Disruption and Deception in Crowdsourcing: Towards a Crowdsourcing Risk Framework

Agnieszka Onuchowska
University of South Florida
aonuchowska@mail.usf.edu

Gert-Jan de Vreede
University of South Florida
gdevreede@usf.edu

Abstract

While crowdsourcing has become increasingly popular among organizations, it also has become increasingly susceptible to unethical and malicious activities. This paper discusses recent examples of disruptive and deceptive efforts on crowdsourcing sites, which impacted the confidentiality, integrity, and availability of the crowdsourcing efforts' service, stakeholders, and data. From these examples, we derive an organizing framework of risk types associated with disruption and deception in crowdsourcing based on commonalities among incidents. The framework includes prank activities, the intentional placement of false information, hacking attempts, DDoS attacks, botnet attacks, privacy violation attempts, and data breaches. Finally, we discuss example controls that can assist in identifying and mitigating disruption and deception risks in crowdsourcing.

1. Introduction

Organizations can use crowdsourcing to take "a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call" [27]. Crowdsourcing has emerged as a viable alternative business model that focuses on problem solving and production provided by the distributed network of individuals [8]. It potentially has many benefits [26][7][12][6]: It can be more cost-effective than having traditional employees perform certain tasks. It enables organizations to get access to a wide and varied collection of opinions and ideas, which can reduce bias in decision-making. It allows organizations and governments to directly engage with customers and citizen. Although crowdsourcing is still evolving as an organizational and societal phenomenon, its potential demonstrated is by data from the crowdsourcing market: 15 major crowd service providers almost tripled their revenues from US$140.80Min 2009 to US$375.70M in 2011 and the global enterprise crowdsourcing market growth rate reported 75% growth in 2011 compared to 53% in 2010 [33].

Yet, several challenges threaten the usefulness of crowdsourcing as a reliable organizational problem solving approach. For example, there are challenges concerning the ownership of crowdsourced products or the perceived lack of quality standards related to crowdsourced goods or services [31][17]. Moreover, recently crowdsourcing sites have emerged with the intention of causing harm online. A rapid increase in malicious crowdsourcing service sites (also known as crowdturfing sites) has been observed in countries like China, the US, and India [50]. Such sites recruit individuals that for a small payment post false negative restaurant reviews, write biased political comments, or post false advertising [51]. There are also examples of legitimate initiatives that have been attacked by individuals seeking to achieve profits from exploiting the crowd-sourcing ventures: In 2016 users posted false reports of blocked road traffic in their neighborhoods on a crowdsourced app Waze to deflect some of the traffic flow from the places where their lived [24].

As crowdsourcing is increasingly becoming one of the ways in which organizations execute projects and support decision-making, disruptive and deceptive use of and responses to crowdsourcing initiatives need to be better understood and mitigated. It is unclear what harm might be caused to individuals and organizations by potential deception in crowdsourcing. For example, scholars are also unsure how disruptive the effects of crowdsourcing pranks and deceptions are on organizations. It is also unclear what are the physical or emotional effects of deceptive or violated crowdsourcing efforts on its contributors or beneficiaries.

Both crowdsourcers and crowdsourcing providers must be aware of existing threats and threats that may develop in the future. Currently, the biggest challenge that crowdsourcing providers and consumers faces concerns the number of crowd participants whose malicious behavior is difficult to detect and control [12]. Thus, the primary motivation for this study is to understand intentional disruptive and deceptive behavior in crowdsourcing contexts. Our main objective is to identify emerging risks that are related to crowdsourcing deception. We discuss recent examples where crowdsourcing websites were attacked or abused by malicious activities. Finally, we propose an organizing framework of risks that result from these security violations and deception cases.

This paper is structured as follows. First, we discuss previous research on disruption and deception

URI: http://hdl.handle.net/10125/50386
ISBN: 978-0-9981331-1-9

in crowdsourcing environments. Next, we present our study's method. We then present the categorization of the identified disruption and deception incidents and analyze the identified cases using defined risk pattern clusters according to the CIA triad. We discuss potential controls ways to mitigate the identified issues. Finally, we discuss the implications of the study, its limitations, and directions for future research.

2. Background

Past research on cyber threats and deception in crowdsourcing has focused on different issues. For example, Dwarakanath et al. [14] focused on crowd trustworthiness in crowdsourced software development initiatives. They proposed a taxonomy of trustworthiness and existing methods to build trust in a crowd. They showed that macro-tasks that require specific skills are related to a high level of untrustworthiness. They also found that workers' poor reputation had a strong impact on trustworthiness. Other findings showed little correlation between monetary benefits and trustworthiness.

Stefanovich et al. [43] analyzed the ability of crowdsourcing systems to cope with attacks. Based on the data collected from DARPA's Shredder Challenge, the researchers identified attack mechanisms and analyzed how users recover from such attacks. They argued that while participants can recover from errors in the long term, the attacks still affected participants: after being attacked, participants develop a notion of not being able to influence malicious behavior. It thus appears that victims of malicious behavior often suffer from motivational challenges to get involved in tasks. Consequently, their task efficiency tends to drop as well. Lasecki et al. [31] found similar motivational challenges as a result of crowdsourcing threats: they found that the more malicious tasks appear in a crowdsourcing environment, the less willing users are to participate in such tasks. Their study further showed that even simple tasks are subject to online manipulation and can be a target for information distraction.

