

Individuals and organizations interested in using social media and crowdsourcing currently lack two key sets of information: a systematic assessment of the vulnerabilities in these technologies and a comprehensive set of best practices describing how to address those vulnerabilities. Identifying those vulnerabilities and developing those best practices are necessary to address a growing number of incidents ranging from innocent mistakes to targeted attacks that have claimed lives and cost millions of dollars.

The trouble with trusting technology is that there are so many ways things can go wrong; problems can occur as a result of weaknesses in the technology, mistakes made by users, or trouble intentionally caused by abusers. This is certainly the case when dealing with crowdsourcing and social media—technologies built on complex, global platforms that rely on the interaction of hundreds of thousands of people every day.

Laying the foundation for trust in social media and crowdsourcing is a three-step process. The first step is to understand the primary ways in which these technologies can be used. The second is to assess the ways in which things can go wrong while using them. The third is to develop best practices to prevent those bad things from happening.

Building trust through understanding technologies, identifying vulnerabilities, and developing best practices does not require that the technologies or the people using them be trustworthy. Instead, it enables users ranging from

Box 1. A Tale of Two Tweets

On January 29, 2013, false messages were posted to Twitter stating that the Department of Justice was investigating the computer hardware manufacturer Audience, Inc. Shortly thereafter, the company's stock value dropped sharply by more than 25 percent. Two days later, similarly faked tweets about a pharmaceutical company coincided with a sudden, 9 percent drop in its value. In both cases, the fabricated messages were sent from accounts impersonating well-known financial analysis firms.¹

Hoax Twitter Account



Real Twitter Account



Building trust through understanding technologies, identifying vulnerabilities, and developing best practices does not require that the technologies or the people using them be trustworthy. Instead, it enables users ranging from individuals to government agencies to trust that they have the information necessary to make informed decisions about when and how to use these technologies effectively.

individuals to government agencies to trust that they have the information necessary to make informed decisions about when and how to use these technologies effectively.

How to Use Social Media

The social media ecosystem includes a growing number of online platforms, such as Facebook, Twitter, and Sina Weibo (i.e., a Chinese social networking site similar to Twitter and Facebook), that enable individuals and organizations to broadcast information, engage in ongoing conversations, or passively collect information posted by users.

Broadcasting, where a message is sent to a group of users, puts the "social" in social media. Twitter is a great example of the broadcast-centric design of social media platforms. Unlike point-to-point communications systems like email or text messages, there's no *To:* field in a Twitter message; each tweet is, by default, broadcast to everyone who is "following" that user's message feed. The confusing format of many Twitter messages is the result of conventions developed to insert common point-to-point messaging features that were intentionally omitted. For example, using an @ followed by a username indicates that, even though a tweet is being shared with everyone, it is intended for a particular user (e.g., "Hey @finkd, loved the chicken picture!").

The broadcast nature of some social media platforms can lead to confusion, as took place when a Christmas dinner photo of the Zuckerberg family, including Facebook founder Mark, was "leaked" to Twitter. Mark's sister Randi posted the photo on Facebook, where it was automatically broadcast to her group of friends and made available to the friends of those friends. One of those friends-of-friends assumed that she was able to see the family photo because the picture was public. She then re-posted it to Twitter, where it quickly went viral, much to the chagrin of Randi Zuckerberg.²

Social media also enables engagement between a wide population of users. It is

Crowdsourcing technology brings together a distributed workforce of individuals in order to collect resources, process information, or create new content.

estimated that 67 percent of adults in the United States use social media, and there are currently more than 1 billion active Facebook users worldwide.³ In some situations, the engagement process is replacing the traditional means by which people communicate. This was certainly the case when a pair of young girls got lost in the storm drains beneath their hometown of Adelaide, Australia. The girls were able to obtain a cell phone signal which they used to post a distress message on Facebook. They were found after a Facebook friend forwarded their information to the local authorities.⁴

Passive collection of social media messages makes it possible to gather information from these platforms without having to broadcast information or engage with others. This can be as simple as browsing through the public content posted online or creating an account that follows other users. Some platforms offer third parties, such as online marketers, the opportunity to pay for direct access to the messages sent by social media users (Twitter refers to their third-party message feed as "The Firehose"). This access is central to the ongoing existence of many social networks that generate revenue through targeted advertising based on the information their users have shared.

