

Erasing Tibet: Censorship on Chinese Video Sharing Site Youku

November 30th, 2016



Tibet Action Institute

བོད་དོན་ལས་འགུལ་བརྒྱུ་གནས་ཁང་།

Open-Source Technology, Training and Strategy

Summary

Internet censorship limits the information Tibetans are able to access, and in particular, creates barriers between the free flow of ideas and information between Tibetans in exile and inside Tibet or China. Social media platforms offer means for Tibetans to express themselves, but these spaces in Tibet and China are heavily constrained by censorship and Tibet-related content is routinely targeted.

This report shows how Youku, one of the most popular video sharing sites in China, censors content related to Tibet. We conducted systematic tests in which videos with a range of content were uploaded to Youku to document how censorship is triggered on the platform. In our tests, any videos related to Tibetan culture and His Holiness the Dalai Lama were blocked, and videos with titles and descriptions in the Tibetan language were almost always blocked. In contrast, videos with content related to nonviolent action and digital security with titles in English or Chinese had lower instances of censorship. These results show that Tibet-related content and the Tibetan language are specifically targeted for censorship on Youku.

Analyzing these patterns can help create a greater understanding of the censorship challenges Tibetans face under China's information controls and potential strategies for communicating with them.

By uploading particular Tibet-related content on Youku, our aim was to begin to develop censorship-resistant strategies that could be used for communication with Tibetans behind the Great Firewall of China based on a nuanced understanding of the current patterns of censorship.

Background

Since 1959, the diaspora of Tibetans fleeing China's occupation of Tibet has grown. A [2009 census](#) by the Planning Commission of the Central Tibetan Administration estimated that there were up to 150,000 Tibetans living in exile. The flow of information over the Internet and via mobile phones between Tibetans in exile and Tibetans inside Tibet or China has become a critical lifeline of communication that is heavily impacted by the Chinese government's censorship.

Information Controls in China and Tibet

China maintains a national filtering system known as the Great Firewall (GFW) that censors content at the network level. [Numerous academic studies](#) have documented how the GFW operates. Tibetan websites and content are routinely blocked by the GFW, creating challenges for Tibetan and Chinese users to access content and for exile-based Tibetans to disseminate information into Tibet and China. While the GFW has become a popular metaphor for describing censorship in China, it does not [adequately describe](#) the complete apparatus of information control in the country.

Many popular social media platforms (e.g., Twitter, Facebook, YouTube, etc.) are blocked in China. Chinese technology companies have developed alternatives to these platforms that have gained wide popularity domestically. Social media companies operating in China are [held liable for content](#) on their platforms and are expected to invest in staff and technology for ensuring compliance with regulations. Failure to comply with regulations can lead to fines or revocation of operating licenses. To stay within regulatory boundaries, social media companies must block prohibited content. Whereas the primary goal of China's national Internet filtering is to restrict foreign websites and information, the goal of social media censorship is to restrict the communications and discourse happening behind the Great Firewall.

Previous research on social media censorship in China has consistently shown that content related to Tibet is systematically targeted. [Research on Sina Weibo](#) (a popular microblogging platform similar to Twitter) found that posts with sensitive words originating from Tibet had a higher deletion rate than posts from other areas such as Beijing. Other work has collected keyword lists used to censor messages on [chat apps](#) and [live streaming services](#). Across the [keyword lists](#) discovered by these studies are references to the Tibetan movement, religion and culture (e.g., “藏独,” “Tibetan Independence,” “达赖,” “Dalai Lama.”) Social media censorship coupled with the GFW creates a highly restrictive communications environment that has been shown to aggressively and routinely censor Tibet-related content. Since Tibetans in Tibet are subject to China's information controls and are as a result more likely to use social media apps developed by Chinese companies, it is important to understand how censorship on these apps works and what impacts it may have on how Tibetans can communicate.



Youku is one of the most popular online video platforms in China with 500 million monthly active users. Youku merged with a competing video site Tudou in 2012 creating Youku Tudou Inc. In 2015, Youku Tudou Inc. was acquired by Alibaba one of China’s largest technology companies.