Harris [22] raised the issue of ethics in crowdsourcing design and analyzed the examples of crowdsourcing initiatives that intentionally would not conform to ethical standards. In his study, he examined how crowdsourcing contributes to population of unethical behaviors and activities such as posting fake online reviews. In the same online reviewing context, Fayazi et al. [16] investigated how to uncover crowdsourced manipulation. The researchers created a sampling method to track down items, which received manipulated online reviews. Their method also aims at identifying and removing users who post false reviews from affected crowdsourcing platforms. Similarly, Chen et al. [10] described crowdsourcing tasks where the provision of new information (exploration) is often distorted and exploited, because the verification part of a crowdsourced task is neglected. The researchers built an agent-based model that helped them balance exploitation and exploration in crowdsourced search tasks where time plays a critical role.

Crowdsourcing users can also create fake identities (Sybils), which, when multiplied, can be used to boost the perpetrators' reputation. Cheng et al. [11] presented a frame-work to assess robustness of reputation mechanism to Sybils. Like flagging fake identities, researchers have also focused on flagging fake or inappropriate content. Kayes et al. [29] built a classifier that detects abusive users of community-based question/answering platforms. Their findings show that flagging suspicious content by the users works effectively as a crowdsourced monitoring function. Moreover, the researchers found that flagged users received more attention but often were not perceived as toxic to the community. On the contrary, flagged users who posted questions received a higher than average level of response compared to questions posted by ordinary users.

Finally, Wolfson et al. [53] analyzed data security-related areas, where crowdsourcing and the law are likely to intersect in the near future. They discuss possible challenges related to the deployment of crowdsourcing and claims that crowdsourcing misuse is highly tied to issues with data security.

As can be seen, research on cyber threats and deception in crowdsourcing is fragmented. Apart from the separate studies on cyber and deception threats to crowdsourcing discussed above, we were unable to identify any study exploring the scope of the problem or addressing the problem from a holistic perspective by taking a full range of threats into consideration. The next section describes our research method to collect and analyze practical examples of cyber threats and deception in crowdsourcing in order to develop a holistic framework of the different types of such threats.

3. Method

The development of a structured overview of cyber threats and deception in crowdsourcing took place in three steps. Each step is detailed below.

3.1 Crowdsourcing organizing framework

We first conceptualized a crowdsourcing effort in terms of the relevant elements that can be distinguished. The purpose of this conceptualization is to collect relevant information about examples of malicious crowdsourcing efforts. Thus, our conceptualization is an example of a Theory for Analyzing, which is "used to classify specific

dimensions or characteristics… by summarizing the commonalities found in discrete observations… when nothing or very little is known about the phenomenon in question" [21]. Our conceptualization is based on past definitions of crowdsourcing (e.g. [27]), previous models (e.g. [38]), and past research (e.g. [14]).

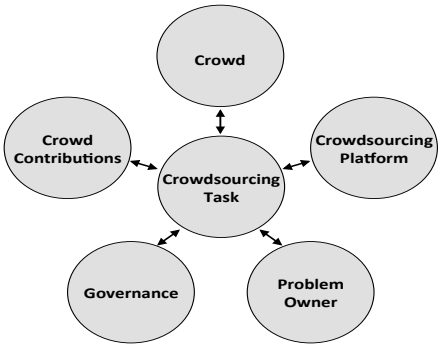

Figure 1. Conceptualization of a crowdsourcing effort.

Our conceptual model of a crowdsourcing effort consists of the following elements (Figure 1):

- Crowdsourcing Task: The work assignment for which contributions are solicited. Crowdsourcing tasks are also referred to as problems or challenges. We captured information that would characterize the task that was at the center of a deceptive crowdsourcing incident.
- Crowd: The individuals (crowd workers) who perform the task. In the deceptive crowdsourcing context, crowd-related issues concern malicious human behavior as well as the identification of mechanisms that might initiate such behavior.
- Crowdsourcing Platform: Connects the crowd and problem owner. When capturing crowdsourcing deception examples, we focus on risks and vulnerabilities related to the platform or its management.
- Problem Owner: Defines a task, posts it on a platform, and provides data and tools for task completion. For our analysis, we focus on information related to the problem-owner's context concerning risks that were initiated by problem owners or caused by their negligence.
- Governance: The policies, reward structures, and moderation of the crowdsourcing effort. We collected information on the governance mechanisms used in the incidents and related vulnerabilities, e.g. lack of quality control.
- Crowd Contributions: Outputs from crowd members when they have completed their tasks. We captured information regarding potentially disruptive or deceptive input, e.g. intentionally false information, and its' effects on the crowdsourcing results.

3.2 Identification of deceptive crowdsourcing incidents

Next, we conducted a search for examples of deceptive crowdsourcing incidents. We scoped our search as follows. First, we focused on examples from academic or practice publications, including web publications, from the last 10 years. As crowdsourcing is a relatively young phenomenon, we did not expect to find examples that go back earlier than 2007. Second, we focused only on legitimate (legal) crowdsourcing efforts that had become target or victim of malicious intent. During our search, we found examples of a distinct type of malicious crowdsourcing called "crowdturfing" [50][51], where the effort is purposefully organized to cause harm. Crowdturfing problem owners reward a crowd to perform malicious activities, e.g. putting negative comments on competitors' products' websites. Yet, our focus is on crowdsourcing efforts that fall victim of malicious activities, such as cybersecurity breaches or deception.

