Journal of Internet Services and Applications

RESEARCH Open Access

A user-driven free speech application for anonymous and verified online, public group discourse

Michael Nekrasov[1]*, Danny Iland[1], Miriam Metzger[2], Lisa Parks[3] and Elizabeth Belding[1]

Abstract

Online social networks are major hubs of communications that are used to share information by societies worldwide. However, the ability to freely communicate on these platforms is increasingly restricted in countries across the globe and existing technological solutions do not fully address the needs of affected communities. In this paper we explore the design process of SecurePost, a novel tool that allows verified group anonymity to those communicating publicly on online social networks. We present survey-based research and ethnographic interviews of communities vulnerable to censorship conducted in Zambia, Turkey, and Mongolia between 2013 to 2016. We use analysis of this data to ground our work. We explore needs and requirements of users such as modes of censorship, resistance to network disruption, and appropriate platform consideration. We outline our technological solution and expand on how on-the-ground research of user communities guides technological requirements.

Keywords: Social media, Internet freedoms, Free speech, Censorship, Anonymity, ICTD

1 Introduction

The Internet is a critical tool for communication across the globe. Internet users share ideas, read news, and engage in dialogue. Online social networks (OSNs), such as Facebook and Twitter, have in particular become major hubs of communication. In June of 2017, Facebook surpassed 2 billion active users [1]. Individuals, companies and bots generate a massive amount of content. Every second, users worldwide post an average of 6000 tweets on Twitter [2] and "like" an average of 9200 items a second on Facebook [1].

Recognizing the importance of online discourse, the United Nations considers that the protection of human rights, especially free speech, should fully extend to the Internet [3, 4]. Despite this, governments, corporations, and individuals often restrict what users can say online and punish those with dissident views. Around the world, Internet freedoms are restricted [5–10]. Even liberal democracies, typical advocates of free speech, increasingly restrict content [11–14]. Globally, suppression of free speech and press freedoms is on the rise [15], accompanying a rise in authoritarianism [16, 17], which threatens the foundations of functioning democracies.

Users often turn to technological solutions to combat this threat to civil liberties. For example, popular anonymity tools such as Tor [18] provide network level anonymity, while person to person messaging tools such as Signal [19] and WhatsApp [20] allow private communication using end-end encryption. Nevertheless, safe public communication remains a challenge. Individuals conversing publicly on OSNs open themselves up to legal and physical dangers, encouraging self censorship and stifling discourse.

Reputation and trust are likewise eroded. In an era of "fake news", users struggle to identify what OSN accounts and posted content can be trusted [21, 22]. Adversaries to open discourse deploy armies of operatives masquerading as legitimate users to sow division [23, 24]. Even trusted news outlets using OSNs can be hacked to spread misinformation [25–27].

Our work aims to understand some of the core issues around freedom of speech online for communities that are particularly vulnerable to censorship, such as journalists,

*Correspondence: mnekrasov@cs.ucsb.edu
[1] Department of Computer Science, UC Santa Barbara, Santa Barbara 93106, CA, USA
Full list of author information is available at the end of the article

political activists and minorities, and develop technological solutions. Over four years we visited communities in Zambia, Turkey, and Mongolia to investigate issues of free speech on the Internet and in particular OSNs. We identified common actors and methods of censorship as well as some technological needs unmet by existing tools. We present an early analysis of this work in [28].

We found that public group discourse while maintaining anonymity and preserving reputation on OSNs was one such unmet need. We used this knowledge to design SecurePost, a novel software tool that provides verified group anonymity on online social networks, in collaboration with these local communities. We first introduced this tool in [29].

This paper is a culmination of our efforts. In this work we describe how we used survey research, combined with ethnographic interviews, to identify common unmet needs in three different communities. We then show how we fuse social and technological research to design a novel tool that satisfies these needs. While our work focuses on specific communities, not populations of countries as a whole, we believe examining the needs of communities particularly vulnerable to censorship provides a lens through which to understand some challenges to overcoming censorship more broadly. The ethics and permissibility of censorship is outside the scope of this work.

We begin this paper by presenting the methodology behind both the social and technical aspects of our work in Section 2. In Section 3, we present the results of our survey analysis that serves as a basis for understanding people's uses of social media and their censorship concerns. We next explore some of these issues more closely, concentrating on findings from interviews and journalistic accounts in Section 4. Based on the social analysis we present our software solution, SecurePost, in Section 6. We then evaluate SecurePost using anonymous in-app surveys and usage statistics in Section 7. Finally, we summarize and draw implications from our work in Section 8.

2 Methodology

To investigate the role of the Internet and OSNs in free speech, we sought out communities vulnerable to censorship. Between 2013 and 2016, we visited three regions: L usaka, Zambia; Istanbul, Turkey; and Ulaanbaatar, Mongolia. These three regions have diverse geopolitical contexts, different levels of socio-economic development and dissimilar cultural and historical backgrounds. In each region we focused on communities particularly vulnerable to censorship. We sought to understand the needs and challenges of these specific communities as a cross-section of global issues surrounding Internet censorship, particularly focusing on free speech on OSNs.

By combining social science and technological research, we developed a software tool that fulfills previously unmet needs to aid at-risk communities. In our work we involved participants in the iterative development of a technological solution that is well-suited to their use-case. Given the diversity of the three selected regions, we believe our solution would be applicable to other communities with similar needs, more broadly bolstering freedom of speech online.

Because there is extensive documentation of censorship in China [30–34] and the Middle East [35–37], we chose to focus on countries less studied at the time of data collection: Zambia, Turkey, and Mongolia. This enabled us to examine a broader scope of communities in terms of freedom of speech and censorship. Our goal is to use experiences with these three target communities as a lens to understand global issues surrounding censorship on OSNs and the Internet overall, and to facilitate the development of tools for protecting free speech.

When we first planned our work, Turkey was beginning to increase its censorship efforts, enabling our team to observe responses to increasingly visible and common censorship practices. Like Turkey, Zambia also experienced an increase in government censorship during the course of our study. By contrast, censorship in Mongolia was relatively static and less widespread [38]. Our team identified collaborators in these countries through previously existing contacts.

Our interdisciplinary team consisted of experts from computer science, communication, and media studies, from departments spanning across physical science, social science, and the humanities. Over the course of our project, we conducted 109 interviews and surveyed 526 individuals. We used a combination of ethnographic analysis of the interviews and descriptive statistics on the survey data to understand Internet access patterns, identify barriers to free speech, and assess shortcomings of existing tools that assist groups and individuals in communicating safely. We obtained IRB approval prior to conducting our fieldwork.

We used snowball sampling to recruit respondents for the surveys and in-depth interviews. We calculated descriptive statistics (frequencies) for all variables of interest, including socio-demographics. We compared participants in three countries using chi-square tests for categorical variables and Fisher's exact test for dichotomous variables.

2.1 Survey

The survey based research aimed to better understand use of information and communication technology in our three target communities and to gauge the opinions of members of our sample on issues of Internet freedom, censorship, and media trust. When interpreting the data, it is important to note that convenience sampling was used for the respondents, so these data may not accurately represent either the population of the respective countries as a whole or residents of the surveyed cities.

The demographics of survey respondents are provided in detail in Table 1. The gender of all survey respondents is roughly equally split between male and female, with slightly more female respondents overall. Across the three countries, we note the high prevalence of responses from individuals with university education and those between the ages of 20–39. This is likely due to our affiliations with universities and organizations, such as LGBTQ centers, with university aged staff. We note that because our focus was on studying communities vulnerable to censorship, the demographics and responses are not necessarily representative of the full populations of these countries.

2.1.1 Zambia

The Zambian survey was distributed in the capital city of Lusaka, using the Open Data Kit survey software on Nexus tablets between December 7 and 18, 2013. Open Data Kit allowed us to securely store encrypted versions of the survey response data without the need for Internet access. The total number of completed surveys was 106.

The convenience sample consisted primarily of individuals who work in media-related fields (e.g., radio stations, newspapers, news websites, blogs, and technology). The research team also recruited respondents from a media institute and a computer lab and technology hub where people can learn computer skills (e.g., game design and coding). The surveys were also distributed at ConnectForum, a technology conference for women hosted jointly by the government and local companies. The women attending the conference were individuals who either worked at technology companies or aspired to work in the field of Internet communication technologies.

Participants in our sample were highly educated, with 98% of participants completing secondary school or higher, compared to the overall Zambian population where about 26% of females and 44% of males have only completed secondary school [39]. The sample was of a higher income level than average for Zambia and contained fewer unemployed. Students, journalists, and people working in the banking industry and in IT were also overrepresented. However, the sample did reflect the population in terms of age, religion, and ethnicity with the exception that it contained more Europeans. This is likely due to conducting the study in Lusaka, which has a higher number of expatriate workers [40] than in other areas of the country.

2.1.2 Turkey

The Turkish survey was distributed in Istanbul using both a paper survey and an Internet-based interface between December 7 and 19, 2014. Using this combination of in-person (paper and pencil) and mediated (online) administration helped researchers reach the largest possible sample with the resources available to the survey team for this part of the project. The total number of completed surveys was 166.

Table 1 Demographics of survey respondents from the three sampled countries

	Zambia		Turkey		Mongolia		All	
	%	(#)	%	(#)	%	(#)	%	(#)
Gender								
Male	50%	(52)	52%	(81)	40%	(100)	46%	(233)
Female	50%	(51)	48%	(74)	60%	(151)	54%	(276)
Age								
Under 20	21%	(22)	7%	(10)	13%	(27)	13%	(59)
20–39	71%	(74)	78%	(120)	74%	(162)	75%	(356)
40 or more	8%	(8)	15%	(23)	13%	(27)	12%	(58)
Education								
Less than primary school	1%	(1)	0%	(0)	1%	(2)	1%	(3)
Primary School	1%	(1)	1%	(2)	2%	(5)	2%	(8)
Secondary School	37%	(39)	31%	(48)	31%	(77)	32%	(164)
Higher Education / University	61%	(64)	68%	(106)	66%	(163)	65%	(333)
Total # surveyed	(n=106)		(n=166)		(n=254)		(n=526)	
Dates surveyed	Dec 7–18, 2013		Dec 7–19, 2014		Jun 16- Jul 4, 2015			

The sample consisted primarily of young people, including students and people working in media-related fields. The gender of survey respondents was relatively balanced. However, many survey respondents were of a lower income level and significantly higher education level than the general population in Turkey [41]. This is likely because of the number of undergraduate and graduate university students in the sample. Furthermore, the majority of respondents indicated they were not affiliated with the most powerful political party in Istanbul at the time the survey was administrated (AKP), suggesting that the sample may be more representative of political minority opinions on the topic of Internet freedom than the average Turkish population. The research team relied on academic contacts at a University in Istanbul to recruit respondents.

2.1.3 Mongolia

The Mongolian survey was distributed in Ulaanbaatar, Mongolia's capital, using paper surveys in the Mongolian language between June 16 and July 4, 2015. Surveys were administered primarily by Mongolian undergraduate students who attend Mongolia National University (MNU), a project partner. These students were supervised by MNU professors. The total number of completed surveys was 254.

The sample consisted primarily of young people, including students and others who were relatively highly educated. In 2011 64.2% of the population of Ulaanbaatar was under 35 years old, suggesting that the number of younger respondents to the survey is somewhat representative of demographic trends in Ulaanbaatar [42]. The sample included slightly more female than male respondents. The survey was administered by Mongolian undergraduate students, who had greater access to individuals of similar characteristics, such as education level.

2.2 Interviews

Ethnographic interviews provided a qualitative dimension to our research on the underlying issues associated with Internet and OSN free speech and censorship experiences. We engaged in both in-depth interviews and informal conversations with a wide range of individuals and organizations, including journalists, political activists, ethnic minorities, law makers, educators, LGBTQ center employees, gender-based violence center employees, government watchdogs, and others affected by censorship.

We reviewed contextual data such as political histories and news media reports about online censorship concerns in each country and identified key concerns and individuals. We then consulted with local partners, who were affiliated with universities and non-governmental organizations, to develop lists of potential informants. Working with our partners, we reached out to and scheduled interviews with many of these informants. We also used the snowball method to expand our informant lists while in each country.