This combination of uses—broadcasting, engagement, and collection—creates a wide variety of opportunities for both individuals and organizations seeking to take advantage of social media. Broadcasting creates the opportunity to build a digital audience measured in the millions. Justin Bieber, who was an unknown Canadian teenager in 2008, now has more Twitter followers than there are people in Canada.⁵

The engagement capability enables everyone from individuals to well-orchestrated brand campaigns to interact with millions of people almost anywhere in the world. Finally, the collection capability gives both users and outsiders access to the wealth of content created and distributed on these platforms.

How to Use Crowdsourcing

Crowdsourcing technology brings together a distributed workforce of individuals in order to collect resources, process information, or create new content. The implementation of a crowdsourcing system can vary widely, from complex online websites that coordinate a million simultaneous workers to low-tech, ad hoc approaches that use a shared spreadsheet.

Crowdsourced collection uses a group of workers to find and gather resources. The recent rise of crowd-funding websites like Kickstarter.com and Indigogo are crowdsourced collection systems used to find and gather money for new projects. Crowdsourced collection can also be used to identify useful information and may be performed with or without the knowledge of the individuals that make up the workforce. During the early days of the Libyan revolution, response agencies outside of the conflict zone used the photos, videos, and other messages posted on social media to build and update a shared situational awareness map.⁶ Similar efforts have sought to collect information from citizens during contested elections in countries such as Egypt, Sudan, and Kenya.

In crowdsourced information processing, a workforce takes an existing set of data and converts, identifies, or extracts information that is useful for a specific task. Crowdsourced information processing can be combined with social media to distinguish useful pieces of information (e.g. identifying damaged buildings during a natural disaster) from the significant volume of messages being posted every moment. Those tasks that are difficult or currently impossible for computers to do are particularly suited for crowdsourced information processing.

For example, online dating sites are using San Francisco-based CrowdFlower's million-person global workforce to identify risqué photos that violate their site's acceptable use policies.⁷ The task is difficult for computers to do, but can be performed in seconds by workers around the world. On the scientific front, the University of Washington has created Foldit, a crowdsourced game in which individuals identify ways that complex protein molecules are folded together. In September of 2011, members of their workforce identified the structure of a protein central to the spread of AIDS. That task, which had been unsolved for the last decade, was accomplished by the Foldit crowd in just over 10 days.⁸

Crowdsourced content creation uses the crowd to produce entirely new information. This can be accomplished by bringing together teams of individuals to tackle a problem or distributing a problem to many different individuals in search of a person capable of developing a solution. The data competition company Kaggle and the U.S. government's Challenge.gov website post challenges for users from around the world to solve problems ranging from predicting the progression of HIV infection to building better lightbulbs.⁹

Many well-known crowdsourcing efforts use a combination of collection, processing, and creation. The articles posted to Wikipedia combine original writing created by its users and a collection of citations to external sources. Both the writing and the sources are put through an editorial review process to ensure that the articles meet the site's standards. The management approaches developed by successful crowdsourcing and social media operations enable the effective use of these technologies. Understanding why these techniques are successful and how to leverage them in other situations requires an evaluation of the underlying vulnerabilities in these technologies.

Understanding What Can Go Wrong

Every negative incident involving social media or crowdsourcing can be attributed to one or more vulnerabilities in the platforms themselves, the ways in which people use them, or the technologies on which they are built. Examining several of these incidents can help to identify a few of these vulnerabilities and make it possible to understand how the interaction between multiple vulnerabilities can lead to negative consequences.

On August 25, 2011, a series of false messages were posted to Twitter and Facebook warning locals in Veracruz, Mexico, that drug gangs were kidnapping people near local schools. Multiple accounts were used to "confirm" the false information—adding further credence to the fabricated scenario. As a result, worried citizens rushed across town to find their children, causing more than 20 car accidents. The two Mexican citizens believed to have started the rumor were arrested on charges of terrorism.¹⁰ In this incident, the combination of fabricated information and false corroboration worked together to amplify and add credence to the inaccurate information.