To find examples of incidents, we performed a wide search using academic (Google Scholar, ABI/INFORM Global) and general (Google) online search engines. We determined search terms in two steps: We first identified search terms that we felt would best reflect our phenomenon of interest. Then, from the relevant publications we found, we identified additional synonyms and other content search terms. We performed a final search with the additional terms to find further publications. Our final search terms list includes terms such as "crowdsourcing deception", "crowdsourcing threats", "cybersecurity crowd-sourcing", "open innovation challenges", "online labor market challenges", "public participation gone bad", "online contest gone bad", "hacking of crowdsourcing webpages", "crowdsourcing controversy", and combi-nations of constituent terms. For each example that we identified, we collected available information per the elements of our conceptual model. As our objective was not to create a comprehensive library of every reported incident, we did not capture information on incidents that could be considered identical in nature as others that we recorded already. For example, descriptions of malicious prank attempts on crowdsourcing websites are relatively common so we only included a few in our set of incidents.

3.3 Extracting deceptive crowdsourcing types

For the final step we looked for commonalities among the different incidents that we recorded to identify distinct types of cyber threats to crowdsourcing. Next, we used the Confidentiality-Integrity-Availability (CIA) triad to sort the threat types in terms of what part of the triad was at risk. We used the CIA triad as this one of the most popular cybersecurity frameworks.

Next, we defined specific risks involved in each type of crowdsourcing threat and for outlining examples of controls that can assist in preventing the risk or mitigating the risk's impact when it materializes.

4. Results

4.1 Crowdsourcing incidents

Our search of publications and online sources yielded a variety of real-case incidents that are detailed in Appendix A. We included 18 incidents in this overview; similar incidents were left out due to page limitations. For each incident we captured relevant information from the publication or online source according to the elements of our conceptual model.

The overview of the incidents shows that half of the Confidentiality-related incidents were related to privacy violations caused by the placement of sensitive data online. All incidents related to Integrity breaches were caused by the placement of false information intending to gain profit, willingness to make other participants lose, or disruption of crowdsourcing initiative using prank information. The only malicious incident type identified as an Availability-issue was related to DDoS attacks, which intend to disrupt the secure flow of information. Table 1 gives an overview.

We realize that the identified collection of malicious and deceptive activities is unlikely to be exhaustive. Companies that fell victim to deceptive crowdsourcing efforts may be reluctant to publicly share information about the experience. Still, the collected incidents show an interesting pattern. We observe that all cases were either related to the placement of false information, sharing of sensitive information, the disruption of secure information flow or intentional deceptive action aimed at undermining crowdsourcing initiatives. User deception by provisioning of false information, disrupting the content of correct information or denying access to correct information appeared to be most common.

4.2 Confidentiality risks and controls

A further analysis of the incidents that are related to confidentiality resulted in two types of malicious crowdsourcing risks. The first concerns the hacking of crowdsourcing sites. Crowdsourcing sites appear to get hacked for the same reasons as other online services, e.g., theft of sensitive data such as user names, email addresses, and shipping addresses. For example, criminals hacked Kickstarter to collect sensitive data including passwords, phone numbers, email addresses, and credit card details. The crowdfunding site Patreon was hacked for user names and their addresses. When considering the risks related to hacking crowdsourcing sites, there are at least two consequences to the disclosure of sensitive and private data. First, there may be financial implications due to penalties that the site owner might need to pay if the information gets hacked. Second, there will be reputational damage, as the hack will cause a loss of trust among existing and future crowd members and problem owners. Controls for this risk should aim to prevent successful hacking attempts and, if a hack has taken place, limit its impact. To mitigate hacking-related risks, the crowdsourcing field should focus on technological measures (e.g. regular platform patching, encryption of information stored on website servers), behavioral measures (e.g. implement strong password policies), and governance measures (e.g. recovery plans, communication plans, and limited collection of personal information).

The second type of Confidentiality risks is related to privacy violations through voluntary sharing of personal data. For example, Netflix released anonymized user records as part of its "Netflix Prize" Contest. The PatientsLikeMe website users' personal data on taken medications and illness symptoms was scraped by Nielsen and used for further business analysis. Both incidents refer to the disclosure of sensitive data, which can cause risks such as the violation of privacy of crowd members, harm the site's reputation, and result in penalties imposed on website owners.

Examples of controls to prevent privacy violation cases include the verification of information that is sourced from non-official channels, regular education of users on ways to protect sensitive data or enabling corrective forms on the site so that users who spot inappropriate data can flag it accordingly.

4.3 Integrity risks and controls

Our analysis of the integrity-related incidents revealed four distinct types. The first type concerns the intentional placement of false information with the goal to increase the perpetrator's gains and expose victims to potential harm. In this situation, victims are often unaware that they have been wronged. For example, when someone enters false information on a traffic site to move traffic away from their neighborhood, others' trips may be rerouted causing unnecessary delays that they are unaware of. Risks related to this type include the crowdsourcing site being perceived unreliable over time causing users to cease contributing to it. Controls should primarily focus on preventing false information being submitted and minimizing the consequence of false information. Examples of such controls include the analysis of recurring patterns for data input, analysis of outliers or inconsistent data, cross referencing of entered information with other data sources, and policies focus on swift removal of user profiles who were caught entering false information.

