STATE-SPONSORED TROLLING

How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns

INSTITUTE FOR THE FUTURE

DIGITAL INTELLIGENCE FUTURES LAB

ABOUT THE AUTHORS

Carly Nyst is a human rights lawyer, privacy and data protection expert, and independent consultant working on technology and human rights. She was previously the legal director of Privacy International, a London-based charity that defends the right to privacy across the world.

Nick Monaco is a research affiliate at the Digital Intelligence Lab at the Institute for the Future and at the Computational Propaganda Project at the Oxford Internet Institute, University of Oxford. His expertise spans the political use of social media bots, online disinformation, foreign affairs, and linguistics.

EDITOR IN CHIEF

Samuel C. Woolley is the current director of the IFTF Digital Intelligence Lab and the former director of research of the Computational Propaganda Project at the University of Oxford. His work explores how socially-oriented online automation tools (bots, algorithms, etc.) are used to enable both democracy and civic control.

ABOUT INSTITUTE FOR THE FUTURE

Institute for the Future (IFTF) is celebrating its 50th anniversary as the world's leading non-profit strategic futures organization. The core of our work is identifying emerging discontinuities that will transform global society and the global marketplace. We provide organizations with insights into business strategy, design process, innovation, and social dilemmas. Our research spans a broad territory of deeply transformative trends, from health and health care to technology, the workplace, and human identity. IFTF strives to comply with fair-use standards and publish only materials in the public domain under the Creative Commons 4.0 International License (CC BY-NC-ND 4.0). IFTF is based in Palo Alto, California. For more, visit www.iftf.org.

ABOUT THE IFTF DIGITAL INTELLIGENCE LAB

The Digital Intelligence Laboratory at Institute for the Future is a social scientific research entity conducting work on the most pressing issues at the intersection of technology and society. We examine how new technologies and media can be used to both benefit and challenge democratic communication.

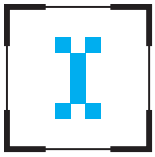
SPECIAL ACKNOWLEDGMENTS

This work is based upon a previous collective effort involving a number of people who have been active in developing this field of research and directed by Camille Francois. The original methodology for this research was developed by Francois, Javier Luque, Ellery Biddle, and Ivan Sigal, with additional input from a number of Global Voices and International Press Institute contributors. Much of the original research for the paper was also contributed by Marianne Diaz, Gülsin Harman, Dağhan Irak, Simin Kagar, and other affiliates of Global Voices and IPI. We would also like to thank Maria Ressa and Rappler for their help and input.

INSTITUTE FOR THE FUTURE
201 Hamilton Avenue
Palo Alto, CA 94301
www.iftf.org

Table of Contents

Executive Summary



In this paper, we examine the emergence of a new phenomenon: state-sponsored trolling. We define this phenomenon as the use by states of targeted online hate and harassment campaigns to intimidate and silence individuals critical of the state. There is evidence that governments around the world, leveraging the surveillance and hacking possibilities afforded by a new era of pervasive technology, are using new digital tactics to persecute perceived opponents at scale. These campaigns can take on the scale and speed of the modern internet with pinpoint personalization from troves of personal data afforded by cheap surveillance technologies and data brokers.

Though state-sponsored trolling occurs in a variety of countries and polities, several commonalities are evident, especially in the strategies and tactics used to carry out attacks. These include but are not limited to making death and rape threats, using bots and automated agents to amplify vitriolic attacks at scale, making accusations of treason or collusion with foreign intelligence agencies, using "black" public relations firms to disseminate hyperpartisan or libelous disinformation about targets, spreading doctored images and memes, unlawfully using spyware and hacking to gather actionable intelligence against targets, and sowing acrimonious sexism. The goal of these attacks is the intimidation and silencing of targeted individuals—most often journalists, activists, human rights defenders, and vocal members of opposition coalitions.

Starting in late 2015, we spent more than eighteen months examining this phenomenon. During this time, we partnered with prominent thinkers and nongovernmental organizations (NGOs), including the Vienna-based International Press Institute (IPI), which advocates for press freedom worldwide, and Global Voices, a citizen-journalism outfit. We conducted a thorough literature review, interviewed targets of state-sponsored trolling, and conducted quantitative analyses of attacks where possible. This paper is the result of these efforts.

Here we explore state-sponsored trolling in multiple sections:

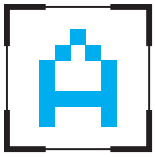
- First, we present new research on the phenomenon from around the world.
- Next, we offer a new framework for attributing attacks in cyberspace, inspired by the Atlantic Council's Cyber Attribution Framework, conceptualizing attacks as fitting into one of four categories: state-executed, state-directed, state-incited, or state-leveraged (Healey 2012).
- We then examine in depth particularly illustrative campaigns in seven countries: Azerbaijan, Bahrain, Ecuador, the Philippines, Turkey, the United States, and Venezuela.
- Finally, we offer a series of recommendations for solving this problem from a policy standpoint, which we hope will inform future conversations and solutions. Some of the policies we suggest are based in law and can be implemented by states, while others are business practices that can be implemented by technology companies.

Changes in law are unlikely to effectively stem the practice of state-sponsored trolling in the short term. As a result, technology companies bear not only the shared responsibility but also the sole ability to curb the practice and effects of state-sponsored harassment campaigns. Here are our policy recommendations, in brief:

- Under international human rights law, require social media platforms to detect and, in some cases, remove hate speech, harassment, and disinformation; and implement such requirements in a transparent and accountable manner that respects due process and reinforces human rights.
- Under US law, adapt the First Amendment, perhaps by building upon existing hate speech prohibitions that are permitted by the First Amendment, such as the federal cyberstalking statute (18 USC § 2261A); create exceptions and add possible new regulations to Section 230 of the Communications Decency Act of 1996, a provision that shields social media platforms from legal liability for the actions of third-party users of their services; and amend and evolve electoral regulation.
- Within technology companies, develop business practices to detect and identify state-linked accounts, detect and identify bots, and improve reporting mechanisms and responsiveness.

It is our hope that describing the phenomenon and proposing policy solutions constitutes an important first step in remedying what we see as a new form of human rights abuse.

Introduction



As a former congresswoman and the daughter of a former president, Martha Roldós was familiar with the reputational affronts and underhanded tactics that accompany political ascendancy in Ecuador. But the attack upon the investigative journalist that began in January 2014 was like nothing Roldós had previously encountered. Its vehicle was the publication by a state newspaper, *El Telégrafo*, of private emails between Roldós and the US National Endowment for Democracy concerning potential philanthropic funding for Roldós's investigative journalism outfit. The newspaper claimed that Roldós was effectively an agent of the CIA, with the aspiration of overthrowing democratic governments in the region. Heavily laden with historical import, the sensationalist claims of collusion with American intelligence were an archetypal example of Latin American disinformation.

The article was just the opening salvo of the attack on Roldós. Following its publication, Roldós was immediately besieged by a tidal wave of tweets and messages, including memes and disfigured representations, claiming not only that Roldós was an American agent but also that she had been involved in the alleged assassination of her own parents. This online trolling campaign was accompanied by an offline one, in which the Radio Pública and government television channels reinforced the veracity of the false claims against Roldós. A week after the original publication, in his weekly television address, Ecuadorian president Rafael Correa congratulated *El Telégrafo* on its publication of Roldós's correspondence (and by implication validated the newspaper's illegal acquisition of private communications) and repeated the newspaper's claims (Presidencia de la República del Ecuador ©SECOM 2014).

The attack on Roldós was not simply an instance of disinformation amplified through digital platforms. Categorizing the onslaught in this way understates its significance. Rather, Roldós's experience is better understood as a state-sponsored trolling campaign against an outspoken critic of the Ecuadorian government. The publication of false claims against Roldós acted as a trigger for a sustained and coordinated government-backed operation against her. Such operations were later explicitly avowed by then-president Correa, who, speaking generally of his intention to deploy trolls in response to criticism and dissent, declared: "People cannot insult or defame in the name of freedom of expression . . . if they send out a tweet, we will send 10,000 tweets calling you a coward" (BBC News 2015).

As the Ecuadorian experience illustrates, disinformation is often only one element of a broader politically motivated attack on the credibility and courage of dissenting voices: journalists, opposition politicians, and activists. While disinformation may exploit inherent characteristics of digital infrastructures, emerging as a unique and perverse by-product of the business models of major digital platforms, it is also a phenomenon that can be exploited. As this paper shows, in many instances disinformation is a tool deployed by governments as part of state-sponsored digital campaigns levied at government critics, campaigns that use disinformation within a sustained, coordinated effort to harass and silence critics. These campaigns mobilize ordinary internet users as well as amateur and professional "cyber militia" to defend state interests, using disinformation in tandem with online harassment. Such attacks appear organic by design, both to exacerbate their intimidation effects on the target and to distance the attack from state responsibility. However, in the cases we studied, attributing trolling attacks to states is not only possible, it is also critical to understanding and reducing the harmful effects of this trend on democratic institutions.

About the Research

We examined the phenomenon of state-sponsored trolling for more than eighteen months beginning in late 2015. During this time, we partnered with prominent thinkers and NGOs, including the Vienna-based International Press Institute (IPI), which advocates for press freedom worldwide, and Global Voices, a citizen journalism outfit. As many of the campaigns we examined happened before we began investigating the phenomenon, we were not always able to acquire data or perform quantitative analyses on the campaigns we studied.

We did, however, conduct formal interviews with many subjects. In addition, we conducted a thorough literature review of the topic throughout the coverage period. The goals of our investigation were (1) to rigorously define a new form of human rights abuse—state-sponsored trolling—with an eye to helping citizens, civil society, private-sector entities, and governments to identify these campaigns in the wild, (2) to develop a new framework for describing state-sponsored trolling attacks with an eye to holding responsible parties responsible, even in cases when attribution is not straightforward, and (3) to begin a conversation about what types of strategies—including but not limited to public and private regulation—may be useful in combatting this issue.

About This Paper

This paper begins, in "How State Information-Control Practices Have Shifted Over Time," by surveying the landscape of state control of information, which has provided fertile breeding ground for the development of trolling as a state tool for suppression of dissenting ideas. We observe the tactical move by states from an ideology of information scarcity to one of information abundance, which sees "speech itself as a censorial weapon" (Wu 2017). This era of information abundance has enabled states to sponsor and execute trolling attacks using

ordinary internet users as well as volunteer, amateur, and professional trolling institutions. Then, in "The Anatomy of State-Sponsored Trolling," we isolate the features of state-sponsored trolling campaigns: language, tools, and tactics.

We next argue, in "Mechanisms of State Responsibility," that attribution is critical to elucidating remedies to state-sponsored trolling. It is often the arm's-length character of state-sponsored trolling attacks—their purported organic nature and seeming distance from state control—that enhances their impact on the targets. Accordingly, we elaborate a framework for conceptualizing state responsibility that seeks to establish that whatever the mechanism of state involvement in the trolling attack (executing it, directing it, inciting it, or leveraging it for state aims), states bear responsibility for the human rights impacts of these trolling campaigns.

We present more than fifteen case studies across seven countries that illustrate how and where states are deploying such attacks. In analyzing instances of state-sponsored trolling in Azerbaijan, Bahrain, Ecuador, the Philippines, Turkey, Venezuela, and the United States, we establish the existence of a broader trend within which national variations occur. We conclude, in "Developing Policy Interventions," by offering some preliminary proposals for policies that can be enacted by states in the long term and by technology companies in the shorter term. We hope that this paper will prompt a further debate about effective and necessary interventions.

Special Terms Used in This Paper

In this paper, we use several terms that must be explicitly defined at the outset to avoid any confusion or ambiguity.

- **State-sponsored trolling:** The use of targeted online hate and harassment campaigns to intimidate and silence individuals critical of the state.

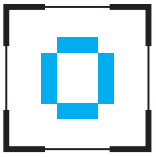
- **Troll:** An online account (operated by an individual or a bot) that deliberately targets an individual with messages of hate and harassment.

- **Disinformation/misinformation:** For this word pair, we defer to the Data & Society Research Institute's definitions in its report "Lexicon of Lies" (Jack 2017) and use them correspondingly. *Disinformation* denotes information that is deliberately false or misleading, while *misinformation* is information whose inaccuracy is unintentional. We also agree with the authors of that report that "the intentions behind any given piece of media content are rarely clear." We predominantly use the term *disinformation* in this report, as we believe it is the intent of states discussed in this report to deliberately smear targets of state-sponsored trolling.

- **Black PR firms:** PR firms that deliberately engage in disinformation and/or harassment campaigns against perceived opponents of a regime, whether with explicit instruction from or tacit approval of the governments they work for. Examples of such firms are discussed in the Bahrain section of this paper. Ong and Cabañes (2018) also discuss the role PR firms play in disinformation and political messaging in the Philippines. The role of such firms in Africa has also been well documented (Newman 2011; G. York 2012).


store id
product id
(DIM_DATE)
id
year = 1997
(FACT SALES)

How State Information-Control Practices Have Shifted Over Time



Others have sought to understand the current digital landscape and the phenomena to which it has given birth (such as disinformation) from the perspectives of the media, technology, or citizenry, but viewing it through the lens of state and political control enables unique insights. Understanding how states have sought to control and harness the information revolution catalyzed by the advent of the public internet and digital technologies from the 1990s allows for a more complete explanation of the current information environment and its corollaries, in particular state-sponsored trolling campaigns.

Throughout history, the powerful have sought to manipulate and control information in order to mold public opinion, garner support, and isolate and discredit outlying ideologies and their proponents. The Soviet Union's Cold War disinformation tactics were an extension of the Catholic Church's seventeenth-century efforts to propagate its ideology, which gave birth to the term *propaganda*. Although the modern practice of propaganda is more comfortably attributed to authoritarian regimes in North Korea or nonstate actors such as ISIS, democratic states equally seek to shape public discourse through the promotion of ideologies that reinforce entrenched power structures; the liberal interventionist narrative practiced by the early-millennium US and UK governments in support of the Iraq War is a clear example thereof. The wartime dissemination of information by allied troops in Iraq and Afghanistan, including through air-dropped leaflets, also illustrates that propaganda is a tool of democratic states just as much as undemocratic states (Shanker and Schmitt 2003).

The advent of the internet posed unparalleled challenges to the state pursuit of information control. The very nature of information—its velocity, volume, and diversity—changed dramatically, demanding new forms of information control. In the three subsequent decades, we have witnessed two generations of state information-control practices.

The pursuit of information scarcity was the first iteration of information control in a digitally connected world. States adopted offensive approaches to restrict access not only to certain information online but also, in some cases, to the internet itself (Goldsmith and Wu 2006). Examples of the information scarcity approach abounded in the early 2000s, with India blocking Yahoo! Groups; Middle Eastern countries such as Bahrain, the United Arab Emirates, Qatar, Oman, Saudi Arabia, Kuwait, Yemen, Sudan, and Tunisia blocking websites that provided skeptical views of Islam, secular and atheist discourse, and sexual content; and China instituting its famous "Great Firewall" (BBC News 2010; Noman and York 2011; Orlowski 2003). In parallel, states pursued the adoption of broadly drawn cybercrime laws designed to prevent the dissemination of certain content and advocated for the adoption of filters designed to block obscene material, in some cases extending regulation in the online environment beyond that applicable offline.

Although such practices continue in many countries, democratic and otherwise, the past decade has seen the emergence of a different state mentality vis-à-vis the internet: that of information abundance. States have shifted from seeking to curtail online activity to attempting to profit from it, motivated by a realization that the data individuals create and disseminate online itself constitutes information translatable into power. The proliferation of the commercial surveillance-technology industry has enabled even the poorest governments to equip themselves with the technical capabilities to monitor their citizens, revealing new and more effective possibilities for state control (Deibert 2013; Granick 2017). At the same time, states have realized that the internet offers new and innovative opportunities for propaganda dissemination that, if successful, obviate the need for censorship. This approach is one of "speech itself as a censorial weapon" (Wu 2017).

Governments today are increasingly in the business of information generation. Equipped with an expanding digital insight into individuals' online behavior, states are seizing upon declining public trust in traditional media outlets and the proliferation of new media sources and platforms to control information in new ways. States are using the same tools they once perceived as a threat to deploy information technology as a means for power consolidation and social control, fueling disinformation operations and disseminating government propaganda at a greater scale than ever before (Weedon, Nuland, and Stamos 2017).

The new digital political landscape is one in which the state itself sows seeds of distrust in the media, fertilizes conspiracy theories and untruths, and harvests the resulting disinformation to serve its own ends (Ball 2017; Marwick and Lewis 2017). Those ends chiefly include straightforwardly political ones: Freedom House reports that online disinformation tactics

have been deployed in elections in eighteen countries over the past year, with states deploying digital tools to fabricate grassroots support for government policies, "a closed loop in which the regime essentially endorses itself, leaving independent groups and ordinary citizens on the outside." But these tools are also being deployed in pursuit of societal and cultural objectives. States are not only advancing their own agenda but also silencing the agendas of others, particularly those belonging to progressive or liberal causes.

It is out of this landscape that state-sponsored trolling campaigns have emerged. Governments have sought to deploy tools of digital repression to silence critical voices altogether, rather than to merely observe and contribute to online environments in which conspiracy theories, disinformation, hostility, and incivility marginalize such voices. In this incarnation of the information abundance strategy, states harness online hate mobs to harass, intimidate, and discredit journalists, activists, and academics perceived to be a threat to state power. The approach is uniquely designed to take advantage of the current digital ecosystem, leveraging the virality and familiarity of social media to amplify state messaging, and deploying bots, hashtags, and memes to disguise industrial campaigns as organic groundswells.

State-sponsored trolling combines several problems that digital rights circles have been viewing in isolation for years—cyberattacks, hacking, invasion of privacy, computational propaganda, disinformation, political bots, and the like—into a larger phenomenon that is in a class by itself.


function
previousElements
EL NOTICIERO

The Anatomy of State-Sponsored Trolling

Existing analyses of the phenomenon of state-sponsored trolling tend to take a one-dimensional view of what Freedom House calls "online content manipulation," which sees disinformation and harassment campaigns tied together in an untidy knot that technology companies, states, and individual citizens all bear the burden of untangling. We surmise, however, that a distinct set of campaigns rises beyond general exploitation of digital infrastructures to the level of state-sponsored attack. Others have used the term "patriotic trolling" to refer to these campaigns, in order to capture the shape of such campaigns, which often obscure, by design, the state's role therein (ABS-CBN News 2018; Geybulla 2016). The term mirrors that used to describe the state hacking campaigns carried out under the guise of independent hackers in an effort to mask the provenance of the attacks (Deibert and Rohozinski 2010).

In our analysis, these state-sponsored trolling attacks share common features and anatomy, despite occurring in vastly different countries and cultural contexts. Below, we describe these features, drawing on the examples of more than seventeen cases studied by the authors over the course of eighteen months beginning in late 2015.

