University of Kentucky
UKnowledge

Information Science Faculty Publications | Information Science

9-2014

Online Deception in Social Media

Michail Tsikerdekis
University of Kentucky, tsikerdekis@uky.edu

Sherali Zeadally
University of Kentucky, szeadally@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/slis_facpub

 Part of the Library and Information Science Commons

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Repository Citation

Tsikerdekis, Michail and Zeadally, Sherali, "Online Deception in Social Media" (2014). *Information Science Faculty Publications*. 12.
https://uknowledge.uky.edu/slis_facpub/12

This Article is brought to you for free and open access by the Information Science at UKnowledge. It has been accepted for inclusion in Information Science Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

Online Deception in Social Media

Digital Object Identifier (DOI)
http://dx.doi.org/10.1145/2629612

Notes/Citation Information
Published in *Communications of the ACM*, v. 57, no. 9, p. 72-80.

© ACM, 2014. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in *Communications of the ACM*, Vol. 57, No. 9, (September 2014) http://doi.acm.org/10.1145/2629612

This article is available at UKnowledge: https://uknowledge.uky.edu/slis_facpub/12

Online Deception in Social Media

Abstract

The explosive growth of social media applications has revolutionized the way we interact with one another. However, the emergence and use of this online environment has also created new opportunities for deception. We present a brief comparison between traditional (i.e., offline) deception and online deception with a focus on social media. Furthermore, we explore some of the factors that can affect the difficulty in achieving deception in social media and we use a deception model to classify different online deception techniques. We also discuss the ease of deployment and success of these techniques. Finally, we highlight some challenges that social media designers must address in the future to protect social media users from online deception.

1. Introduction

The rapid proliferation of Web-based technologies have revolutionized the way content is generated and exchanged over the Internet leading to an explosive growth in social media applications and services. Social media enable the creation and the exchange of user-generated content and the design of a wide range of Internet-based applications. This growth has been fueled not only by the increase in the number of services but also by the rapid rate of their adoption by users. Between 2005 and 2013, we have witnessed a 64 percent increase in the number of people using social media[1]. For instance, Twitter usage increased by 10 percent in the period 2010-2013. A total of 1.2 billion users connect through Facebook and Twitter with their accounts[24]. However, the ease of getting an account also makes it easier for individuals to deceive others. Previous work on deception has found that people in general lie daily and several past efforts have attempted to detect and understand deception[20]. Throughout history, deception has been used in various contexts along with technology (Second World War, Trojan War, etc.) to enhance an attacker's deceptive action(s). Social media provide new environments and technologies for potential deceivers. There are frequent examples of people that have been deceived through the use of social media and some with devastating consequences in their personal lives.

In this paper, we consider deception as a deliberate act with the intent to mislead others while the recipients are not made aware or expect that such an act is taking place and that the goal of the deceiver is to transfer that false belief to the deceived ones[2,9]. This perspective on deception becomes particularly relevant when examining social media services in which the boundary between protecting one's privacy and deceiving others becomes blurry. Furthermore, we also argue that these false beliefs are transferred through verbal and non-verbal communication[14] and deception is measurable and identifiable through verbal (e.g., audio or text), non-verbal (e.g., body movement) and physiological cues (such as heartbeat).

One may argue that training and raising awareness such as the one provided to security personnel[7] may be an effective avenue for protecting users of their social media. However, people who are trained to detect deception perform worse in detection accuracy than people that do not[17] and evidence of a "privacy paradox" point to individuals sharing detailed information even though they are aware of privacy concerns[26] making themselves more vulnerable to attacks. To make things worse, social media, as a set of Internet-based applications, is also broadly defined term with multiple categories that include virtual environments that are vastly different from one another[15,16].

This work aims to present the concept of deception and explores its use in social media in particular. We focus on the motivations for deception in social media and we explore various deception techniques that have been used recently and their impact on social media users. Finally, we discuss some of the challenges that we need to address in the future in the area of deception in social media. While detecting and preventing deception are important aspects that relate to the topic of deception, understanding online deception and classifying techniques used in social media is the first step in fighting it. Our future publications will explore aspects of online deception detection and prevention because the strict length constraint imposed on this paper does not permit us to do.

2. Online Deception

Nature favors deception as a mechanism for gaining a strategic advantage. For example, viceroy butterflies deceive birds by looking alike with monarch butterflies (which have a bitter taste) thereby ensuring their survival as long as there are not too many in a system[8]. Similarly, humans have been using deception in connection to a *benign* or *hostile* intent[3]. In warfare, Sun Tzu[29] argued that "all warfare is based on deception."

Social media services can be classified based on social presence/media richness and self-representation/self-disclosure[16]. Social presence can also be influenced by the intimacy and immediacy of the medium in which the communication takes place while media richness describes the amount of information that can be transmitted at a given point in time. Self-representation determines the control that users have in representing themselves whereas self-disclosure determines revealing one's information whether willingly or unwillingly. Using the aforementioned characteristics, a table was developed by Kaplan and Haenlein[16] that included the following social media: *blogs*, *collaborative projects* (e.g., Wikipedia), *social networking sites* (e.g., Facebook), *content communities* (e.g., Youtube), *virtual social worlds* (e.g., Second Life) and *virtual game worlds* (e.g. World of Warcraft).We present an expanded classification of social media (shown in Table 1) that also includes *microblogging* (e.g., Twitter) and *social news sites* (e.g., Reddit). We place microblogging between blogs and social networking sites[15] and social news sites above microblogging given their similarity to microblogging in terms of social presence/media richness (limited content allowed to be communicated through the medium and average immediacy as news come in) and low self-presentation/self-disclosure due to their nature as content-oriented communities.