Consider the false messages (described on page 2) that caused Audience's stock price to drop. The information fabrication was not the only vulnerability leveraged in that incident, identity impersonation played a role as well. The Twitter account from which the messages were posted, @Mudd1waters, was designed to look like the Twitter account of a well-known financial analysis firm, Muddy Waters; its staffers tweet from the account @muddywatersre. The imposter account even went as far as listing the owner of the account as Conrad Block, the founder of Muddy Waters, and using the company's logo as their profile picture.¹¹

Although these two incidents involved the willful creation of false information, actions do not have to be intentionally malicious. Take the accidental distribution of the Zuckerbergs' Christmas Day photo. In that case, Randi Zuckerberg knowingly posted the photo to Facebook, and the photo was shared according to her account's distribution control settings. Things went wrong because those controls did not apply to the people who received the message. The redistribution of the photo to Twitter removed all of Facebook's access control restrictions allowing the photo to spread across the entire Internet. The issues of distribution and redistribution control are not the only examples of vulnerabilities from that incident. The photo itself disclosed the location of several people, including billionaire Mark Zuckerberg, and placed them in a specific location at a specific time.

Although the disclosure of the Zuckerberg family Christmas photo was innocuous, several prominent U.S. Congressmen have lost their jobs following the accidental distribution of photos that were less wholesome. Their stories are chronicled on the website "The Facebook Fired," along with the stories of dozens of others who have made similar mistakes.¹² Identifying sensitive information can also be automated, as with a project from Rutgers University that uses the geographic location tags embedded in Instagram photos to identify the time and places, including private residences, where the photos were taken. The operators of the site did prevent searches for tags such as "underwear," but did not filter potentially revealing terms such as "bikini."¹³

Disclosing location and timing can have serious financial and operational consequences. In Iraq, a highly targeted mortar strike destroyed a set of Apache helicopters shortly after they had arrived at a remote operating base. The Army believes the strike became possible after soldiers uploaded geo-tagged photos of the aircraft to the Internet where they were discovered by insurgents, thereby informing the attackers where to aim their munitions.¹⁴

Information disclosure can be extremely dangerous when it reveals a user's identity. In the border town of Nuevo Laredo, a group of Mexican citizens have come together to track the activity of drug cartels operating in the area by posting on several online websites. In retaliation, four of those citizens have been tracked down, murdered, and their bodies left in public locations around the city. The bodies were accompanied by signs listing the websites on which they had been posting messages and, in two cases, the usernames they had been using.¹⁵ The Nuevo Laredo murders highlight the challenges of crowdsourcing in areas with actively hostile organizations. Taking part in the crowdsourced collection of cartel activity made the citizens a target.

The vulnerabilities in the crowdsourcing process itself can also be targeted for attack.

During the Russian parliamentary elections of 2011, the country's only independent election-monitoring organization built and deployed a crowdsourced information collection system to track reports of fraud. During the election, a video was circulated on YouTube attacking the credibility of the site and the reports it contained. The video's narrator, a young woman, introduced the system and was shown submitting a series of false reports that were then made public on the election-monitoring organization's website. The narrator used the publishing of obviously false reports as proof that none of the reports on the site could be trusted, labeling the crowdsourcing system yet another attempt by hostile nations to slander the country's leadership.16

The platforms themselves also can be directly attacked. In February of 2013, Facebook, Twitter, and numerous other high-profile technology companies disclosed that their internal systems had been compromised in a series of attacks attributed to Eastern European criminals.¹⁷ Although the companies state that no user information was lost during the breaches, another security researcher recently posted a flaw in Facebook's authentication process that could be used to gain full access to a user's account information.¹⁸

Vulnerabilities in users' computers can further exacerbate the risks of using crowdsourcing and social media. In Syria, there have been a series of cyberattacks using custom-made viruses that target activists' computers. Once infected, the computers allow attackers to Informed decision-making allows users to choose actions that take advantage of a technology's capabilities despite the risks associated with its vulnerabilities.