Critics in the Crosshairs

State-sponsored trolling attacks can first be identified by their targets and the actions that trigger them. Journalists, activists, and others who criticize the government, government affiliates, or status quo institutions are the prime targets of states using digital platforms and tools. Journalists Marc Owen Jones, Martha Roldós, Arzu Geybulla, and David French have all been subjected to trolling campaigns connected with the Bahraini, Ecuadorian, Azerbaijani, and American governments respectively. Media figures are also often the targets of campaigns waged by the Turkish government. Often, the media figures subjected to state-sponsored harassment are those reporting on the use of state-sponsored trolling itself: this was the case for Maria Ressa, founder of Filipino media outlet Rappler, who became the victim of state-sponsored trolling after reporting on the government's misuse of social media (Etter 2017). Human rights defenders and activists, such as Bahraini activist Maryam Al-Khawaja, are also targeted by state-sponsored trolls.

The Language of Trolls

Although state-sponsored trolling attacks represent an innovative manipulation of new technologies in pursuit of old aims, they largely fall back on well-established messaging tactics to seed distrust in mainstream media and turn public opinion against journalists and activists. These include:

» **Accusations of collusion with foreign intelligence agencies.** Martha Roldós was accused of CIA affiliation, while Azeri journalist Arzu Geybulla was called an Armenian spy. Bahraini activist Maryam Al-Khawaja and her family were labeled as terrorists and Iranian agents by government spokesmen, and Selin Girit was called an English agent by Turkish trolls.

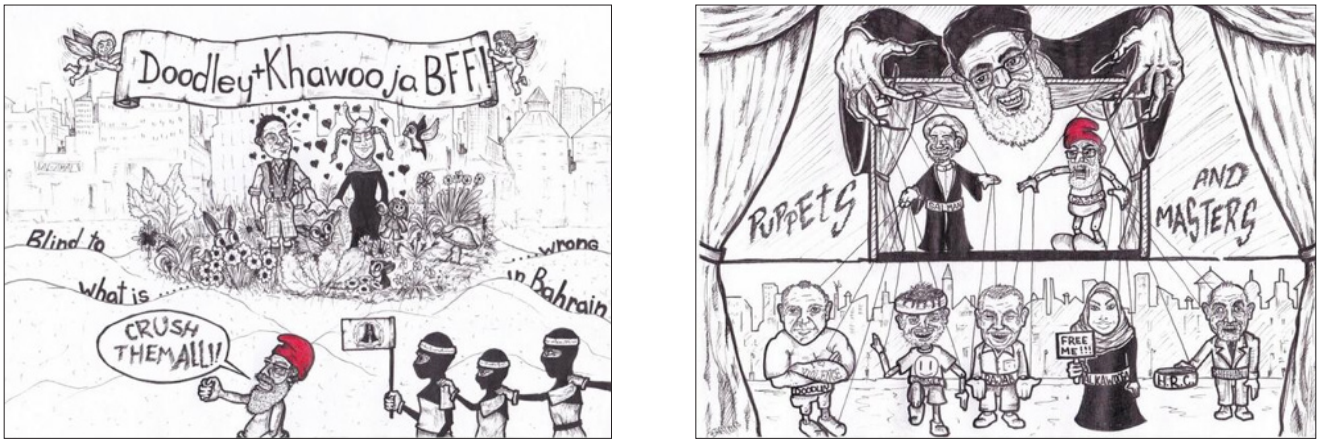


Figure 1. Examples of cartoons used in state-sponsored trolling attacks on Maryam Al-Khawaja and Brian Dooley in Bahrain.
Photos courtesy of Dr. Marc Owen Jones.

» **Accusations of treason.** Venezuelan trolls labeled businessman Lorenzo Mendoza a traitor who was leading an economic war against the country. Government-backed bloggers in the Philippines attempted to trend #ArrestMariaRessa on Twitter after Rappler published a transcript of the first phone conversation between US president Donald Trump and Philippines president Rodrigo Duterte (Posetti 2017). The campaign mirrored that previously waged against Senator Leila de Lima, recognized by Amnesty International as a "human rights defender under threat," who was ultimately arrested after an online campaign urging #ArrestLeiladeLima (Etter 2017).

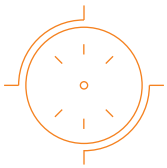
» **Use of violent hate speech as a means of overwhelming and intimidating targets.** Every female target of government-backed harassment receives rape threats and is subjected to sexist and misogynistic language. Turkish journalist Ceyda Karan received explicit rape threats. Filipino journalist Maria Ressa received, on average, ninety hate messages an hour during one attack, including a call for her to be raped repeatedly until she died.

» **Creation of elaborate cartoons and memes.** Those used in attacks on Maryam Al-Khawaja and Brian Dooley in Bahrain are shown in Figure 1. This is a pattern seen in nearly all cases and across all countries.

Finally, in an interesting illustration of the high degree of manipulation embodied by state-sponsored attacks, trolls often accuse targets of the very behaviors the state is engaging in. In numerous countries, for example, trolls make claims that targets are affiliated with Nazism or fascist elements. Politicians and their proxies use claims of "fake news" as a form of dog whistling to state-sponsored trolls, which claims are then repeated and amplified by supporters.

Bots and Algorithms

Demonstrating a savvy appropriation of emerging technical tools, state-sponsored trolling campaigns have used political bots and gamed algorithms to amplify the effect of attacks. Bots, which serve not only to amplify attacks but also to change their character, making a campaign seem more organic and widespread, have come to feature heavily in state-sponsored trolling attacks and are broadly deployed by political parties and movements to attack or drown out critics, boost follower numbers, and magnify the messages of political candidates (Confessore et al. 2018; Howard and Woolley 2016). In Mexico, political bots were so commonly deployed by President Enrique Peña Nieto's government that they were labeled Peñabots. Indeed, they were part of the campaign against journalist Martha Roldós. Bots also feature in campaigns in Turkey, where at least eighteen thousand bot accounts tweet in favor of President Recep Erdoğan (Poyrazlar 2014).



Trolls appropriate and game the algorithms of social media sites in order to increase the prominence and pervasiveness of their messaging. Gaming of algorithms is the deliberate exploitation of a platform's underlying automated process to achieve an outcome not intended by the platform. For example, trolls will flag legitimate social media accounts as fake accounts in order to have targets' accounts temporarily suspended until they can prove their identity. In one form of algorithm gaming, trolls hijack hashtags in order to drown out legitimate expression. For example, trolls have co-opted hashtags at events where Maryam Al-Khawaja was speaking. This most notably happened at the Oslo Freedom Forum in Norway (Halvorssen 2011). This tactic was also used against Arzu Geybulla when she spoke at an Organization for Security and Cooperation in Europe (OSCE) event in Warsaw (Geybulla 2016).

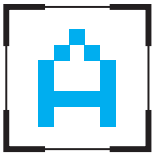
Election Antecedents

The infrastructure and mechanisms for state-sponsored trolling attacks in numerous countries have grown out of, or been built upon, infrastructure and mechanisms established during election campaigns. Candidates and parties develop resources such as databases of supporters, committed campaign volunteers, social-media-influencing arms, and dedicated communications channels that are deployed during elections to advance a party's platform and undermine the opposition. Once a candidate or party is successful, these same resources are often deployed in pursuit of consolidating and extending power; like a muscle that has been trained to perform a particular task, once in government politicians continue to campaign, using the same aggressive and often harmful tactics.

We have observed this pattern in a number of countries, chief among them the Philippines and Ecuador. In Ecuador, former president Rafael Correa's 2012 reelection campaign saw the candidate's first foray into social media manipulation, with the campaign establishing a dedicated email address and communication channel to communicate to supporters how to amplify campaign messages on social media platforms. The "Correístas" email list was part of a "social media manipulation plan" devised by a private public relations firm contracted by Correa, Inteligencia Emocional. Leaked Inteligencia Emocional documents establish that Correa intended to use social media supporters to propagate positive media items and target those spreading undermining messages (Ecuador Transparente 2016). After his reelection, Correa continued to use Correístas, along with another social media channel, Somos+. In announcing the channel, the president indicated his intention to respond at scale to online dissent, saying: "People cannot insult or defame in the name of freedom of expression . . . if they send out a tweet, we will send 10,000 tweets calling you a coward" (BBC News 2015).

Filipino president Rodrigo Duterte himself admitted to paying trolls during his election campaign, though he denies having used them while in office (Ranada 2017b). However, analysis conducted by Filipino media outlet Rappler demonstrates that of twenty-six troll accounts key to Duterte's election campaign, many have remained active during his presidency; twelve million internet users have been co-opted into amplifying pro-Duterte trolling campaigns as a result (Etter 2017). Indeed, the Duterte government has even elevated bloggers and social media influencers acting as trolls to positions within the government.

Mechanisms of State Responsibility



Attributing responsibility for actions that occur in the online realm is at best imperfect and at worst impossible. As the Atlantic Council has recognized, because "the Internet enables anonymity more than security," policy makers struggle to determine the source of cyberattacks, and analysts fall into the trap of "attribution fixation" (Healy, 2012). This attribution problem is exacerbated in the context of political harassment campaigns that take place primarily on social media platforms: such campaigns are designed to appear spontaneous and organic, camouflaged by the chaotic ephemera. Because of this, even identifying the occurrence of a state-sponsored trolling attack is a challenge, let alone isolating its origin and attributing responsibility for it to a particular actor.

We agree with the Atlantic Council that for the purpose of policy making, the question of who did it should be trumped by the question of who is to blame. In that regard, we prefer to categorize state-sponsored trolling attacks along a spectrum of state responsibility. We see four often-overlapping mechanisms by which governments become responsible for online harassment campaigns.

Category 1: State-Executed

In many contexts, harassment campaigns against critics and dissenters originate directly from state apparatuses. State-funded and -directed "cyber militia" execute strategies designed by the government to disseminate propaganda, isolate dissenting views, and drown out or remove anti-government sentiment.

We see three broad forms of cyber militia being deployed by governments: volunteer, amateur, and professional. Most commonly, governments use volunteers to undertake social media messaging and campaigns in exchange for social capital and the protection of government allegiance. For example, in Azerbaijan, party-affiliated and government-funded youth groups act as a front for state-sponsored trolling initiatives. Ireli ("Forward"), one such organization, aims to "produce young people who can take an active part in the information war," and volunteer youth group participants seek a form of "quantitative success" from participating in trolling and propaganda dissemination in the belief that posting a large amount of content will increase the likelihood of advancing into government positions (Geybulla 2016; News.Az 2011).

While it can be tempting to dismiss the influence of pro-government youth groups in online trolling, it is important to note they have been widely used in China, Russia, and Turkey, states that are increasingly regarded as dark paragons of disinformation (Henochowicz 2015). Experts have noted the importance of Russian youth groups such as Nashi ("Ours") in carrying out state-sponsored or -encouraged cyberattacks and trolling campaigns. This was notably the case when a member of Nashi confessed to the *Financial Times* that it had carried out the 2007 cyberattacks on Estonia (Clover 2009). Other Russian conflicts with neighboring states in the 2000s, notably Lithuania and Georgia, were accompanied by similar cyberattacks (Soldatov and Borogan 2015).

Journalist Noah Shachtman has noted the utility of appearing to keep these youth groups' cyberaffairs at arm's length: "Part of the ingenuity of using Nashi as cyberwarfare arm is the group's nominally independent status: while the group does the Kremlin's bidding, its funding comes from pro-business owners looking to ingratiate themselves with the regime. Even if they claim credit for the attacks, they are still one level removed from the Russian government—however implausible that seems" (Shachtman 2009).

Similarly, the Turkish government maintains a volunteer group of six thousand "social media representatives" spread across Turkey who receive training in Ankara in order to promote party perspectives and monitor online discussion (Albayrak and Parkinson 2013). Filipino president Rodrigo Duterte groomed a cyber militia of around five hundred volunteers during his election campaign, eventually promoting key volunteers to government jobs after his election.

Some countries provide remuneration to their cyber militia, although members are still drawn from the general public; in China, for example, members of the "50 Cent Army" are paid nominal sums to engage in nationalistic propaganda (King, Pan, and Roberts 2016). India's Bharatiya Janata Party (BJP) established its own "information technology" cell. The BJP IT cell, a mix of volunteer and paid amateur trolls, tasks members daily with a messaging task and maintains a "hit list" of mainstream journalists who must be attacked (Chaturvedi 2016).

In countries such as Russia, the practice of state-sponsored trolling has been professionalized, with "troll farms" operating in a corporatized manner to support government social media campaigns. In Russia, the most well-known troll farm is the Internet Research Agency (IRA), but there are reportedly scores of such organizations all around the country (Chen 2015; Soldatov and Borogan 2015). One need not look far for links between the IRA and

the youth group Nashi. A former head of Nashi, Aleksei Soskovets, admitted to using Nashi trolling methods when he moved to the IRA, now notorious for its involvement in spreading disinformation during the 2016 presidential election in the United States (Garmazhapova 2014).

Category 2: State-Directed or -Coordinated

In both Ecuador and Venezuela, we see governments directing or coordinating, but not executing, state-sponsored trolling attacks. State-coordinated campaigns involve the use of coordination channels to disseminate signals and messaging to committed supporters and volunteers, and to outsource harassment campaigns to private actors. Venezuela is an example of the former approach; the Venezuelan Ministry of Communications and Information and its dependent office the *Sistema Integrado Bolivariano de Generación de Contenido en Venezuela* (SIBGECOV, the Bolivarian Integrated System of Content Generation in Venezuela) deploy Telegram channels as a central messaging service that instructs participants and subscribers to disseminate certain messages, memes, and hashtags. For example, in the case of a campaign against Lorenzo Mendoza, CEO of Empresas Polar, the *Chavez en Red* Telegram channel directed supporters to troll Mendoza using the hashtag #LorenzoEsEscasez ("Lorenzo is scarcity").

The Ecuadorian government has similarly used social media channels, such as Somos+, to counter what the state cast as a "systematic smear campaign" by users who "abuse the anonymity and freedom that the social networks provide." Ecuador also outsourced social media campaigns to private entities; one investigation revealed that private company Ribeney Sociedad Anonima was awarded a government contract for the operation of a troll center charged with both attacking and monitoring people expressing opposition to Correa online (Morla 2015).

Category 3: State-Incited or -Fueled

Perhaps the most pernicious of state-sponsored trolling campaigns are those in which the government maintains an arm's-length distance from the attack but nevertheless both instigates and profits from it. Such methods rely on the manipulation of internet users' psychology to ignite and sustain a campaign and on the autovirality of online campaigns. Governments use high-profile proxies and other government stand-ins to signal state support for a particular attack, having long ago planted the seed in the minds of citizens that trolling is a method supported, or at least not opposed, by the government.

The strategy of inciting or fueling trolling campaigns has been witnessed in the United States, where hyperpartisan news outlets such as Breitbart—formerly chaired by Steve Bannon, former White House chief strategist under President Trump, and funded by Robert Mercer, Trump's largest donor—and sources

Figure 2. Official Telegram account for Diosdado Cabello's TV show *Con el Mazo Dando*, promoting attacks on Luis Florido with the hashtag #FloridoEresUnPajuo, which alleges Florido has a falsified graduate degree.

Screenshots courtesy of Marianne Diaz.

close to Trump signal to trolls who to target. This was the case with respect to Erick Erickson, who after being called "a major sleaze and buffoon" by Trump on Twitter was the subject of a Breitbart article that triggered an online trolling campaign (Grove, 2016). In Venezuela, former vice president Diosdado Cabello, who currently hosts the TV show *Con el Mazo Dando (Hitting with the Sledgehammer)* on the Venezuelan state-owned TV channel VTV8, used his TV show and a Telegram channel associated with it to encourage Twitter attacks on opposition politician Luis Florido using the hashtag #FloridoEresUnPajuo ("Florido, you're a lying idiot"). Attacks on Florido lasted for days; they were vitriolic and crude and frequently accused him of being a traitor to Venezuela. A screenshot of a government official participating in the attacks is shown in Figure 2.

In Turkey, journalist Ceyda Karan was subjected to a three-day-long trolling campaign in which two high-profile media actors played a key role: pro-Erdoğan journalist Fatih Tezcan, who has more than 560,000 followers, and Bayram Zilan, a self-declared "AKP journalist" with 49,000 followers. Tezcan and Zilan were central players in a campaign that involved 13,723 tweets against Karan sent by 5,800 Twitter users (see Figures 3–7).

Category 4: State-Leveraged or -Endorsed

As state-sponsored trolling attacks become a more familiar and commonplace methodology for silencing online dissent, such attacks are becoming seemingly more remote from state institutions. In perhaps the most cynical manipulation of online behaviors, governments point to the existence of seemingly independent groundswells of public opinion to justify and legitimate state positions. We have seen this tactic employed in China, for example, where the Chinese state pointed to the online abuse of a French journalist to justify a conclusion that the journalist was "hurting the feelings of the Chinese people" and should not have her visa renewed (Phillips 2015; Su 2016). In doing so, it signaled to internet users its tacit approval of harassment campaigns and implicitly promised impunity for state-sponsored trolls.

The Twitter account of Indian prime minister Narendra Modi follows at least twenty-six known troll accounts, and the prime minister has hosted a reception attended by many of the same trolls (Chaturvedi 2016; The Quint 2015). Similarly, Filipino president Rodrigo Duterte has given bloggers active in online harassment campaigns accreditation to cover presidential foreign and local trips (Ranada 2017a).