In public filings, Alibaba acknowledges the business risk posed by Chinese government authorities penalizing the company for hosting prohibited content. The company explains that it is prohibited from disseminating any content that is in “violation of PRC laws and regulations, impairs the national dignity of China or the public interest, or is obscene, superstitious, fraudulent or defamatory”.

Previous Youku Tudou Inc. disclosures describe systems for monitoring and filtering prohibited and copyright infringing content. Content is automatically screened by a “video fingerprint technology,” a text filtering system that “screens content based on pre-set keywords,” and video processing that “screens out pornographic and obscene content based on colors and images.” Content that passes through these screenings is then reviewed by employees. Youku references periodic phone calls and written notices from Chinese regulatory authorities that provide orders to delete content deemed sensitive, and provide lists of prohibited content.

Documenting Censorship on Youku

Despite the popularity of Youku, there has been no systematic study of censorship on the platform. This report provides a first look at how Youku filters content with the goal of determining which languages and content types trigger censorship.

We developed five content types that we hypothesized would have varying probabilities of being blocked (see Table 1).

Table 1

Category	Description	Probability of blocking
Funny Videos	Generic videos with animals and humor	Very low
Tibetan Culture	Tibetan songs, dance, etc.	Low
Digital Security	Instructional videos on increasing online security	Medium
Nonviolent Action	Stories of nonviolent social change efforts not related to the Tibetan community	High
HHDL	Teachings from His Holiness the Dalai Lama	Very high

The probability of blocking escalated with the level of political content in the videos. The “Funny Videos” category represented the average innocuous video online and acted as a control. The “Tibetan Culture” category had Tibetan songs and was not political. The “Digital Security” category included instructional videos on increasing security online aimed at Tibetan audiences. While the content was not overtly political, we ranked the probability of blocking as medium since the topic may still be seen as sensitive. The “Nonviolent Action” category included videos about nonviolent social change efforts outside of the Tibetan context. Due to the explicitly political content of these videos we expected a high probability of blocking. Finally, we included videos about His Holiness the Dalai Lama which we suspected had a very high probability of blocking.

The Communist Party of China routinely denounces the Dalai Lama, the spiritual leader of the Tibetan people and Tibetan Buddhism. In Tibet, displaying any sign of loyalty to him, including possession of his photo (even on a cell phone), can be met with detention, arrest, torture, violent crackdowns and political ‘re-education’ programs. [Past research](#) on social media censorship in China has frequently found references to the Dalai Lama blocked. We ran searches for Dalai Lama-related keywords on Youku which did not return any relevant results, further suggesting that this is a category specifically targeted for censorship.

Testing Method

We created 15 Youku accounts for testing purposes and uploaded 15 videos from our five content categories with titles and descriptions in three different languages: Tibetan, Chinese and English. This combination of content type with language was used to determine if videos were being censored based on content or the written descriptions and metadata.

We uploaded the videos through a virtual private network (VPN) connection and a browser in incognito mode to change our network location each time we posted a new video. The purpose of this step was to ensure that Youku censorship was not based on location. After posting the videos, we disseminated links to them on WeChat to increase the number of views.

Using a script, we automated daily checks of the posted videos to determine if they were accessible. The script checked for censorship warning text, took a screenshot of the page and collected metadata (e.g., the number of videos, comments, etc.).

Creating accounts on Youku is a challenge. In 2014, the Chinese government introduced regulations that require any users who upload videos to Chinese video sharing sites to register with their real name. This policy was presented as an attempt to prevent “vulgar content” that would have a “negative effect on society,” though it was widely seen by rights groups and dissidents as an attempt to suppress anti-government sentiment.

Up until early 2016, users could still log in to Youku with a Renren account (the Chinese Facebook-like service), effectively bypassing the name verification process. However, by the end of February 2016, users were required to link existing accounts to an email or cell phone. Now, a cell phone number is needed for all new accounts, making it essentially a [real-name registration system](#). This restriction on registration impacted our research by limiting the number of accounts we could generate, and also greatly limits people in general who are posting to Chinese video sites.