We conducted more than 35 interviews in each country. Sometimes these interviews were with individuals and sometimes with small focus groups. In total we interviewed more than 150 people across the three countries. Theoretical saturation was reached when we encountered informants identifying the same or similar censorship agents, concerns, sites, and sources. We conducted some interviews in English and worked with translators to conduct others. Since many of our informants have been on the front lines of free speech struggles, we anonymized their identities and securely stored all interview data.

We used grounded theory [43] to analyze our interview data and extrapolated informants' censorship concerns and tactical responses to them based on our close analysis of the transcribed interview data.

2.3 Application development

Based on qualitative and quantitative data, we identified previously unmet challenges to the use of online social networks for these target communities. We then built a software solution, called SecurePost [44], which addresses these challenges by enabling new ways for balancing anonymity and trust in OSNs.

We developed software in parallel to our field visits, and conducted user testing of prototypes while in the field. During each visit partners were able to try our most recent prototypes and provide feedback. In this manner we received iterative feedback from our partners that helped us refine our application and further understand the requirements of these communities. We evaluated our software using a pair of optional anonymous surveys administered through the SecurePost application.

3 Survey and interview results

From the survey data we evaluated how vulnerable populations use and access the Internet, and in particular OSNs. We present the survey results and provide discussion supported by the ethnographic interviews. Note that an anonymity agreement in our IRB precludes us from making direct quotes of the interviews. For our analysis, we examined popular OSN platforms and the types of activities users engage in when using social media. We then looked at the difficulties that people experience when accessing the Internet and OSNs, including access disruptions and censorship. Finally, we compared how free users felt when using the Internet and the types of behaviors they engaged in when faced with censorship.

3.1 Internet and online social network usage

As the basis of planning a technical solution, we examined usage modality and preferred platforms. We asked respondents about their Internet and OSN usage, including frequency of usage and their preference of social network. We summarize the key findings in Table 2.

Across all three countries, the majority of respondents stated they use the Internet every day or more (with Mongolia to a lesser extent than the other countries). For all respondents, 94% stated they use the Internet at least once a week, indicating a high utilization of the Internet in these communities.

Further, the majority of users stated they use OSNs every day or more. Facebook, Twitter, and Google+ were used by the largest number of respondents (it is unclear how many people differentiated Google+ from Google search and other Google services; interviewees did not report significant usage of, or interest in, the Google+ OSN). Other OSNs included YouTube and WhatsApp, and to a lesser frequency Viber, Snapchat, Vimeo, WeChat, Tumbler, Pintrest, and Vine.

The number of users stating they use OSNs daily slightly out-paces the number of Internet users. When looking at individual responses, the same respondents claim higher OSN usage than total Internet usage. This is likely because some users perceive mobile applications and OSNs as something separate from web browsing as a whole.

The results from these communities fit into global trends for Internet and OSN usage. As of 2017, the ITU estimated 3.6 billion people, roughly 47% of the world's population, use the Internet [45]. Young people are at the forefront of adoption with 70% of people between 15 and 24 years old online [46]. Globally, online social networks are some of the most visited websites [47]. As of October 2017, Facebook had over 2 billion active users (1.3 billion daily), while Twitter had over 328 million active users [48]. These statistics highlight the importance of OSNs as communication platforms when confronting censorship.

We asked respondents what device they used to access the Internet. The most used devices were smart phones and then laptops. Globally, as of 2017, it is estimated that 58% of the world's population had a mobile-broadband

Table 2 Internet and online social network usage

	Zambia		Turkey		Mongolia		All	
	%	(#)	%	(#)	%	(#)	%	(#)
Use internet								
Less than once a week (or never)	2%	(2)	1%	(2)	11%	(27)	6%	(31)
Once a week	4%	(4)	1%	(1)	8%	(19)	5%	(24)
2–3 times a week	13%	(14)	5%	(9)	31%	(77)	19%	(100)
Every day or more	81%	(85)	93%	(154)	50%	(125)	70%	(364)
All devices used to access internet (can select multiple)								
Desktop	26%	(27)	48%	(79)	49%	(121)	44%	(227)
Laptop	80%	(85)	80%	(132)	53%	(131)	67%	(348)
E-reader/Tablet	17%	(18)	31%	(52)	14%	(35)	20%	(105)
Smart phone	74%	(78)	84%	(140)	72%	(177)	76%	(395)
Basic/Feature phone	13%	(14)	1%	(1)	14%	(34)	9%	(49)
Social network use								
Never	4%	(4)	0%	(0)	0%	(0)	1%	(4)
Less than once a week	1%	(1)	1%	(2)	5%	(11)	3%	(14)
Once a week	4%	(4)	1%	(1)	7%	(17)	4%	(22)
2–3 times a week	9%	(9)	8%	(13)	29%	(70)	18%	(92)
Every day or more	83%	(86)	90%	(150)	60%	(145)	74%	(381)
All social networks used (can select multiple)								
Facebook	91%	(96)	87%	(145)	97%	(237)	93%	(478)
Twitter	52%	(55)	76%	(126)	34%	(82)	51%	(263)
Google+	52%	(55)	57%	(95)	58%	(143)	57%	(293)
Instagram	3%	(3)	72%	(119)	35%	(86)	40%	(208)
LinkedIn	34%	(36)	31%	(51)	7%	(16)	20%	(103)

subscription, while only 13% had fixed broadband subscriptions [46]. Further, the annual growth rate of global mobile broadband subscriptions (20%) is out-pacing that of fixed broadband subscriptions (9%) [46]. These statistics highlight the current and future needs for mobile-device-based solutions in both the cases of our target countries, and in the larger global context.

In order to identify a suitable operating system for development, we asked responders about their device model and operating system. In 2013, survey responses from Zambia indicated that Blackberry was the most popular type of phone (n=33, 34%), followed by Android (n=24, 25%). Nokia phones were next in popularity (n=20, 21%), followed by the iPhone (n=10, 10%). Since then, Android has become dominant in Zambia, accounting for close to 50% of the market share [49]. This is in line with anecdotal evidence from respondents during our return visit in 2016. Responses from the 2015 Mongolia survey indicated that Android was dominant (n=140, 59.8%), followed by iOS (n=76, 28.6%). Relatively few people had other operating systems, including Blackberry (n=10, 4.3%).

We did not collect data on device model and operating systems for participants in Turkey. According to Stat Counter, a website that tracks browser and mobile OS usage worldwide, as of September 2017 Android accounted for 80% of the mobile operating market share in Turkey, followed by iOS with 18%, while all other operating systems accounted for less then 2% of the market share [50].

These trends are reflected globally. As of end of May 2017, Android made up 86% of the global market share [51]. Android's popularity, especially in the developing world, may be due to the availability of cheap phones, from a variety of manufacturers running the Android OS, saturating markets in Africa and the Middle East [52]. Many of these phones run older versions of the Android OS, have limited storage and computational power, and do not receive software upgrades. Nonetheless, when developing tools that work in the developing world, Android applications allow access to a large population.

3.2 Activities on online social networks

Respondents engage in a wide variety of activities in online social networks. In particular, respondents most commonly consumed content about their hobbies and entertainment, things happening in their community and the world, as well as information about their friends and family. Respondents most commonly produce content such as photos and information about hobbies and entertainment. A full breakdown is shown in Fig. 1.

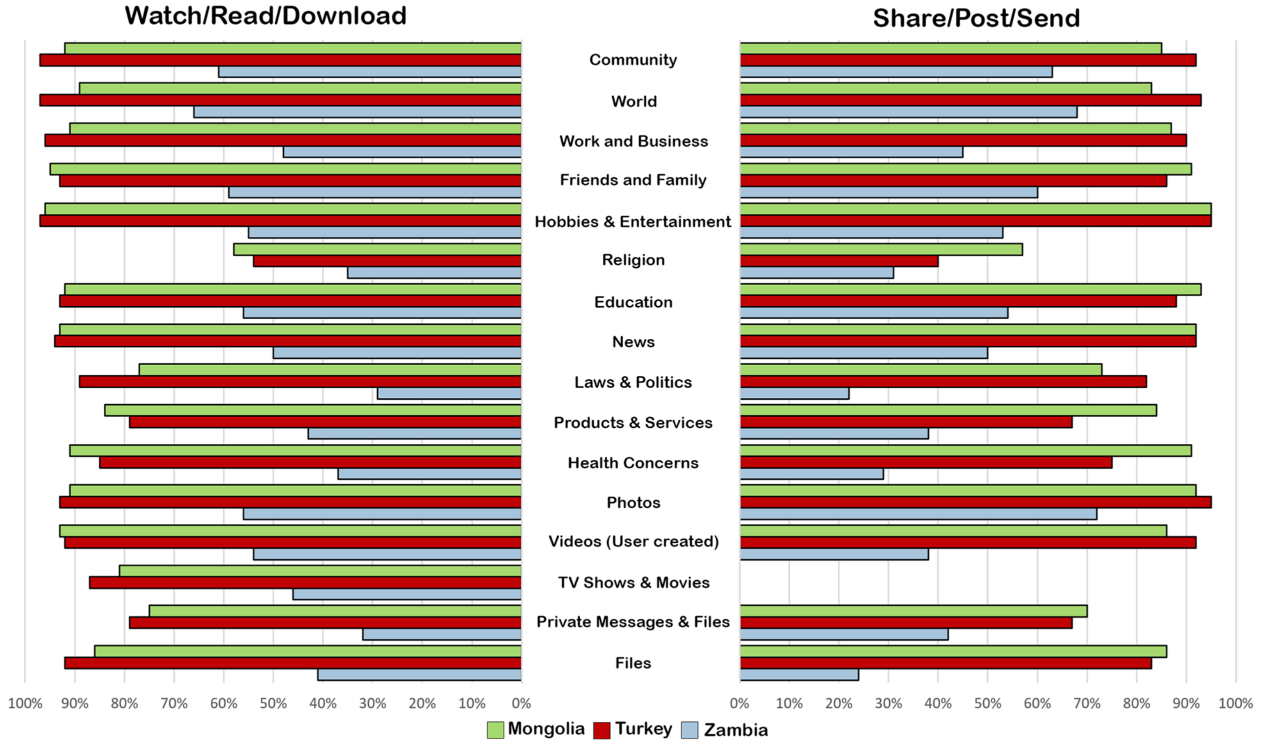


Fig. 1 Types of activities users engaged in on OSNs. Respondents could pick any number of activities. The left part of the chart represents activities that consumed content and on the right activities that generated content

In Turkey and Mongolia, we also asked about the frequency of information consumption and generation. Respondents in Turkey indicated that they most frequently share, post, or send information about things happening in the community and around the world, as well as hobbies, news, photos. Content was less frequently shared or consumed on topics such as religion, products, and health.

In Mongolia, respondents indicated that they most frequently share and post information about hobbies or entertainment, as well as friends and family, education, photos, news, health, and work. Content was less frequently shared or consumed on topics such as religion, laws and politics, and things happening around the world.

Some activities in which users engage, such as posting about religion, health, and politics, have some relevant topics of discourse that could endanger the user. We explore this in Section 4.

3.3 Internet disruption

In the survey, we examined the frequency and perceived reason for disruptions that respondents experienced in attempting to use the Internet and OSNs. A summary of our results is presented in Table 3. Across the three countries, 78% of respondents reported at least sometimes experiencing disruption to their activity, most commonly in Zambia (96%).

In Zambia and Mongolia, power outage was reported as a major reason for disruptions. This was less of a problem in Turkey, which has better infrastructure. In all three countries, particularly in Zambia, poor network reliability was a major reason for access disruptions. Government or other censorship which was not frequently cited as a common cause of disruptions in Zambia (7%), was reported more common in Mongolia (19%), and was cited as a major cause of disruptions in Turkey (44%).

Such temporary disruptions emphasize the need for tools that are tolerant to networking delays when fetching and delivering content. Permanent disruptions, like those imposed by government censorship, require circumvention tools to bypass those blocks.

3.4 Censorship

In addition to investigating how and for what purposes respondents use OSNs, we investigated how their use is impacted by censorship. We summarize these results in Table 4.

Across all three countries, only 11% described feeling “very free” to express themselves on the Internet and OSNs. The majority (53%) felt only “a little free” or “not free at all”. In Turkey especially, 37% of responders said they did not at all feel free to express themselves on OSNs.