Table 1 Identified risks types and controls objectives

Type	Risk description	Control objective	Control examples
C1. Hacking crowdsourcing site (e.g., theft of private information)	Loss of trust among stakeholders; financial loss; reputational loss impacting future investors decisions; penalties if the information gets stolen	Reduction of successful hacking attempts. Reduction of hacking impact/recovery planning effort	Regular patching of platform software and user applications that support the crowdsourced service; strong password policy; continuous education of all stakeholders on avoiding hacker attacks; recovery plans/ communication plans, regularly reviewed; encryption of private information; reduce amount of PII data collected
C2. Privacy violation	Legal privacy violation due to sharing of unverified or sensitive information between the crowd members; reputational loss; regulatory or legal penalties	Prevention of privacy violation cases	Verification of information sourced from non-official channels; prevention of sensitive data sharing; providing forms to report inappropriate data; regular user education on sensitive data protection
I1. Placement of false information (e.g., rerouting of the traffic) – perpetrator gains, victim is exposed to potential harm or might not be aware that has been victimized	Website becomes unreliable, users might not be willing to use the service any more	Preventing the spread of false information; Minimization of the damage brought by false information Keeping the data consistent and accurate – in case the issues are found the data is rectified	Analysis of recurring patterns for data input; analysis of outliers/ inconsistent data; identification of instances where the entry of false data is possible; monitoring and assessment of crowd behavior; flagging type of controls – people can report false information; cross referencing of entered information with other data sources, e.g., traffic reports; introduction of policies that allow the removal of users who were caught with entering false data; introduction of delay between collecting the information and posting it on the website; limiting the number of entries from a single user/ single area and reviewing the information that comes from the same source
I2. Placement of false information to make others lose or to cause harm to others (e.g., competitions)	Website becomes unreliable, users might not be willing to use the service anymore; Financial losses are likely to follow	Preventing the input and/or the dissemination of false information; minimization of disinformation efforts; keeping the data consistent and accurate	Monitoring and assessment of crowd behavior; monitoring of inconsistencies in information inputs; online monitoring – statistical sampling techniques; allowing participants to raise concerns in case they spot problems on a webpage; introduction of immediate action that follows false information entry, e.g., perpetrators are banned and their profiles are removed from the system; verification of identity before users post information or take part in competition
I3. Botnet attack to bring false results, e.g. placement of false traffic information	Website becomes unreliable, users might not be willing to use the service anymore; Financial losses are likely to follow	Keeping the data consistent and accurate	Continuous automated data checks, recognition of suspicious patterns; implementation of technological solutions that identify or prevent automatic scripting; use of analytics to recognize patterns, to analyze where the site is approached from and to monitor Internet traffic
I4. Prank activity in crowdsourcing contests	The crowd that places wrong or erroneous information 'for fun' puts the organizer at risk of needing to retract the contest. Financial losses and damage of the organizer's credibility are likely to occur	Keeping the data consistent and accurate	Monitoring and assessing crowd behavior; verification of user's identity before they can post information on a website or take part in an online competition; remove false information; banning people; enabling of possibility to flag suspicious content by users
A1. Botnet attack (DDoS)	DDoS attacks negatively impact availability and performance of crowdsourcing sites and can lead to negative financial consequences as well as destroy the image of the site	Prevention of DDoS attacks; early detection of DDoS attacks; reduction of impact of DDoS attacks	Reactive: Event monitoring of hosting infrastructure, rerouting of unwanted internet traffic, continuous automated data checks, recognition of suspicious network flow patterns; Proactive: installation of firewalls, monitoring the history of network flows, deployment of multiple Internet Service Providers, recovery plan in place in case the attacks cannot be prevented

The second type concerns the placement of false information to make other crowd members lose or to cause harm. This type occurred several times at DARPA crowdsourced competitions. The competition results were affected as perpetrators entered false information to frustrate other competitors' work on the crowdsourced tasks. The key risk related to such action concerns crowdsourcing efforts becoming unreliable so that users become reluctant to use the service. Controls to mitigate this risk focus on preventing the dissemination of false information, minimizing disinformation efforts, and maintaining consistency and accuracy of information that has already been uploaded. Control activities include monitoring and assessing crowd behavior, monitoring for inconsistencies in contributions, and assessing selected contributions using statistical sampling techniques. Also, controls could be considered that allow crowd members to raise concerns when they spot problems. Individuals making false contributions should be banned immediately and their profiles and contributions should be removed from the system. Another control concerns identity verification before users are allowed to post information or enter in a competition.

The third type concerns the placement of false information on a crowdsourcing site through botnet attacks. This type is different from the previous two because of the technological nature of this type of threat. Waze was hit by a botnet attack when students created false automated profiles to influence traffic information and reroute the traffic. The main risk related to malicious automated behavior is that websites become unreliable so that users no longer use the services anymore. Possible controls to mitigate botnet attacks include analytics to recognize patterns in registration and data input and technological solutions that identify or prevent automatic scripting.