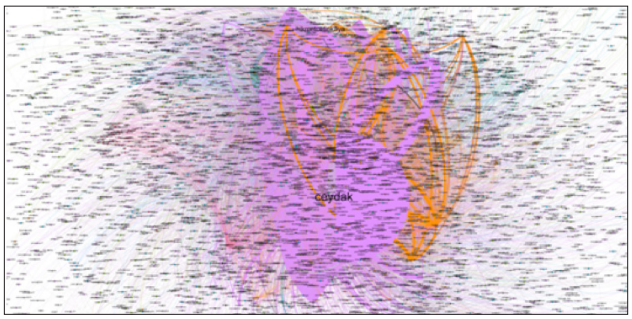


Figure 3. NodeXL representation of the 13,723 interactions among 5,800 Twitter accounts collected in analysis of the state-sponsored trolling campaign against Turkish journalist Ceyda Karan.*

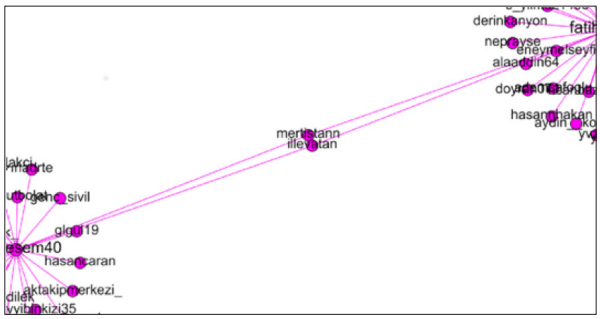


Figure 4. Depiction of "bridges"—users who connect two otherwise unconnected clusters and act as conduits, passing messages from one cluster to another. Bridges are integral to viral spread.*

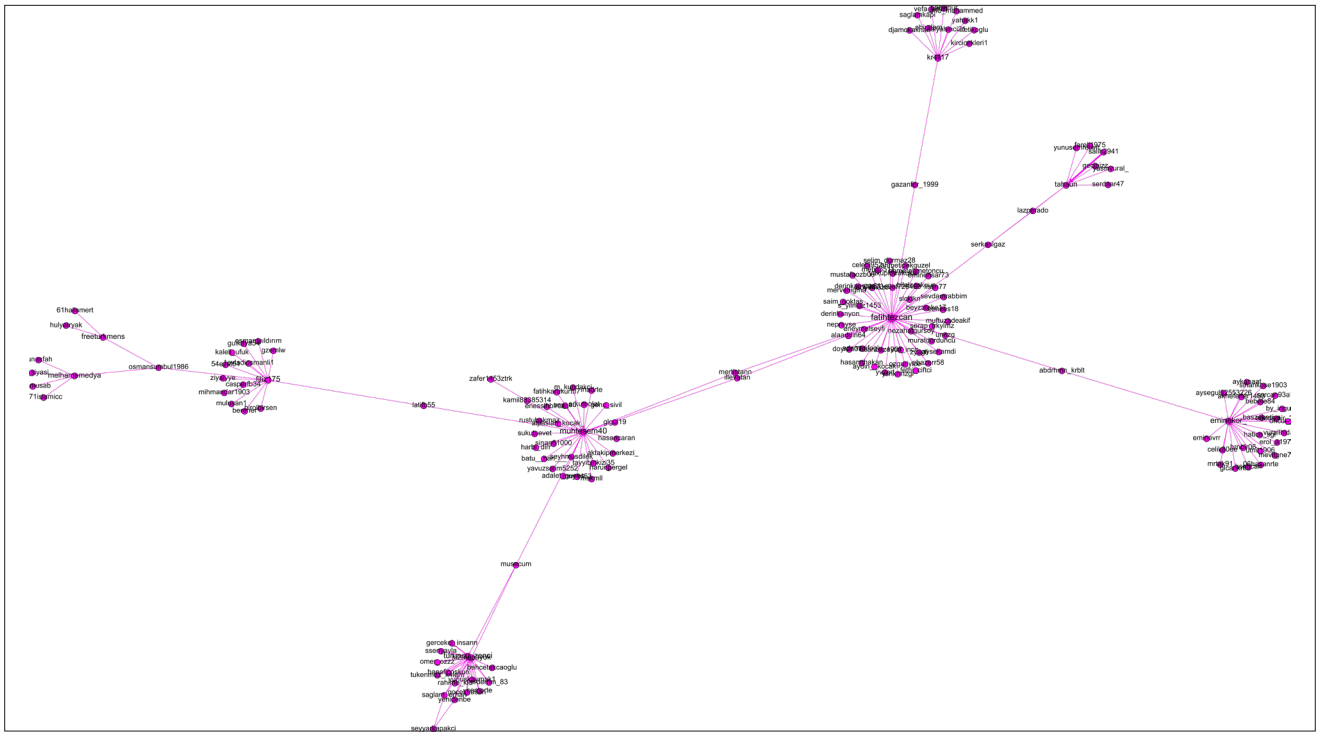


Figure 5. The entire network of eight clusters of users involved in the attack on Ceyda Karan. IPI connected the dots using Gephi.*

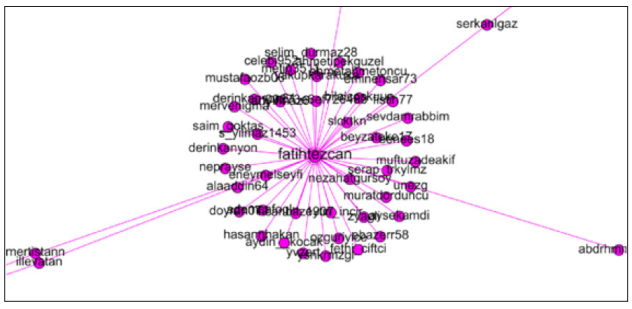


Figure 6. The most influential cluster in distributing the intimidating messages on Twitter, and at the center, Fatih Tezcan, the most influential user.*

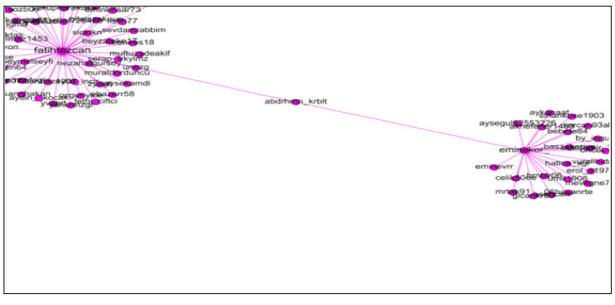
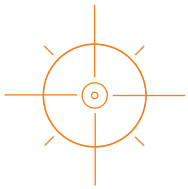


Figure 7. Connection between @fatihtezcan and another influential Twitter account, @eminekor_.*

*All Images: International Press Institute, 2016

Case Studies

Following are country-specific case studies of state-sponsored trolling. Several of the cases included here were selected based on both completed and ongoing analysis from the people and organizations central to developing the methodology for this report. These people and organizations are noted in the acknowledgments section at the beginning of this paper. Additional cases were selected by the authors of this report for comparative political reasons or via grounded research on the phenomenon of state-sponsored trolling.



AZERBAIJAN

Azerbaijan is among the most repressive of the post-Soviet countries, ranking in the bottom twenty countries of the Reporters Without Borders 2017 World Press Freedom Index. President Ilham Aliyev's tenure has seen an increase in persecution of journalists and opponents, including a more repressive online sphere and physical persecution offline (Reporters Sans Frontières 2017a). Multiple journalists have been targeted by state-sponsored trolling campaigns in Azerbaijan; researchers such as Katy Pearce have thoroughly documented the strategies used in these campaigns, such as coordinated hashtags and hashtag hijacking (Pearce 2014, 2015).

"I've been called many things; a slut, a dog, a pig—you name it. These insults involved my ill mother and deceased father. She was a whore; he was a traitor who slept with an Armenian slut. I have been publicly shamed for writing columns for *Agos*, a Turkish-Armenian weekly, while living in Istanbul."

"Being a woman is enough. If you're a vocal woman opposed to the authorities, the harassment knows no limits."

—Journalist Arzu Geybulla
on her experience with state-sponsored trolling

The most vitriolic and enduring state-sponsored trolling campaign in Azerbaijan has been against Arzu Geybulla, a journalist who has written for Al-Jazeera, *Foreign Policy*, and the Turkish-Armenian weekly *Agos.* Geybulla has been continually targeted since 2014 (Geybulla 2016; PEN International 2014; Tan 2015). For her independent reporting on Azerbaijan's human rights abuses, Geybulla has been a frequent target of both online and state-media harassment in Azerbaijan. In campaigns against her, Geybulla has received death and rape threats, has been accused of treason and of working as a spy for Armenia and the West, has been the target of elaborate memes and cartoons, and has received threats on her family's safety (Arzu Geybulla, personal communication, 2 February 2018).

It is possible to draw a direct line from attacks on journalists and others to the Azeri government; digital forensic investigations in 2017 revealed that distributed denial of service (DDoS) attacks on independent online media outlets originated from Azerbaijan's Ministry of Transport, Communications, and High Technologies (Qurium Media Foundation 2017). However, more frequently, state-sponsored trolling campaigns are coordinated and conducted by entities at arm's length from the state. In the case of the attacks on Geybulla, pro-government youth groups have been the main propagators of these attacks.

One of the main youth groups involved in initial attacks on Geybulla is Ireli ("Forward"). According to the director of the group, Rauf Mardiyev, Ireli's goals are "education of young people and the protection of Azerbaijan's interests in the virtual world . . . . Our objective is to produce young people who can take an active part in the information war." Mardiyev also

described Ireli's blogs and websites as being "dedicated to Azerbaijani truth" and having a network of 25,000 accounts on Facebook (News.Az 2011). Mardiyev has been particularly vocal about successful state-sponsored trolling campaigns on Facebook, as can be seen in Figure 8.

The Azeri government funded and sponsored Ireli's founding. Experts on the outfit have also noted the substantial government resources that the group received up to mid-2014, as well as the strong presidential ties the group benefited from (Diuk 2012; Nikolayenko 2012).

Researcher Katy Pearce, who is familiar with both Russian and Azeri manipulation of social media, has also noted the similarities between Ireli and the Russian youth group Nashi: "Ireli is not unlike Nashi in Russia . . . as it claims independence from the state but receives a great deal of government support . . . and its members make no effort to mask their support for the ruling regime" (Pearce 2015).

It is likely that state-sponsored trolling attacks like those experienced by Arzu Geybulla are complemented by state surveillance of journalists and other critics. Citizen Lab has confirmed that Hacking Team's remote control system (RCS) spyware has endpoints in Azerbaijan. The research institution suspects the Azeri government is behind the use of this spyware, as Azerbaijan is a known customer of the Italian spyware firm (Marczak et al. 2014). Investigations by experts at Amnesty International also found that several human rights activists and opponents of the regime were victims of spear-phishing and malware attacks (Guarnieri, Franco, and Anderson 2017).

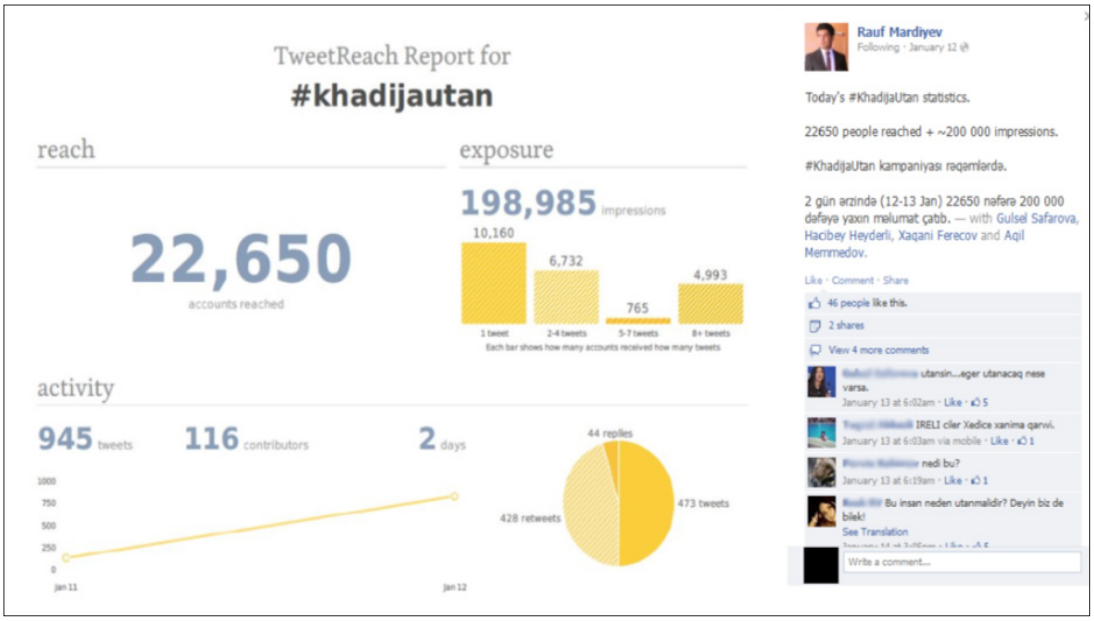
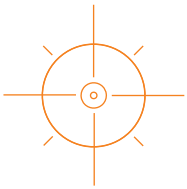


Figure 8. Rauf Mardiyev, former secretary-general of Ireli, vaunting the success of a coordinated hashtag (#khadijautan, meaning "shame on Khadija") against outspoken journalist Khadija Ismayliiova (Pearce 2014).



BAHRAIN

Bahrain has garnered notoriety as one of the most digitally repressive regimes on Earth. With a population of only 1.38 million citizens, a vast amount of oil wealth, and a repressive monarchy, Bahrain has absolute power to surveil and restrict the communications of its citizens. Bahrain's internet penetration rate is among the highest in the world—it was 98 percent in 2016—and this connectivity offers an unparalleled infrastructure for pervasive surveillance and repression (Freedom House 2017c). The government has indirect control of internet service providers within the country through its Telecommunications Regulation Authority (TRA) (Freedom House 2016b).

Several journalists and activists have been targeted with vitriolic state-sponsored trolling campaigns. Journalists Brian Dooley and Nick Kristof and researcher Jillian C. York have documented or suffered political trolling in the country (Dooley 2011; Larsen 2011; J. C. York 2011). There have been numerous reports of journalists and bloggers being jailed or tortured for expressing opinions that run counter to the government. Ali al-Dairi, founder of the news outlet *Bahrain Mirror,* and Ali Abdulemam, a popular activist blogger, both had their citizenship extralegally revoked in early 2015 for their online activities (Abdulemam 2015; Freedom House 2016b).

Prominent human rights activist Maryam Al-Khawaja and academic Marc Owen Jones have suffered state-sponsored trolling campaigns in Bahrain. Jones has highlighted the innovation and uptick in trolling that occurred during and after the Bahraini uprisings that coincided with the Arab Spring in 2011. Particularly troubling is the Hareghum account (@7areghum). (*Hareghum* is an Arabic term meaning "the one that burns them.") This account functioned as a mass identity-revealing and doxing account, predominantly during the 2011 Bahrain uprising (Marc Owen Jones, personal communication, 10 February 2018). It would post photos of Bahrainis at anti-government protests, release their personal details (address, name, family members, phone number, and the like), and call on other users to reveal the identities of other protesters. The account even allegedly advertised a Ministry of the Interior hotline where one could report protesters engaging in anti-government activity directly to the government (Bahrain Independent Commission of Inquiry 2011, p. 391).

The Bahrain Independent Commission of Inquiry (BICI) found that people named by the account would avoid sleeping at home in fear of their safety and detailed its nefarious activities. "Harghum [sic] openly harassed, threatened, and defamed certain individuals, and in some cases placed them in immediate danger. The Commission considers such harassment to be a violation of a person's right to privacy while also amounting to hate speech and incitement to violence" (BICI 2011, p. 401).

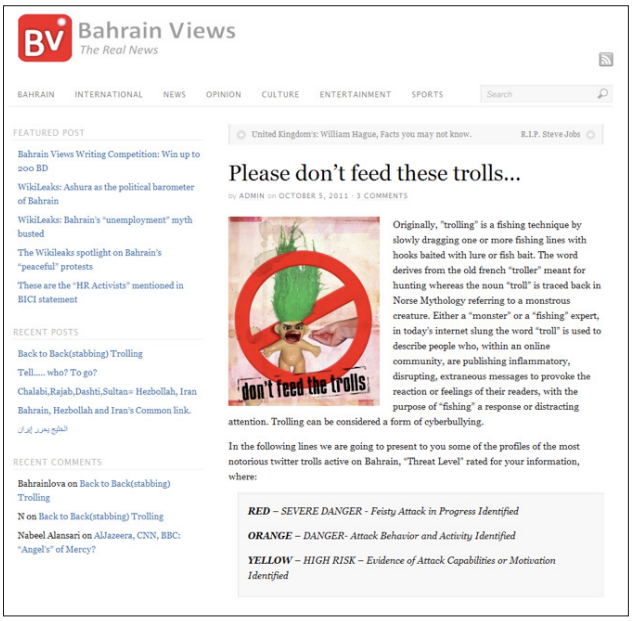
Though the BICI found that this account had violated both Bahraini and international law, the Bahraini government never did anything about it. The account no longer exists, but the chilling effect that it had in a country of only 1.38 million citizens cannot be overstated. It inspired imitation dox trolls that attacked prominent human rights activists such as Maryam Al-Khawaja (Jones 2013). In state-sponsored trolling campaigns, Al-Khawaja was subjected to various forms of harassment, including being targeted with violent rape and death threats, accusations of treason and working for Iran, hashtag hijacking, and in-person heckling at events where she was speaking.

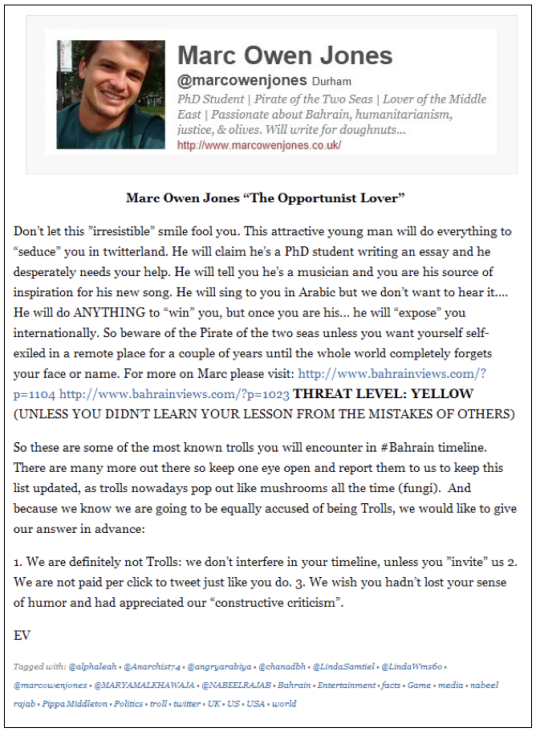
Bahrain has been a hotbed for elaborate digital libel campaigns against targets. Disinformation has been pervasive and multimodal in Bahrain, particularly against targets of state-sponsored trolling campaigns. PR bloggers posing as

journalists on pro-government propaganda blogs such as Bahrain Views and Bahrain Independent have written libelous stories. Liliane Khalil is one such fake journalist who was exposed by Marc Owen Jones: Khalil was revealed to have ties to Task Consultancy, a company that was awarded a PR tender from Bahrain's government in June 2011 (Al Jazeera 2011; Desmukh 2011a; Freedom House 2016b; Jones 2011, 2013).

In the same vein, "hit blogs" have been published accusing political activists of being trolls. The irony here is that many of these accused "trolls" have in fact been targeted with state-sponsored trolling campaigns (shown in Figure 9).

