for Falun) did not trigger censorship because the server was not checking for the regular expression at first, only for the literal string, “法轮”. However, after posting “法轮” and triggering censorship, posts containing “法ccc轮” would then also trigger censorship, indicating that the Weibo server apparently remembers users who have triggered a keyword and performs more aggressive pattern matching to prevent variations on the keyword.

One major difference between Weibo and other typical social microblogging sites such as Twitter is that Weibo allows users to post pictures. It is common to post blog posts as pictures. Sometimes these pictures contain sensitive words, but often the reason for posting a post as a picture is not to evade censorship but rather to over-

come the 140-UNICODE-character limit of Weibo (140 UNICODE characters in Chinese can express roughly 3-5 times as much information as 140 ASCII characters in English characters for a Twitter post can). Another interesting difference of Weibo is that it allows comments on microblog posts. This recently caused Weibo and another Chinese microblogging site to be shut down temporarily because the comments were not being censored as aggressively as the posts, leading to rumors, *e.g.*, that tanks were rolling into Beijing [26].

Another interesting observation can be found in Figure 1, which shows some potentially interesting dynamics of a particular censored keyword and neologism. Day 0 for the graph is 15 November 2011, and the *y*-axis is on a log-scale plot where 1 corresponds to no posts for that particular day. The events of the Wukan village protests are described on Wikipedia [24]. On 9 December 2011 Xue Jinbo was arrested, and he died in police custody on 11 December 2011. The first peak of the original word (乌坎 Wukan, in blue with squares) occurs on 12 December 2011, the day that protests began. By 14 December 2011, when posts for Wukan nearly hit zero, potentially due to censorship, thousands of police had laid siege to the village. The neologism (乌坎 Niaokan, in red with circles) came into existence on 13 December 2011. The peak for wukan on 21 December 2011 corresponds with an announcement that the government and the villagers had reached a peaceful agreement. It appears that censorship was applied only long enough for the news about Wukan to change from sensitive news to a story of successful government intervention to reach a peaceful resolution to the problem.

4 Chat programs in China

To obtain a broader picture of chat program censorship in China, we reverse-engineered SinaUC chat to obtain the URL and encryption key for downloading the various censorship lists it uses on a daily basis. We then compared our results to the results of Knockel *et al.* [14] for TOM-Skype. SinaUC and TOM-Skype are the only two major chat programs that we know of that download their censorship and surveillance keyword lists into the client. Other programs such as QQChat perform censorship on the server side.

SinaUC, unlike TOM-Skype, does not perform surveillance. SinaUC has five lists that are built in but are also updated via the web. One list censors incoming and outgoing text chat and usernames (replacing the usernames with an ID number instead), one censors only usernames (again, replacing offending usernames with an ID number), a third censors incoming and outgoing text chat, and two more have unknown purposes.

The most salient observation in comparing the SinaUC and TOM-Skype lists is that they do not appear to have

common origins. Going by unique strings, a snapshot of both lists from 6 April 2012 shows 1130 keywords in TOM-Skype and 1490 in SinaUC. Only 21 words are common to both lists, and all 21 are very common stock keywords that we would expect to see on any censorship list in China, such as 法轮 (falun), fuck, and 大纪元 (The Epoch Times).

The Tom-Skype list was frequently updated based on current events. Knockel *et al.* [14] reported specific protest locations and other keywords related to the Jasmine protests in the censorship and surveillance list and many keywords related to the Beijing demolitions when they first obtained the decryption keys and blacklist URLs for TOM-Skype in March 2011. We used these keys and URLs to track TOM-Skype in addition to SinaUC, for comparison, and in the TOM-Skype lists we witnessed several changes related to current events (but the changes were relatively infrequent, sometimes months apart rather than every day). Examples of black-listed keywords for both censorship and surveillance from TOM-Skype included:

- 薄熙来 (Bo Xilai), who first appeared on the blacklist on 16 May 2011, but was removed the next day. It is not clear why his name appeared on this day. His name was featured more prominently on the blacklist starting 21 March 2012 because of the political issues surrounding him.
- 荣守京 (Rong Shoujing), which is the pseudonym of a reporter who alleged police torture of Ai Weiwei in a user-submitted article published in April 2011 by Human Rights in China (HRIC) [30]. There is no confirmation of the veracity of these claims and HRIC emphasized in a statement that this was a user-submitted article and not their own reporting. Rong Shoujing's name first appeared on the blacklist after the article was published for a period of time. For an unknown reason, as of 27 March 2012 this name is the only keyword on some versions of the blacklist. For a reason that is also unknown to us, on 27 March 2012 and persisting for one day the keyword blacklist had 1134 identical keywords, all of which were the same string, “荣守京”.
- 占领长安街 (Occupy Chang An Street), 占领温州 (Occupy Wenzhou), 占领上海 (Occupy Shanghai), 占领山东临沂 (Occupy Shangdong), *etc.* These keywords appeared soon after the emergence of the Occupy Wall Street and related Occupy movements in the United States.

Note that the TOM-Skype blacklists come in many versions, so the above dates are simplifications based

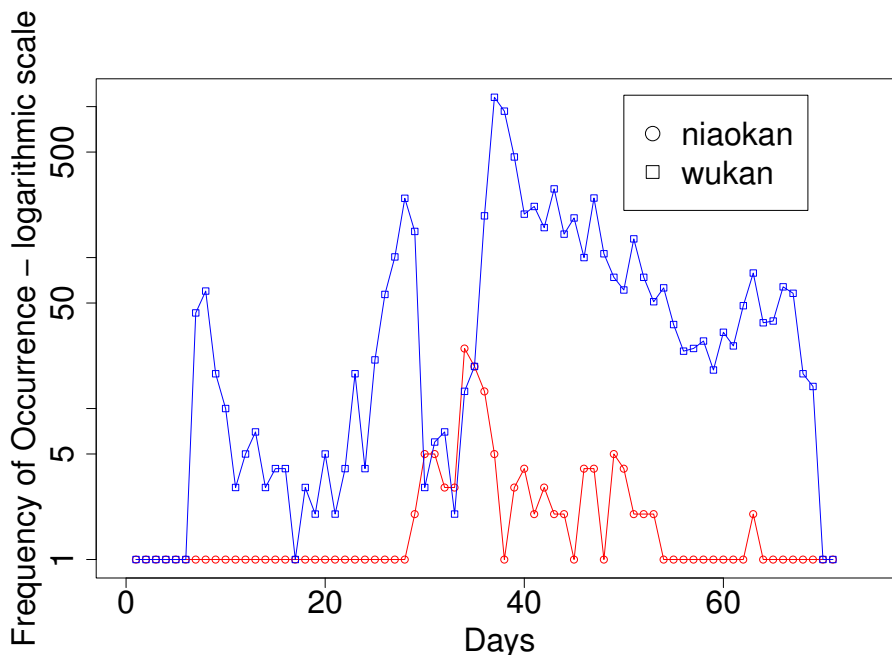


Figure 1: Bigram trends for 乌坎 (Wukan) and 鸟坎 (Niaokan). 鸟坎 is a neologism for 乌坎 (Wukan), a village in southern China that was the site of considerable social unrest in 2011. Day 0 is 15 November 2011, and the y -axis is on a log scale with 1 corresponding to zero posts for that particular day.

on various lists. The SinaUC blacklists, in contrast to the TOM-Skype blacklists, are not updated with current events and appear to be more targeted at spam removal than at censorship (though there are many keywords that are clearly censorship, including censorship of political content, pornography, and other categories that are typically censored in China).