The last Integrity-type concerns prank activities. Crowd members may contribute wrong, erroneous, or even objectionable information "for fun". This may force the problem owner to retract the contest. Consequently, financial losses and damages to the organizer's credibility are likely to occur. The following controls can assist in keeping data consistent and accurate: monitoring and assessment of crowd behavior, user identity verification before allowing participation in the contest or making contributions, swift removal of false information, and enabling users to flag suspicious content by other users.

4.4 Availability risks and controls

The only availability type that emerged from our collection of incidents concerns DDoS attacks caused by botnets. An example is project on Crowdsourced Satellite Imagery, which was taken down by a botnet DDoS attack. Risks related to DDoS botnet attacks include limited availability and performance of the crowdsourcing site resulting in financial and reputational loss. Reactive controls include event monitoring of the hosting infrastructure, rerouting unwanted Internet traffic, continuous automated data checks, and monitoring suspicious network flow patterns. Proactive control examples include firewalls, monitoring the history of network flows, deploying multiple ISPs, and creating recovery plans in case of successful attacks.

5. Discussion and conclusions

The purpose of this study was to identify emerging risks related to disruptive and deceptive behavior in crowdsourcing contexts. Through an extensive search of academic and online sources, we gathered recent examples of incidents where crowdsourcing sites were attacked or abused. We identified the CIA element at risk for each incident. We observed that most unique real-life incidents were related to intentional deceptive actions such as placement of prank information, placement of false information, unauthorized disclosure of sensitive information or disruption of the secure flow of information. Based on the collected incidents, we derived an organizing framework of distinct risk types concerning disruptive and deceptive crowdsourcing. The framework consists of two risk types for Confidentiality issues, four risk types for Integrity issues, and one type for Availability issues. The framework further outlines possible controls to mitigate the risk types in terms of preventing the risks or containing the damage once the risks materialize.

Our findings have implications for research in crowdsourcing deception and cybersecurity. Crowdsourcing threats appear to be predominantly related to social engineering attacks. Descriptions of deceptive activities dominated in our collection of incidents. This shows a need for more in-depth research on human factors related to the motivations and behaviors of cybercriminals and cyber perpetrators that target crowdsourcing. For this purpose, the framework that we developed provides an initial outline for a structured research program into the nature and effects of malicious and deceptive crowdsourcing activities. The framework can assist categorizing past research and identifying gaps to be addressed in future studies. Furthermore, it can be used as a starting point for researchers theorizing about antecedents to malicious and deceptive behavior in crowdsourcing. Furthermore, through the future collection of crowdsourcing incidents, the framework can help identify commonalities between certain types of incidents in terms of behavioral and contextual factors.

From a practical perspective, the incident examples and organizing framework can provide guidance for

organizations that are considering using crowdsourcing for their business needs and for crowdsourcing providers to assess and strengthen their control framework to mitigate deception threats. For instance, results indicate that risks are predominantly linked to reputational and/or financial loss. Both Confidentiality risk types refer to privacy violation risks that are likely to damage a crowdsourcing platform's reputation and result in fines. The Integrity risk types are related to the loss of reliability and reputation that crowdsourcing platforms face when their sites are flooded with false information. This will ultimately cause users to abandon them. The Availability risk type impacts platform performance and directly leads to loss of profits due to the unavailability of a crowdsourcing site. Another practical implication relates to the need for the crowdsourcing industry to consider setting up a mechanism through which they can (anonymously) share cyber and deception incidents. An organized library of incidents will support learning about potential risks and will ultimately strengthen the security of the industry as a whole. For instance, a secured platform to collect crowdsourcing deception incidents would assist businesses to learn to recognize and avert deceptive actions.

A key limitation of our study is that we likely were not able to collect an exhaustive set of incidents. First, crowdsourcing is still a relatively new phenomenon so additional risk types may emerge in the near future. Second, organizations may be aware of the occurrence of deceptive actions so incidents go unreported. Finally, some organizations are likely to be reluctant in sharing detailed information on incidents out of competitive and reputational considerations.

Future research directions include addressing the limitations of this study by expanding the collection of incidents through interviews with crowdsourcing service providers. Also, future research may focus on issues such as the extent to which risk awareness deters organizations from employing crowdsourcing, whether certain tasks are more vulnerable to deception than others, and whether crowdsourcing efforts are more vulnerable to deception in certain cultures than others. Finally, theoretical research could investigate the underlying cognitive mechanisms why people who aim to deceive others online, e.g., why do people post false information and what are the reasons for choosing crowdsourcing crowds as target victims? Such research could be informed by criminology theories such as General Strain Theory or Routine Activities Theory [25].