Results

Our results show that content related to Tibet and videos with titles and descriptions in the Tibetan language were almost always blocked. The aggressive blocking of Tibet-related content surpassed our original expectations.

We expected that videos with content on Tibetan culture that was not overtly political would be less likely to be censored than videos on digital security and nonviolent action. We found the inverse to be true. All videos in the Tibetan culture category were consistently blocked including versions with titles and descriptions in each of the three languages. Digital security and non-violence videos with titles and descriptions in English were never blocked, and the same videos with titles and descriptions in Chinese were only blocked 20% of the time. Similarly videos in the Dalai Lama category in any of the three languages were consistently blocked. These results show that any Tibet-related content is more likely to be blocked than non-Tibet related content.

Language also proved to be a reliable trigger for censorship. Of the videos with descriptions and titles in Tibetan, 96% of the uploads, regardless of content, resulted in either a “Shielded” or “Transcoding Issue” error message, even though the videos were the same as the ones uploaded with English and Chinese text. At first, we suspected the error messages may have been due to Youku simply not supporting the Tibetan language. However, we found that a video on nonviolent action in English with titles and description in Tibetan was not censored, which shows that the platform can support Tibetan text. It is unclear why this video was not blocked, but for the other videos in the category, it appears that use of Tibetan language in the titles and descriptions triggered censorship. In comparison, only 48% of videos with titles and descriptions in Chinese or English were blocked. It is unclear why 40% of the Funny Videos uploaded in English were blocked.

video	status	heat	operating
<input type="checkbox"/>  Pangen Metok 2015-11-25 22:03	 Transcoding fails see details	 0  0  0	delete
<input type="checkbox"/>  Pangen Metok (Tibetan Song) 2015-11-24 22:03	 Shielded, other processing according to a audiovisual regulations see details	 1  0  0	delete

A Tibetan Song video about Tibetan Culture in the same format showing a shielded with one upload and then transcoding error.

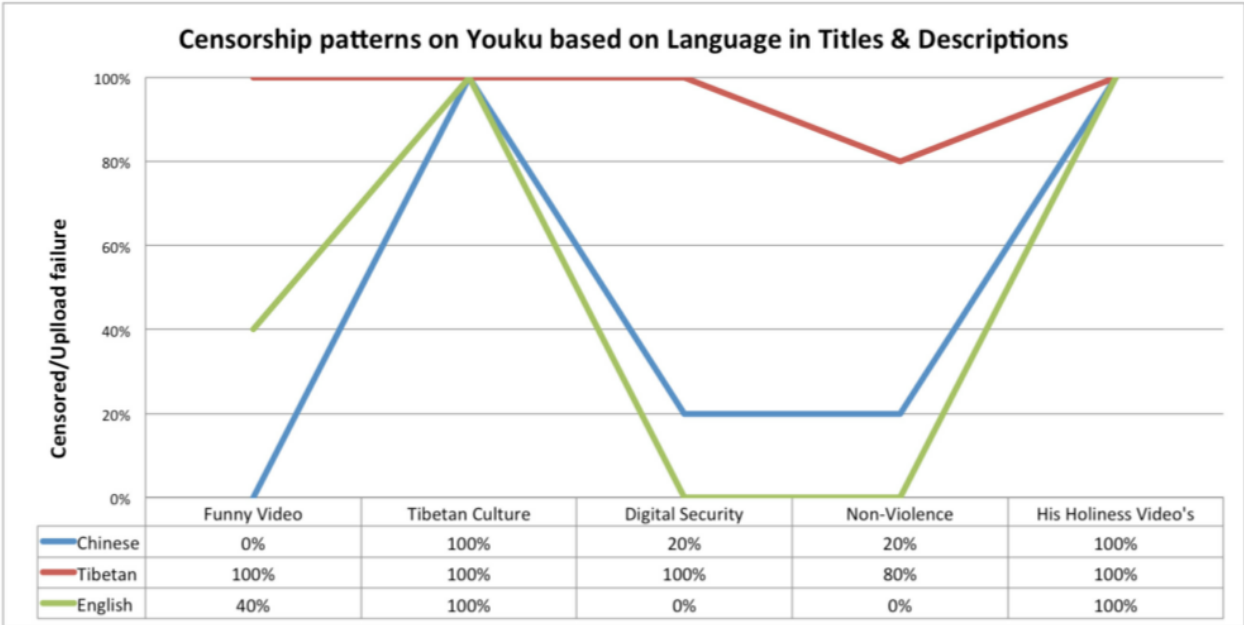
In a supplementary test, we uploaded a video on Tibetan culture with titles and description in Chinese describing non-Tibet related content. We found in this instance, the video was not censored. We attempted the same test with Dalai Lama videos but found regardless of the language or accuracy of the title and description that these videos were always blocked.