Ample evidence exists of governmental involvement in spyware campaigns against activists and governmental links to accounts calling for violence against protesters on social media, most notably in studies published by Bahrain Watch (Marczak 2013a, 2013b). The outfit's IP Spy Files unveiled 120 accounts





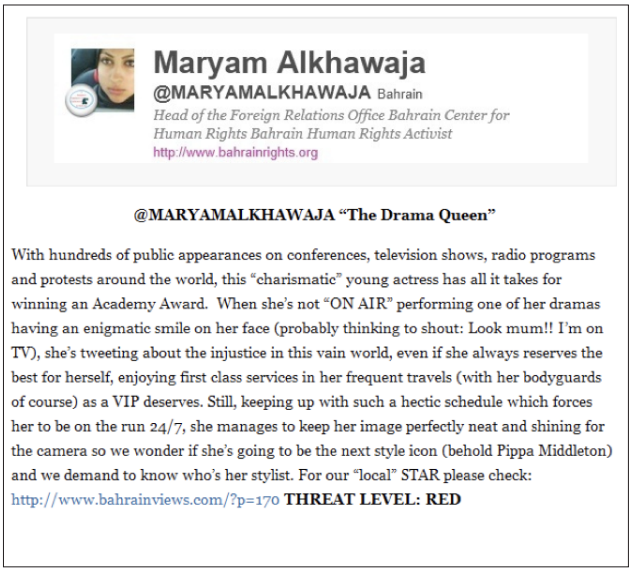


Figure 9. Pro-government propaganda blog Bahrain Views encouraging targeting of prominent activists and journalists as trolls.

Image credit: Mark Owen Jones, 2017

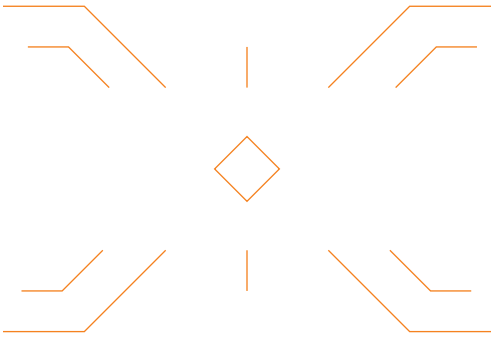
(both pro- and anti-government) that were targeted with phishing links that led back to the government (Bahrain Watch 2013a).

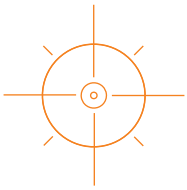
With respect to state-sponsored trolling attacks, Bahrain has deployed distancing tactics, notably using individuals with close ties to the government and black PR firms. Bahraini individuals outside the government with close ties to the regime have also engaged in attacks on the same targets, which we consider another distancing tactic. For instance, Najeb Y Alhamer (@NajebYAlhamer), the chairman of the newspapers *AlAyam* and the *Daily Tribune*, has frequently trolled Maryam Al-Khawaja and her sister online. In addition to having a powerful position as a media mogul, Alhamer is close to the ruling family in Bahrain. The strategy of delegating harassment from governmental officials to hyperpartisan pro-government media officials has also been seen in Turkey.

Black PR firms have also played a notable role in Bahrain. Many of these firms offer "reputation management" or "reputation laundering" services, which can take many forms, including blogs maintained by fake personalities, fake social media accounts, and hyperpartisan blogs and op-eds. The government has spent $32 million in contracts for at least eighteen public relations firms in the United Kingdom and the United States to improve its image domestically and abroad (Bahrain Watch 2013b).

It can be tempting to dismiss the role of reputation laundering as old hat—lobbyists and PR firms have lobbied for countries with abysmal human rights records in the West for decades (Brogan 1993). However, their role in the online era is vastly more insidious. Black PR firms' work for governments can involve libelous attacks on perceived opponents, which in turn can provide fodder for state-sponsored trolling attacks at scale. In the modern era, these campaigns can take on the scale and speed of the modern internet with pinpoint personalization from troves of personal data afforded by cheap surveillance technologies and data brokers.

Olton, a British firm that has marketed itself as "specialis[ing] in the exploitation, collection, collation, and fusion of Open Source Information" is known to have contracted with Bahrain (Desmukh 2011b; Messieh 2011). It is also known that at least one of its employees has contracted with Bahrain's Ministry of the Interior, the office responsible for the country's domestic security apparatus (Jones 2013). We surmise that this trend will continue in the future, with black PR firms' attacks on targets growing ever more invasive and precise, given the availability of cheap surveillance technology, the ease of publishing online, and the ever-increasing pool of data available on individuals. Experts have already highlighted the dangers that will exist in the future with the exploitation of publicly available data (Hu 2016). ▣





ECUADOR

Freedom of the press saw a precipitous decline in Ecuador from 2002 to 2015—the country dropped eighty-five ranks in Reporters Without Borders' World Press Freedom Index during that period. During Rafael Correa's presidency (2007–2017), the country dropped forty-nine spots total (Reporters Sans Frontières 2017a). While press freedom and democratic governance suffered greatly under Correa, the current president, Lenín Moreno, has showed signs of improving the country, namely with reforms aimed at putting limits on executive power and restoring press freedom (Ayala and Rochabrún 2018; Committee to Protect Journalists 2018; *The Economist* 2017d).

Under former president Correa, press freedom and freedom of expression online were stifled, even as governmental propaganda and trolling online thrived. In his weekly address to the nation, Enlace Ciudadano (Citizen Link), Correa regularly called for attacks on government critics, revealed the identities of critical Twitter accounts, and defamed online satirist Crudo Ecuador and journalist Martha Roldós, the two victims of state-sponsored trolling in Ecuador examined here. In one address, he said, "Do not kid yourselves with all of these infamous social media campaigns . . . we have to confront them and we are already getting prepared for it. If they are a thousand, we are a hundred thousand, we are more, many more" (Presidencia de la República del Ecuador ©SECOM 2015a).

This decline in freedom has coincided with an increase in persecution. The Associated Whistleblowing Press and Ecuador Transparente reported with conclusive proof that at least eight opposition activists, politicians, and journalists were targeted with spyware purchased from the notorious Italian firm Hacking Team by the Ecuadorian intelligence agency SENAIN from 2012 to 2014 (Associated Whistleblowing Press and Ecuador Transparente 2015; *PanAm Post* 2015).

The anonymous Ecuadorian political satirist Crudo Ecuador was a popular figure in Ecuadorian society for his creation of memes satirizing the political situation in Ecuador. In early 2015, after being mocked by Crudo Ecuador, in one of his weekly addresses to the nation, Correa defamed Crudo Ecuador and called on citizens to reveal his identity: "We are going to identify this person to see if he is so funny when we find out who he is. We have our Communications Law. Not only the government, the president—each of you can defend the truth, can defend the honor, the dignity of the people" (Presidencia de la República del Ecuador ©SECOM 2015b).

> **"Whether you like it or not, you self-censor, you are very careful about your words and the headlines, often we would even ask each other how to redact a tweet."**
>
> —Anonymous Ecuadorian journalist (Freedom House, 2016a)

An anonymous account, @elpatriotaec, doxed Crudo, revealing his phone number, ID number, and address, as well as the names of his parents, and photos of him that had apparently been obtained by stalking. @elpatriotaec also published password-protected documents to which only Crudo Ecuador's lawyer and the Ecuadorian Intellectual Property Institute

had access. Crudo, who was revealed to be a thirty-year-old by the name of Gabriel González, received online and offline death threats during this time as well. In the wake of this harassment, he shut down his Facebook page and social media accounts and left a message for the president: #UstedGanó ("You won") (Viñas and Alarcón n.d.).

As mentioned in the introduction to this paper, investigative journalist Martha Roldós experienced a similar tidal wave of online abuse and harassment. The daughter of Jaime Roldós, a former Ecuadorian president who died tragically in a plane crash shortly after taking office in 1979, Roldós has served in Ecuador's parliament and currently works as an investigative journalist. In January 2014, her email was hacked, and her correspondence with the National Endowment for Democracy (NED) was published in a state newspaper, *El Telégrafo*. This state newspaper claimed that NED was funded by the CIA and that its goals were to destabilize governments that opposed US policies (*El Telégrafo* 2014).

This story in turn fueled an online state-sponsored trolling campaign against Roldós in which trolls disparaged her physical appearance, threatened her with rape and death, and accused her of being a CIA agent and even being complicit in her parents' death. Selected screenshots of this campaign are shown in Figure 10.

Correa explicitly called on citizens on national TV to dox and harass Crudo Ecuador, and state media outlets participated in smears on

Figure 10. Screenshots from patriotic state-sponsored trolls that attacked Martha Roldós on January 6, 2014—the same day her emails were published in *El Telégrafo.*

Image credit: screenshot from author/collaborators

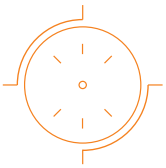
Roldós. Leaked documents released by Ecuador Transparente have also shown that the National Intelligence Secretariat has targeted journalists, politicians, and activists with surveillance and data collection, including the interception of phone calls and emails (Associated Whistleblowing Press and Ecuador Transparente 2015).

This evidence tallies with Citizen Lab's findings in its PackRat investigation. PackRat was a seven-year hacking campaign that targeted opposition activists, politicians, and journalists throughout South America, particularly in countries that are members of the intergovernmental organization ALBA , from 2008 to 2015. Though the ultimate perpetrators behind PackRat remain unknown, Citizen Lab speculated that the most likely offender was a government-backed entity (Scott-Railton et al. 2015). Roldós and Crudo were both found to be among the targets of PackRat (Janowitz 2015; Scott-Railton et al. 2015).

In addition, a 2016 leak, the Godwin Papers (Los Papeles de Godwin), revealed contracts between Ecuadorian governmental officials and private companies such as Inteligencia Emocional and Kronopio. These contracts included proposals to spread propaganda online with fake accounts and attack government critics (Ecuador Transparente 2016). Notable proposed targets were the former secretary of communications, Mónica Chuji; local press watchdog Fundamedios; and the Inter-American Commission on Human Rights and its Special Rapporteur for Freedom of Expression, Catalina Botero (Freedom House 2016d). Multiple sources have also noted the Correa regime's habit of underhandedly exploiting the US Digital Millennium Copyright Act (DMCA) to remove critical content from YouTube, often through a proxy Spanish company, Ares Rights (Ball and Hamilos 2015; Sutton 2014; Tegel 2015). Correa also founded online groups such as Somos+ and Correístas to organize messaging campaigns online. ▣

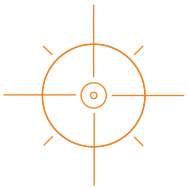
"[T]his was a new kind of harassment . . . [In the past] I was denied my political rights, I had to appeal, I had armed men outside my house pointing a gun [at] my daughter . . . but not cyber harassment. Since I . . . became a sponsor of investigative journalists, my time of cyber harassment began, and it was from the president of Ecuador."

—Martha Roldós's testimony on her experience with state-sponsored trolling, RightsCon, 2016



"[The] troll manual in Ecuador is the same [as the] troll manual in Russia. They do the same kind of things."

—Martha Roldós, target of state-sponsored trolling in Ecuador, 2016



THE PHILIPPINES

While Freedom House still ranks the Philippines' online sphere as "free," the country has had a risky atmosphere for local journalists for decades. According to Reporters Without Borders, the Philippines "continues to be one of the most dangerous countries for the media. Private militias, often hired by local politicians, silence journalists with complete impunity" (Reporters Sans Frontières 2017b). Indeed, the Committee to Protect Journalists found that forty-eight reporters have been killed there in the past decade (Wichtel 2017).

The election of President Rodrigo Duterte in May 2016 further exacerbated the situation. Duterte demonstrated an adept use of social media and digital tools to silence critics and undermine mainstream journalists throughout his election and thereafter (Etter 2017). In his first press conference, President-elect Duterte claimed corrupt journalists deserved to be killed. He has continued to attack journalists and critics throughout his administration (Freedom House 2017b). Multiple former paid trolls have, on the condition of anonymity, come forward to speak about their experience working in the 2016 presidential campaign (Almario-Gonzalez 2017; Caruncho 2016). Three high-profile women were the target of state-sponsored trolling attacks in the Philippines in 2016 and 2017: Vice President Leni Robredo, journalist Maria Ressa, and Senator Leila de Lima. Attacks on these women have all involved character assassination, threats of rape and violence, misogyny, and disinformation (Maria Ressa, personal communication, 5 January 2018).

> **"They [Facebook] haven't done anything to deal with the fundamental problem, which is they're allowing lies to be treated the same way as truth and spreading it. . . . Either they're negligent or they're complicit in state-sponsored hate."**
>
> —Maria Ressa, journalist, founder and CEO of Rappler

Ressa is one of the most accomplished journalists in the Philippines, having served for nearly two decades in the top ranks at CNN's Asia Bureau, doing groundbreaking work on terrorist networks in Southeast Asia, and most recently founding Rappler, an independent Filipino news agency. Ressa moved into the crosshairs of the Duterte trolling apparatus for her coverage of its use of disinformation and state-sponsored trolling on social media both before and after Duterte's election in late 2016. Campaigns against Ressa and Rappler have been ongoing since the publication of Rappler's series of articles "Propaganda War: Weaponizing the Internet." In February 2018, Duterte banned Rappler from covering events at the presidential palace (Regencia 2018).

At the height of the attacks, Ressa and Rappler experienced an average of ninety hate messages per hour (Arsenault 2017). In other attacks, Ressa was threatened with death, including by a user who claimed he wanted Ressa to be "raped to death" (Etter 2017). Other leading opposition figures, such as Leila de Lima, a prominent senator who has challenged Duterte's extrajudicial killings in his war on drugs, have also suffered state-sponsored trolling campaigns (Amnesty International 2017; Chen 2016). In addition to being the target of online hate campaigns, de Lima has been jailed. Amnesty International named de Lima one of its human rights defenders under threat in 2017 (ABS-CBN News 2017b). In de Lima's case, state-sponsored trolling attacks laid the groundwork for her arrest. Online smears discredited and attacked her under the coordinated hashtag #ArrestLeiladeLima. In 2017, de Lima was arrested on politically motivated charges and remains in prison (Amnesty International 2017).

Ressa has suffered ongoing attacks since 2016, including campaigns with organized hashtags

such as #ArrestMariaRessa (see Figure 11), a calque on the hashtag that preceded de Lima's arrest. Facebook has been the dominant arena for the attacks as the most popular social media platform in the country, though Twitter and other media have been used as well. After years of attacks, in early January 2018, the Securities and Exchange Commission of the Philippines revoked Rappler's license to do business (CNN Philippines 2018). The Philippine Center for Investigative Journalism and the National Union of Journalists of the Philippines have both challenged the move as a politically motivated attack on press freedom (Elemia 2018).

State use of disinformation, paid commentators, and trolls has been documented by several sources in 2016, 2017, and 2018 (Almario-Gonzalez 2017; Bradshaw and Howard 2017; Caruncho 2016; Etter 2017; Freedom House 2017b, 2017d; Ong and Cabañes 2018; Reyes and Millari 2016). The use of state-sponsored trolling and disinformation in the Philippines is particularly insidious and pernicious given that Filipinos lead the world in social media use (ABS-CBN News 2017a).

In addition to the president's own admission of paying trolls during his campaign, the head of the Armed Forces of the Philippines, General Edward Año, publicly apologized to Maria Ressa for active military members spreading false news stories about her and participating in attacks on her during Duterte's presidency (Figure 12).

Troublingly, participants in these attacks and prominent disseminators of disinformation have been promoted to positions within the government itself. Most notably, blogger and actress Mocha Uson has been promoted to assistant communications secretary, and R. J. Nieto, who runs the influential pro-Duterte site Thinking Pinoy, has been hired as a consultant to the Department of Foreign Affairs (Etter 2017). As of August 2018, the Presidential Communications Operations Office has enacted an interim policy allowing "social media users" and "social media publishers" to be accredited with full press credentials to cover Duterte's events. The policy will give credentials to any Filipino citizen at least eighteen years old who has at least five thousand followers on any social media platform, and covers only bloggers who "generate news and information regarding the activities of the President" (Morales 2017; Ranada 2017c).

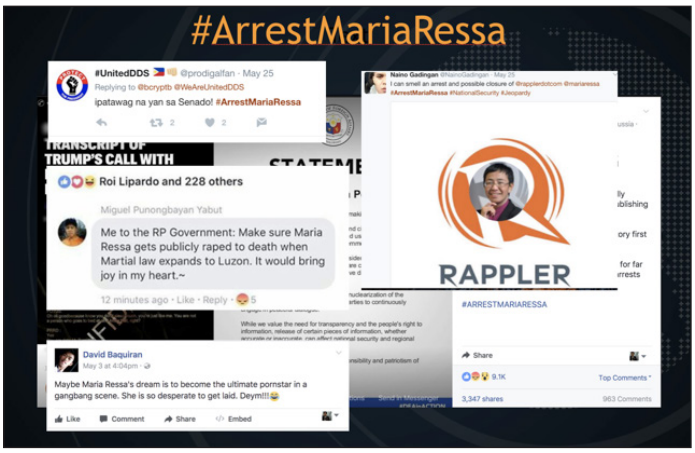
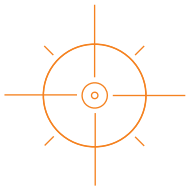


Figure 11. Screenshots of participants in the #ArrestMariaRessa campaign. This was one of the hashtags used in state-sponsored trolling campaigns against Ressa.

Photo provided by Maria Ressa.

Figure 12. Screenshot of public apologies from the chief of the Armed Forces of the Philippines to journalist Maria Ressa. Military members serving under General Edward Año participated in trolling attacks on Ressa.

Photo provided by Maria Ressa.



TURKEY

After a brief period of liberalization, Turkey has steadily grown more authoritarian under President Recep Erdoğan since his Justice and Development Party (AKP) took power in 2003 (Karaveli 2016; *The Economist* 2017a). A failed coup d'état in July 2016 made the slide into authoritarianism even more precipitous: emergency powers declared by the government vastly reduced freedom of speech, resulting in a post-coup purge. In the following months, nearly 100,000 academics, journalists, and activists were dismissed or arrested (*The Economist* 2016). In both 2016 and 2017, Turkey jailed more journalists than any other nation on Earth, and it has continuously carried out new waves of arrests (Beiser 2017). More than 190 media outlets have been closed during the government's consolidation of power since the coup attempt (European Commission for Democracy Through Law 2017, p. 23).

"Of course [attacks have an impact on journalists], I mean, you can't expect otherwise. During Gezi, everyone was aware that journalists could actually get attacked, online or out on the streets, and in the last three years it's gotten worse. Every time you put out a report, you expect something might happen, everyone does, especially if it's on a controversial subject."

—Selin Girit on the state-sponsored trolling attacks in 2013

The three Turkish cases examined here are from an upcoming report from the International Press Institute (IPI) on state-sponsored trolling in Turkey. IPI conducted quantitative social network analyses of the attacks as well as interviews with many of the affected parties on the ground. The Turkish campaigns are unique in that they are the most longitudinal view we have within one country: the cases span three years, from the Gezi Park protests of 2013 to the post-coup atmosphere in late 2016.

In these cases, we see a refining of the Turkish state's trolling apparatus with each successive attack. Each case grows more sophisticated, lasts longer, and has less evident direct involvement from the government.

Selin Girit was a correspondent for the BBC World Service who covered the Gezi Park protests. On 23 June 2013, Ankara's mayor, Malih Gökçek, a member of the ruling AKP party, attacked Girit with seventeen tweets on Twitter for her reporting on the protests, accusing her of "betraying her country" and being an "English agent."

