MEASURING QQMAIL’S AUTOMATED EMAIL CENSORSHIP IN CHINA

Jeffrey Knockel and Lotus Ruan
Citizen Lab, University of Toronto
❌ 法轮大法 (Falun Dafa)
❌ 维吾尔人失踪 (Uyghurs are missing)
❌ 暗杀 + 习近平 (Assassination + Xi Jinping)
刘士辉 (Liu Shihui, a human rights defender)
一个光明的民主中国 (bright democratic China)
法轮大法 (Falun Dafa)
维吾尔人失踪 (Uyghurs are missing)
暗杀 + 习近平 (Assassination + Xi Jinping)
…censored unless it contains…
刘士辉 (Liu Shihui, a human rights defender)
一个光明的民主中国 (bright democratic China)
Legal environment in China

- Information controlled via intermediary liability
- Fines
- License revocations
- And more
Tencent’s QQMail

- Tencent operated (WeChat, QQ, …)
- Very popular in China
- gmail.com blocked in China
- Bound by Chinese laws
Want to measure how email message content is censored

- Without requiring QQMail accounts
- Without requiring vantage points in China
Sender

SMTP port 465

DNS server

MX record lookup

Recipient

Sender’s SMTP server

SMTP port 25

Recipient’s MX server
SMTP Error:
Mailbox unavailable

Bounce email
How can we tell if a message is censored?

- “Mail content denied” → message censored
- Even if the recipient address does not exist
- Use nonexistent recipient address
- “Mailbox unavailable” → message not censored
SMTP Error: Mailbox content denied
How do we generate test messages?

1. Randomly compiled from other keyword lists (hundreds of thousands of keywords)
2. Spliced from Chinese-language Tweets
簧薇片zcpx.com几八全体力量型武师攻击增加405点日GM马驰www.82yyy.com肾源中介qq關税纳税元述:我也想试试。黄师傅, 请。鴿派91jqw花貝涛aa.beijingxx7.com政治幼女高中毕业之销魂初夜錦锋牌B8*塞你公*中国领导人P的冷(8)线漳各种门种子52旅金镶玉艳8jiufengboAV情色周yongkang瞧鹅投象3lu7f萝زوا逼学生运动修頓球场涩莠骡烷少妇操b图操农村少妇少妇操尿图橘琉璃东方鸿运娱乐城BRRN.tk批发苯乙酸帮人清债无忧国际娱乐2饲蛤 refute红色御林军中国是不讲理q2686965186投^&公司七灵派传播开房被監控角色瞄沃www.321hack.com民用全套无敌霸王想赚零花钱伽个签国会操锦涛吃过几个男人的免费淫色影院我爱狐狸精毛meimei免费电影秘书
wxdkfp.netyuwangtiantang.com100999999.orgadminhd.com兼\\.职长春下午两点茉莉花成人av看黄片揭黑打貪本站10006多部激情电影全部免费试看三天欲火焚身Unbalanceアンバランスa.aa88567.cn廿陆年前天an门唐人街情色导航杨*巍老菊花回良玉*
⑧八九八*八瘟假鸨零扒宪璋代开日月图库硝铵炸药的制作方法大陆FuckHead磅解码器你妈个臭亲自指挥176小康目标不谨慎qq卖自製手狗供应餐饮定额发票陈宗興農產品悉近平两会東北獨立迷换药茼城約 athletes|两会报道007間諜最新版二十年前担坦克|二脚交性免费电影|性香港ghb水qq六四清場代刷7钻參選鲷鰤鰟530惨審{2}叶{2}电{2}器letiantang1.com供应仿真槍模sixfour运动儿子顶到姐姐的花心80suncitynet学生妹兼职成人高清电影下载算你色视频孙政才
دوختور白莲教又出现了, 这下只好派人去查看了！
What’s triggering the censorship?

- Ruohan et al. (FOCI 2019)
- Efficiently determine which combination of keywords censored a message
- Using as few test messages as possible
Keyword combination

E.g., 习主席 + 辞职 + 提出
(Chairman Xi + resignation + proposed)

Each of the keyword combination’s components must be present somewhere in the message, in any order.
天赋人权 + 勿忘六四

(natural rights + don’t forget June 4)
173 censored keyword combinations

- Political System: 33
- Xinjiang: 35
- Mobilization: 16
- Dissident: 9
- Tiananmen: 6
- Power Struggle: 29
- Falun Gong: 5
- Tibet: 1
- General: 2
- Xi Jinping: 32
- Party leader: 3
- Tencent: 2
Plurality Xinjiang-related

- 东突同胞解放 (the liberation of East Turkistan compatriot)
- 维吾尔人失踪 (Uyghurs are missing)
- 低劣种族 + 新疆 (low-class ethnicity + Xinjiang)
Related to Tencent CEO Ma Huating

- 女子 + 马化腾 (bitch + Ma Huating)
- 马化腾 + 人渣 (Ma Huating + scum)
38 extenuating combinations

- Keyword combinations whose presence disables censorship of one or more censored keyword combinations
- None of them seem intuitive
- 刘士辉 (Liu Shihui) extenuates 122 / 173 censored keyword combinations
- 8 extenuating combinations only extenuate one censored keyword combination
Some seem a bit spammy

- 淘宝兼职 (Taobao part-time)
- 官网注册 + 棋牌游戏 (official website registration + board games)
- 绩 + 成 (achievement + success)
- URL shorteners (e.g., bit.ly/, baid.us/, t.cn/)
Others seem like they should be censored themselves

- 刘士辉 (Liu Shihui, a human rights defender) extenuates 不会颁发和平奖的奖章 (no Peace Prize Medal will be awarded)
- 一个光明的民主中国 (bright democratic China) extenuates 兵变 + 北京出事 (mutiny + incident in Beijing)
- 法輪功 (Falun Gong, in traditional Chinese) extenuates 法轮大法 (Falun Dafa, in simplified Chinese)
Hypotheses

● Deliberate behavior
  ○ Balance requirements of mainland Chinese censorship with expectations of overseas users
  ○ Allow some criticism of government

● Misconfiguration
  ○ E.g., extenuating are a different category of keyword, such as spam
  ○ Once a keyword is discovered, all other keyword checks are short circuited
  ○ Groups such as Voice of America often spam Chinese email accounts
Extenuating combinations ruin censorship measurements

Assume that both “a + b” and “x” are censored keyword combinations.

Assume that keyword combination “Z” extenuates “x”.

スター: censored (by “a + b”)
スター: no longer censored
スター+b: still censored (by “a + b” AND “x”)
スター+b: still censored (by “x”)
スター+b: no longer censored
スター+b
Significance of understanding extenuating combinations

- Discovering them gives us a more complete picture of censorship
- Knowledge of them presents evasion opportunities
- If you don’t consider their existence, then your censorship measurements can be subtly wrong
- Their intended purpose in this case is perplexing
For a deep dive into our data:

https://github.com/citizenlab/chat-censorship/qqmail
Acknowledgements

We thank the anonymous reviewers for their helpful feedback. This project was supported by Open Society Foundations. This project was supervised by Ron Deibert, Citizen Lab Director.
the end