During our tests we did not find evidence that uploading a video that was subsequently censored had any effect on how other videos uploaded by the same account were treated.

Financial disclosure documents from Youku describe a text filtering system that “screens content based on pre-set keywords” before videos are manually reviewed. Our results suggest that this system may specifically target the Tibetan language, which would explain why we consistently found that videos with titles and descriptions in Tibetan were blocked regardless of the content. The consistent blocking of the Dalai Lama videos regardless of metadata could be due to these videos already being identified in the system and tagged as prohibited content potentially by the “video fingerprint” technology described by the disclosure documents.

These results are a first look at how Youku censors content. Further research is needed to gain a more comprehensive picture. However, what is already clear is that Tibet-related content is aggressively and systematically targeted by Youku.

Graph of Censorship Findings



Strategies for Disseminating Content on Youku

Since Youku has such a wide reach inside Tibet and China, it is an important platform to disseminate Tibet-related content. The censorship barriers we document show that getting this content on the the platform has to be done strategically.

As almost all of the videos with Tibetan language in the descriptions and titles were blocked, simply uploading videos on Youku using Chinese or English text is significantly less likely to trigger Chinese government censors. Also, including any references to the Dalai Lama or Tibetan culture in the title, description or content is likely to result in videos being censored.

It is possible that using mundane and unrelated descriptions and titles may keep a video from being censored, depending on the video content, though more research would be needed to determine this for certain. Also, if a video is censored, it seems possible to have it published by changing the description and/or title, even if it is uploaded using the same account. However, this may not always be the case, so using multiple accounts when uploading remains the best option.

Finally, when uploading videos to Youku, it is critical to maintain best practices for staying anonymous online, including using a VPN along with incognito mode in the browser. These steps will ensure that uploads happen from different locations and prevent IP-based censorship.

Conclusion

The most striking discovery from our research is that censorship on Youku is based on both content and language. Any video about the Dalai Lama was blocked as were nearly all the videos with titles and descriptions in the Tibetan language.

Given that Youku is the highest viewed Chinese video sharing site, it is a potentially powerful platform for publishing Tibet-related content. Therefore, it is important to develop nuanced, censorship-resistant strategies for communicating to Tibetans behind the Firewall. This research project represents the beginning steps of deeper analysis into censorship on Youku and developing such strategies.

Research into how Tibet-related information is actively censored on Youku and other Chinese social media presents an important opportunity to highlight more broadly China's criminalization of Tibetan culture and language. With more extensive research, it can be an avenue for holding Chinese internet companies accountable and challenging them on their discriminatory practices with regard to Tibetan language and content allowed online.

ENDNOTES:

1. Listed below are the categories of videos uploaded. As the videos themselves and the accounts used to upload them are publicly visible, we have not shared the actual titles and video descriptions so that we can use them for future testing.

Category 1: Funny Videos

Content language: English

Video title and description: Available in Tibetan, Chinese and English.

Category 2: Tibetan Culture

Content language: Tibetan

Video title and description: Available in Tibetan, Chinese and English

Category 3: Digital Security

Content language: Tibetan

Video title and description: Available in Tibetan, Chinese and English

Category 4: Nonviolent Action

Content language: English

Video title and description: Available in Tibetan, Chinese and English

Category 5: His Holiness the Dalai Lama

Content language: Tibetan

Video title and description: Available in Tibetan, Chinese and English