When asked how often users modify their behavior to protect against government or other monitors seeing things that they post in OSNs, respondents in Mongolia and Zambia demonstrated no statistically significant differences from each other. In those two countries, the majority of users (64% in Zambia, 56% in Mongolia) reported at least sometimes modifying behavior in OSNs (p=0.225). On the other hand, in Turkey, 88% of users reported never or almost never modifying behavior, a statistically significant difference when examining all three countries (p=0.000). This is interesting because out of the three countries, Turkey has a higher incidence of government censorship [38], and - as observed earlier - more users reported feeling not free.

Table 3 Disruption of internet and OSN usage

	Zambia		Turkey		Mongolia		All	
	%	(#)	%	(#)	%	(#)	%	(#)
Frequency of internet disruption								
Almost never	4%	(4)	25%	(41)	27%	(68)	22%	(113)
Sometimes	42%	(44)	48%	(79)	61%	(152)	53%	(275)
Often	25%	(26)	19%	(31)	10%	(25)	16%	(82)
Very often	29%	(30)	8%	(13)	2%	(5)	9%	(48)
Reason for internet disruption (can select multiple)								
Power outage or electrical problems	48%	(51)	15%	(24)	51%	(128)	39%	(203)
Unpaid fees for Internet service	29%	(31)	7%	(11)	32%	(80)	23%	(122)
Poor network connection or service	91%	(96)	68%	(112)	44%	(110)	61%	(318)
Government or other censorship	7%	(7)	44%	(73)	19%	(47)	24%	(127)
Unknown reason	3%	(3)	15%	(25)	10%	(25)	10%	(53)
Other reason	5%	(5)	5%	(8)	8%	(20)	6%	(33)

Table 4 User behavior and perceived freedom on OSNs

	Zambia		Turkey		Mongolia		All	
	%	(#)	%	(#)	%	(#)	%	(#)
How free users feel on internet & OSNs								
Not free at all	32%	(34)	37%	(60)	19%	(47)	27%	(141)
A little free	19%	(20)	21%	(34)	31%	(78)	26%	(132)
Somewhat Free	40%	(42)	30%	(48)	38%	(96)	36%	(186)
Very Free	9%	(9)	12%	(20)	12%	(29)	11%	(58)
How often users modified behavior								
Never	19%	(18)	71%	(114)	12%	(29)	32%	(161)
Almost never	18%	(17)	17%	(28)	32%	(79)	25%	(124)
Sometimes	47%	(46)	9%	(15)	34%	(83)	29%	(144)
Often	11%	(11)	1%	(2)	11%	(26)	8%	(39)
Very often	5%	(5)	1%	(2)	12%	(30)	7%	(37)
How users modified behavior (can select multiple)								
Limit/censor what to post	77%	(82)	59%	(98)	67%	(164)	66%	(344)
Use a secure connection	25%	(26)	20%	(33)	24%	(59)	23%	(118)
Do not use a real name	4%	(4)	8%	(14)	10%	(24)	8%	(42)
Avoid internet or social network	3%	(3)	8%	(14)	9%	(22)	8%	(39)

When asked about methods users took in modifying behavior, Mongolia, Zambia, and Turkey had no statistically significant differences for most strategies. About a quarter used secure connections (p=0.554), a few hid their names (8%, p=0.166), and some avoided using the Internet or OSNs entirely (8%, p=0.119). For self-censorship, however, respondents in Turkey again differed in behavior. While the majority of users in Zambia (77%) and Mongolia (67%) self-censored what they posted online (p=0.893), users in Turkey (59%) reported significantly less self-censorship (p=0.0076).

A possible explanation for these behaviors comes from our interviews. In Turkey, many of the journalists and activists we interviewed conveyed a sense of obstinance to censorship efforts. Those we interviewed emphasized the importance of exercising free speech and were willing to continue to do so, even after arrest. And, as noted earlier, a proportionately larger number of our survey respondents in Turkey were members of political opposition groups.

4 Categorizing barriers to free speech

As we observed in the previous section, while most respondents do not feel free to express themselves on OSNs, very few (8% across all three countries) reported posting anonymously. Despite this, in the ethnographic interviews, respondents repeatedly brought up issues concerning anonymity, as well as the ways anonymity impacts trust. To understand how a tool could address these concerns, we focused on four central elements: (1) The negative consequences of the use of real-world identity. (2) Difficulties individuals experienced in retaining anonymity. (3) The impact of anonymity on reputation. (4) Problems with retaining reputation and trust in spite of active censorship. (5) Additional methods adversaries use to limit free speech. We use this discussion as the basis of our software based solution.

4.1 Using real-world identity on social media

Authoritarian governments pass strict laws curtailing free speech. In Turkey, for example, following a 2005 restructure of the Turkish penal code, Turkey passed Article 301, which prohibits the “denigration of the Turkish nation”, and Article 216, which bans “inflaming hatred and hostility among peoples” [53]. These laws have been used to target journalists, artists, and unaffiliated individuals for criticizing government, policy, and religion in any medium [54, 55]. Zambia likewise recently saw multiple arrests including an opposition leader [56] as well as an engineering student critical of the president [57], on the grounds of defaming Facebook posts. Globally, 27% of all Internet users live in countries where individuals have been arrested for posting, sharing, or liking a post on social media [58].

Libel laws are also a weapon that companies, organizations, and wealthy individuals can use to curtail freedom. These adversaries sue for defamation or insult - wrongfully in many cases - in response to stories on OSNs and other media [59]. Expensive legal fees and threat of financial ruin silences those who do not have the monetary resources to defend themselves. This creates a chilling effect on free speech [60].

This is especially a problem in Mongolia where libel laws are frequently used as a way to silence the press [61]. In Mongolia, the burden of proof for libel rests with the defendant and libel constitutes a criminal charge. Journalists reporting on topics such as corporate corruption must prove their reporting is true and accurate and evidence not containing original copies and notarized documents may be thrown out as inadmissible [62]. During the Mongolian interviews, self-censorship due to threat of a libel lawsuit was a frequent topic of concern. In 2015, after our visit, multiple individuals were arrested in separate cases over posts on Twitter [63].

Revealing identity and personal details can make an individual a target by members of their physical or online community. Expressing views or interests that go against societal norms can impact job availability and interpersonal relations [64].

From our interviews, we heard how journalists reporting on topics opposing dominant political identities, such as on Kurdish issues in Turkey, face constant barrage of hateful posts. In Zambia, we heard how elements of a user's identity, such as ethnicity, gender, and past posting record, stereotypes the user to the point of overshadowing the discussion of substantive content.

Threats may extend into an individual's home. When interviewing members of a gender-based violence prevention center in Mongolia, we heard stories of threats coming from a person's own family. Family members of some of these women would monitor posts and private messages on social media and punish perceived affronts with physical violence.

4.2 Balancing reputation and anonymity on OSNs

With so many threats and such serious consequences, many users self-censor their online posts. Yet, as we observed in Section 3, some users still remain adamant about speaking their minds, putting them in potential jeopardy. While some post anonymously or use fake aliases, relatively few reported doing so in our surveys. In part, this may be due to the difficulty of being anonymous online. Major social networks, like Facebook, impose a real name policy as part of their terms of service [65, 66]. Other sites, like Twitter, may require identifying information, such as a phone number to create or verify an account.

Even if information is not provided by the user, OSNs still log the IP address of the requests. This information can be subpoenaed by governments [67, 68]. Requiring all users to use anonymity services like VPNs or Tor [18] for each post is unrealistic, since groups can be composed of posters with varying technological expertise, and a single mistake can be costly.

If adversaries lack the ability to identify users based on IP, they can still de-anonymize users based on the content they post. Users of social media regularly post identifying information. An account posting a personal photo can identify an individual. Other information posted by a user can inadvertently allow adversaries to guess identities. For example, revealing details about education, past residences, and events might be enough to uniquely identify an individual. A dedicated attacker could use information exposed by the account over a period of time to establish identity, exposing the user to threat. Adversaries will *dox* users, publishing identifying information, which may escalate digital conflict to physical confrontation [69, 70]. Even if meticulous discipline is followed, self-censoring all identifying information limits an individual's ability to communicate about their personal experience.

4.3 Impact of anonymity on reputation

Hiding identity likewise has an impact on trust of the posted content. Readers of social media use the reputation of an account to gauge trustworthiness of the content. The reputation of and past experience with an organization, such as a newspaper, leads readers to believe in the content of that account and differentiate it from fake news promulgated by bad actors. The reputation of a source and the identity and history of the author are the first things fact checkers and librarians recommend that readers check when evaluating if a story is fake [71, 72].

In the case of the Zambian Watchdog (ZWD) online news service, we interviewed journalists who said they maintained anonymous blogs to avoid government intervention when discussing sensitive topics. Others we talked to in Zambia noted that the lack of real-world identities by ZWD journalists and lack of a physical address for the organization weakened accountability and verification of content. Respondents said this led ZWD to lose credibility among some of its readers [73].

Increasingly, hostile governments and other adversaries use fake accounts to orchestrate attacks. Governments, like Russia, hire paid *trolls* to carry out legitimate sounding discourse online [23, 24, 74]. They automate attacks using *bots* to generate a massive number of messages that flood OSNs [75, 76], and set up anonymous *sybils*, which are fake accounts posing as legitimate users [77], as vessels for these agents. These accounts post spam [78] and fake-news [79] in coordinated attacks to steer conversation and drown out competing ideas. Users attempting genuine discourse, especially those using anonymous accounts that are hard to distinguish from the attackers, can get lost in the noise. Our interview pool included victims of these false reporting attacks, who reported their accounts being banned due to third party reports.

Journalists and public figures whose careers are tied to reputation may find it difficult to utilize the protections offered by anonymity in light of the need to build and maintain a reputation that garners trust with readers. For example, a lawyer and LGBTQ activist who was politically active against Turkey's president Recep Tayyip Erdoğan was arrested, convicted, and fined over a tweet [80] criticizing the president. Subsequently, he was arrested again for unrelated charges, including membership to HDPIstanbul, a WhatsApp group belonging to the Peoples' Democratic Party [81, 82]. Despite harassment by the Turkish government, he continues to actively use social media and maintains his real name and identity online.

From our interviews, after facing legal jeopardy, some reported adopting anonymity in order to maintain jobs unrelated to their online activities. However, several journalists and activists stated that, despite past and future threats, it is important to publicly stand up using their real names and identities as an act of opposition, personal pride, as well as to garner trust.

4.4 Maintaining reputation after censorship

An additional dimension to online identity and the decision to practice anonymity is the ability for governments and individuals to censor content on OSNs. Governments may block entire websites [5] or ask OSNs to censor a particular account or post. For example, Twitter maintains an entire system for withholding posts that are censored in a particular country [83].

Aside from legal requests, governments and groups exploit weaknesses in mechanisms built to protect users. Most OSNs have a reporting functionality that users can employ to report cyber-bullying and flag content that may not be appropriate, such as pornography, for a particular communication channel, such as a university's homepage on Facebook. Free speech adversaries exploit these reporting tools to falsely flag dissenting opinions. Governments and other adversaries deploy bots and trolls to report posts. This usually leads to an automatic account ban, and may take a long time for OSNs to rectify. Russia, for example, deploys trolls to censor popular accounts by reporting content as threatening violence or containing pornographic material [84].

In our interviews, people reported similar instances, where their account would be banned after crowds of users repeatedly mis-flagged content. This led to repeatedly creating new accounts. When legitimate OSN accounts (anonymous or not) are compromised and users switch to new accounts, the trust of the readers may be imperiled. New accounts have to demonstrate continuity with previous ownership, and re-establish readership. This can be even more challenging for anonymous accounts that can't rely on real-world identify to demonstrate continuity. Adversaries can take this opportunity to impersonate accounts to confuse readers and further degrade trust.

4.5 Geographical censorship

In some instances, instead of targeting web sites or individuals, governments or organizations censor entire geographical areas by restricting all Internet access. In some areas access is restricted permanently, such as in the Za'atari refugee camp, where access was not provided in order to discourage refugees from encroaching on the local labor market [85]. Other times access can be cut off in response to events such as protests, which is frequently the case in Turkey - especially in Kurdish regions [6, 86–88].

From our interviews, citizen journalists in Kurdish areas of Turkey reported encountering these types of tactics [89]. When reporting in areas where Internet is severed, they record content and store it on their device for later publication. However they expressed concerns that sometimes their device will be seized and searched. Similar reports emerged during our interviews of journalists in Zambia as well [73]. Much like the case of access disruption due to electricity, this scenario is a notable consideration for the design of technology to protect freedom of speech.