5 Discussion

Motivation is perhaps the most surprising element to Internet censorship that we have encountered, because we had assumed that censors are fully motivated to block content and the censored are fully motivated to disseminate it. In our tracking of chat programs in China, we noticed a phenomenon that we call the “intern effect.” At times, we often suspected that a keyword blacklist was being typed up by an over-worked college intern who was given vague instructions to filter out anything that might be against the law. There probably is a more rigorous legal process involved in a company deciding what they should block, but we have not been able to understand this process. The company developing the blacklist relatively independently from the government would be consistent with Link’s concept of the “Anaconda and the Chandelier” [16], in which companies are given free reign to censor whatever they deem to be sensitive, with the understanding that the government will come down

hard on companies that do not do a good job of this. At other times, however, we witnessed very specific protest locations and times. It is still not clear to us who makes the lists that are used for censorship of chat programs, and if their motivations are directly related to state security or are simply based on a broad definition of spam. Further raising doubts about the existence of a central blacklist for chat programs in China, from which the different companies derive their own lists, is the fact that there was very little overlap between SinaUC and TOM-Skype.

Motivation in Weibo censorship seems more clear. Very specific keywords cause posts to be flagged for moderation, and the posts that are allowed *vs.* those that are deleted are consistent with China’s Internet censorship policies. Perhaps one factor that explains the apparent difference in motivation between chat program censorship and microblogging censorship is that, according to Mr. Tao [19], censorship of communications (such as telephony) and censorship of news (such as blogs) are handled by two different government bodies.

Motivation plays an interesting role in the censorship that we observed for open wireless networks in Albuquerque. What was censored in the library was consistent with federal law, and the censorship of peer-to-peer content in the coffee shop was probably an effort to defend against abuses of network resources. Why is pornogra-

phy, gambling, and content related to drugs and alcohol filtered in a fast food restaurant, though? If this fast food chain has had problems with abuses of their network for these purposes, why do other fast food chains not also filter these types of content? Who is being targeted with the censorship of pornography at the Albuquerque Sunport airport, the passengers or the employees of the airport?

Our measurements for open wireless networks in Albuquerque raised more questions about motivation than they answered about the censorship itself. Also, it is assumed but not explicitly measured in our results that the censorship is relatively static and simply contains a blacklist of keywords or websites that is updated but is not adapted to block new content related to current events. In order to detect censorship related to current events, a measurement technique must explicitly look for this.

The **resources** that the censor must expend is another interesting element of Internet censorship. This was most apparent in the Weibo post and search censorship. One hypothesis about why searching posts is more stringently censored than creating new posts is that it requires less resources. In other words, censoring posts may be more important than censoring searches, but censoring posts consumes many more resources so the censorship is not applied as broadly for posts as it is for searches. Searches are much shorter than posts, and searching for extremely obfuscated versions of a keyword will only cause a search to fail, whereas, with a post human readers will still be able to read obfuscated posts. Weibo's post censorship had the optimization mentioned in Section 3 where it would only check for regular expressions for obfuscations of keywords after a user had triggered censorship with straightforward versions of the keyword. Also, our experiments and the results of others [1] indicate that a large amount of human effort is involved in Weibo post censorship. Until recently, comments on posts went largely un-moderated. This all points to the fact that in some cases censors are very resource-constrained and must apply censorship with discretion.

Finally, **time** is an important element of Internet censorship, and is perhaps the most challenging to measure objectively. The censored have some timely information that they want to disseminate, and the censor attempts to slow down this dissemination. What is interesting about the Wukan example in Section 3 is that many of the censored may have changed their mind about the information to disseminate. The censors stopped Wukan from becoming a trend for a few days, and in that time the government found a peaceful resolution to the unrest in Wukan. Thus when the ban on the keyword Wukan was lifted, Wukan did become a trend but probably from microbloggers spreading the good news of the peaceful resolution, and not from those who had formerly had

grievances about the situation. It is important to understand more advanced forms of media manipulation, such as using Internet censorship to get ahead of a news cycle, because they may be more effective than the long-term restrictions that are more typically cited as Internet censorship.

A very important aspect of Internet censorship, which is not the focus of this paper but is worth mentioning, is surveillance. Surveillance was explicitly visible in our study of chat programs in Section 4. For open wireless networks in Albuquerque from Section 2 and Chinese social media in Section 3, we could only speculate whether surveillance is also part of the capabilities the censors have built into the same systems used for censorship. Surveillance is powerful because it can provide information about what might need to be censored, and when users know that censorship is paired with surveillance, the surveillance can reinforce self-censorship behavior.