The tweets contained the hashtags #İNGİLTEREADINAAJANLIKYAPMASELİNGİRİT ("Don't be an agent for England, Selin Girit") and #BBCTÜRKİYEYİKARIŞTIRMAHABERLERİ DOĞRUVER ("BBC, don't promote chaos in Turkey, report the truth."). Gökçek explicitly called on his followers to make these hashtags trend: "I want all who love their country to make the hashtag a trending topic. That way, our reaction will be heard abroad." Within hours, Gökçek's main hashtag, f#İNGİLTEREADINAAJANLIKYAPMASELİNGİRİT, had been used in more than 35,000 tweets.

During this campaign, Girit received numerous rape and death threats, many of them from bots. In addition to being the user who initiated the campaign, Malih Gökçek was found in IPI's analysis to be the most influential user in the campaign.

Ceyda Karan is a journalist who worked for *Cumhuriyet*, one of Turkey's few remaining independent newspapers, in 2015. In early 2015, Karan published an op-ed in support of victims of the recent Charlie Hebdo attacks in France. The column contained images of the prophet Mohammed and was the centerpiece of a trial against Karan a year later. A court in Istanbul found Karan and her co-author guilty of "inciting hatred and public enmity via media" and sentenced both to two years in prison. Shortly after the verdict, on 28 April 2016, Karan posted a tweet on her sentence (Figure 13).

Just seventeen minutes after Karan posted this tweet, a pro-AKP TV commentator, Fatih Tezcan, posted: "Hikmet Çetinkaya and Ceyda Karan who published the Charlie Hebdo cartoon that insults our prophet are both sentenced to two years in jail. Yes, but not enough!" (Figure 14). After this tweet, Karan was subjected to a massive state-sponsored trolling campaign against her that lasted nearly three days.

Ceyda's attackers called for her hanging and the reinstatement of Sharia law—a jab at her publication's tendency to promote secularist ideas. Sexism and misogyny were frequent in the attacks, which were amplified by bots. Acerbic death and rape threats featured prominently. Attackers also impugned Karan's journalistic integrity. As was the case with Selin Girit, attacks continued after the hate campaign with every new story Karan published or tweeted about. Like Girit, Karan was targeted by bots during the 2013 Gezi Park protests.

This case represents a refinement of the Turkish state's trolling apparatus. No office-holding AKP politician called for attacks, but a high-profile pro-AKP commentator with more than 400,000 followers drove the campaign. Analyses showed that Fatih Tezcan was the most influential user driving the hate campaign.

Figure 13. Tweet from Ceyda Karan posted in the wake of her prison sentence for "inciting hatred and public enmity via media." The tweet reads: "Our sentence, two years in prison, is dedicated as a gift to our liberal fascists . . . #JeSuiCharlie."

Image credit: International Press Institute, 2016

Figure 14. Tweet from pro-AKP TV commentator Fatih Tezcan arguing that Ceyda Karan's sentencing was not enough. After this tweet was posted, a vast three-day hate campaign against her ensued on Twitter.

Image credit: International Press Institute, 2016

The third victim of state-sponsored trolling in Turkey, Nevşin Mengü, was a television correspondent for CNN Türk, an independent affiliate of CNN International, until her recent resignation in late 2017. On 15 July 2016, CNN Türk was the first television station to interview President Recep Erdoğan. Mengü covered the unfolding coup attempt as the night went on. While her coverage was mostly positive, Mengü did take a moment to note the alternate dangers to democracy that continued AKP rule posed to the society, and questioned the motives of some of the protesters in the streets.

This commentary angered AKP supporters and provided impetus for a state-sponsored trolling attack against Mengü that would last nearly a week. After her segment, out-of-context quotes, tweets, and clips from former interviews with Mengü circulated online, painting her as a coup supporter. Reactions to this misleading misinformation campaign included accusations of treason, calls for her to be hanged, and other acrimonious death and rape threats. The portrayal of Mengü as a supporter of the failed coup is highly significant. Under the state of emergency imposed in Turkey at that time, Mengü could have faced severe charges had the authorities considered her a coup supporter.

A low-profile pro-AKP user with 3,500 followers, @drisavuz, posted the tweet that launched the state's trolling campaign against Mengü (Figure 15). Despite the user's relatively low influence on Twitter, the post spread rapidly on the social media platform, drawing the attention of influential pro-government figures, such as pro-Erdoğan journalist Fatih Tezcan and *Milet* newspaper editor-in-chief Bayram Zilan (@bayramzilan), a self-declared "AKP journalist"

"Even though you suffered great hardships, do you still keep this pro-coup woman?"

"Take no offense; I am no longer objective. I declare I am a journalist of AK PARTY!"

Figure 15. Screenshot of the tweet that sparked the hate campaign against Mengü from a low-profile user, later picked up by an influential AKP supporter, Bayram Zilan.

Image credit: International Press Institute, 2016

with 49,000 followers. Both seized on the message, actively helping to spread it.

Mengü's case represents a final optimization of state-sponsored trolling in Turkey. It demonstrates that the need for coordination and organization of a successful hate campaign has dropped dramatically: a low-profile user was able to initiate an attack that was later picked up by more influential pro-government figures. Governmental officials no longer need to directly initiate an attack or explicitly announce their intentions for a campaign to succeed.

Several facts point to state involvement in these Turkish trolling attacks. These include direct involvement and participation of high-profile politicians, leaked recordings of one of Erdoğan's close advisors discussing trolling, a documented history of governmental bot usage, and quantitative evidence of prominent politicians and pro-AKP commentators at the center of troll campaigns and networks (Hafiza Kolektifi 2015; International Press Institute forthcoming).

As mentioned above, the mayor of Ankara explicitly called on users to amplify an attack that he initiated on the journalist Selin Girit and was shown by IPI's quantitative analyses to be the most influential user in the campaign against her (International Press Institute forthcoming). In addition, Mustafa Varank, a close advisor to Erdoğan, was shown, in Hafiza Kolektifi's quantitative analysis of a troll network from 2015, to be implicated with responses by politicians and anonymous pro-AKP trolls (Hafiza Kolektifi 2015) (see below).

Leaked telephone conversations and emails from the RedHack leaks in Turkey also have shown Varank and Erdoğan to be involved in discussions about propaganda strategy and trolling online. A recording of Erdoğan's daughter emerged in which she asked Varank to boost her social media presence with "AK trolls" (Hoyng and Es 2017; Kizilkaya 2015; Sozeri 2016).

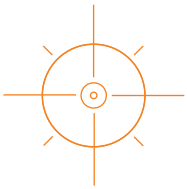
"One would see a new tweet appear each second on your timeline. . . . I could not believe it, the insults and threats were horrendous. Bot accounts were continuously targeting throughout the month of August. I think it was one of the first organized attacks. They threatened to penetrate me with a broken bottle."

—Ceyda Karan on state-sponsored trolling attacks suffered during the Gezi Park protests in 2013

Pro-AKP trolls have also had a suspicious tendency to get involved in Turkey's foreign affairs and larger geopolitical issues, especially when affairs grow more strained. This was notably the case with a massive Twitter hack and messaging campaign during a period of intense tension between Turkey on the one hand and Germany and the Netherlands on the other in 2016 and during a 2015 Russian-Turkish troll war that followed the downing of a Russian jet by Turkish forces (Sozeri 2015; Toor 2017).

The impact of state-sponsored trolling in Turkey is clear. In their 2017 comprehensive review of social media manipulation in Turkey for the past four years, scholars Ergin Bulut and Erdem Yörük wrote that "trolling has impacted the language of politics itself. As citizens, we increasingly find ourselves asking whether we are being trolled by our leaders. . . . Politicians endorse trolls' discriminatory language on Twitter to appeal to the masses. Similarly, pro-[AKP] journalists disseminate fake news just as trolls do." They went on to add: "Twitter . . . is [now] a medium of government-led populist polarization, misinformation and lynching." (Bulut and Yörük 2017).

In the same vein, Sedat Yılmaz, a Turkish journalist, finds the impact of state-sponsored trolling undeniable: "All of this constitutes a wide, vast, overwhelming atmosphere of persecution" (International Press Institute, 2016). ▣



THE UNITED STATES

The election of Donald Trump in late 2016 was a harbinger of a decline in civil liberties and freedom of expression in the United States. Freedom House and Reporters Without Borders both downgraded the freedom of the press ratings of the United States in 2017, explicitly citing Donald Trump as an influencing factor (Freedom House 2017a; Reporters Sans Frontières 2017a).

Trump himself has described journalists with various terms having nuances of abhorrence, categorizing them as "scum," "slime," "disgusting," and "enem[ies] of the people" (Tashman 2017). Trump's closest confidants and staff in the White House have struck the same tone: former chief strategist and former head of Breitbart News Steve Bannon described the mainstream press as "the opposition party," and one of Trump's communications directors explicitly stated he wanted to "fucking kill all the leakers" (McCaskill 2017; Stein 2017).

The freedom of the online sphere in the United States was also downgraded by Freedom House: "Fake news and aggressive trolling of journalists both during and after the presidential election contributed to a score decline in the United States' otherwise generally free environment" (Freedom House 2017d). In July 2017, the Trump administration filed a request to compel DreamHost to hand over the IP addresses of all users who had visited a website that helped coordinate inauguration protests (Wong and Solon 2017). US Customs and Border Protection asked Twitter to reveal the identity of a user who opposed Trump's immigration policy online (Abramowitz 2017). Researchers at the Oxford Internet Institute also showed that pro-Trump bots accrued positions of high influence—interrupting communication flows during the election—on Twitter during the 2016 presidential campaign (Woolley and Guilbeault 2017).

There have been several reports of organized trolling of those who have questioned President Trump online, both before and after he took office. During Trump's campaign for the presidency in 2015 and 2016, many conservative opponents faced online trolling for critical comments, columns, and essays about the campaign. Erick Erickson, the former head of RedState, a conservative blog, revoked Trump's invitation to an event held by the blog after the candidate's controversial comments about a female journalist, Megyn Kelly. After this, in addition to being called a "major sleaze and buffoon" by Trump himself on Twitter, Erickson was the subject of a Breitbart story and suffered threatening online trolling for his remarks and actions. Erickson also received offline threats, including threatening mail to his home and family and organized phone calls to his employer requesting that he be fired.

Rick Wilson, a Republican political consultant, faced attacks from Breitbart and trolls after a CNN appearance in which he criticized Trump and Breitbart. Trolls harassed him with photoshopped photos of his daughter and threats of gang rape against her. Wilson's home address and phone were leaked, and offline harassment proceeded in the form of prank calls and mass deliveries of pizza, the Quran, and moving boxes (Grove 2016).

A young aide to Jeb Bush's presidential campaign, Lauren Batchelder, also suffered intense trolling, including death and rape threats, after asking Trump a critical question at a town

hall event in 2015 and saying he was not "a friend to women." Trump himself later referred to Batchelder as an "arrogant young woman" on Twitter and accused her of being a plant from Bush's campaign. Trump's director of social media, Dan Scavino Jr., joined the fray and posted screenshots of Lauren Batchelder's social media accounts, designating her as a target for trolls (Figure 16). Batchelder also received threatening emails and voicemails, as well as ongoing sexist and obscene trolling, for more than a year after the event (J. Johnson 2016).

State-sponsored trolling attacks continued after Trump's inauguration. Rosa Brooks, a law professor at Georgetown University, was targeted with trolling after publishing a column in *Foreign Policy* on 30 January 2017. Brooks noted the possibility of military advisors disobeying orders from Trump during his tenure, remarks that were characterized by Breitbart as calling for a military coup. Other outlets, such as Alex Jones's conspiracy-theory outfit Infowars and the white-supremacist website the Daily Stormer, also joined in the attack on Brooks, accusing her of treason and sedition. Trolls attacked Brooks with obscene death threats and harassment. Emails and phone calls calling for her to be fired were received at her university (Brooks 2017).

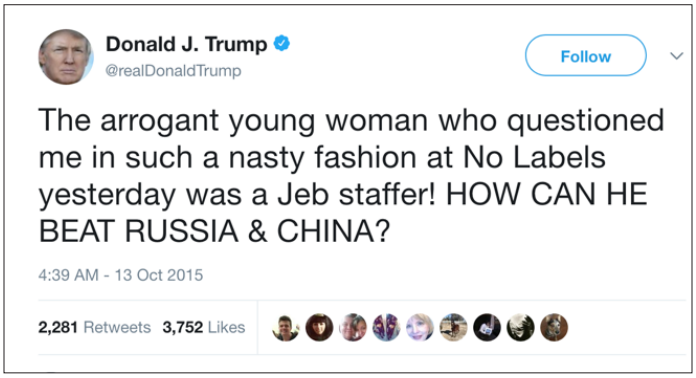


Figure 16. Tweets in October 2015 from then-candidate Donald Trump and his social media manager, Dan Scavino Jr., attacking Lauren Batchelder, who at a town hall event had asked Trump about his views.

Screenshots taken by authors.

Patterns we have seen in other countries engaging in state-sponsored trolling also emerge in the United States: the involvement of hyperpartisan news outlets and sources close to the president (Breitbart and social media manager Dan Scavino Jr.) to reveal targets, the evolution from an electioneering trolling machine to an incumbent government's apparatus, and statements tantamount to a coded condoning of vitriolic harassment online from high officials.

Dan Scavino Jr., one of the key instigators of attacks on Lauren Batchelder during Trump's campaign in 2015, is now the White House director of social media. Scavino even sometimes types tweets for Trump's accounts (E. Johnson 2017; Ohlheiser 2017). Scavino has continued to attack critics and point targets out to trolls. The US Office of Special Counsel found that one of Scavino's tweets from April 2017—calling for the defeat of a congressman who opposed one of the attempts to repeal Obamacare, a key part of Trump's political agenda (Figure 17)—violated the Hatch Act. The Hatch Act is meant to prevent political activity by government employees (Lipton 2017). Scavino's hiring as the White House social media director is similar to patterns of promotion for prominent trolls in the Philippines, notably Mocha Uson's appointment to the post of assistant communications secretary in the Philippines.

Figure 17. Tweet from the White House director of social media calling for the defeat of one of the first Republicans to oppose Trump's initial efforts to repeal Obamacare. The Office of Special Counsel found that with this tweet, Scavino violated the Hatch Act, an act passed to prevent political activity by government officials.

Screenshot from authors

As a key former White House chief strategist and former head of Breitbart, Steve Bannon is also worthy of special attention for his role in encouraging online trolling. Bannon was one of the original heads of Breitbart, an online news outlet he has referred to as "a platform for the alt right" (Posner 2016). Bannon's personal remarks have revealed a coded endorsement of vitriolic trolling: "If a guy comes after our audience—starts calling working-class people vulgarians and brownshirts and Nazis and post-literate—we're going to leave a mark. We're not shy about it at all. We've got some lads that like to mix it up" (Brooks 2017).

Statements like these are remarkably similar to former Ecuadorian president Rafael Correa's comments about having 10,000 accounts to respond to every single account that criticized his government, or Indian defense minister Manohar Parrikar's coded endorsement of attacks on Indian movie star Aamir Khan (Chaturvedi 2016). India and Ecuador are also among the states in which the government's trolling apparatus was initially incubated as an electioneering propaganda attack machine. ▣

VENEZUELA

The situation in Venezuela in recent years has been extremely dire. Food and medicine shortages have been widespread since 2014 (Vidal and Diaz 2016). Viewed from 2012 onward, Venezuela's current economic collapse is the steepest in modern Latin American history (*The Economist* 2017b). In this context, President Nicolás Maduro has successfully consolidated control—packing the courts with loyalists and usurping the powers of the parliament through a puppet body called the Constituency Assembly, established in July 2017 (*The Economist* 2017c). His autocratic rule has become dictatorial, and he has promoted his government's viewpoints and persecuted opponents of his regime both offline and online. Online attacks have taken the form of particularly overt state-sponsored trolling attacks on perceived opponents of the regime.

Marianne Diaz of Global Voices, who is also director of the nongovernmental organization Acceso Libre, has thoroughly documented several cases of state-sponsored trolling in Venezuela. In the first, Luis Florido, a congressman in the National Assembly (Venezuela's unicameral legislature) and leader of one of the opposition parties, Voluntad Popular (VP), claimed that members of his party were being tortured in prisons controlled by Diosdado Cabello. Cabello holds significant clout in the country—he was vice president of Venezuela under Hugo Chávez and formerly president of the National Assembly. Cabello currently hosts the TV show *Con el Mazo Dando (Hitting with the Sledgehammer)* on the Venezuelan state-owned TV channel VTV8. Freedom House has noted Cabello's use of his personal website to attack and discredit human rights defenders and journalists more generally (Freedom House 2016c).

After Florido's remarks, Cabello used his TV show and a Telegram channel associated with it to encourage Twitter attacks on Florido using the hashtag #FloridoEresUnPajuo ("Florido, you're a lying idiot"). Attacks on Florido lasted for days; they were vitriolic and crude and frequently accused him of being a traitor to Venezuela. Governmental officials participated in the attacks (Figure 18).

Diaz's second documented state-sponsored trolling case concerns Lorenzo Mendoza. Mendoza is the billionaire owner of Empresas Polar, a food conglomerate in Venezuela and the largest privately owned corporation in the nation (Kurmanaev 2016a). As the economic crisis has deepened, Maduro has frequently accused the private sector of waging "economic war" on Venezuela, scapegoating prominent companies and businessmen for the ongoing food and medicine shortages. Among the main targets of these accusations has been Lorenzo Mendoza.

Figure 18. Tweet by Johan Acevedo, government coordinator of communication and social media, Venezuela.

Screenshot courtesy of Marianne Díaz

Offline attacks against Mendoza have been ongoing since Chávez's ascendancy to the presidency. Chávez himself accused Mendoza of being a *pelucón* ("bigwig conservative") and said that Mendoza would go to hell. He also threatened to expropriate Mendoza's business (Forero 2016; Schipani 2017). These attacks have intensified as political and economic turmoil has gripped Venezuela. Maduro has publicly called Mendoza a "parasite" and a "bandit, thief, oligarch, and traitor" and has continued to threaten to expropriate his company. He has also blamed Mendoza on multiple state TV stations for waging economic war on the country (Forero 2016).

These offline attacks have metastasized into ongoing state-sponsored trolling campaigns online since March 2016. One of the trolling campaigns against Mendoza blamed him for the food shortages in the country with the hashtag #LorenzoEsEscásez ("Lorenzo Is scarcity"). Several official governmental accounts participated in these attacks (Figure 19).

The Venezuelan government has been singularly overt about attacking critics and spreading propaganda online. It has explicitly announced its plans to train "digital guerillas," and governmental ministers and ministries have participated in state-sponsored trolling campaigns. The Digital Guerilla and the Guerrilla Comunicacional are civilian forces that receive training to spread the regime's viewpoints online and attack opponents (Figures 20 and 21).