5 Outlining requirements for a new tool

The quantitative survey results complemented with the qualitative interviews, global statistics and journalistic reports, provide an outline to understand and address some of the threats and needs of users when protecting freedom of speech online.

OSNs Highly Utilized: From Section 3 we saw that the Internet, and in particular OSNs, are heavily utilized by our partner communities. OSNs are used daily for posting and consuming content on a variety of activities. However, the majority of users report feeling only "a little free" or "not free at all" when using them and, consequently, users modify their behavior. This results in some self-censorship of content, limiting what they discuss on these platforms to reduce the threat of legal and physical harm from governments, corporations, and sometimes even family members.

Smart Phones Are a Primary Mode of Access: In these communities, smart phones are the dominant way for individuals to access the Internet and OSNs, with this trend continuing to increase globally. When designing tools for this type of community, support for smart phones running Android (the most used OS type) gives widest user coverage. For our own technological solution, in the context of limited developer resources, we chose Android phones as the primary platform for developing and deploying our tool. Given the popularity of Facebook and Twitter (which shares similar text based posting behaviors with Facebook), and based on feedback from our local partners, we elected to support these two platforms for our work.

Backwards Compatibility: Given the wide abundance of older devices - especially in Zambia and Mongolia, we sought to develop a tool that was backwards compatible with older operating systems. This puts restrictions on the types of both native and external APIs that developers can utilize during development. Often developers target newer versions due to limitations of older software and hardware. For example, while some of our respondents used devices running Android 2.x, Signal [19] limits it support to Android 4.0 and above.

Network Disruption: Respondents reported disruptions to power, communication, and instances of government restrictions on Internet access to geographic areas. Tools catering to people who may live in areas where network and power are unreliable have to account these limitations. Applications should limit power and data consumption and provide a user experience that does not rely on continuous connectivity. In our tool we implemented local caching of content that was tolerant of network delay.

Confiscation and Search: Due to temporary or permanent lack of Internet connectivity, as we discuss in Section 4.5, journalists sometimes ferry information back to Internet connected sites. During this time devices could be seized, which suggests a need for local data encryption.

Anonymity Protection: As discussed in Section 4, we observed that when groups use social media to spread news and ideas, there is a tension between anonymity and trust. While other tools, such as WhatsApp[20] and Signal [19], address aspects of individual security and *private* group communications, there are still unmet needs in *public* group communication in the presence of censorship. Users that reveal real-world identities open themselves to both legal and physical threats. However, users face many technical challenges when staying anonymous on OSNs.

Reputation Preservation: Users also find it difficult to retain reputations and trust, especially when posting anonymously. Individual anonymous users may find it difficult to build trust, competing with armies of *bots*, *trolls* and *sybils*. Groups find it difficult to maintain reputation after hacking, infiltration or censorship. There is a need for new tools to address ways of preserving reputation while maintaining anonymity for group discussion on OSNs.

Appropriateness to Intended Users: Lastly, a security tool has to ultimately be usable to the intended population. This extends to both the behaviors of individuals and the local culture where it will be used. A technical solution that works well in one context (for example tested and developed in a western University) might not be suitable for applications in other contexts. In the next section we discuss how our team addressed the aforementioned requirements, including instances where iterative feedback from partner communities steered the technological design of the application to suit local needs.

6 SecurePost: a Tool for Verified group-anonymity

To address the constraints outlined in the previous section we developed an application called SecurePost, that allows individuals to share a single group identity while retaining individual anonymity on OSNs. SecurePost is comprised of three coordinated modules: an Android application that allows group members to post content and manage membership; a proxy server that relays posts to social networks; and a browser extension that allows members of the public to verify those posts. Together the modules provide group-anonymity coupled with an ability to verifying the integrity and authenticity of posts.

Using SecurePost, group members can make posts to shared OSN accounts, while masking their individual identity. The connection is routed through the companion proxy server, hiding the IP address from the OSN platform, which may cooperate with a hostile government. The identity of a poster is likewise hidden from other group members, giving plausible deniability for any given post. For members of the public looking at the posts on social media, it appears as if posts from an account have a single author with no way of identifying individual posters or a group's membership roster. In this manner, a group is able to build a social media presence while retaining anonymity of its members.

Because OSN accounts can be seized or infiltrated, SecurePost provides tools to verify that content is genuine and unmodified. SecurePost allows groups to attach cryptographic signatures to every post in order to verify the authenticity and integrity of messages. A companion browser extension allows anyone, even readers not part of the group, to verify posts directly on the social network web page. Using the extension, readers can know if a signed post came from a trusted member of the group and if its text has been modified, even if the OSN account has been compromised.

In the event that a group has been compromised, SecurePost allows dedicated administrators to retain membership control. If administrators see erroneous posts coming from the group, they are able to reset membership, expelling all other members, and invalidating past posts. This allows group leaders to protect the group in the event of infiltration as well as warn readers, protecting the group's reputation.

Together group-anonymity with a layer of verification provides a mechanism for groups to balance personal anonymity with building a trusted group identity.

6.1 The mobile application

At the center of SecurePost is the mobile application. Through the application, users are able to form groups, manage membership, as well as make and view posts to social media. As mentioned in the previous sections, due to resource constraints, we developed our application exclusively for Android, as it is the most common mobile operating system. Mindful of the prevalence of cheaper devices in the developing world running older operating system versions, we support Android API level 10. This makes our application backwards compatible with devices running Android 2.3.3 and above. As of October 2017, this accounts for 99.9% of Android devices registered with Google [90].

6.1.1 Forming a SecurePost group

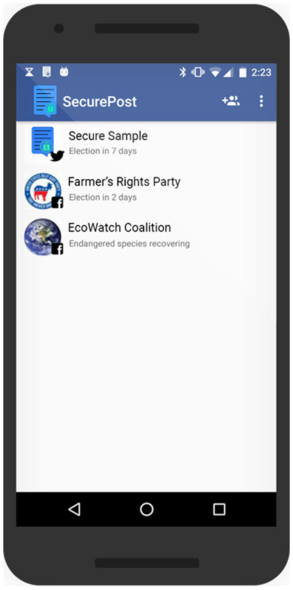
Each SecurePost group is tied to a corresponding OSN account. Currently SecurePost supports Facebook and Twitter, but its modular design can be extended to include any similar platform. SecurePost allows users to take part in multiple groups, without the need to switch accounts. Figure 2 shows the group overview screen of the application, where the user is a member of three groups.

To set up a SecurePost group, users require a corresponding OSN account on the platform of their choice. They can either use an established account (such as a Twitter handle for a newspaper) or set up an anonymous account. The application presents the platform specific API based login web page to the user. After a user logs on, the OSN platforms returns an access token. This is a common design pattern used by other social media applications. The access token is forwarded to the proxy server (discussed in Section 6.2) and discarded by the application.

The group creator is granted administrative rights to the group, and can post, invite others, and manage the group. As this initial step requires direct contact with the social media platform, to avoid associating an IP address to the user, we recommend that this step is done through an anonymity service such as Tor [18]. Other than this initial creation step, all application traffic is routed via the proxy server. Notably the user name and password are only needed during this initial group creation process. Inviting new members does not require the sharing of login credentials (a method of sharing social media accounts frequently cited by interviewees).

6.1.2 Group membership

Instead of relying on password sharing, SecurePost uses public key cryptography to authorize users. When creating an account, the app generates two asymmetric 2048 bit RSA key pairs. The app stores the private keys, and transmits the public keys to the proxy server. One key pair grants *posting credentials* while the other pair grants *administrative credentials.*

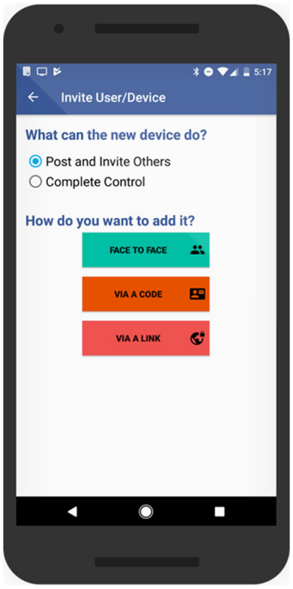


Secure Sample
Election in 7 days

Farmer's Rights Party
Election in 2 days

EcoWatch Coalition
Endangered species recovering

Fig. 2 List of SecurePost groups



What can the new device do?

● Post and Invite Others
○ Complete Control

How do you want to add it?

Fig. 3 Invite Wizard

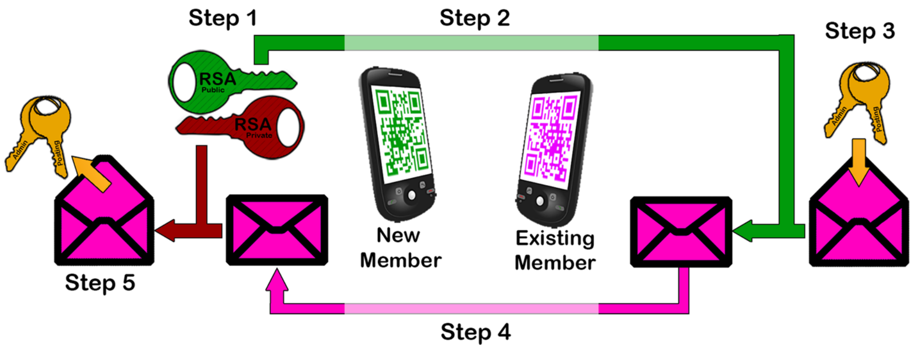


Fig. 4 Process Flow of Visual Invitation Scheme. Step 1: generate ephemeral asymmetric key pair. Step 2: ephemeral public key sent as QR code. Step 3: ephemeral public key encrypts group credentials. Step 4: send encrypted invite as QR code. Step 5: ephemeral private key decrypts group credentials

These two keys signify two classes of users: *administrators* (holding both posting and administrative keys) and *posters* (holding only the posting key). Anyone with a private key to the group can transfer a copy of that key via the application to grant access at the same privilege level or lower. In other words, an administrator can recruit administrators and posters while posters can only recruit other posters.

The complexity of the key exchange is hidden from the user via a graphical user interface (shown in Fig. 3). The application provides a number of methods of recruitment: Face-to-Face credential exchange, a sharable token, and a link.

Face-to-Face Visual Recruitment: When in close geographic proximity, SecurePost offers a secure key exchange via visual scanning of QR codes, as shown in Fig. 4. The exchange happens in two steps between a **recruit** (someone who wants to join the group) and an **existing group member**:

1 The **Recruit** initiates join process by visually showing an **existing group member** a join request containing a QR code.

   (a) The **recruit** seeking to join the group generates a 2048 bit RSA ephemeral key pair *(EP_PUB & EP_PRV)*.
   (b) The **recruit** shows *EP_PUB* in the form of a QR code *(QR_PUB)* to the **existing group member**.
   (c) The **existing member** scans *QR_PUB* and extracts *EP_PUB*.

2 The **existing group member** responds by sharing encrypted group credentials with the **recruit** in the form of a QR code.

   (a) The existing group member takes the group credentials of a desired privilege level *(CRED)* - i.e. the group's private keys - and encrypts them using *EP_PUB* as *EN_CRED*.
   (b) The **existing member** then display *EN_CRED* back in the form of a new QR code *(QR_INVITE)* to the **recruit**.
   (c) The **recruit** scans *QR_INVITE* and extracts *EN_CRED*.
   (d) The **recruit** decrypts *EN_CRED* using *EP_PRV*.
   (e) The **recruit** now has *CRED* and is part of the group.
   (f) Both the **recruit** and **existing member** discard the ephemeral keys *EP_PUB* and *EP_PRV* as they are no longer needed.

Once the recruit possesses the private keys, they are a part of the group: they can authenticate with the proxy and are ready to post. By using a two step process, an adversary visually observing the exchange would be unable to decrypt the group credentials without the recruit's ephemeral private key.

Alternative Recruitment: When physical proximity is not possible or unsafe, SecurePost allows alternate recruitment strategies via the use of an invite code or link. These strategies allow users to transmit keys via secure communication channels established outside the application.