10. References

[1] Abhinav K (2015): Trustworthiness in Crowdsourcing. A Thesis Report submitted in partial fulfilment for the degree of MTech Computer Science 29 July 2015

[2] BBC News (2015). Crowdfunding site Patreon hacked data leaked. Retrieved from http://www.bbc.com/news-/technology-34423932

[3] BBC News (2014): Kickstarter crowdfunding website hacked. http://www.bbc.com/news/business-26222113

[4] Biersdorfer, J.D. (2016), How Waze Tries to Keep Its Crowd Honest. http://nyti.ms/2cLlTkY

[5] Bode Karl (2015): Miami Cops Flood Waze With Bogus Speed Trap Data, Don't Understand How Crowd Sourcing Works. https://www.techdirt.com/articles/20150209-/12580529961/miami-cops-flood-waze-with-bogus-speed-trap-data-dont-understand-how-crowd-sourcing-works.shtml

[6] Bommert, B. (2010): Collaborative innovation in the public sector. International Public Management Review, 11, 15-33

[7] Bonabeau, E. (2009): Decisions 2.0: the power of collective intelligence. Sloan Management Review, 50, 45–52

[8] Brabham Daren C. (2008): Crowdsourcing as a Model for Problem Solving An Introduction and Cases. Convergence: The International Journal of Research into New Media Technologies, Vol 14(1): 75–90

[9] Buchanan, M. (2013). The Web's Failed Hunt for the Boston Bomber - The New Yorker. Retrieved from http://www.newyorker.com/tech/elements/the-webs-failed-hunt-for-the-boston-bomber

[10] Chen, H., Rahwan, I. and Cebrian, M. (2016): Bandit strategies in social search: the case of the DARPA red balloon challenge. EPJ Data Science

[11] Cheng, A., Friedman, E. (2005): Sybilproof Reputation Mechanisms. SIGCOMM'05 Workshops, August 22–26, 2005, Philadelphia, PA, USA

[12] Cox Landon P. (2011): Truth in Crowdsourcing. IEEE September/October 2011

[13] Dewey, C. (2014). The many problems with SketchFactor, the new crime crowdsourcing app that some are calling racist - The Washington Post. Retrieved from https://www.-washingtonpost.com/news/the-intersect/wp/2014/-08/12/the-many-problems-with-sketchfactor-the-new-crime-crowdsour-cing-app-that-some-are-calling-racist/?utm_term=.61c78f9b0d8e

[14] Dwarakanath ,A., Shrikanth N.C., Abhinav, K., Kass, A. (2016): Trustworthiness in Enterprise Crowdsourcing: a Taxonomy & evidence from data. ICSE '16 Companion, May 14-22, 2016, Austin, TX, USA

[15] Edwards, J. (2010). PatientsLikeMe Is More Villain Than Victim in Patient Data "Scraping" Scandal - CBS News. Retrieved from http://www.cbsnews.com/news/-patientslikeme-is-more-villain-than-victim-in-patient-data-scraping-scandal/

[16] Fayazi, A., Lee, K., Caverlee, J., Squicciarini, A. (2015): Uncovering Crowdsourced Manipulation of Online Reviews. SIGIR'15, August 09 - 13, 2015, Santiago, Chile.

[17] Floren Cindy (2012): The Advantages and Disadvantages of Using Crowdsourcing to Improve Your Ecommerce Business. http://myecommerce.biz/blog/2012/-07/the-advantages-and-disadvantages-of-using-crowd-sourcing-to-improve-your-ecommerce-business/

[18] Franceschi-Bicchierai, L. (2015). Crowdfunding Site Patreon Gets Hacked - Motherboard. Retrieved from https://motherboard.vice.com/en_us/article/crowdfunding-site-patreon-gets-hacked

[19] Fuller J (2012): Die Gefahren des Crowdsourcing. http://www.harvardbusinessmanager.de/blogs/a-840963.html