Figure 19. Official Twitter account of the Ministry of Habitat and Housing for the Venezuelan state of Yaracuy, here using a hashtag propagated in a state-sponsored trolling attack on Lorenzo Mendoza. The translated text is "Polar monopolizes, deflects and stops producing regulated and government-subsidized food products, therefore #LorenzoIsScarcity."

Screenshot courtesy of Marianne Díaz

Figure 20. Tweet and photos of Digital Guerilla training from INCES, an institute that belongs to the Venezuelan Ministry of Work. Translation: "Showing the achievements of the revolution and strengthening the communicational guerrilla was part of the social networks workshop."

Screenshot courtesy of Marianne Diaz.

According to Marianne Diaz, Telegram channels associated with the Ministry of Communications—mainly the now-defunct @SIBGECOV and @comunicaciondigital—have disseminated hashtags, memes, and content to be used in the state's trolling attacks on targets. The Ministry's Telegram channels are openly operated by the Ministry's Digital Communications Directorate.

In addition to the digital guerrillas, governmental ministers have also promoted and participated in state-sponsored trolling. In late April 2017, Ernesto Villegas, the minister of communication and information in the country, explicitly announced the government's plans to set up physical "Candanga points" around the country to help train citizens to promote the regime's viewpoints and attack opponents online: "We're going to open Twitter, Facebook and Instagram accounts for them [Venezuelan private citizens] and give basic instructions to each Venezuelan so that they can become digital militants. The Digital Militia is born today." Villegas added: "We must be clear and valiant, . . . one of the new weapons is the use of these technologies: they are the new weapons of combat on social networks and the spaces we have to conquer" (Infobae 2017).

Villegas has also personally encouraged and participated in campaigns against DolarToday, a popular website that estimates the true value of the Venezuelan bolívar, the nation's basic unit of currency (Kurmanaev 2016b). In December 2016, Villegas tweeted a video instructing users how to get the app taken down from the Google Play store (Figure 22).

Figure 21. Tweet from VTV showing photos of a gathering of the Guerrilla Comunicacional. VTV is owned by the Venezuelan Ministry of Communications. Translation: "#InPhotos. Here is the Communicational Guerilla camp from the reserved areas of the ZooPark in the #ItsTimeToDefendTheHomeland."

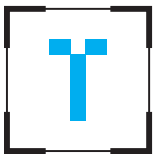
Screenshot courtesy of Marianne Diaz

Figure 22. Venezuelan minister of communications Ernesto Villegas posting a video on how to remove DolarToday from the Google Play store in late 2016. For years, DolarToday has been one of the main tools available for citizens and enterprises to estimate the true value of Venezuela's currency, the bolívar (Kurmanaev 2016b).

Screenshot taken by authors


STILL
WE
RISE

Developing Policy Interventions



This paper has shown that among the generalized disinformation, content manipulation, and extremist speech that exists today online and in digital technologies, it is possible to identify instances in which states are weaponizing online information to take targeted action against specific individuals. In articulating a conceptual framework for assigning state responsibility, we have sought to help researchers and public commentators transcend the frequent outright denials by states and begin to assert state liability for online harassment campaigns. But establishing that states are in the business of state-sponsored trolling, a significant obstacle in and of itself, is only one part of a much larger challenge: prescribing policy solutions to address state-sponsored digital harassment campaigns.

We believe that this phenomenon should be addressed through policy interventions originating in a diverse range of policy communities. Specifically, we see three main avenues for formulating effective policy responses: (1) international human rights law, (2) US law, and (3) policies of major technology companies.

International Human Rights Law

Understandings of international human rights law need to expand and evolve to recognize that state-sponsored trolling attacks amount to a violation of states' obligations. The fact that such attacks happen online, and occasionally across national borders, does not mean that human rights law has no relevance to them; the major international human rights policy-making bodies have recognized that all human rights apply equally online and offline (Human Rights Council 2012; OSCE 2011).

Indeed, for many people around the world, the internet has become the key medium through which their free speech rights can be exercised. The weaponization of information in the form of state-sponsored trolling attacks thus constitutes an interference with individuals' right to freedom of expression and opinion, enshrined in Article 19 of the Universal Declaration of Human Rights and of the International Covenant on Civil and Political Rights, as well as the European Convention on Human Rights (Article 10), the European Union Charter of Fundamental Rights (Article 11), the American Convention on Human Rights (Article 13), the African Charter on Human and People's Rights (Article 9), and the ASEAN Human Rights Declaration (Article 23). This protected right encapsulates a right not only to impart, but also to seek and receive, information and ideas of all kinds, regardless of frontiers (UN Human Rights Committee 2011).

International human rights law is not a rigid legal code, though, and it permits restrictions on the right to freedom of expression in accordance with strict conditions. Permissible limitations on free expression are those that are provided by law, necessary to meet a legitimate objective, and proportionate to that objective. This test has been restated in numerous international human rights instruments, most notably in the UN Human Rights Committee's General Comment No. 34. Under international law, the only legitimate objectives toward which restrictions can be aimed are (1) respect of the rights or reputations of others or (2) protection of national security or of public order, or of public health or morals (International Covenant on Civil and Political Rights 1966 Article 19(3)).

International human rights law does not permit states to restrict individuals' right to freedom of speech and access to information in order to levy online campaigns designed to minimize and silence dissenting speech or to remove critics from the public stage. It does not permit the purposeful dissemination of disinformation and the harnessing of bots and other digital tools to drown out progressive information and to intimidate journalists and activists. It does not allow states to harass and intimidate individuals through the use of violent speech and imagery.

On the contrary, human rights law requires states to take positive measures to protect individuals' human rights, including their right to freedom of expression and access to information. Somewhat controversially, international human rights law also requires states to take action to prohibit, by law, forms of expression generally known as hate speech. Article 20 of the International Covenant on Civil and Political Rights states that "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law." The Inter-American Convention on Human Rights contains a similar provision (Article 13), but the European Convention on Human Rights (ECHR) does not. Hate speech is nevertheless equally prohibited under European human rights law, and the European Court of Human Rights has dealt with its conflict with freedom-of-expression rights by deploying Article 17 of the Convention, which prohibits the destruction of human rights (Seurot v. France 2004).

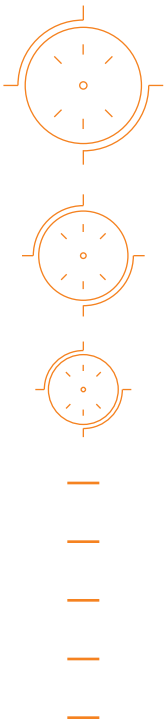
The purpose of Article 17, the Council of Europe has argued, is "to prevent the principles enshrined in the ECHR from being embezzled by [purveyors of hate speech and others], at their own advantage, whose actions aim at destroying those same principles" (Council of Europe 2007). To this end, the European Court (and its predecessor, the European Commission on Human Rights) has found that Article 17 excludes from human rights protection the establishment of totalitarian political doctrine (B.H, M.W., H.P. and G.K. v. Austria 1989) and expression that constitutes the denial or justification of crimes against humanity, such as the Holocaust, linked with incitement to religious discrimination (Lehideux and Isorni v. France 1998), incitement to racial discrimination (Glimmerveen and Hagenbeek v. the Netherlands 1979), and incitement to religious discrimination (Norwood v. United Kingdom 2004).

The European Court's approach lays bare an important distinction: not all forms of hate speech are unlawful. The term is too vague to use in any meaningful way, given lack of agreement about what constitutes hate speech, its frequent situation-specific redefinition, and evolving societal attitudes toward equality and discrimination; in this context, general prohibitions on hate speech could be used to silence or censor legitimate speech. Rather, the prevention and restriction of hate speech must only take place in contexts in which the speech rises to the level of incitement to discrimination, hostility, or violence (International Covenant on Civil and Political Rights 1966).

Although a comprehensive exploration of what constitutes prohibitable hate speech under international human rights law is beyond the scope of this paper, generally speaking, the definition of hate speech that may be suppressed or prohibited excludes speech that is offensive, disturbing, or shocking (Handyside v. UK 1976); blasphemy or "defamation of religion"; and defamation (Article 19 2015). The fundamental elements of hate speech include intent (the perpetrator must have intended to incite hatred), incitement (there must be a nexus between the statements and the prescribed result), and context (a critical element; what was the likely impact of the statement in the particular context in which it was made?) (Mendel 2010).

Even under this higher threshold of hate speech—one that requires a connection between hate speech and incitement to violence or discrimination—there is a strong argument that the types of expression embraced by states in state-sponsored trolling attacks should not enjoy the protection of freedom of expression, but rather that they constitute hate speech that should be prohibited. This is particularly the case in state-sponsored campaigns that embrace incitement to violence against targets on the lines of race, religion, gender, or sexual orientation.

The European Union has recently taken steps to curtail the proliferation of online hate speech by developing a Code of Conduct pertaining to illegal online hate speech, according to which a number of tech companies and platforms have made a series of commitments (European Commission 2016). These commitments include putting in place effective processes to review notifications regarding illegal hate speech on platforms in order to remove or disable content expeditiously; the review of notifications of illegal hate speech within twenty-four hours and removal or disabling of content; the establishment of "trusted flagging" mechanisms, whereby experts and civil society organizations have an elevated ability to flag illegal hate speech; identifying and promoting counternarratives and encouraging critical thinking; and countering hateful rhetoric and prejudice at scale. The initiative, while well



intentioned, illustrates the difficulty of regulating hate speech online and has garnered widespread criticism from free-expression advocates in Europe (Article 19 2015; Jeppesen 2016).

We agree with some, but not all, of these criticisms. It is certainly true that placing responsibilities on private-sector entities to remove or disable content according to ill-defined definitions of illegal hate speech, definitions that differ across jurisdictions and cultures, could incentivize the regulation and restriction of legitimate online speech with negative consequences for free-expression rights. However, it seems to us that the old adage of fighting hate speech with more speech is rendered ineffective by modern social media platforms, whose algorithms do not provide an equal playing field for all online speech. Those platforms do not constitute an empty page on which every internet user has an equal right to write, but rather they manipulate the dissemination of information according to commercial imperatives, prioritizing high-engagement, often controversial material.

Measures designed to rectify the imbalance could include requiring platforms to detect and, in some cases, remove hate speech, harassment, and disinformation. This seems to us to be a legitimate demand on social media platforms. Provided such measures are implemented in a transparent and accountable manner that respects due process and reinforces human rights, they could make the online sphere more hospitable to a plurality of voices.

US Law

It is no accident of jurisdiction that the major technology companies are domiciled in the United States. Social media platforms are both a product and a beneficiary of the First Amendment, one of the world's most permissive free-speech regimes. The US Constitution "demands that content-based restrictions on speech be presumed invalid" (Ashcroft v. American Civil Liberties Union 2004).

At the risk of simplifying the status of hate speech under US law (with respect to which there is a rich and extensive jurisprudential history not the subject of this paper), expression cannot be prohibited even when it advocates the use of force or violence, except where such speech is directed to inciting or producing imminent lawless action and is likely to incite or produce such an action (Brandenburg v. Ohio 1969). This amounts to a far higher threshold for prohibiting hate speech than that which exists under international human rights law, as it requires a link between the speech in question and immediate injury or harm; expression that can be restricted includes "conduct that itself inflicts injury or tends to incite immediate violence" (R.A.V. v. The City of St. Paul, Minnesota, 1992).

The authors are not US legal experts and do not seek to opine on the possible legal routes for bringing state-sponsored trolling that occurs on US-based social media platforms and other intermediaries within the scope of exceptions to the First Amendment. Rather, we only highlight possible options for reconciling the First Amendment with online harassment campaigns, as suggested by others.

In his essay "Is the First Amendment Obsolete?" Tim Wu addresses head-on how "the rise of abusive online mobs who seek to wear down targeted speakers . . . directly employed by, loosely associated with, or merely aligned with the goals of the government or particular politicians" renders the First Amendment and its jurisprudence "a bystander in an age of aggressive efforts to propagandize and control online speech" (Wu 2017). Wu suggests two opposing ways past this impotence:

- Accept a limited First Amendment and advocate instead for increased liability on the part of technology companies, "the most important speech brokers of our time," equivalent to the norms and policies traditionally associated with twentieth-century journalism.

- Find a way for the First Amendment to adapt to twenty-first-century challenges such as state-sponsored trolling.

Regarding the latter route, Wu sees a few possible adaptations, including these:

- Utilize the First Amendment's accomplice-liability doctrine to establish that online harassment campaigns that involve governments or politicians are a form of state action.

- Expand the state-action doctrine to encompass the conduct of major speech platforms, an option that strikes Wu as unpromising and undesirable.

- Build upon existing hate speech prohibitions that are permitted by the First Amendment, such as the federal cyberstalking statute (18 USC § 2261A).

In line with Wu's final suggestion, Tim Hwang argues for a "well-calibrated modification" of Section 230 of the Communications Decency Act of 1996 (CDA 230), a provision that shields social media platforms from legal liability for the actions of third-party users of their services (Hwang 2017). Whereas under European human rights law internet intermediaries become liable for the speech of their users under certain circumstances (Delfi AS v. Estonia 2015), no such obligation exists under US law, an omission that has been seen as a driver of innovation in online services. Hwang, considering how the active spreading of political disinformation (including, but not exclusively, by state-sponsored actors) can be countered, discounts efforts such as requiring disclosure and verification of real identities on platforms, or restricting the access of perpetrators of political disinformation to advertising platforms, as short term—and ultimately ineffective—salves. He also advocates against exempting the dissemination of falsehoods, defamatory statements, or invasions of privacy from CDA 230. Rather, he supports creating exceptions to CDA 230 for a number of existing laws, such as portions of the Federal Election Campaign Act that prohibit foreign interests from engaging in activities to shape elections, and for fraudulent activity in order to target unlabeled bots or paid agents purporting to be genuine users. And he advocates adding possible new regulations to CDA 230, such as

- requiring data brokers to enable citizens to scrutinize and opt out of their personal data being used for microtargeting, and

- requiring those involved in the collection of voter data to disclose data processing to individuals.

By removing the application of CDA 230 in these and other limited circumstances, platforms would be placed under a legal obligation to ensure compliance by users with the aforementioned laws. Such modifications, Hwang argues, "may go a long way in helping to give the public and civil society a fighting chance by encouraging platforms to stabilize and balance the marketplaces of ideas they own and operate. Of particular importance is the reduction or elimination of techniques of distribution that—regardless of the truth or falsity of the messages channelled through them—erode trust in public discourse and democratic processes."

The prospect of amending and evolving electoral regulation holds particular promise, even outside of the realm of CDA 230. Given the prominence of trolling attacks during and in the aftermath of elections, targeted policy making in the field of electoral regulation could have a significant impact on the prevalence of state-sponsored harassment campaigns, particularly those that occur cross-border. Critically, this would require ensuring that activities conducted on social media platforms that cannot be easily categorized as political advertising are brought within the ambit of regulation that restricts the amount of investment in political campaigning and that speaks to the origin and destination of campaigning funds.

Debates are already under way about how electoral regulation both within the United States and outside of it may evolve to take into account the new realities. In the US Congress, the Honest Ads Act, a bipartisan bill, is aimed at ensuring political ads sold online comply with the same rules and transparency obligations that apply to television and radio advertisements (Romm 2017). The British Information Commissioner's Office has already announced an investigation into "the use of data analytics for political purposes," responding to concerns raised about the role of foreign actors and companies in the Leave campaign for Brexit (Booth 2017).

Policies of Technology Companies

The slow pace of legal change means that the possible changes in law and regulation suggested above are unlikely to effectively stem the practice of state-sponsored trolling in the short term. In the long term, it is likely that any regulatory adaptations will once again be outpaced by technological advancements. If the law catches up, states will find new ways to weaponize digital technologies against critics and dissenters. As a result, technology companies bear not only the shared responsibility but also the sole ability to curb the practice and effects of state-sponsored harassment campaigns.

Social media platforms have long resisted the imposition of liability, and when they have voluntarily assumed responsibilities, they have done so begrudgingly. Defenders of online freedoms have been reluctant to pressure platforms to take a more proactive role in moderating and shaping the content they host, fearing that platforms will take either a heavy-handed or a too-cautious approach to content moderation, or will become compromised as a tool for state control or censorship. But as this report illustrates, social media networks are already captured, curated, and controlled—by the algorithms that underpin them and by actors who are able to operationalize them for pernicious ends. Whether they like it or not, platforms are no longer intermediaries; they take a position on the types of behavior and information they promote or suppress, through either their acts or their omissions.

As social media networks acknowledge their transformation from neutral platform to publisher and grapple with the attendant responsibilities, they have an opportunity to ensure their position is defined no longer by their omissions but instead by their acts. Those acts should include measures designed to identify and deamplify state-sponsored harassment and hate campaigns. To this end, online media companies should consider the following steps:

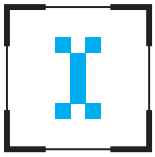
» **Detect and identify state-linked accounts.** Platforms could develop the capability to detect when an attack has its origin in a government actor or government proxy, or when a certain set of activities has links to political actors or resembles similar events, and flag such attacks for users. This would disable a key feature of state-sponsored trolling campaigns—their seemingly organic and informal nature, which both co-opts unsuspecting internet users into supporting the campaign and amplifies the effect of the attack on the target, who perceives a seemingly spontaneous groundswell of public opinion against her or him.

» **Detect and identify bots.** Detecting and identifying the existence of bots on their networks would be a simple but effective means of diluting the impact of state-sponsored trolling campaigns. Bot detection, though an inexact science, is technically possible and holds great promise for liberating online platforms from the grasp of those who wish to weaponize them. Furthermore, the sheer volume of messages is a tool that silences targets of such campaigns, and if that volume could be reduced by development of means of filtering out bots and automated messages, the impact of state-sponsored attacks would be further limited.

In the same vein, another simple fix would be for platforms to design their infrastructure to require bots or automated accounts to be identified as such by the user. Under such a proposal, bots would have a marker or warning that they are automated accounts. This would have minimal negative impact on the free flow of information, while equipping social media users to take a critical approach to content shared by an automated account.

» **Improve reporting mechanisms and responsiveness.** Social media platforms are eternally under pressure to improve mechanisms for reporting inappropriate and illegal content, and we wish to add to that pressure by reiterating that targets of state-sponsored trolling attacks are reliant on the actions of social networks to remove expeditiously content that has been flagged. We recognize that automatic removal of flagged content is not consistent with supporting internet users' free-expression rights, and that there is necessarily a lag between reporting and removal. However, we think platforms could go further by identifying content as "flagged" or "reported" immediately, so that other users can identify it as such during the period between reporting and removal. Such a mechanism would also assist in countering, for example, disinformation; platforms could develop a means for allowing users to contest the veracity of online content that would immediately notify other users that there had been a claim of falsity that must be verified.