For these strategies the existing group members use the graphical interface to initiate recruitment. The application encodes the the corresponding private keys into an invite code or link. The group member can then manually copy this code or use the Android share intent to paste it into a secure application of their choice, for example an end-end encrypted messaging client. The recruit pastes this code into their SecurePost application (or clicks the link) which decodes the private keys, granting group membership.

However, if this link or code is intercepted an adversary may be let into the group.

This option gives group members more flexibility but also more responsibility. We added this option in response to user testing, as users wanted a way to asynchronously invite users without physical access. Note that in both recruitment schemes the group members retain custody of the group's private keys. The proxy server only has access to the public keys.

6.1.3 Group administration

Authentication of group membership is verified by the proxy server using the *administrative* and *posting* public keys. These key are unique to the group and not the user. This approach intentionally omits a user registry. From the perspective of the proxy server and OSN platforms, each group appears as a single entity. The number and identities of group members is only known out of band through social interaction and is not retrievable from any part of the system.

As any group member can invite others, by compartmentalizing recruitment history from other group members, it is possible to hide the full membership roster from any single group member. In this manner, groups can enlist confidential contributers.

This structure imposes limitations on group administration. As there is no user registry or unique identifier, SecurePost lacks the ability to rescind membership to an individual user. Instead, if the group is compromised, an administrator must reset membership entirely. In this process the administrator performing the reset generates new key-pairs and transmits the public keys to the server (much like the initial process of group creation). All prior members are expelled (including other administrators) and have to be re-invited - this time hopefully with a higher level of scrutiny. In addition past posts are invalidated; they continue to remain on social media but are no longer marked as verified by the browser extension.

6.1.4 Social media posts

Once users join a group they are able to post directly to the associated OSN account. To members of the public (and other group members) it appears as if all a group's posts come from a single entity.

In addition to providing anonymity, SecurePost also offers verification for posts. Before transmission, each post is signed using the *posting key* of the group using SHA-256. This signature can then be used to verify the integrity and authenticity of the post. The application automatically verifies and displays posts from other group members as shown in Fig. 5. The general public can verify the posts on OSNs via the use of the browser extension described in Section 6.3. To handle situations where connectivity is disrupted - such as power failure or regional government censorship - posts are stored locally and forwarded to the proxy server when Internet connectivity is reestablished.

Several other algorithms have been proposed to provide group anonymity. Ring signature cryptography schemes, as proposed by Rivest, Shamir, and Tauman [91], provide a similar group anonymity guarantee. Signed messages can be authenticated as being authored by the owner of one the public keys, but it is not possible to determine which one.

However, the ring signature verification process requires the public keys of all group members. This is problematic for both design and implementation reasons in the context of SecurePost. Ring signatures used publicly in this application would leak metadata about the group membership. Specifically, an adversary could always determine the number of members, and could monitor the public key set for changes to determine when members join the group. Additionally, the amount of data that would need to be encoded into the account's profile image for the plugin verification to work would scale linearly with group size, potentially placing a cap on group size.

Group signatures, as proposed by Chaum and van Heyst [92], are similar to ring signatures, except they would allow the group owner to de-anonymize posts by members using the group owner's secret key. This would potentially resolve the problem of 'bad actors' joining the group, by allowing the group owner to identify and expel users who post content that the group owners do not agree with. This would allow the group administrator to identify and expel posters of problematic content from the group without completely purging group membership. However, this kind of cryptosystem reduces the strength of deniability, since there exists a person who can prove that a particular group member wrote the post. This puts the group owner at risk from external actors, who may be motivated to threaten or harm the group owner if they do not de-anonymize a particular post author.

Multimedia Posts: Initially, we only planned to support posting and verification of text. However, in our interviews, respondents stressed the importance of posting images as well as audio and video. Respondents found multimedia content, such as images of police impropriety, is an effective tool for reader motivation. This was particularly important to respondents in areas where Internet connectivity is disrupted and risk of device confiscation is high. In this case, users wanted to store images in an encrypted manner and queue them for posting when connectivity is available.

As a response we added the ability for users to post images. Unfortunately, as of now, SecurePost does not allow the verification of validity and authenticity for the posted images as OSNs compress and alter images prior to display. Thus the bitwise verification system we use for text post does not translate to multimedia posts. In the current implementation images are marked as unverifiable. Verification of images as well as support for other multimedia content are future work.

6.1.5 Secure storage

The application works even without continuous Internet access. Previous posts are cached, and new posts are stored locally and transmitted when Internet connectivity is reestablished. As devices can be searched, lost, stolen, or confiscated, all user data are encrypted. In scenarios where a region temporarily or permanently lacks Internet access, users can utilize SecurePost to prepare and ferry posts for delivery when they re-establish connectivity.

Previous posts, group memberships, keys, and other identifying information are stored using SQLCipher [93] (an encrypted database for Android). When the application is first launched, users choose an application-wide password for unlocking the app. This is needed each time the app is started in order to decrypt the database. While running, the application displays a persistent notification reminding the user that the database is unlocked. Dismissing this notification rapidly locks the database and closes the application.

The application password is unique to the device and not shared with the proxy server. In the event of password loss, the data are not recoverable. Users who forget the password are offered the option to wipe the data, allowing them to start over. This does unfortunately erase a user's group memberships as credentials are all locally stored.

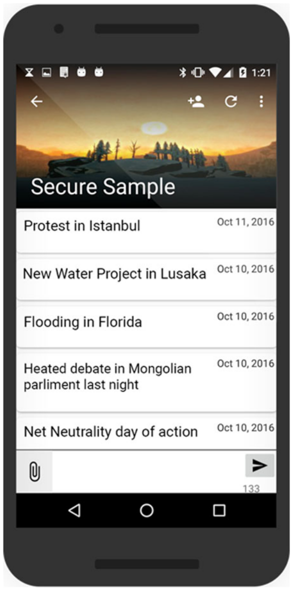
Our application is tailored to populations that often use cheaper devices running older operating systems. Older devices lack support for modern features such as full disk encryption or a secure enclave [94]. Our application-level encryption is done independently of any operating system-level encryption that the device may support. It can be used in conjunction with other security methods, and adds a layer of security for older devices. If the phone is confiscated by an adversary, the adversary would need to perform a costly attack to decrypt data, which would still not expose group membership.

Option to Self-Destruct: To handle the case of device confiscation, raised as a concern in the interviews, SecurePost implements an optional "self-destruct" password. This is set up in addition to the regular application password and is shown in Fig. 6. The self destruct password can be revealed if the user is under duress. Entering the password is visually presented as an incorrect password attempt while in reality it irretrievably wipes the content of the application.

6.1.6 Matching cryptographic design with needs

Our current invitation process is a reflection of feedback from our partners. In the initial designs of SecurePost, we envisioned scenarios where users could form short-term groups at physical gatherings such as protests.

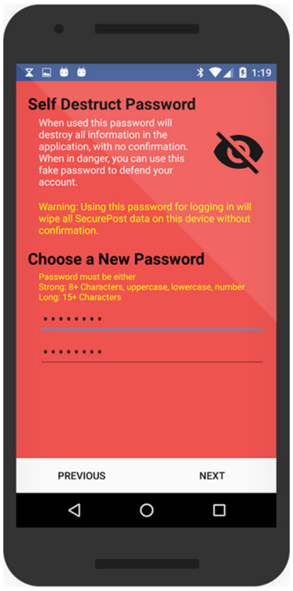
In this initial design users shared a secret pass phrase to join a group. Anyone with the pass phrase could post. Administration was only possible by the owner of the OSN



Secure Sample

Protest in Istanbul	Oct 11, 2016
New Water Project in Lusaka	Oct 10, 2016
Flooding in Florida	Oct 10, 2016
Heated debate in Mongolian parliment last night	Oct 10, 2016
Net Neutrality day of action	Oct 10, 2016

Fig. 5 View of an Individual Group



Self Destruct Password

When used this password will destroy all information in the application, with no confirmation. When in danger, you can use this fake password to defend your account.

Warning: Using this password for logging in will wipe all SecurePost data on this device without confirmation.

Choose a New Password

Password must be either
Strong: 8+ Characters, uppercase, lowercase, number
Long: 15+ Characters

•••••••••

•••••••••

Fig. 6 Setting a Self-destruct Password

account using the OSN account username and password. When a user joined a group, the app generated a time-synchronized hash chain using the pass phrase as a seed. To post, the group member signed the post using the hash from the current time slot which was verified by the proxy server. Using the browser extension, anyone from the public could verify any past post by using the signature of the most recent one. As time passed the hash chain shrank until it expired, at which point the group was dissolved.

Initially we were optimistic about this solution as it allowed an easy way for groups to spontaneously form. However through interviews and user testing we were forced to reconsider our approach. Users explained the importance of long-lasting groups that build trust and credibility over time. Despite short lived activities like protests, once users bonded together, they disliked the idea of auto-expiring membership. When recounting Gezi park protests in Istanbul, protesters noted the importance of continuing to grow activist groups after the event.

Sharing a long password also proved a usability challenge. We required users to come up with and share a long password or phrase, independent of the OSN account and unique for each group. Users had a hard time remembering passwords and resorted to using simple easy to guess passwords. The situation was complicated further by having an application password, a self-destruct password and the possibility of adding multiple groups.

Lastly, as mentioned, the hash chain approach used posts from the latest time block to validate posts from previous time blocks. This meant that : (1) the most recent post could not be validated and (2) posts in the current time block could not be validated. In testing, users stressed that the most recent post is often the most important as it is the most timely, and were confused why there was no mechanism for validation.

From this feedback we settled on the key based approach described in detail earlier in this section. The major design change is described here to highlight the importance of understanding the user community in system design. Through the course of this project, our team was able to move from preconceived understanding of the technical needs of users to a tailored approach through social analysis and iterative user testing.

6.2 The proxy server

The Android application communicates with a companion proxy server. Aside from the initial group creation (that uses the native social media platform authorization API) all content flows through this proxy. The proxy keeps group-level state and masks the individual user's IP address from the OSNs.

As discussed in Section 6.1, when groups are first created through the SecurePost application, the OSN platform issues an access token that the app forwards to the proxy. The app also creates and forwards the group's public *posting key* and the public *administrative key*. The proxy stores the OSN access token and uses it to make posts and change banner and profile images. The proxy also stores the public keys and uses them to verify posting and administrative rights. If the group is reset by the administrator through the application, the keys are updated but the OSN access token remains consistent. Notably the proxy does not ever receive the private keys for the group.

Since OSNs log the IP addresses of users, they may be compelled by governments to identify users and produce access logs. The proxy masks individual users' IP addresses. The proxy itself does not log IP addresses or keep any individual user metrics. Because the server may become compromised, it does not store private keys for groups. Adversaries who gain access to the proxy would be able to make posts using the OSN access token but not sign them. Any posts made by adversaries using the OSN access token from the proxy server would lack signatures and show up as invalid. The OSN access token could still be rescinded by a group member with the login information to the OSN account.

The server consists of a standalone Java application running Jetty utilizing a MongoDB No-SQL database for storage. Interaction with the application is implemented through a JSON REST API running over HTTPS. If the proxy needs to be scaled, multiple instances of the proxy can be spun up synchronizing via the MongoDB.

Currently, we run an instance of the proxy on Amazon Web Services. By default, users of the SecurePost application utilize this instance. As users may want to audit the source code and run their own instance, the project is open source and we allow easy configuration of the application to point to a different proxy server instance.

Unlike OSN platforms that have financial incentive to cooperate with adversaries, anyone can run a SecurePost proxy on any machine of their choosing. While some may choose to run it in a data center (possibly in another country outside the jurisdiction of their government) others may choose to do so on anonymous machines. This flexibility allows a particular group to retain control of their security and the point of failure.

If a proxy is blocked, the posts made by the group will still be visible and verifiable for the general population. There would be a brief disruption in posting content for group members, but as the group administrators have full access to the app and proxy, they could migrate to a new IP or new machine to bypass this block. Further, as we elaborate in Section 6.3, as long as the verification protocol is not altered, the same browser extension can run irrespective of the back-end proxy with no reconfiguration.

6.3 The browser extension

SecurePost is designed to be used to disseminate information to the public. Posts made from the application are posted directly to social media. The browser extension allows any member of the public, irrespective of group membership to verify any post made with SecurePost. Currently the browser extension is implemented as a cross-browser extension and is compatible with Chrome, Firefox, and Opera. It runs independent of the proxy server, and uses only the contents of the OSN web page to verify posts.