access the user's usernames and passwords to social media sites, Skype, and other online platforms. The stolen credentials are being used to impersonate users online in order to spread the virus and compromise other unsuspecting members of the activists' social networks.¹⁹

The incidents involving social media and crowdsourcing range from benign misunderstandings to intentional, sophisticated attacks. Each of the incidents involved one or more vulnerabilities exposed through the construction and use of these technologies. Those vulnerabilities include information fabrication. identity impersonation, audience reaction, redistribution control, account access control, information persistence, information verification, and numerous forms of information disclosure (association, physical location, online activity, possession, and identity). These are only some of the many vulnerabilities that exist in social media and crowdsourcing technologies. Although they are numerous, these vulnerabilities can be used for more than just enabling bad things to happen.

Leveraging Vulnerabilities

Understanding vulnerabilities in social media and crowdsourcing is key to identifying ways to use these technologies safely and effectively. Simply knowing the vulnerabilities associated with a technology can enable users to make informed decisions about how and when to use it. That information can be further leveraged to create best practices to guide those using the technologies through the steps necessary to systematically eliminate or mitigate those vulnerabilities—thereby preventing attacks and avoiding negative consequences.

Informed decision-making allows users to choose actions that take advantage of a technology's capabilities despite the risks associated with its vulnerabilities.

Consider the incident involving geo-tagged photos of Apache helicopters in Iraq. Not posting a geo-tagged photo would certainly have eliminated the risk; however, there are a number of other options. Removing the geo-tagged information from the photo before posting it online (mitigating the location disclosure vulnerability) would have removed the precise targeting information needed for the attack. Similarly, the successful attack could have been avoided by posting the geo-tagged photos after the helicopters had left the base, thus eliminating the timing disclosure vulnerability associated with the Apaches.

Professional crowdsourcing organizations are successful because of the extensive

work they have done to identify and mitigate the vulnerabilities associated with their technologies. These approaches range from periodically testing workers by asking them questions for which the answer is known and only considering a task complete when multiple, independent workers agree on the answer. For example, CrowdFlower's photo-vetting system optimizes judgments from its workforce using a verification algorithm where five independent workers must identify if an image violates one or more of the platform's rules for objectionable content and each worker's decision is weighted according to how accurate he or she has been in the past.²⁰ Similarly, Wikipedia maintains the ongoing accuracy of its articles in part through a process that alerts reviewers immediately after an article they oversee has been changed so they can verify the updates are accurate.²¹

Although there is certainly a wide range of vulnerabilities associated with each technology, that number is finite, and the same vulnerabilities come up again and again in many different types of deployments. The limited number and tendency toward repetition mean that organizations need not rediscover those vulnerabilities each time they set out to use these technologies.

From Vulnerability Assessment to Best Practices

Like any technology, crowdsourcing and social media have a variety of vulnerabilities. The challenge is not to avoid these technologies because of their vulnerabilities, but understand those vulnerabilities on order to identify the steps that can be taken to effectively address the risks.

Many of the steps taken to avoid commonly found vulnerabilities can be converted into best practices in social media and crowdsourcing. The best practices would be a set of existing, ready-to-use guidelines that include information about both potential vulnerabilities and the actions that can be used to avoid them.

The best practices for social media would include guidance on ways to identify sensitive types of information that should not be disclosed and ways to detect and respond to inaccurate or fabricated information. Best practices for crowdsourcing would include

Like any technology, crowdsourcing and social media have a variety of vulnerabilities. The challenge is not to avoid these technologies because of their vulnerabilities, but understand those vulnerabilities on order to identify the steps that can be taken to effectively address the risks. mechanisms for ensuring accurate judgments from workers and could leverage many of the practices already in use by professional crowdsourcing organizations.

The two sets of information-vulnerabilities and best practices-can be used in tandem. Each best practice can be categorized according to which vulnerabilities it addresses. Users who have used the information from the vulnerability assessment to identify the vulnerabilities in their operations can cross-reference that list with the best practices they should use.