5.1 Open questions

The questions we would like to discuss at the workshop, from both the social sciences and technical measurement perspectives, include:

- How can measurement methodologies incorporate motivation, resources, and time in a meaningful way?
- Would it be helpful to document Internet censorship measurements in a way that matches some legal definition of a trade law, human rights abuse, *etc.*? For example, could the motivation, resources, and time be tied to many specific instances of stifling the right to free assembly over a period of time? Would this be helpful for affecting policy?
- What tools can we use to separate the layers of censorship in measurements? For example, in a keyword blacklist that is tracked over time, is it possible to know which terms the company censored because of general laws and policies and which terms were the result of more specific instructions related to current events?
- For censorship that is not government-sponsored, such as the censorship we measured at Wendy's, what elements of motivation, resources, and time should we be measuring for? In other words, what aspects of motivation, resources, and time would be the most likely to have policy implications, if any?

6 Conclusion

We have given examples from Internet censorship measurements in different contexts where motivation, resources, and time were important elements in the application and potential uses of Internet censorship. The

challenge for the Internet censorship research community is to develop methods for measuring these three elements explicitly when performing measurement studies. The first step in addressing this challenge will be to determine exactly how we expect measurement studies to influence policy. We hope that these issues will be topics of discussion at the workshop.

Acknowledgments

We would like to thank the FOCI anonymous reviewers for valuable feedback. This material is based upon work supported by the National Science Foundation under Grant Nos. #0644058, #0844880, #0905177, and #1017602. Jed Crandall is also supported by the Defense Advanced Research Projects Agency.

When possible, all source code, experiments, and results used to support our position in this paper will be made available. Some experiments (such as the regular expression matching experiments for Weibo) were performed informally. Our Weibo data may have privacy concerns (for example, if users have deleted posts for privacy reasons and those posts are still in our database) that restrict its release. For any questions or requests concerning data, code, methods, etc., please email the corresponding author.

Notes

¹Corresponding author, e-mail address: crandall@cs.unm.edu.

References

- [1] BAMMAN, D., O'CONNOR, B., AND SMITH, N. A. Censorship and deletion practices in chinese social media. *First Monday* 17, 3 (2012).
- [2] CHASE, M. S., AND MULVENON, J. C. *You've Got Dissent! Chinese Dissident Use of the Internet and Beijing's Counter-Strategies*. RAND Corporation, 2002.
- [3] CHINA DIGITAL TIMES. Regularly updated list of keywords censored on weibo. Available at https://docs.google.com/spreadsheet/ccc?key=0Aqe87wrWj9w_dFpJWjZoM19BNkFfV2JrWS1pMEtYcEE#gid=0.
- [4] CLAYTON, R. Failures in a hybrid content blocking system. In *Privacy Enhancing Technologies* (2005), pp. 78–92.
- [5] CLAYTON, R., MURDOCH, S. J., AND WATSON, R. N. M. Ignoring the great firewall of china. *I/S: A Journal of Law and Policy for the Information Society* 3, 2 (2007), 70–77.
- [6] CRANDALL, J. R., ZINN, D., BYRD, M., BARR, E., AND EAST, R. ConceptDoppler: a weather tracker for Internet censorship. In *Proc. of 14th ACM Conference on Computer and Communications Security (CCS)* (2007).
- [7] DANEZIS, G., AND ANDERSON, R. The economics of resisting censorship. *IEEE Security and Privacy* 3, 1 (2005), 45–50.
- [8] DEIBERT, R. J., PALFREY, J. G., ROHOZINSKI, R., AND ZITTRAIN, J. Access denied: The practice and policy of global internet filtering. *The MIT Press* (2007).
- [9] DORNSEIF, M. Government mandated blocking of foreign web content. In *Security, E-Learning, E-Services: Proceedings of the 17. DFN-Arbeitstagung über Kommunikationsnetze* (2003), J. von Knop, W. Haverkamp, and E. Jessen, Eds., Lecture Notes in Informatics, pp. 617–648.
- [10] ESPINOZA, A., AND CRANDALL, J. R. Work-in-progress: Automated named entity extraction for tracking censorship of current events. In *FOCI 11: Proceedings of the USENIX Workshop on Free and Open Communications on the Internet* (2011).
- [11] FIAT, A., AND SAIA, J. Censorship resistant peer-to-peer content addressable networks. In *Proceedings of the Thirteenth ACM Symposium on Discrete Algorithms (SODA)* (2002).
- [12] The Great Firewall Revealed. Whitepaper released by the Global Internet Freedom Consortium in December of 2002.
- [13] “Race to the Bottom”: Corporate Complicity in Chinese Internet Censorship. In *Human Rights Watch* (August 2006). <http://www.hrw.org/reports/2006/china0806>.
- [14] KNOCKEL, J., CRANDALL, J. R., AND SAIA, J. Three researchers, five conjectures: An empirical analysis of tom-skype censorship and surveillance. In *FOCI 11: Proceedings of the USENIX Workshop on Free and Open Communications on the Internet* (2011).
- [15] LIANG, C. Red light, green light: has China achieved its goals through the 2000 Internet regulations? *Vanderbilt Journal of Transnational Law* 345 (2001).