6.3.1 *Post signature*

The application automatically generates signatures for each post using the private *posting key* for the group. If the SecurePost group enables the option to use verification, the proxy appends an image containing the cryptographic signature when posting to the OSN, as shown by Fig. 7. As OSNs compress images, typically as JPGs, we use a compression resistant encoding for the image based signature. The signature is encoded as a series of monochromatic 3x3 pixel squares aligned to the row/column boundaries that match the JPG encoding algorithm. The encoded signature presents to the user similar to TV "snow" at the bottom of an image that informs the reader how to verify the post.

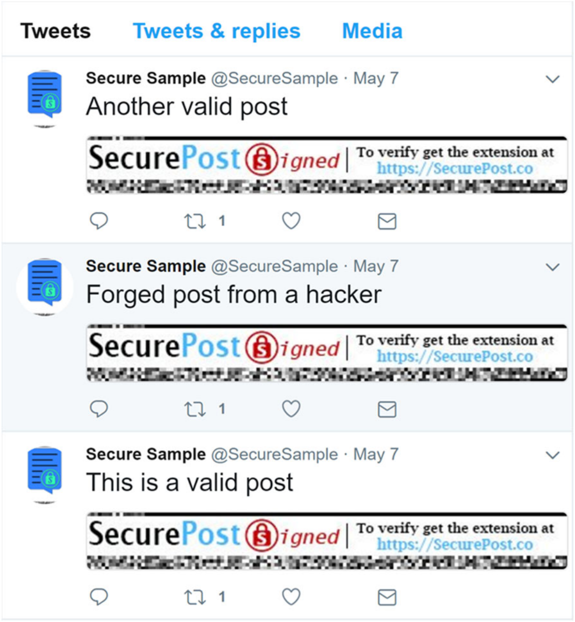
The browser extension processes this banner, decoding the signature back to binary. To verify the signature, the extension requires the *posting public key*. This key is encoded by the proxy, in the same manner as the signature, into the profile image on Twitter and the cover image on Facebook. This occurs when the group is first created and whenever the images are changed. The public posting key is likewise decoded by the browser and, with the aid of the signature, is used to verify authenticity and integrity of the post.

Notably this technique does not require co-operation with the social media platform or the proxy server. The same extension can verify messages from multiple groups which may be using different proxy servers provided they use the encoding protocol for message signatures. This removes the need for non group members to access the proxy, reducing the risk of traffic analysis by an entity with a sufficient view of the network. Additionally this approach reduces the server load on the proxy as the number of readers requiring verification would likely be magnitudes greater then content creators.

6.3.2 *Verifying posts*

When installed, the extension automatically verifies posts when on a supported OSN web page. The extension first determines whether the account uses SecurePost by reading a pixel pattern encoded in the corner of the profile image, and if so goes on to validate the posts in the feed. The extension uses a mix of color and symbols to inform the user of the validity of a post, as shown in Fig. 8. It also hides the signature images from the user to improve the user experience.

Posts made directly to social media, without the use of the SecurePost app, do not have access to the *private posting key* and are marked as unverified. Similarly, as discussed earlier, multimedia posts, made through the app are not able to be verified. As the signature is based on



Tweets
Tweets & replies
Media

Secure Sample @SecureSample · May 7
Another valid post

SecurePost Signed | To verify get the extension at https://SecurePost.co
1

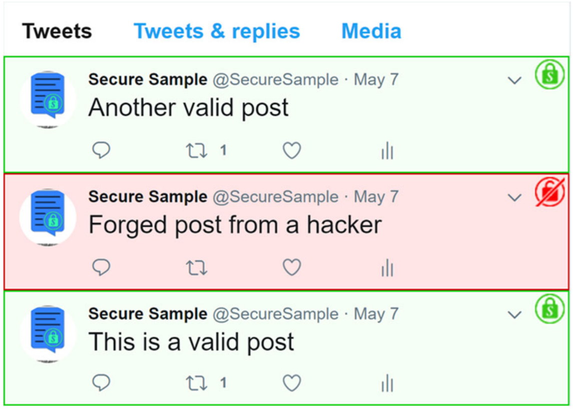
Secure Sample @SecureSample · May 7
Forged post from a hacker

SecurePost Signed | To verify get the extension at https://SecurePost.co
1

Secure Sample @SecureSample · May 7
This is a valid post

SecurePost Signed | To verify get the extension at https://SecurePost.co
1

Fig. 7 Unmodified Twitter Feed. Example of a Twitter feed using the optional verification feature of SecurePost. Each post has an associated signature image



Secure Sample @SecureSample · May 7
Another valid post

Secure Sample @SecureSample · May 7
Forged post from a hacker

Secure Sample @SecureSample · May 7
This is a valid post

Fig. 8 The Browser Extension. Example of the same feed as Fig. 7, but now using the browser extension. The signature image is automatically hidden and posts are highlighted to show validity. Mousing over the image in the top right corner of a post shows the user the reason that the post was marked

the content of the post, copying another post's signature image is insufficient to falsely validate a post. Users are graphically presented with the reason that a post is not valid or not able to be validated by hovering over an icon in the corner of the post.

If somehow an adversary took full control of the OSN account and set up a new SecurePost group, they could change the public key and make new posts that are verifiable. However all previous posts (that readers trusted) will show up as no longer verified to any member of the public using the browser. There is no way for them to create new verified posts that use the previous signing key. This would be a strong indicator to the public that there was a serious problem and perhaps a change of ownership has taken place.

6.3.3 Design process for image-based signatures

The ability to verify posts is one of the key facets of our work. As our partners wanted both Twitter and Facebook support, we needed a solution that was sensitive to the character limits of these OSNs, which at the time, was 140 characters on Twitter.

In our initial design, we experimented with text-based signatures using 21-bit CJK Unified Ideographs. This character set has high bit-density per character, which allowed us to maximize bit count while minimizing the number of characters. An example of this approach is shown in Fig. 9.

While doing user testing and interviews in Mongolia, participants expressed that this approach was unsuitable due to local cultural norms. In Mongolia there are strong tensions with China, and social ramifications for perceived affinity for one of its neighbors. By using characters associated with China or Korea, even if the characters did not correspond to actual or Korean text, users exposed themselves to perceptions of siding with these countries. This was particularly problematic for groups already marginalized.

In response to this social constraint, we moved to an image-based signatures which fits within cultural norms while still satisfying character limits. This change in design highlights the value in understanding the social context for which software is developed.

7 Usage and evaluation

SecurePost is freely available on the Google Play Store. The browser extension, which allows users to verify posts, is available on the Google Chrome Web Store for free. Our app is available in 7 languages: Arabic, English, French, Mongolian, Russian, Spanish, Turkish. All modules, including the Android application, proxy server, and browser extension are open source and available through BitBucket and on our website [44].

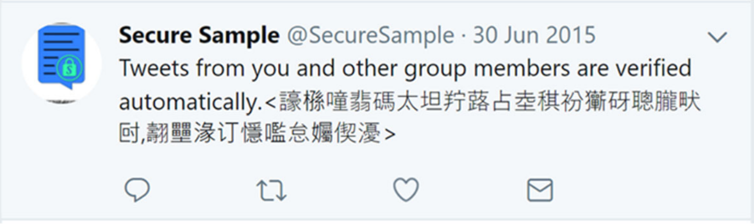
7.1 Satisfying design requirements

As *OSNs are highly utilized* by our partner communities, we focused our work on the two highly utilized social networks: Twitter and Facebook. Since *smart phones are the primary method of Internet access*, our technical solution was designed to work with Android smart phones, while providing *backwards compatibility* for older devices.

In our work we identified public group communication that *protected anonymity* while *preserving reputation* as an unmet need for our partner communities. We address the need for *anonymity* by allowing groups to share a single OSN account while maintaining individual anonymity. By sharing access keys, there is no group roster. IPs are kept hidden from OSNs via the use of a proxy that groups can retain full ownership and control over.

To *maintain reputation* we provide a method of verifying post authenticity with the use of cryptographic signatures. Members of the public can verify that a post came from a group using a companion browser. Posts that are modified are no longer verified, and accounts that are seized or hacked are unable to post verifiable posts. If a group is ever compromised administrators can reset membership and invalidate past posts to signal to readers that there is a problem and they should re-evaluate the truthfulness of current posts. In this manner groups can build reputation, and signal if that reputation might be compromised.

In future work we are interested in exploring ways of migrating groups between OSN accounts using the same signature keys and proxy servers. This would allow groups to switch to a new OSN account in light of censorship while preserving reputation, and could allow verification of identity across services. We are also exploring ideas of



Secure Sample @SecureSample · 30 Jun 2015

Tweets from you and other group members are verified automatically.<譹楙噇翡碼太坦羘蕗占垄likewise tiers of user class so that posts from a designated core group of users can retain verification after group reset.

As participants, such as journalists, expressed a need to report from areas with periodic or permanent *network disruption*, SecurePost maintains a local cache of posts. Users can view past posts without connectivity, and compose posts that are delivered when connectivity is reestablished. In this manner users can ferry information from disconnected geographic regions.

As storing content locally leaves users vulnerable to *device confiscation and search*, SecurePost encrypts all data using an application-wide password. In the event that a user is under duress they can provide a false password that wipes the application data including posting keys.

While designing the application we received iterative feedback from participants that tailored design decisions to be *appropriate for our intended users*. As outlined in the previous section we modified our initial approaches to fit cultural norms (such as switching from CJK text-based signatures) and fit with user social behavior (such as abandoning the initial hash-chain approach).

7.2 Usage survey

As anonymity is at the core of the design philosophy, SecurePost does not collect usage data beyond that which is necessary for functionality of the app and server, as well as basic install statistics collected by the Google Play Store. Based on data from the Google Play Store, as of October 2017, we have had over 400 installs of our application. As we show in Fig. 10, our users reside not only the countries we were explicitly targeting: USA, Mongolia, Turkey, Zambia, but in other countries where censorship is a problem (59 countries in total). SecurePost users have so far formed 68 groups and made 336 posts.

When the application is first run, users are given the option to provide optional demographic information. The data to date are summarized in Table 5. As of October 2017, we received 329 responses. From this survey, we found the majority (62%) of users identified as male. Most were 30 years old or younger (91%) and had some higher education (87%). Because we led user studies and presented our application at universities and organizations where college-age students are likely to work, we expected our users to fall into this demographic. In developing countries, younger generations and men are also more likely to use the Internet as a whole [46].

For language, few respondents (15% of total responses) stated a preferred language. Those that did largely preferred English, the dominant language in Zambia and the United States, which makes up a large portion of all installs.

Our survey also automatically registered the language to which the application was currently set. The application

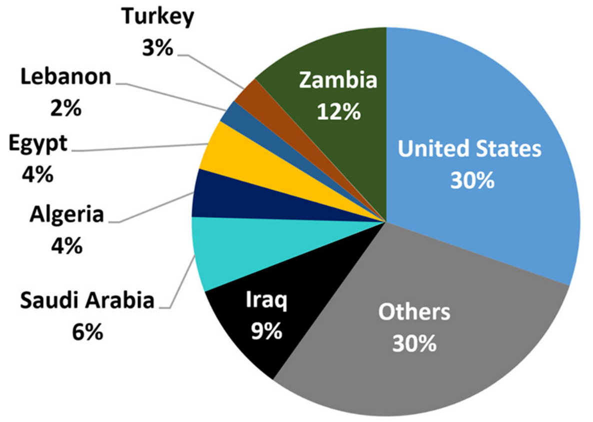


Fig. 10 Installs by Country. SecurePost application installation by country as of October 2017

Table 5 Results of initial demographic survey. Collected by the application when it is first run

	%	(#)
Gender		
Male	62%	(95)
Female	35%	(54)
Other	3%	(4)
Age		
≤ 20 years old	42%	(67)
21–30 years old	49%	(79)
31–40 years old	6%	(9)
41–50 years old	2%	(3)
≥ 51 years old	1%	(1)
Education		
Primary school	1%	(2)
Secondary school / High school	12%	(18)
Higher education or university	87%	(130)
Preferred Language		
Arabic	2%	(1)
English	90%	(43)
Spanish	8%	(4)
Application/OS Language		
Arabic	0.3%	(1)
English	95%	(314)
Spanish	2%	(8)
French	1%	(3)
Russian	0.3%	(1)
Turkish	0.6%	(2)
Total # Surveyed	(n=329)	

language is a match-up between operating system defaults and the seven supported languages of the application. Languages not supported by the application would be reported as the next likely language, usually English. The exact implementation varies by Android version and is detailed in [95]. In this metric, English is again the dominant language (95%), suggesting that our translations are not heavily utilized.