[20] Goolsby, Rebecca (n.d.): On Cybersecurity, Crowdsourcing, and Social Cyber-Attack. Office of Naval Research, https://www.wilsoncenter.org/sites/default/files-/127219170-On-Cybersecurity-Crowdsourcing-Cyber-Attack-Commons-Lab-Policy-Memo-Series-Vol-1.pdf
[21] Gregor, S. (2006): "The nature of theory in information systems," MISQ (30:3), pp. 611-642
[22] Harris Christopher G. (2011): Dirty Deeds Done Dirt Cheap. A Darker Side to Crowdsourcing. IEEE International Conference on Privacy, Security, Risk and Trust
[23] Harris, M. (2015): How A Lone Hacker Shredded the Myth of Crowdsourcing, backchannel.com/how-a-lone-hacker-shredded-the-myth-of-crowdsourcing-d9d0534f1731#.vkn21poy6
[24] Hendrix, S. (2016): Traffic-weary homeowners and Waze are at war, again. Guess who's winning? www.washington-post.com/local/traffic-weary-homeowners-and-waze-are-at-war-again-guess-whos-winning/2016/06/05/c466d-f46-299d-11e6-b989-4e5479715b54_story.html?utm_term=.442e5763ecb2
[25] Holt T.J. And Bossler, A.M., (2016): Cybercrime in Progress: Theory and prevention of technology-enables offenses, first edition, Routledge, ISBN: 9781138024168
[26] Howe, J. (2006): The rise of Crowdsourcing, Wired Magazine, Issue14/06, June
[27] Howe, J. (2006): Crowdsourcing: A Definition. Crowd-sourcing.typepad.com/cs/2006/06/crowdsourcing_a.html
[28] Kaplan, D. (2013). Reddit site downed by DDoS attacks. Retrieved from https://www.scmagazine.com/reddit-site-downed-by-ddos-attacks/article/543507/
[29] Kayes E., Kourtellis N., Quercia D., Iamnitchi A., Bonchi F. (2015): The Social World of Content Abusers in Community Question Answering, WWW 2015, May 18–22, 2015, Florence, Italy
[30] Lamere, P. (2009). Inside the precision hack | Music Machinery. Retrieved from https://musicmachinery.com-/2009/04/15/inside-the-precision-hack/
[31] Lasecki Walter S., Teevan Jaime, Kamar Ece (2014) Information Extraction and Manipulation Threats in Crowd-Powered Systems. CSCW '14, February 15–19, 2014, Baltimore, Maryland, USA
[32] Lasecki Walter S., Teevan Jaime, Kamar Ece (n.d.): The Cost of Asking Crowd Workers to Behave Maliciously
[33] Lionbridge (2013): The Crowd in the Cloud: Exploring the Future of Outsourcing White Paper. Massolution Crowd Powered Index. http://www.lionbridge.com/files/2012/11/-Lionbridge-White-Paper_The-Crowd-in-the-Cloud-final.pdf
[34] Medina, D.A. (2014): Crowdsourcing Competitions Encourage Malicious Behavior, Study Finds. www.next-gov.com/big-data/2014/09/crowdsourcing-competitions-encourage-malicious-behavior-study-finds/93410/
[35] Naroditskiy V, Jennings NR, Van Hentenryck P, Cebrian M. (2014): Crowdsourcing contest dilemma. J. R. Soc. Interface 11: 20140532. http://dx.doi.org/10.1098/rsif.2014.0532
[36] Niller, E. (2016). DoS Attack Crashes Website Monitoring North Korea's Nuclear Test Site | WIRED. Retrieved from https://www.wired.com/2016/09/dos-attack-crashes-website-monitoring-north-koreas-nuclear-test-site/
[37] Paganini, P. (2014, June 23). Largest DDoS attack hit PopVote, Hong Kong Democracy voting site. Retrieved from securityaffairs.co/wordpress/26030/cyber-crime/popvote-largest-ddos-attack.html
[38] Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., et al. (January, 2013). Foundations of Crowdsourcing: A Review of IS Research, Proceedings of the 46th HICSS. IEEE Computer Society Press.
[39] Pescatore, J. (2013): How DDoS Detection and Mitigation Can Fight Advanced Targeted Attacks. A SANS Whitepaper. Retrieved from https://www.sans.org/reading-room/whitepapers/analyst/ddos-detection-mitigation-fight-advanced-targeted-attacks-35000
[40] Rheenen, van E. (2016): 15 Polls Hijacked by the Internet. Retrieved from mentalfloss.com/article/52524/15-polls-hijacked-internet
[41] Roberts, E. (n.d.). How Automatic Voting Bots Manipulate Online Polls | Distil Networks. Retrieved from https://resources.distilnetworks.com/all-blog-posts/how-to-manipulate-an-online-poll-with-a-bot
[42] Rosenfeld E (2012): Mountain Dew's 'Dub the Dew' Online Poll Goes Horribly Wrong. TIME, Aug. 14, 2012. http://newsfeed.time.com/2012/08/14/mountain-dews-dub-the-dew-online-poll-goes-horribly-wrong/
[43] Stefanovitch N, Alshamsi A, Cebrian A, Rahwan I (2014): Error and attack tolerance of collective problem solving: The DARPA shredder challenge. EPJ Data Science, 3(1):1–27, 2014 epjdatascience.springeropen.com/articles/-10.1140/epjds/s13688-014-0013-1
[44] Sterling, T. (2017). Dutch voting guide sites offline in apparent cyber attack. Reuters. Retrieved from http://www.reuters.com/article/us-netherlands-election-cyber-idUSKBN16M1C6
[45] The Washington Post. (2014): SketchFactor controversy showcases challenges of crowdsourcing. Retrieved from washingtonpost.com/opinions/sketchfactor-controversy-show-cases-challenges-of-crowdsourcing/2014/08/20/18348bea-2311-11e4-958c-268a320a60ce_story.html?utm_term=.408c547b40fe
[46] Tian Tian, Jun Zhu, Fen Xia, Xin Zhuang, Tong Zhang (2015): Crowd Fraud Detection in Internet Advertising. WWW 2015, May 18–22, 2015, Florence, Italy.
[47] Tufnell Nickolas (2014): Students hack Waze, send in army of traffic bots. http://www.wired.co.uk/article/waze-hacked-fake-traffic-jam
[48] VanPutte, Michael A. (n.d.): Challenges in Securing Crowdsourcing Solutions.
[49] Wang Gang, Wang Bolun, Wang, Tianyi, Nika Ana, Haitao Zheng, Ben Y. Zhao (2016): Defending against Sybil Devices in Crowdsourced Mapping Services', MobiSys'16, June 25-30, 2016, Singapore, Singapore
[50] Wang Gang, Christo Wilson, Xiaohan Zhao, Yibo Zhu, Manish Mohanlal, Haitao Zheng and Ben Y. Zhao (2012): Serf and Turf: Crowdturfing for Fun and Profit, WWW 2012, April 16–20, 2012, Lyon, France.
[51] Wang, G., Konolige, T., Wilson, C., Wang, X., Zheng, H., Zhao, B. Y. (2013): You are how you click: Clickstream analysis for sybil detection. In Proc. of USENIX Security, Washington, D.C., August 2013
[52] Wang Tianyi, Wang Gang, Li Xing, Haitao Zheng, Ben Y. Zhao (2013): Characterizing and Detecting Malicious Crowdsourcing. SIGCOMM'13, August 12–16, 2013, Hong Kong, China. ACM 978-1-4503-2056-6/13/08
[53] Wolfson S., Lease M. (2011): Look Before You Leap: Legal Pitfalls of Crowdsourcing. ASIST 2011, October 9–13, 2011, New Orleans, LA, USA.