Conclusion



In this report, we have sought to describe the emergence of a new form of human rights abuse: state-sponsored trolling. We define this as the use of targeted online hate and harassment campaigns to intimidate and silence individuals critical of the state. We have illustrated that phenomena previously discussed in isolation—such as the use of political bots to amplify campaigns, concerns over privacy, extralegal hacking of opposition, and viral disinformation—can combine and metastasize into vitriolic campaigns at scale that target individuals with pinpoint personalization afforded by modern digital technologies. We also prescribe a new attribution framework for holding responsible parties accountable for attacks, even in the absence of hard forensic attribution techniques.

We move to discussing how state-sponsored trolling fits into the ambit of existing legal structures and talk about potential policy prescriptions to combat the issue. Some possibilities include the expansion of current understandings of unprotected hate speech under international and US law to include the types of online harassment and hate speech deployed as part of state-sponsored trolling attacks; the reconsideration of intermediary liability regulation to reinforce the role played by platforms in facilitating trolling campaigns; and the evolution of responsibility for technology companies to detect and identify state-linked accounts, bots, and hateful content online. We acknowledge that none of these suggestions in and of themselves addresses the entire phenomenon of state-sponsored trolling, nor are they without their problems. Nevertheless, we see value in starting a conversation about how to build a more hospitable online sphere free from state manipulation and weaponization.

This report is the first comprehensive attempt to describe the phenomenon of state-sponsored trolling from a qualitative and quantitative standpoint. While it is impossible to always tie the threads back to the ultimate perpetrators of these attacks, we humbly hope that this report is a first step toward empowering individuals, researchers, and policy makers to spot this phenomenon in the wild and attempt to combat it.

Bibliography

Abdulemam, A. (2015, February 20). Ali Abdulemam: "I Have Not Lost My Identity. I Am Bahraini." Global Voices. Retrieved November 21, 2017, from https://globalvoices.org/2015/02/20/ali-abdulemam-i-have-not-lost-my-identity-i-am-bahraini/

Abramowitz, M. J. (2017). Hobbling a Champion of Global Press Freedom. *Freedom of the Press 2017*. Retrieved from https://freedomhouse.org/report/freedom-press/freedom-press-2017#usa-essay

ABS-CBN News (2017a, January 25). Filipinos Lead the World in Social Media Use: Survey. Retrieved January 6, 2018, from http://news.abs-cbn.com/business/01/25/17/filipinos-lead-the-world-in-social-media-use-survey

ABS-CBN News (2017b, May 19). Amnesty Int'l Cites De Lima in "Human Rights Defenders Under Threat." Retrieved January 6, 2018, from http://news.abs-cbn.com/news/05/18/17/amnesty-intl-cites-de-lima-in-human-rights-defenders-under-threat

ABS-CBN News (2018, January 30). Rappler Boss Condemns "Patriotic Trolling" on Social Media. Retrieved April 9, 2018, from http://news.abs-cbn.com/video/news/01/30/18/rappler-boss-condemns-patriotic-trolling-on-social-media

Al Jazeera (2011). *The Hunt for #lilianekhalil - Marc Owen Jones*. Retrieved from https://www.youtube.com/watch?v=TgCp15kVggI

Albayrak, A., and Parkinson, J. (2013, September 16). Turkey's Government Forms 6,000-Member Social Media Team. *Wall Street Journal*. Retrieved from https://www.wsj.com/articles/turkeys-government-forms-6000member-social-media-team-1379351399

Almario-Gonzalez, C. (2017, January 20). Unmasking the Trolls: Spin Masters Behind Fake Accounts, News Sites. Retrieved April 4, 2018, from http://news.abs-cbn.com/focus/01/20/17/unmasking-the-trolls-spin-masters-behind-fake-accounts-news-sites

Amnesty International (2017, May 16). Human Rights Defenders Under Threat—A Shrinking Space for Civil Society. Retrieved from https://www.amnesty.ie/wp-content/uploads/2017/05/HRD-briefing-12-May-FINAL.pdf

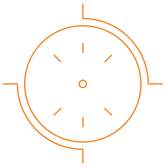
Arsenault, A. (2017, April 27). "Democracy as We Know It Is Dead": Filipino Journalists Fight Fake News. Retrieved January 6, 2018, from http://www.cbc.ca/news/world/democracy-as-we-know-it-is-dead-filipino-journalists-fight-fake-news-1.4086920

Article 19 (2015). "Hate Speech" Explained: A Toolkit. Retrieved February 4, 2018, from https://www.article19.org/resources/hate-speech-explained-a-toolkit/

Ashcoft v. American Civil Liberties Union (2002). 535 U.S. 564.

Associated Whistleblowing Press and Ecuador Transparente (2015, August 4). Ecuadorian Intelligence Agency Spied Systematically on Politicians and Activists. Retrieved November 20, 2017, from https://data.awp.is/ecuadortransparente/2015/08/04/29.html

Ayala, M., and Rochabrún, M. (2018, February 4). Ecuador Votes to Bring Back Presidential Term Limits. *New York Times*. Retrieved from https://www.nytimes.com/2018/02/04/world/americas/ecuador-presidential-term-limits.html



Bahrain Independent Commission of Inquiry (BICI) (2011, November 23). *Report of the Bahrain Independent Commission of Inquiry*. Retrieved from http://www.bici.org.bh/BICIreportEN.pdf

Bahrain Watch (2013a). The IP Spy Files: How Bahrain's Government Silences Anonymous Online Dissent. Retrieved December 11, 2017, from https://bahrainwatch.org/ipspy/viewreport.php

Bahrain Watch (2013b, October 14). Western PR Firms Compete for Fresh Multi-Million Dollar Contract with Bahrain Govt. Retrieved January 9, 2018, from https://bahrainwatch.org/blog/2013/10/14/pr-firms-compete-for-multi-million-dollar-contract-bahrain/

Ball, J. (2017). *Post-Truth: How Bullshit Conquered the World*. London: Biteback Publishing.

Ball, J., and Hamilos, P. (2015, September 24). Ecuador's President Used Millions of Dollars of Public Funds to Censor Critical Online Videos. Retrieved November 20, 2017, from https://www.buzzfeed.com/jamesball/ecuadors-president-used-millions-of-dollars-of-public-funds

BBC News (2010, March 23). Timeline: China and Net Censorship. Retrieved from http://news.bbc.co.uk/2/hi/8460129.stm

BBC News (2015, January 30). President Correa's Troll Warfare. Retrieved from http://www.bbc.com/news/blogs-trending-31057933

Beiser, E. (2017, December 13). Record Number of Journalists Jailed as Turkey, China, Egypt Pay Scant Price for Repression. Retrieved December 26, 2017, from https://cpj.org/reports/2017/12/journalists-prison-jail-record-number-turkey-china-egypt.php

B.H, M.W., H.P., and G.K. v. Austria (1989). Application no. 12774/87, European Commission on Human Rights.

Booth, R. (2017, May 17). Inquiry Launched into Targeting of UK Voters Through Social Media. Retrieved January 26, 2018, from http://www.theguardian.com/technology/2017/may/17/inquiry-launched-into-how-uk-parties-target-voters-through-social-media

Bradshaw, S., and Howard, P. N. (2017). Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation. Project on Computational Propaganda. Retrieved from http://comprop.oii.ox.ac.uk/publishing/troops-trolls-and-trouble-makers-a-global-inventory-of-organized-social-media-manipulation/

Brandenburg v. Ohio (1969). 395 U.S. 444.

Brogan, P. (1993). The Torturers' Lobby: How Human Rights-Abusing Nations Are Represented in Washington. The Center for Public Integrity.

Brooks, R. (2017, February 6). And Then the Breitbart Lynch Mob Came for Me. Retrieved October 30, 2017, from https://foreignpolicy.com/2017/02/06/and-then-the-breitbart-lynch-mob-came-for-me-bannon-trolls-trump/

Bulut, E., and Yörük, E. (2017). Digital Populism: Trolls and Political Polarization of Twitter in Turkey. *International Journal Of Communication* 11. Retrieved from http://ijoc.org/index.php/ijoc/article/view/6702/2158

Caruncho, E. S. (2016, August 28). Confessions of a Troll. Retrieved April 4, 2018, from http://lifestyle.inquirer.net/236403/confessions-of-a-troll/

Chaturvedi, S. (2016). *I Am a Troll: Inside the Secret World of the BJP's Digital Army.* New Delhi: Juggernaut.

Chen, A. (2015, June 2). The Agency. *New York Times*. Retrieved from http://www.nytimes.com/2015/06/07/magazine/the-agency.html

Chen, A. (2016, November 14). When a Populist Demagogue Takes Power. *New Yorker*. Retrieved from https://www.newyorker.com/magazine/2016/11/21/when-a-populist-demagogue-takes-power

Clover, C. (2009, March 11). Kremlin-Backed Group Behind Estonia Cyber Blitz. *Financial Times*.

CNN Philippines (2018, January 16). SEC Cancels Rappler's License to Do Business. Retrieved January 26, 2018, from http://cnnphilippines.com/news/2018/01/15/sec-revokes-rappler-license-to-operate.html

Committee to Protect Journalists (2018, March 16). Ecuador Pledges to Reform Repressive Media Law. Retrieved April 4, 2018, from https://cpj.org/2018/03/ecuador-pledges-to-reform-repressive-media-law.php

Council of Europe (2007). National Law. Retrieved June 22, 2018, from https://www.coe.int/t/pace/.../stopviolence/Source/Legislation_2007_1_vol1_E.pdf

Confessore, N., Dance, G.J.X., Harris, R., and Hansen, M. (2018, January 27). The Follower Factory. *New York Times*. Retrieved January 27, 2018, from https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html

Deibert, R. J. (2013). *Black Code: Inside the Battle for Cyberspace*. Toronto, ON: Signal.

Deibert, R., and Rohozinski, R. (2010). Liberation vs. Control: The Future of Cyberspace, *Journal of Democracy* 21(4), 43–57.

Delfi AS v. Estonia (2015). Application no. 64569/09, European Court of Human Rights.

Desmukh, F. (2011a, October 12). An Epilogue for Liliane Khalil. Retrieved January 9, 2011, from https://web.archive.org/web/20111114172637/http://chanad.posterous.com:80/an-epilogue-for-liliane-khalil

Desmukh, F. (2011b, November 10). British Intelligence Gathering Firm Assists Bahraini Regime Amidst Crackdown. Retrieved January 9, 2018, from https://web.archive.org/web/20111113091354/http://chanad.posterous.com:80/british-intelligence-gathering-firm-assists-b

Diuk, N. (2012). *The Next Generation in Russia, Ukraine, and Azerbaijan Youth, Politics, Identity, and Change*. Lanham, MD: Rowman & Littlefield.

Dooley, B. (2011, November 17). "Troll" Attacks on #Bahrain Tweets Show Depth of Government Attempts to Silence Dissent. Retrieved from https://www.huffingtonpost.com/brian-dooley/troll-attacks-on-bahrain_b_1099642.html

Ecuador Transparente (2016, November 14). Los Papeles de Godwin. Retrieved November 20, 2017, from https://ecuadortransparente.org/papeles-godwin/

El Telégrafo (2014, January 6). La NED de EE.UU. Financiará Proyecto Mediático en Ecuador. Retrieved November 20, 2017, from http://tinyurl.com/jbuhoyv

Elemia, C. (2018, January 16). Journalist Groups Hit SEC Decision vs. Rappler. Retrieved January 26, 2018, from https://www.rappler.com/nation/193707-nujp-focap-rappler-registration-sec-decision

Etter, L. (2017, December 7). What Happens When the Government Uses Facebook as a Weapon? *Bloomberg Businessweek*. Retrieved from https://www.bloomberg.com/news/features/2017-12-07/how-rodrigo-duterte-turned-facebook-into-a-weapon-with-a-little-help-from-facebook

European Commission (2016, May 31). European Commission and IT Companies Announce Code of Conduct on Illegal Online Hate Speech. Retrieved April 9, 2018, from http://europa.eu/rapid/press-release_IP-16-1937_en.htm

European Commission for Democracy Through Law (2017). *On the Measures Provided in the Recent Emergency Decree Laws with Respect to Freedom of the Media*. Strasbourg: Council of Europe. Retrieved from https://www.humanrights.ch/upload/pdf/170321_Turkey_Media.pdf

Forero, J. (2016, June 3). Venezuelans, Facing Food Shortages, Rally Behind Vilified Conglomerate. *Wall Street Journal*. Retrieved from http://www.wsj.com/articles/venezuelas-biggest-private-company-fights-for-survival-1464964360

Freedom House (2016a). Freedom on the Net 2016. Retrieved June 14, 2017, from https://freedomhouse.org/report/freedom-net/freedom-net-2016

Freedom House (2016b, November 9). Freedom on the Net 2016: Bahrain. Retrieved November 21, 2017, from https://freedomhouse.org/report/freedom-net/2016/bahrain

Freedom House (2016c, November 10). Freedom on the Net 2016: Venezuela. Retrieved October 24, 2017, from https://freedomhouse.org/report/freedom-net/2016/venezuela

Freedom House (2016d, November 11). Freedom on the Net 2016: Ecuador. Retrieved November 20, 2017, from https://freedomhouse.org/report/freedom-net/2016/ecuador

Freedom House (2017a, April 18). Freedom of the Press 2017. Retrieved January 13, 2018, from https://freedomhouse.org/report/freedom-press/freedom-press-2017

Freedom House (2017b, April 27). Freedom on the Net 2017: Philippines. Retrieved January 6, 2018, from https://freedomhouse.org/report/freedom-press/2017/philippines

Freedom House (2017c, November 14). Freedom on the Net 2017: Bahrain. Retrieved December 13, 2017, from https://freedomhouse.org/report/freedom-net/2017/bahrain

Freedom House (2017d, November 14). New Report—Freedom on the Net 2017: Manipulating Social Media to Undermine Democracy. Retrieved January 13, 2018, from https://freedomhouse.org/article/new-report-freedom-net-2017-manipulating-social-media-undermine-democracy

Garmazhapova, A. (2014, May 22). (Pa)trolling the RuNet | openDemocracy. Retrieved December 5, 2017, from https://www.opendemocracy.net/od-russia/alexandra-garmazhapova/patrolling-runet

Geybulla, A. (2016, November 22). In the Crosshairs of Azerbaijan's Patriotic Trolls. Retrieved December 5, 2017, from https://www.opendemocracy.net/od-russia/arzu-geybulla/azerbaijan-patriotic-trolls

Glimmerveen and Hagenbeek v. the Netherlands (1979). Application nos. 8348/78 and 8406.78, European Commission on Human Rights.

Goldsmith, J., and Wu, T. (2006). *Who Controls the Internet? Illusions of a Borderless World.* New York: Oxford University Press.

Granick, J. S. (2017). *American Spies: Modern Surveillance, Why You Should Care, and What You Can Do About It*. Cambridge, UK: Cambridge University Press.

Grove, L. (2016, March 1). How Breitbart Unleashes Hate Mobs to Threaten, Dox, and Troll Trump Critics. *The Daily Beast*. Retrieved from https://www.thedailybeast.com/articles/2016/03/01/how-breitbart-unleashes-hate-mobs-to-threaten-dox-and-troll-trump-critics

Guarnieri, C., Franco, J., and Anderson, C. (2017, March 10). False Friends: How Fake Accounts and Crude Malware Targeted Dissidents in Azerbaijan. Retrieved January 8, 2018, from https://medium.com/amnesty-insights/false-friends-how-fake-accounts-and-crude-malware-targeted-dissidents-in-azerbaijan-9b6594cafe60

Hafiza Kolektifi (2015). AK Troller Kimler? ["Who Are the AK Trolls?"]. Retrieved from http://hafizakolektifi.org/drupal-8.0.2/2015/10/25/ak-troller-kimler

Halvorssen, T. (2011, May 19). PR Mercenaries, Their Dictator Masters, and the Human Rights Stain. Retrieved from https://www.huffingtonpost.com/thor-halvorssen/pr-mercenaries-their-dict_b_863716.html

Handyside v. UK (1976). Application no. 5493/72, European Commission on Human Rights.

Healey, J. (2012, February 22). Beyond Attribution: Seeking National Responsibility in Cyberspace. Retrieved April 2, 2018, from http://www.atlanticcouncil.org/publications/issue-briefs/beyond-attribution-seeking-national-responsibility-in-cyberspace

Henochowicz, A. (2015, April 13). Youth Volunteers to Spread Sunshine Online. Retrieved December 7, 2017, from https://chinadigitaltimes.net/2015/04/translation-youth-volunteers-to-spread-sunshine-online/

Howard, P. N., and Woolley, S. (2016). Political Communication, Computational Propaganda, and Autonomous Agents. *International Journal of Communication* 10 (Special Issue), 20.

Hoyng, R., and Es, M. (2017). Conspiratorial Webs: Media Ecology and Parallel Realities in Turkey. *International Journal of Communication* 11, 4219–4238.

Hu, E. (2016, November 14). Responsible Data Concerns with Open Source Intelligence. Retrieved January 11, 2018, from https://responsibledata.io/responsible-data-open-source-intelligence/

Human Rights Council (2012). U.N Human Rights Council: First Resolution on Internet Free Speech. Retrieved June 22, 2018, from http://www.loc.gov/law/foreign-news/article/u-n-human-rights-council-first-resolution-on-internet-free-speech/

Hwang, T. (2017). *Dealing with Disinformation: Evaluating the Case for CDA 230 Amendment* (SSRN Scholarly Paper No. ID 3089442). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=3089442

Infobae (2017, April 28). El Régimen Chavista Lanzó la Milicia Digital para "Ganar la Batalla" en las Redes Sociales. Retrieved October 24, 2017, from https://www.infobae.com/america/venezuela/2017/04/28/el-regimen-chavista-lanzo-la-milicia-digital-para-ganar-la-batalla-en-las-redes-sociales/

International Covenant on Civil and Political Rights (1966, December 16). United Nations General Assembly Resolution 2200A (XX1).

International Press Institute (forthcoming). Report on State-Sponsored Trolling in Turkey.