After using the application for three days, users were asked if they would take an optional survey based on their experiences using SecurePost. This follow-up survey is also available through an in-app menu. A subset of the survey questions is summarized in Table 6. A total of 22 users have so far completed this survey, which is a small sample. We are working to increase the response rate by re-offering the survey.

Responses to the follow-up survey indicate that the majority of respondents found SecurePost to be at least somewhat useful (73%). When asked to what extent they thought that using SecurePost has improved their confidence in using online social networks, most said it improved their confidence a little, somewhat, or a lot (91%). When asked to what extent users feel freer in their ability to express themselves on Twitter and/or Facebook without concerns about surveillance or security of the messages sent when using SecurePost, most said they feel freer to some extent (73%).

Table 6 Results of follow-up survey

	%	(#)
How useful is SecurePost to user		
Not useful at all	4%	(1)
A little useful	23%	(5)
Somewhat useful	41%	(9)
Very Useful	32%	(7)
SecurePost Improved confidence with OSNs		
Not improved confidence	9%	(2)
Improved confidence a little	41%	(9)
Improved confidence somewhat	45%	(10)
Improved confidence a lot	5%	(1)
Feel Freer on OSNs		
A lot less free	4%	(1)
A little less free	9%	(2)
Equally as free	4%	(1)
A little more free	50%	(11)
A lot more free	23%	(5)
Total # Surveyed	(n=22)	

Collected by the application after three days of use

8 Conclusion

Our research seeks to understand and address barriers to free speech on the Internet for vulnerable communities. While there are many significant differences across these diverse communities, we observed particular patterns in free speech challenges among Internet and OSN users across these contexts. Social research revealed user concerns ranging from account disruptions to online credibility issues to equipment seizures. We built a novel tool to address such challenges in partnership with affected communities.

SecurePost allows users greater control of anonymity through group-based communication on OSNs. Through verified group anonymity, users build trust and reputation as a collective, without exposing the identities of individuals. By allowing communities to set up their own instances of the SecurePost proxy server, users do not have to trust OSNs to protect IPs and identities of individual members.

Using SecurePost, an administrator can share OSN accounts without sharing the account passwords and hence maintain control of the account. If the account is seized or hacked, the browser extension can still identify fraudulent or edited posts using the cryptographic signature.

We developed SecurePost to support the most popular OSN platforms (Facebook and Twitter) and device types (Android smart phone), providing compatibility for older devices lacking some of the security features (e.g. encrypted storage) of newer phones. The application also provides a means of storing messages for later delivery to counter network disruption due to power loss or government censure. Because the project is open source and designed for modularity, other similar platforms and systems can be incorporated in the future.

Like other anonymity applications, individual users can still be revealed by posting personal information in the contents of a message. Given time and access to the file structure of the device, it is also possible for the encrypted storage to be decrypted, for example via a brute force attack. Additionally, adversaries with a sufficient view of the network may still implement de-anonymization through timing analysis. We hope to address these vulnerabilities in future work.

A frequent concern for the development of security applications is the potential misuse by ill-meaning organizations, like terrorist cells. Because the data are all posted publicly, our app does not expand the capabilities of malevolent secret communication. While our tool allows users to remain anonymous, it does not prevent OSNs from censoring accounts or content. We leave the decision of what constitutes a danger to the OSN.

Because all elements of our software are open source, communities do not have to trust us (the developers) or

OSNs to protect the identities of individual members. They can audit and improve the code, and can set up their own instances of the SecurePost proxy server that is isolated from developers.

Finally, our work provides insight into community collaborations. By partnering with locals and understanding social context, specific needs, and user behaviors, we were able to come up with a novel method of adding verification to non-cooperative online social networks.

Abbreviations

OSN: Online social network; ZWD: Zambia watchdog

Acknowledgments

We would like to thank all the other people who have helped with this work. Many thanks to Hannah Goodwin, Kristin Hocevar, Lisa Han, Ariel Hasell, and Rahul Mukherjee for conducting interviews and surveys as well as providing analysis. Thanks to Ben Zhao, Divya Sambasivan, and Pritha Narayanappa for helping develop SecurePost. Thanks to Irina Artamonova for helping with the statistical analysis of our survey data. And finally thanks to our many overseas partners for the hospitality, patience, and many hours of work.

Funding

This work was funded by the US Department of State.

Availability of data and materials

Please contact author for data requests.

Authors' contributions

MN was the primary author of this manuscript, and a large contributer to the application. DI contributed heavily to application design, particularly the cryptography. MM was the lead for survey based work. LP was the lead for interview based work. EB was the lead for application development. All authors read and approved the final manuscript.

Ethics approval and consent to participate

We obtained IRB approval from UC Santa Barbara prior to conducting our fieldwork.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Department of Computer Science, UC Santa Barbara, Santa Barbara 93106, CA, USA. [2]Department of Communication, UC Santa Barbara, Santa Barbara 93106, CA, USA. [3]Department of Comparative Media Studies, MIT, Cambridge 02139, MA, USA.

Received: 11 May 2018 Accepted: 1 August 2018
Published online: 02 November 2018

References

1. Facebook. Two Billion People Coming Together on Facebook. 2017. https://newsroom.fb.com/news/2017/06/two-billion-people-coming-together%-on-facebook/. Accessed on 23 Apr 2018.
2. Stats IL. Twitter Usage Statistics. http://www.internetlivestats.com/twitter-statistics/. Accessed on 17 Oct 2017.
3. United Nations. Universal Declaration of Human Rights. 1948. http://www.un.org/en/universal-declaration-human-rights/. Accessed on 16 Feb 2017.
4. United Nations. Resolution 32/13: The promotion, protection and enjoyment of human rights on the Internet. 2016. https://http://undocs.org/A/HRC/RES/32/13. Accessed Oct 2017.
5. Peterson A. Turkey strengthens Twitter ban, institutes IP level block. 2014. https://www.washingtonpost.com/news/the-switch/wp/2014/03/22/turkey-strengthens-twitter-ban-institutes-ip-level-block. Accessed June 2016.
6. Turkey Blocks. New internet shutdown in Turkey's Southeast: 8% of country now offline amidst Diyarbakir unrest. 2016. https://turkeyblocks.org/2016/10/27/new-internet-shutdown-turkey-southeast-offline-diyarbakir-unrest/. Accessed June 2016.
7. Dainotti A, Squarcella C, Aben E, Claffy KC, Chiesa M, Russo M, Pescapé A. Analysis of Country-wide Internet Outages Caused by Censorship. In: Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference. IMC '11. Berlin; 2011. p. 1–18. http://doi.acm.org/10.1145/2X00000.2068818. Accessed Oct 2017. https://doi.org/10.1145/2068816.2068818.
8. Lee TB. Here's how Iran censors the Internet. 2013. https://www.washingtonpost.com/news/the-switch/wp/2013/08/15/heres-how-iran-censors-the-internet. Accessed Oct 2017.
9. Kelly S, Earp M, Reed L, Shahbaz A, Truong M. Privatizing Censorship, Eroding Privacy. 2015. https://freedomhouse.org/sites/default/files/FH_FOTN_2015Report.pdf. Accessed Oct 2017.
10. Lim K, Danubrata E. Singapore seen getting tough on dissent as cartoonist charged. 2013. http://www.reuters.com/article/us-singapore-dissent-idUSBRE96P0AF20130726. Accessed Oct 2017.
11. Breindl Y, Wright J. Internet Filtering in Liberal Democracies. In: Proceedings of the 2nd USENIX Workshop on Free and Open Communications on the Internet. Bellevue: USENIX; 2012. https://www.usenix.org/conference/foci12/workshop-program/presentation/Breindl. Accessed Oct 2017.
12. Goldman D. Donald Trump wants to 'close up' the Internet. 2015. http://money.cnn.com/2015/12/08/technology/donald-trump-internet/. Accessed Oct 2017.
13. Riley C. Theresa May: Internet must be regulated to prevent terrorism. 2017. http://money.cnn.com/2017/06/04/technology/social-media-terrorism-extremism-london/index.html. Accessed Oct 2017.
14. Norris P. It's not just Trump. Authoritarian populism is rising across the West. Here's why. 2016. https://www.washingtonpost.com/news/monkey-cage/wp/2016/03/11/its-not-just-trump-authoritarian-populism-is-rising-across-the-west-heres-why. Accessed Oct 2017.
15. House F. Freedom in the World 2017. 2017. https://freedomhouse.org/sites/default/files/FH_FIW_2017_Report_Final.pdf. Accessed on 25 Oct 2017.
16. Kasparov G, Halvorssen T. Why the rise of authoritarianism is a global catastrophe. 2017. https://www.washingtonpost.com/news/democracy-post/wp/2017/02/13/why-the-rise-of-authoritarianism-is-a-global-catastrophe. Accessed Oct 2017.
17. Williams R. The Rise of Authoritarianism. 2016. https://www.psychologytoday.com/blog/wired-success/201603/the-rise-authoritarianism. Accessed Oct 2017.
18. Syverson P, Dingledine R, Mathewson N. Tor: the second generation onion router. In: Proceedings of the USENIX Conference on Security Symposium. USENIX Association. SSYM'04. Berkeley: USENIX Association. p. 21–21. http://dl.acm.org/citation.cfm?id=1X00000.1251396. Accessed May 2017. http://dl.acm.org/citation.cfm?id=1251375.1251396.
19. Open Whisper Systems. Signal. https://itunes.apple.com/us/app/signal-private-messenger/id874139669. Accessed Feb 2017.
20. Inc W. WhatsApp. https://www.whatsapp.com/. Accessed on 27 Oct 2017.
21. Carey B. How Fiction Becomes Fact on Social Media. 2017. https://www.nytimes.com/2017/10/20/health/social-media-fake-news.html?r=0. Accessed Oct 2017.
22. Earle S. Trolls, Bots and Fake News: The Mysterious World of Social Media Manipulation. 2017. http://www.newsweek.com/trolls-bots-and-fake-news-dark-and-mysterious-world-social-media-manipulation-682155. Accessed Oct 2017.
23. Benedictus L. Invasion of the troll armies: 'Social media where the war goes on'. 2016. https://www.theguardian.com/media/2016/nov/06/troll-armies-social-media-trump-russian. Accessed Feb 2017.
24. Chen A. The Agency. 2015. https://www.nytimes.com/2015/06/07/magazine/the-agency.html. Accessed Feb 2017.