Appendix A. Disruption and Deception Incidents in Crowdsourcing

Type & Task description	Crowd	Crowdsourcing Platform	Problem owner	Governance	Crowd Contributions	CIA Triad	Source
I1. Sybil attack – people posting false data on crowdsourced map system Waze	Users posted false reports of a wreck, speed trap or other blockage in their neighborhoods to deflect some of the traffic flow	The app detected a saboteur only after two weeks of daily false information posts				Integrity – wrong data was put into the system to redirect the traffic away from the local neighborhood	[24]
I2. DARPA Red Balloon Challenge – false locations of balloon placements were submitted across teams	False locations were submitted across the teams				False locations' goal was to confuse other participants	Integrity – wrong data strongly influenced the quality of information	[10]
I2. DARPA Shredder Challenge required participants to put shredded documents together. Some of the crowd workers sabotaged the initiative by repeatedly undoing the work delivered by other crowd workers	A crowd of attackers piled up the pieces of the jigsaw and sabotaged genuine users' work making it much more complex, as they had first to unstack the pieces and then search for correct matches				Inserted input contributed to creation of a wrong result, long term impact of the attacks resulted in decrease in participation	Integrity violation – provided data computed false results	[43]
I4. Prank designs submitted to Henkel challenge – the designs got voted to the top ranks and consequently the company had to retract public voting	Intentionally pranked designs confused the voting crowd		Henkel decided to sort out some suggestions		Users had malici-ously influenced the vote; wrong input was placed on site	Integrity – data provided resulted in the cancellation of the voting	[19]
C2. Netflix released hundred million anonymized user records as part of its "Netflix Prize" Contest. User records became known after they were combined with other information to find the identities of the users in the dataset				Data, when combined with exter-nal informa-tion disclos-ed user identities		Confidentiality – data containing identities of the contest participants was disclosed after being combined with publicly accessible information	[53]
C1. Hacking attack on crowdsourcing page to collect sensitive data incl. passwords, phone numbers, email addresses and credit card numbers (Kickstarter)				Lack of suf-ficient secu-rity measures to hacking attempt		Confidentiality – Kickstarter user data was hacked	[3]
I4. Mountain Dew Campaign poll was bombarded with unusable names proposals	The crowd started posting offensive proposals in the contest for a green-apple infused soft drink name				Offensive campaign results shut initiative down	Integrity – provided prank data forced the organizers to close the campaign	[42]
C1. Crowdfunding site Patreon was hacked for users' names, email, posts, and shipping addresses				Site's user database got hacked		Confidentiality - hackers accessed users' names, email, posts, and shipping addresses	[18] [2]
I2. Police officers flooding Waze app with false information on their activity to make the app's information less useful to drivers	The crowd (police) entering fake data on Waze app					Integrity – false information provided aiming to spread disinformation	[5]

I3. Botnet attack on Waze	Two students created fake Waze accounts to create fake traffic jams					Integrity – false information provided aiming to spread disinformation	[47]
I4. Internet polls steered off track by the voting crowd into outrageous results	E.g. prank responses on polls related to naming ships, bridges, products, planets or voting for concert spots for musicians					Integrity – prank poll responses aiming to spread disinformation	[40]
I3. Robo-voting that skewed public polls, e.g. Time 100 Most Influential People		Polls sabotaged by automated scripts (bots) that submitted multiple votes				Integrity – voting results were skewed	[41] [30]
A1. DDoS attack on PopVote, Hong Kong Democracy voting website in June 2014		PopVote hosted by Amazon Web Services, Cloudflare and UDomain. All hosting services hit by large scale DDoS attacks				Availability – the goal was to take the website down	[37]
A1. Project on Crowdsourced Satellite Imagery taken down by DDoS attack		Site sourcing images of sensitive geographical sites taken down by a DDoS attack				Availability – the website stopped working	[36]
A1. Reddit site became unavailable because of DDoS attacks		The website was hit with DDoS attack				Availability - Reddit site experienced an outage	[28]
I2. Failed hunt for the Boston Bomber	After Boston bombing attack, the crowd attempted and failed to accurately identify the offenders				Wrong information on suspects on Reddit and Twitter resulted in crowd searching for innocent people	Integrity – wrong input data created wrong output and resulted for the chase of innocent people	[9]
I2. SketchFactor application crowdsourced crime and public safety reporting perceptions, not facts		Platform for users to browse reports of “sketchy” behavior and crowdsource their own stories was accused of racism and profiling				Integrity - the data available on SketchFactor compared with actual crime data revealed few clear overlaps	[13]
C2. PatientsLikeMe users’ personal data on taken medications and illness symptoms was gathered by Nielsen	Nielsen opened personal accounts on the app and copied personal data from chatrooms and on bulletin boards.					Confidentiality – although there was no privacy violation, Nielsen found a way to extract app’s users’ data	[15]