The cumulative result of identifying vulnerabilities and developing best practices will be a body of knowledge capable of protecting users in each of the many ways these technologies are being used.

The Path Forward

One of the best ways to ensure that the use of complex technologies goes smoothly is to address all of the things that can go wrong. Social media and crowdsourcing offer new opportunities for everyone from individuals to large organizations, but a growing number of incidents, ranging from benign misunderstandings to millions of dollars in damages and the loss of life, demonstrate the risks involved in using these technologies.

Despite those risks, it is possible to develop a form of trust in these technologies. Doing so requires identifying and understanding the vulnerabilities in the technologies and establishing best practices to reduce or eliminate the chance that these vulnerabilities will be exploited. Much work remains to be done to arrive at that point—both in the assessment of the vulnerabilities and the creation of best practices. This work is essential to the safe and effective use of these powerful new technologies.

Notes

- Eleazar Melendez, "Twitter Stock Market Hoax Draws Attention of Regulators," *Huffington Post*, February 2, 2013, http://www.huffingtonpost.com/2013/02/01/ twitter-stock-market-hoax_n_2601753.html.
- "Mark Zuckerberg's sister Randi complains of privacy breach after photo she posted of her family jokingly reacting to new 'Poke' application gets leaked," *Daily Mail Reporter*, December 27, 2012, http://www. dailymail.co.uk/news/article-2253345/Zuckerbergfamily-photo-leaked-Randi-Zuckerberg-need-update-Facebook-privacy-settings.html.
- Joanna Brenner, "Social Networking," Pew Internet, February 14, 2013, http://pewinternet.org/ Commentary/2012/March/Pew-Internet-Social-Networking-full-detail.aspx; Somini Sengupta and Nick Bilton, "A Billion Users Raise Stakes at Facebook for Revenue," *The New York Times*, October 4, 2012, http://bits.blogs.nytimes.com/2012/10/04/facebookpasses-1-billion-active-users/.
- "Trapped girls call for help on Facebook," ABC News, September 9, 2009, http://www.abc.net.au/ news/2009-09-07/trapped-girls-call-for-help-onfacebook/1420352.
- Misty Harris. "Justin Bieber now has more Twitter followers than Canada has people," *The Province*, February 25, 2013, http://www.theprovince.com/sports/ Justin+Bieber+more+Twitter+followers+ than+Canada+people/8015168/story.html.
- Neal Ungerleider, "Here's a Map of the Humanitarian Crisis Hotspots in Libya (Don't Tell Gaddafi)," Fast Company, March 9, 2011, http://www.fastcompany. com/1736822/heres-map-humanitarian-crisis-hotspotslibya-dont-tell-gaddafi.
- "Crowdsourced Image Moderation: Real Time Foto Moderator is a simple API for content moderation,"

CrowdFlower, Accessed March 27, 2013 http:// crowdflower.com/rtfm.

- Elizabeth Armstrong Moore, "Foldit game leads to AIDS research breakthrough," CNet, September 19, 2011, http://news.cnet.com/8301-27083_3-20108365-247/ foldit-game-leads-to-aids-research-breakthrough/.
- Alina Dizik, "Kaggle's Anthony Goldbloom Helps Companies Crunch Data With Crowdsourcing for Quant Geniuses," Fast Company, October 31, 2011, http://www.fastcompany.com/1789736/ kaggles-anthony-goldbloom-helps-companies-crunchdata-crowdsourcing-quant-geniuses; Wyatt Kash, "Government Challenge Programs Foster New Wave Of Low Cost Innovations," AOL Government, October 5, 2011, http://gov.aol.com/2011/10/05/governmentchallenge-programs-foster-new-wave-of-low-costinnova/.
- Julian Miglierini, "Mexico 'Twitter terrorism' charges cause uproar," BBC News, September 6, 2011, http:// www.bbc.co.uk/news/world-latin-america-14800200.
- "Twitter in muddy waters: hoax moves stock price 25%," Peerreach, February 4, 2013, http://blog.peerreach. com/2013/02/twitter-in-muddy-waters-hoax-movesstock-price-25/.
- 12. *The Facebook Fired*, Accessed March 27, 2013 https:// thefacebookfired.wordpress.com/.
- Jamie Condliffe, "This Instagram-Street View Mash-Up Is a Stalker's Wet Dream," Gizmodo, January 2, 2013, http://gizmodo.com/5972425/this-instagram+streetview-mash+up-is-a-stalkers-wet-dream.
- Cheryl Rodewig, "Geotagging poses security risk," *The* Official Homepage of the United States Army, March 7, 2012, http://www.army.mil/article/75165/Geotagging_ poses_security_risks/.