25. Jackson D. AP Twitter feed hacked; no attack at White House. 2013. https://www.usatoday.com/story/theoval/2013/04/23/obama-carney-associated-press-hack-white-house/2106757/. Accessed Oct 2017.
26. Bowden G. BBC Northampton Twitter Account Issues Donald Trump Shot Tweet After 'Hack'. 2017. http://www.huffingtonpost.co.uk/entry/bbc-northampton-twitter-account-donald-trump_uk_588340fae4b0f94bb303e768. Accessed Oct 2017.
27. Ingram M. Twitter Hack Takes Over Accounts to Spread Fake News. 2017. http://fortune.com/2017/06/09/twitter-hack-fake-news/. Accessed Oct 2017.
28. Nekrasov M, Parks L, Belding E. Limits to Internet Freedoms: Being Heard in an Increasingly Authoritarian World. In: Proceedings of the Third Workshop on Computing Within Limits. ACM LIMITS '17. Santa Barbara; 2017.
29. Nekrasov M, Iland D, Metzger M, Zhao B, Belding E. SecurePost: Verified Group-Anonymity on Social Media. In: Proceedings of the 7th USENIX Workshop on Free and Open Communications on the Internet FOCI. USENIX; 2017.
30. Zittrain J, Edelman B. Internet filtering in China. IEEE Internet Comput. 2003;7(2):70–77.
31. King G, Pan J, Roberts ME. How Censorship in China Allows Government Criticism but Silences Collective Expression. Am Polit Sci Rev. 2013;107(2): 326–43.
32. Crandall JR, Zinn D, Byrd M, Barr ET, East R. ConceptDoppler: a weather tracker for internet censorship. In: ACM Conference on Computer and Communications Security; 2007. p. 352–65.
33. Winter P, Lindskog S. How the Great Firewall of China is Blocking Tor. In: Proceedings of the 2nd USENIX Workshop on Free and Open Communications on the Internet. Bellevue: USENIX; 2012. https://www.usenix.org/conference/foci12/workshop-program/presentation/Winter. Accessed June 2016.
34. Kou Y, Semaan B, Nardi B. A Confucian Look at Internet Censorship in China. In: Bernhaupt R, Dalvi G, Joshi A, Balkrishan DK, O'Neill J, Winckler M, editors. Human-Computer Interaction. Cham: Springer International Publishing; 2017. p. 377–98. https://doi.org/10.1007/978-3-319-67744-6_25. Accessed Oct 2017.
35. Aryan S, Aryan H, Halderman JA. Internet Censorship in Iran: A First Look. In: Proceedings of the 3rd USENIX Workshop on Free and Open Communications on the Internet. Washington; 2013. https://www.usenix.org/conference/foci13/internet-censorship-iran-first-look. Accessed June 2016.
36. Chaabane A, Chen T, Cunche M, De Cristofaro E, Friedman A, Kaafar MA. Censorship in the Wild: Analyzing Internet Filtering in Syria. In: Proceedings of the 2014 Conference on Internet Measurement Conference. IMC '14; 2014. p. 285–98.
37. Nabi Z. The Anatomy of Web Censorship in Pakistan. In: Proceedings of the 3rd USENIX Workshop on Free and Open Communications on the Internet. Washington; 2013. https://www.usenix.org/conference/foci13/workshop-program/presentation/Nabi. Accessed June 2016.
38. House F. Freedom in the World 2016 Table of Country Scores. 2016. https://freedomhouse.org/report/freedom-world-2016/table-scores Accessed 23 Oct 2017.
39. Internet Monitor. Zambia. http://thenetmonitor.org/countries/zmb/access. Accessed June 2014.
40. High Commissioner of the Republic of Zambia. Demography. http://www.zambiapretoria.net/demography/. Accessed 09 April 2018.
41. OECD Better Life Index. Turkey. http://www.oecdbetterlifeindex.org/countries/turkey/. Accessed June 2015.
42. Chilkhaasuren B, Baasankhuu B. Population and economic activities of Ulaanbaatar. 2012. https://www.ubstat.mn2Fupload2Freports2Fub_khotiin_khun_am_ediin_zasag_angli_ulaanbaatar_2012-08.pdf. Accessed June 2016.
43. Strauss A, Corbin J. Grounded theory methodology. Handb Qual Res. 1994;17:273–85.
44. SecurePost. SecurePost - Safe, Secure, Social Media. https://securepost.co. Accessed May 2017.
45. ITU. 2017 estimates for key ICT indicators. 2017. http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2017/ITU_Key_2005-2017_ICT_data.xls. Accessed Oct 2017.
46. ITU. ICT Facts and Figures 2017. 2017. http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf. Accessed Oct 2017.
47. Alexa. Top 500 Global Sites. https://www.alexa.com/topsites. Accessed 30 Oct 2017.
48. Statista. Global social media ranking 2017. 2017. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/. Accessed Oct 2017.
49. StatCounter. Mobile Operating System Market Share Zambia. 2017. http://gs.statcounter.com/os-market-share/mobile/zambia/#monthly-201302-201709-bar. Accessed 16 Oct 2017.
50. StatCounter. Mobile Operating System Market Share Turkey. 2017. http://gs.statcounter.com/os-market-share/mobile/turkey. Accessed 23 Oct 2017.
51. Statista. Mobile OS market share 2017. 2017. https://www.statista.com/statistics/266136/global-market-share-held-by-smartphone-operating-systems/. Accessed 16 Oct 2017.
52. Bhattacharya A. Android (GOOG) just hit a record 88% market share of all smartphones. 2016. https://qz.com/826672/android-goog-just-hit-a-record-88-market-share-of-all-smartphones/. Accessed Oct 2017.
53. FreedomHouse. Turkey - Country report - Freedom of the Press - 2014. 2014. https://freedomhouse.org/report/freedom-press/2014/turkey. Accessed 02 Oct 2017.
54. Lowen M. Is Gollum good or evil? Jail term in Turkey hinges on answer. 2015. http://www.bbc.com/news/world-europe-32302697. Accessed Oct 2017.
55. Letsch C. Turkish composer and pianist convicted of blasphemy on Twitter. 2013. https://www.theguardian.com/world/2013/apr/15/turkish-composer-fazil-say-convicted-blasphemhy. Accessed 02 Oct 2017.
56. Akwei I. Zambian opposition leader arrested over 'libelous' Facebook post. 2017. http://www.africanews.com/2017/04/14/zambian-opposition-leader-arrested-for-libelous-facebook-post//. Accessed 02 Oct 2017.
57. Times L. Zambia : Police arrest engineering student for 'insulting' President Lungu on Facebook. 2017. https://www.lusakatimes.com/2017/07/25/police-arrest-unza-student-insulting-president-lungu-facebook/. Accessed 25 Oct 2017.
58. FreedomHouse. Freedom Of The Net - 2016. 2016. https://freedomhouse.org/report/freedom-net/freedom-net-2016. Accessed 02 Oct 2017.
59. Sturcke J. Libel laws explained. 2006. https://www.theguardian.com/technology/2006/aug/31/news.politicsandthemedia. Accessed Oct 2017.
60. Krotoszynski Jr RJ. Defamation in the Digital Age: Some Comparative Law Observations on the Difficulty of Reconciling Free Speech and Reputation in the Emerging Global Village. Wash Lee Law Rev. 2005;62(1):339.
61. US Department of State. Mongolia Country Reports on Human Rights Practices. 2016. http://www.state.gov/j/drl/rls/hrrpt/humanrightsreport/index.htm?year=2016&dlid=265356. Accessed 03 Oct 2017.
62. Gardner L. Mongolia's Media Laws Threaten Press Freedom. 2014. http://mediashift.org/2014/04/mongolias-media-laws-threaten-press-freedom/. Accessed 03 Oct 2017.
63. FreedomHouse. Mongolia - Country report - Freedom of the Press. 2015. https://freedomhouse.org/report/freedom-press/2015/mongolia. Accessed 03 Oct 2017.
64. Hebl MR, Foster JB, Mannix LM, Dovidio JF. Formal and interpersonal discrimination: A field study of bias toward homosexual applicants. Personal Soc Psychol Bull. 2002;28(6):815–825.
65. Facebook. What names are allowed on Facebook?. https://www.facebook.com/help/112146705538576. Accessed Feb 2017.
66. Galperin E. Changes to Facebook's "Real Names" Policy Still Don't Fix the Problem. 2015. https://www.eff.org/deeplinks/2015/12/changes-facebooks-real-names-policy-still-dont-fix-problem. Accessed Feb 2017.
67. Facebook. Government Requests Report. https://govtrequests.facebook.com/. Accessed Feb 2017.
68. Twitter. Information requests. https://transparency.twitter.com/en/information-requests.html. Accessed May 2017.
69. Hinduja S. Doxing and Cyberbullying. 2015. http://cyberbullying.org/doxing-and-cyberbullying. Accessed Oct 2017.
70. Ellis EG. Doxing Is a Perilous Form of Justice—Even When It's Outing Nazis. 2017. https://www.wired.com/story/doxing-charlottesville/. Accessed Oct 2017.
71. IFLA. How To Spot Fake News. https://www.ifla.org/publications/node/11174. Accessed 02 Oct 2017.
72. Kiely E, Robertson L. How to Spot Fake News. 2016. http://www.factcheck.org/2016/11/how-to-spot-fake-news/. Accessed 02 Oct 2017.
73. Parks L, Mukherjee R. From platform jumping to self-censorship: Internet freedom, social media, and circumvention practices in Zambia.

Communication and Critical/Cultural Studies. 2017. https://doi.org/10.1080/14791420.2017.1290262. Accessed Oct 2017.
74. Walker S. Salutin' Putin: inside a Russian troll house. 2015. https://www.theguardian.com/world/2015/apr/02/putin-kremlin-inside-russian-troll-house. Accessed Feb 2017.
75. Hess A. On Twitter, a Battle Among Political Bots. 2016. https://www.nytimes.com/2016/12/14/arts/on-twitter-a-battle-among-political-bots.html. Accessed Feb 2017.
76. Miller C. Bots will set the political agenda in 2017. 2017. http://www.wired.co.uk/article/politics-governments-bots-twitter. Accessed Feb 2017.
77. Yang Z, Wilson C, Wang X, Gao T, Zhao BY, Dai Y. Uncovering social network sybils in the wild. ACM Trans Knowl Discov Data (TKDD). 2014;8(1):2.
78. Verkamp JP, Gupta M. Five Incidents, One Theme: Twitter Spam as a Weapon to Drown Voices of Protest. In: Proceedings of the 3rd USENIX Workshop on Free and Open Communications on the Internet. Washington; 2013. https://www.usenix.org/conference/foci13/technical-sessions/papers/verkamp. Accessed Oct 2017.
79. Allcott H, Gentzkow M. Social Media and Fake News in the 2016 Election. J Econ Perspect. 2017;31(2):211–36.
80. FIDH. Turkey: Provisional release of human rights lawyer Mr. Levent Piskin. 2016. https://www.fidh.org/en/issues/human-rights-defenders/turkey-provisional-release-of-human-rights-lawyer-mr-levent-piskin. Accessed Oct 2017.
81. Stockholm Center for Freedom. Demirtaş's lawyer accused of joining HDP's WhatsApp group –. 2017. https://stockholmcf.org/demirtass-lawyer-accused-of-joining-hdps-whatsapp-group/. Accessed 02 Oct 2017.
82. Hatti W. Attorney Levent Pişkin is being charged for meeting his client Selahattin Demirtaş. 2017. https://washingtonhatti.com/2017/04/11/attorney-levent-piskin-is-being-charged-for-meeting-his-client-selahattin-demirtas/. Accessed 02 Oct 2017.
83. Twitter. Twitter Help Center: Country withheld content. https://support.twitter.com/articles/20169222. Accessed 17 Oct 2017.
84. Shevchenko V. Ukrainians petition Facebook against 'Russian trolls'. 2015. http://www.bbc.com/news/world-europe-32720965. Accessed June 2016.
85. Pizzi M. Isolated in Camp, Syrians Desperate to Get Online. 2015. http://america.aljazeera.com/articles/2015/7/16/internet-access-zaatari-camp.html. Accessed June 2016.
86. Richtel M. Egypt Cuts Off Most Internet and Cellphone Service. 2011. http://www.nytimes.com/2011/01/29/technology/internet/29cutoff.html. Accessed June 2016.
87. Chulov M. Syria shuts off internet access across the country. 2012. https://www.theguardian.com/world/2012/nov/29/syria-blocks-internet. Accessed June 2016.
88. Conditt J. Turkey shuts off internet service in 11 Kurdish cities. 2016. https://www.engadget.com/2016/10/27/turkey-internet-shutdown-kurdish-cities/. Accessed June 2016.
89. Parks L, Goodwin H, Han L. "I Have the Government in My Pocket": Social Media Users in Turkey, Transmit-Trap Dynamics, and Struggles Over Internet Freedom; 2017.
90. Google. Android Developers Dashboard. https://developer.android.com/about/dashboards/index.html. Accessed 04 Oct 2017.
91. Rivest R, Shamir A, Tauman Y. How to leak a secret. Advances in Cryptology—ASIACRYPT 2001; 2001, pp. 552–565.
92. Chaum D, Van Heyst E. Group signatures. In: Advances in Cryptology—EUROCRYPT'91. Springer; 1991. p. 257–265.
93. Zetetic. SQLCipher. https://www.zetetic.net/sqlcipher/sqlcipher-for-android/. Accessed 13 May 2017.
94. Apple. iOS Security; 2017. https://www.apple.com/business/docs/iOS_Security_Guide.pdf. Accessed Oct 2017.
95. Google. Language and Locale. https://developer.android.com/guide/topics/resources/multilingual-support.html. Accessed 25 Oct 2017.