- [16] LINK, P. China: The anaconda in the chandelier. *The New York Review of Books*, April, 2002.
- [17] MACKINNON, R. Flatter world and thicker walls? blogs, censorship and civic discourse in china. *Public Choice* 134, 1-2 (2007), 31–46.
- [18] MACKINNON, R. China’s censorship 2.0: How companies censor bloggers. *First Monday* 14, 2 (2009).
- [19] MR. TAO. China: Journey to the heart of Internet censorship. Investigative report sponsored by Reporters Without Borders and Chinese Human Rights Defenders, Oct 2007.
- [20] PARK, J. C., AND CRANDALL, J. R. Empirical study of a national-scale distributed intrusion detection system: Backbone-level filtering of html responses in china. In *Proceedings of the 2010 IEEE 30th International Conference on Distributed Computing Systems* (Washington, DC, USA, 2010), ICDCS ’10, IEEE Computer Society, pp. 315–326.
- [21] SFAKIANAKIS, A., ATHANASOPOULOS, E., AND IOANNIDIS, S. CensMon: A Web Censorship Monitor. In *USENIX Workshop on Free and Open Communications on the Internet* (San Francisco, CA, USA, 2011), USENIX Association.
- [22] TOKACHU. The not-so-great firewall of China. *2600 Magazine*, Winter 2006–2007.
- [23] VILLENEUVE, N. Breaching trust: An analysis of surveillance and security practices on China’s TOM-Skype platform. Available at <http://www.infowar-monitor.net/breachingtrust/>.
- [24] WIKIPEDIA. Protests of wukan. <http://zh.wikipedia.org/wiki/乌坎事件>.
- [25] WRIGHT, J., DE SOUZA, T., AND BROWN, I. Fine-Grained Censorship Mapping: Information Sources, Legality and Ethics. In *USENIX Workshop on Free and Open Communications on the Internet* (San Francisco, CA, USA, 2011), USENIX Association.
- [26] XINHUA NEWS. China’s major microblogs suspend comment function to “clean up rumors”. 31 March 2012, Available at http://news.xinhuanet.com/english/china/2012-03/31/c_131500416.htm.
- [27] XU, X., MAO, Z. M., AND HALDERMAN, J. A. Internet Censorship in China: Where Does the Filtering Occur? In *Passive and Active Measurement Conference* (Atlanta, GA, USA, 2011), Springer, pp. 133–142.
- [28] ZHU, T., BRONK, C., AND WALLACH, D. S. An analysis of chinese search engine filtering. *CoRR abs/1107.3794* (2011).
- [29] ZITTRAIN, J., AND EDELMAN, B. Internet filtering in China. *IEEE Internet Computing* 7, 2 (2003), 70–77.
- [30] 荣守京 (RONG SHOUJING). 艾未未受酷刑认罪背后的惊天阴谋 (Ai Weiwei Tortured Until Confessing to Tax Evasion. Human Rights in China Biweekly article, available at <http://biweekly.hrichina.org/article/985>.