- "Facts also fall victim in Mexico 'social media' killings," Los Angeles Times, November 11, 2011, http:// latimesblogs.latimes.com/world_now/2011/11/socialmedia-mexico-killings-victim-confusion.html.
- "Exposing the 'Voting Violations' of the 'Voices," YouTube, uploaded on January 8, 2012, http://www. youtube.com/watch?v=Xgu38tM47XY.
- Michael Riley and Adam Satariano, "Malware Attack on Apple Said to Come From Eastern Europe," Bloomberg, February 19, 2013, http://www.bloomberg.com/ news/2013-02-19/apple-says-a-small-number-of-maccomputers-infected-by-malware.html.
- Christopher Brook, "Facebook Patches OAuth Authentication Vulnerability," Threat Post, February 26, 2013, https://threatpost.com/en_us/blogs/facebookpatches-oauth-authentication-vulnerability-022613.
- 19. Ben Brumfield, "Computer spyware is newest weapon in Syrian conflict," CNN, February 17, 2012, http:// edition.cnn.com/2012/02/17/tech/web/computervirus-syria/index.html and Nicole Perlroth, "Software Meant to Fight Crime Is Used to Spy on Dissidents," The New York Times, August 30, 2012, http://www. nytimes.com/2012/08/31/technology/finspy-software-istracking-political-dissidents.html.
- 20. Liz Gannes, "CrowdFlower Heads Downmarket With New Photo Moderation Tools," May 7, 2012, http:// allthingsd.com/20120507/crowdflower-headsdownmarket-with-new-photo-moderation-tools/.
- "Editorial Oversight and Control," Wikipedia, January 21, 2013, http://en.wikipedia.org/wiki/Wikipedia:Editorial_ oversight_and_control.

GEORGE CHAMALES is a security expert with over a decade of experience in the defense of computer systems for government, corporate, and humanitarian groups. His security work includes vulnerability assessments of nationally significant critical infrastructure systems with the DOE, development of new techniques to address malicious hackers with DARPA and evaluating the security of international corporations with Deloitte. His work in the humanitarian sector includes supporting the UN during the Libyan revolution, data fusion activities in the Afghanistan, and the creation of crowdsourcing capabilities for the US Navy. In addition, he has provided technical leadership to teams in Pakistan, Sudan, Somalia, Haiti and Egypt.

Email: George@roguegenius.com Website: http://roguegenius.com

STUDY DIRECTOR: Lea Shanley, Commons Lab, Wilson Center **EDITORS:** Lea Shanley and Aaron Lovell, Wilson Center

THE WILSON CENTER, chartered by Congress as the official memorial to President Woodrow Wilson, is the nation's key non-partisan policy forum for tackling global issues through independent research and open dialogue to inform actionable ideas for Congress, the Administration and the broader policy community.

Science and Technology Innovation Program

One Woodrow Wilson Plaza 1300 Pennsylvania Ave. NW Washington, DC 20004-3027 (202) 691-4000

THE COMMONS LAB advances research and non-partisan policy analysis on emerging technologies that facilitate collaborative, science-based and citizen-driven decision-making. New tools like social media and crowdsourcing methods are empowering average people to monitor their environment, collectively generate actionable scientific data, and support disaster response.



http://CommonsLab.wilsoncenter.org



http://bit.ly/CommonsLabVideo



@STIPCommonsLab



/CommonsLab



http://bit.ly/CommonsLabReports



W | Wilson Center 🛛 🏈 COMMONS LAB