Jack, C. (2017). "Lexicon of Lies: Terms for Problematic Information." Data & Society Research Institute. Retrieved from https://datasociety.net/output/lexicon-of-lies/

Janowitz, N. (2015, December 17). The Hackers Targeting Dissidents Throughout Latin America May Be State Sponsored. Retrieved November 20, 2017, from https://news.vice.com/article/the-hackers-targeting-dissidents-throughout-latin-america-may-be-state-sponsored

Jeppesen, J.-H. (2016, December 12). First Report on the EU Hate Speech Code of Conduct Shows Need for Transparency, Judicial Oversight, and Appeals | Center for Democracy & Technology. Retrieved April 9, 2018, from https://cdt.org/blog/first-report-eu-hate-speech-code-of-conduct-shows-need-transparency-judicial-oversight-appeals/

Johnson, E. (2017, June 10). Dan Scavino Is the Other @realdonaldtrump. Retrieved January 14, 2018, from https://www.politico.com/story/2017/06/10/dan-scavino-trump-social-media-profile-239381

Johnson, J. (2016, December 8). This Is What Happens When Donald Trump Attacks a Private Citizen on Twitter. *Washington Post*. Retrieved from https://www.washingtonpost.com/politics/this-is-what-happens-when-donald-trump-attacks-a-private-citizen-on-twitter/2016/12/08/a1380ece-bd62-11e6-91ee-1adddfe36cbe_story.html

Jones, M. O. (2011). Busted! Journalist Liliane Khalil Exposed. Retrieved from https://web.archive.org/web/20111108051431/http://www.marcowenjones.byethost2.com/?p=364

Jones, M. O. (2013). Social Media, Surveillance, and Social Control in the Bahrain Uprising. *Westminster Papers in Communication and Culture* 9(2), 69–92. Retrieved from https://doi.org/http://dx.doi.org/10.16997/wpcc.167

Karaveli, H. (2016, November/December). Erdogan's Journey. *Foreign Affairs*. Retrieved from https://www.foreignaffairs.com/articles/turkey/2016-10-17/erdogan-s-journey

King, G., Pan, J., and Roberts, M. E. (2016). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument. Retrieved from http://gking.harvard.edu/files/gking/files/50c.pdf?m=1464086643

Kizilkaya, E. (2015, May 14). Turkey's Ruling AKP Fields New "Digital Army." Retrieved December 26, 2017, from http://www.hurriyetdailynews.com/opinion/emre-kizilkaya/turkeys-ruling-akp-fields-new-digital-army-82384

Kurmanaev, A. (2016a, February 2). Venezuela's Biggest Firm Says Country Needs Foreign Aid. *Wall Street Journal*. Retrieved from http://www.wsj.com/articles/venezuelas-biggest-firm-says-country-needs-foreign-aid-1454449374

Kurmanaev, A. (2016b, November 20). Venezuela's Nemesis Is a Hardware Salesman at a Home Depot in Alabama. *Wall Street Journal*. Retrieved from http://www.wsj.com/articles/venezuelas-nemesis-is-a-screw-salesman-at-a-home-depot-in-alabama-1479672919

Larsen, S. (2011, February 19). Bahrain: #NickKristof Bullied on Twitter. Retrieved December 11, 2017, from https://globalvoices.org/2011/02/19/bahrain-nickkristof-bullied-on-twitter/

Lehideux and Isorni v. France (1998). Application no. 24662/94, European Court of Human Rights.

Lipton, E. (2017, June 9). White House Official's Political Tweet Was Illegal, Agency Says. *New York Times*. Retrieved from https://www.nytimes.com/2017/06/09/us/politics/dan-scavino-hatch-act-amash.html

Marczak, B. (2013a, July 31). Bahrain Govt Using Fake Twitter Accounts to Track Online Critics. Retrieved December 11, 2017, from https://bahrainwatch.org/blog/2013/07/31/bahrain-govt-using-fake-twitter-accounts-to-track-online-critics/

Marczak, B. (2013b, August 5). Is Bahrain's Government Running Extremist Accounts? Retrieved December 11, 2017, from https://bahrainwatch.org/blog/2013/08/05/is-bahrains-government-running-extremist-accounts/

Marczak, B., Guarnieri, C., Marquis-Boire, M., and Scott-Railton, J. (2014, February 17). Mapping Hacking Team's "Untraceable" Spyware. Retrieved December 5, 2017, from https://citizenlab.ca/2014/02/mapping-hacking-teams-untraceable-spyware/

Marwick, A., and Lewis, R. (2017). *Media Manipulation and Disinformation Online*. Data & Society Research Institute. Retrieved from https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf

McCaskill, N. D. (2017, January 27). Trump Backs Bannon: "The Media Is the Opposition Party." Retrieved April 3, 2018, from https://www.politico.com/story/2017/01/donald-trump-steve-bannon-media-opposition-party-234280

Mendel, T. (2010). *Hate Speech Rules Under International Law*. Center for Law and Democracy. Retrieved from http://www.law-democracy.org/wp-content/uploads/2010/07/10.02.hate-speech.Macedonia-book.pdf

Messieh, N. (2011, November 10). Is a British Firm Helping Bahrain's Ministry of Interior Monitor Online Activity? Retrieved December 12, 2017, from https://thenextweb.com/me/2011/11/10/is-a-british-firm-helping-bahrains-ministry-of-interior-monitor-online-activity/

Morales, Y. (2017, February 24). PCOO to Accredit Social Media Publishers, Users. Retrieved January 7, 2018, from http://cnnphilippines.com/news/2017/02/24/PCOO-to-accredit-social-media-publishers-users.html

Morla, R. (2015, March 25). Correa's Social-Media Troll Center Exposed in Quito. Retrieved December 7, 2017, from https://panampost.com/rebeca-morla/2015/03/25/correas-social-media-troll-center-exposed-in-quito/

Newman, M. (2011, December 6). PR Firm "Attacked" Critics of Rwandan Government. Retrieved December 9, 2017, from https://www.thebureauinvestigates.com/stories/2011-12-06/pr-firm-attacked-critics-of-rwandan-government

News.Az (2011, June 8). Ireli Youth Union Focusing on IT. Retrieved December 5, 2017, from https://news.az/articles/society/38037

Nikolayenko, O. (2012). Tactical Interactions Between Youth Movements and Incumbent Governments in Postcommunist States. *Research in Social Movements, Conflicts and Change* 34, 27–61.

Noman, H., and York, J. C. (2011, March). West Censoring East: The Use of Western Technologies by Middle East Censors, 2010-2011 | OpenNet Initiative. Retrieved January 26, 2018, from https://opennet.net/west-censoring-east-the-use-western-technologies-middle-east-censors-2010-2011

Norwood v. United Kingdom (2004). Application no. 23131/03, European Court of Human Rights.

Ohlheiser, A. (2017, October 4). The (Other) Man Behind the Curtain of Trump's Twitter Account Is Revealed Again. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/the-intersect/wp/2017/10/04/the-other-man-behind-the-curtain-of-trumps-twitter-account-is-revealed-again/

Ong, J. C., and Cabañes, J.V.A. (2018). *Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines*. University of Leeds. Retrieved from http://newtontechfordev.com/wp-content/uploads/2018/02/ARCHITECTS-OF-NETWORKED-DISINFORMATION-FULL-REPORT.pdf

Orlowski, A. (2003, September 24). India Blocks Yahoo! Groups. Retrieved January 26, 2018, from https://www.theregister.co.uk/2003/09/24/india_blocks_yahoo_groups/

OSCE: Organization for Security and Cooperation in Europe (2011). Joint Declaration of Freedom of Expression and the Internet. Retrieved June 22, 2018, from https://www.osce.org/fom/78309

PanAm Post (2015, August 10). Hacking Team Helped Ecuador Spy on Opposition Activist. Retrieved November 20, 2017, from https://panampost.com/panam-staff/2015/08/10/hacking-team-helped-ecuador-spy-on-opposition-activist/

Pearce, K. (2014). Two Can Play at That Game: Social Media Opportunities in Azerbaijan for Government and Opposition. *Demokratizatsiya* 22(1).

Pearce, K. (2015). Democratizing Kompromat: The Affordances of Social Media for State-Sponsored Harassment. *Information, Communication & Society* 18(10), 1–17. https://doi.org/10.1080/1369118X.2015.1021705

PEN International (2014, October 9). Azerbaijan: Journalist and Political Analyst Arzu Geybullayeva Threatened. PEN International. Retrieved January 8, 2018, from http://www.pen-international.org/newsitems/azerbaijan-journalist-and-political-analyst-arzu-geybullayeva-threatened/

Phillips, T. (2015, December 26). French Journalist Accuses China of Intimidating Foreign Press. *The Guardian.* Retrieved January 26, 2018, from http://www.theguardian.com/world/2015/dec/26/china-ursula-gauthier-french-journalist-xinjiang

Posetti, J. (2017, July 13). Online Harassment: Lessons from the Philippines. Retrieved January 26, 2018, from https://gijn.org/2017/07/13/fighting-online-harassment-lessons-from-the-philippines/

Posner, S. (2016, August 22). How Stephen Bannon Created an Online Haven for White Nationalists. Retrieved January 13, 2018, from http://www.theinvestigativefund.org/investigation/2016/08/22/how-stephen-bannon-created-an-online-haven-for-white-nationalists/

Poyrazlar, E. (2014, March 26). Turkey's Leader Bans His Own Twitter Bot Army. Retrieved January 26, 2018, from http://www.vocativ.com/world/turkey-world/turkeys-leader-nearly-banned-twitter-bot-army/

Presidencia de la República del Ecuador ©SECOM (2014). *Enlace Ciudadano Nro. 356 Desde Babahoyo—Los Ríos*. Retrieved from https://www.youtube.com/watch?v=fF00hydWvn8

Presidencia de la República del Ecuador ©SECOM (2015a). *Enlace Ciudadano Nro. 407 Desde el Comité del Pueblo en Quito, Pichincha*. Retrieved from https://www.youtube.com/watch?v=qI7D_BI0RWU

Presidencia de la República del Ecuador ©SECOM (2015b). *Enlace Ciudadano Nro. 408 Desde Gonzanamá, Loja*. Retrieved from https://www.youtube.com/watch?v=ZfIJLEKq4xw

Qurium Media Foundation (2017, March 10). News Media Websites Attacked from Governmental Infrastructure in Azerbaijan. Retrieved January 8, 2018, from /news-media-websites-attacked-from-governmental-infrastructure-in-azerbaijan/

Ranada, P. (2017a, February 9). Duterte Gives Online Defenders Access to Palace Events. Retrieved January 26, 2018, from https://www.rappler.com/nation/160910-duterte-online-defenders-access-palace-events

Ranada, P. (2017b, July 25). Duterte Says Online Defenders, Trolls Hired Only During Campaign. Retrieved January 6, 2018, from https://www.rappler.com/nation/176615-duterte-online-defenders-trolls-hired-campaign

Ranada, P. (2017c, August 9). Andanar Approves "Interim Policy" for Accrediting Bloggers. Retrieved January 8, 2018, from https://www.rappler.com/nation/178219-interim-policy-bloggers-pcoo-andanar

R.A.V v. The City of St. Paul, Minnesota (1992). 505 U.S.377.

Regencia, T. (2018, February 20). Duterte Bans Rappler Reporters from Presidential Palace. Retrieved April 4, 2018, from https://www.aljazeera.com/news/2018/02/duterte-bans-rappler-reporters-presidential-palace-180220130842018.html

Reporters Sans Frontières (2017a). 2017 World Press Freedom Index. Retrieved December 5, 2017, from https://rsf.org/en/ranking/2017

Reporters Sans Frontières (2017b). Philippines: Concern About Duterte. Retrieved January 6, 2018, from https://rsf.org/en/philippines

Reyes, R. R., and Millari, M. R. (2016, November 27). Money and Credulity Drive Duterte's "Keyboard Army." Retrieved January 6, 2018, from https://businessmirror.com.ph/money-and-credulity-drive-dutertes-keyboard-army/

Roldós, M. (2016, March). *State Sponsored Cyber Attacks in Theory and Practice*. San Francisco, CA. Retrieved from https://livestream.com/pemo/rightscon16day2/videos/117641538

Romm, T. (2017, October 31). Tech Titans Support More Political Ad Transparency—but Aren't Embracing This New Senate Bill. Retrieved January 26, 2018, from https://www.recode.net/2017/10/31/16579880/facebook-google-twitter-honest-ads-act-political-ads-russia

Schipani, A. (2017, March 17). Empresas Polar: A Symbol of Resistance Amid Venezuela Crisis. Retrieved October 24, 2017, from https://www.ft.com/content/9f038fae-d368-11e6-b06b-680c49b4b4c0

Scott-Railton, J., Marquis-Boire, M., Guarnieri, C., and Marschalek, M. (2015, December 8). Packrat: Seven Years of a South American Threat Actor. Retrieved October 24, 2017, from https://citizenlab.ca/2015/12/packrat-report/

Seurot v. France (2004). Application no. 57383/00, European Court of Human Rights.

Shachtman, N. (2009, March 11). Kremlin Kids: We Launched the Estonian Cyber War. Retrieved December 7, 2017, from https://www.wired.com/2009/03/pro-kremlin-gro/

Shanker, T., and Schmitt, E. (2003, February 24). Threats and Responses: Hearts and Minds; Firing Leaflets and Electrons, U.S. Wages Information War. *New York Times*. Retrieved from https://www.nytimes.com/2003/02/24/world/threats-responses-hearts-minds-firing-leaflets-electrons-us-wages-information.html

Soldatov, A., and Borogan, I. (2015). *The Red Web: The Struggle Between Russia's Digital Dictators and the New Online Revolutionaries*. New York: PublicAffairs.

Sozeri, E. K. (2015, December 14). Inside the Great Troll War Between Russia and Turkey. *Daily Dot*. Retrieved January 8, 2018, from https://www.dailydot.com/layer8/russia-turkey-missle-turkey-troll-war-twitter/

Sozeri, E. K. (2016, September 30). RedHack Leaks Reveal the Rise of Turkey's Pro-Government Twitter Trolls. *Daily Dot*. Retrieved from https://www.dailydot.com/layer8/redhack-turkey-albayrak-censorship/

Stein, J. (2017, July 27). The White House's New Communications Director Just Vowed to "Kill All" Government Leakers. Retrieved January 13, 2018, from https://www.vox.com/2017/7/27/16053278/scaramucci-anthony-new-yorker

Su, J. (2016, March 11). Zao beijing qu zhu hou zhe ming fa guo nv ji zhe gen wo men tan le tan ["After Being Deported from Beijing, This French Reporter Talked with Us"]. Retrieved January 26, 2018, from https://theinitium.com/article/20160311-international-francereporter/

Sutton, M. (2014, May 15). State Censorship by Copyright? Spanish Firm Abuses DMCA to Silence Critics of Ecuador's Government. Retrieved November 20, 2017, from https://www.eff.org/deeplinks/2014/05/state-censorship-copyright-spanish-firm-abuses-DMCA

Tan, V. (2015, November 3). Project Exile: Death Threats Keep Azeri Journalist Abroad. Retrieved January 8, 2018, from http://globaljournalist.org/2015/11/project-exile-death-threats-keep-azerbaijani-journalist-abroad/

Tashman, B. (2017, October 13). Donald Trump Thinks the Freedom of the Press Is "Disgusting." Retrieved January 13, 2018, from https://www.aclu.org/blog/free-speech/freedom-press/donald-trump-thinks-freedom-press-disgusting

Tegel, S. (2015, January 19). Is Ecuador's "anti-imperialist" president using US copyright law to censor online critics? Retrieved November 20, 2017, from https://www.pri.org/stories/2015-01-19/ecuador-s-anti-imperialist-president-using-us-copyright-law-censor-online-critics

The Economist (2016, September 10). A Conspiracy So Immense. Retrieved from https://www.economist.com/news/europe/21706536-turkeys-post-coup-crackdown-has-become-witch-hunt-conspiracy-so-immense?fsrc=scn/tw_ec/a_conspiracy_so_immense

The Economist (2017a, April 15). Turkey Is Sliding into Dictatorship. Retrieved from https://www.economist.com/news/leaders/21720590-recep-tayyip-erdogan-carrying-out-harshest-crackdown-decades-west-must-not-abandon

The Economist (2017b, July 29). How to Deal with Venezuela. Retrieved October 20, 2017, from https://www.economist.com/news/leaders/21725559-sanctions-should-target-officials-not-country-how-deal-venezuela

The Economist (2017c, July 29). Nicolás Maduro Tries to Make Thugocracy Permanent in Venezuela. Retrieved September 19, 2017, from https://www.economist.com/news/briefing/21725558-unpopular-regimes-attempt-impose-dictatorship-could-end-bloodily-nicol-s-maduro-tries

The Economist (2017d, October 19). The Virtue of Ecuador's Leninism. Retrieved October 23, 2017, from https://www.economist.com/news/americas/21730467-ecuador-shows-presidents-populist-parties-do-not-always-mess-things-up-virtue

The Quint (2015, June 5). Twitter Trolls Among #Super150 Invited by PM Modi. Retrieved January 26, 2018, from https://www.thequint.com/technology/2015/07/06/twitter-trolls-among-super150-invited-by-pm-modi

Toor, A. (2017, March 15). Massive Twitter Hijack Spreads Swastikas and Pro-Erdoğan Propaganda. Retrieved January 8, 2018, from https://www.theverge.com/2017/3/15/14932490/twitter-hack-turkey-nazi-holland-germany-erdogan

UN Human Rights Committee (2011, July). General Comment No. 34.

United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression, and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information. (2012, June 29). Human Rights Council Resolution, The Promotion, Protection, and Enjoyment of Human Rights on the Internet, A/HRC/20/L.13. Human Rights Council Twentieth Session.

Vidal, L., and Diaz, L. C. (2016, June 6). Keys to Understanding the Political and Economic Crisis in Venezuela. Global Voices. Retrieved October 20, 2017, from https://globalvoices.org/2016/06/06/keys-to-understanding-the-political-and-economic-crisis-in-venezuela/

Viñas, S., and Alarcón, D. (n.d.). *Correa vs. Crudo*. Retrieved from http://radioambulante.org/en/audio-en/correa-vs-crudo-2

Weedon, J., Nuland, W., and Stamos, A. (2017). Information Operations and Facebook. Menlo Park, CA: Facebook.

Wichtel, E. (2017, October 31). Getting Away with Murder. Retrieved April 3, 2018, from https://cpj.org/reports/2017/10/impunity-index-getting-away-with-murder-killed-justice.php

Wong, J. C., and Solon, O. (2017, August 15). US Government Demands Details on All Visitors to Anti-Trump Protest Website. *The Guardian*. Retrieved April 4, 2018, from http://www.theguardian.com/world/2017/aug/14/donald-trump-inauguration-protest-website-search-warrant-dreamhost

Woolley, S. C., and Guilbeault, D. (2017, June 19). Computational Propaganda in the United States of America: Manufacturing Consensus Online. Retrieved September 19, 2017, from http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-USA.pdf

Wu, T. (2017). Is the First Amendment Obsolete? Knight First Amendment Institute. Retrieved from https://knightcolumbia.org/sites/default/files/content/Emerging%20Threats%20Tim%20Wu%20Is%20the%20First%20Amendment%20Obsolete.pdf

York, G. (2012, February 1). Buying a Better Image: African Leaders Enlist U.S. Agencies for Pricey Reputation Makeovers. *The Globe and Mail*.

York, J. C. (2011, October 12). Twitter Trolling as Propaganda Tactic: Bahrain and Syria. Retrieved December 11, 2017, from https://jilliancyork.com/2011/10/12/twitter-trolling-as-propaganda-tactic-bahrain-and-syria/