		Social presence / Media richness			
		Low			High
Self-presentation / Self-disclosure	Low	Collaborative projects	Social news sites	Content communities	Virtual game worlds
	High	Blogs	Microblogging	Social networking sites	Virtual social worlds

Table 1: Social media classifications.

As shown in Table 1, social media which provide users with a lot of freedom for presenting themselves are in the second row while social media that force users to adapt to certain roles or have no option for disclosing parts of their identities are in the first row. Moreover, with an increase in media richness and social presence, we note the transition from social media offering just text for communication to rich media aimed to simulate the real world using verbal and non-verbal messages as well as more immediacy in communications for virtual game worlds and virtual social worlds. The differences between these types of social media services affect how deception is implemented and its potential success.

In most social media platforms, most communications are text-based and are done asynchronously. In such environments deceivers have a great advantage for altering content which is a cheap way to deceive others. Zahavi[31] pointed out the difference between assessment signals that are reliable and hard to fake and, conventional signals that are easier to fake. For example, in the real world if an elderly person wants to pass as a younger person, he/she can dress younger or dye his/her hair and this will produce conventional signals. However, it would be much harder to fake a driver's license (an assessment signal). Social media however provide an environment in which assessment signals are not required and are not the norm making deception easier to achieve. For instance, gender switching online requires often only a name change.

3. Difficulty in Achieving Online Deception

It is not surprising that the level of difficulty in achieving online deception is determined by several factors associated with the deceiver, the social media service, the deceptive act and the potential victim. These factors will determine how easy or difficult it is for a deceiver to engage in online deception. High difficulty in achieving deception may deter potential deceivers while low difficulty may be seen as an easy opportunity to deceive others. Figure 1 shows the various entities (deceiver, social medium, victim) involved in online deception.

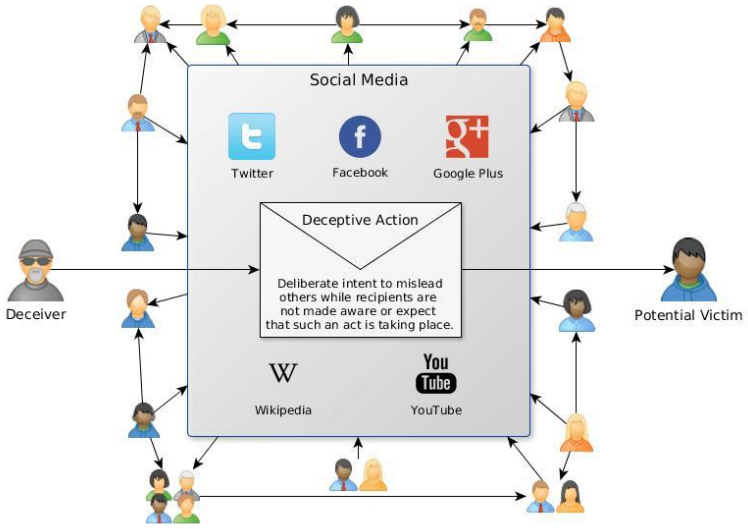


Figure 1: Entities involved in online deception.

3.1 The Deceiver

Several factors associated with the deceiver will determine the difficulty in achieving online deception. These factors include *expectations*, *goals*, *motivations*, his/her *relation to target* and *target's degree of suspicion*[2]. Expectation is a factor that determines the likelihood of success in deception. More complex messages have a higher likelihood to succeed[20]. Goals and motivations also determine the difficulty of deception. Goals are broader and long-term while motivations consist of specific short-term objectives. They directly influence the choice and type of a deceptive act. One motivation taxonomy, developed by Buller and Burgoon[2], described three different motivators for deception: a) *instrumental* where one can identify goal-oriented deception such as lying on one's resume on a social medium to get more job offers, b) *relational* (also known as social capital) such as aiming to preserve social relations typical in online social networks[26], and c) *identity* such as preserving one's reputation from shameful events on their online profile. The aforementioned motivating factors in turn determine the cost (i.e., the level of difficulty in achieving deception) of deception for a deceiver. For example, a deceiver motivated to fake his or her identity will have to put more effort offline to succeed due to the presence of signals that are much more difficult to fake rather than online where many of the identity-based clues (gender, age, etc.) may take the form of just conventional signals (e.g., adding this information to one's profile page without verification). The difficulty in achieving deception is also determined by the deceiver's relation to a target. Familiarity with a target and its close social network make it easier to gain trust and reduces the difficulty in achieving deception. Many users assume that with technology comes enhanced security and are more relaxed in trusting others online[4]. Further, the level of trust that individuals place on the deceiver, will also reduce their degree of suspicion towards him or her thereby increasing their chances of being deceived.

The *moral cost* also increases the difficulty in achieving deception[26]. Morals can heavily influence what deceivers consider as immoral in regards to withholding information or even lying. In the real world the immediacy of interaction may make it much harder to deceive for some individuals. In contrast, in the case of online environments distance and anonymity[28] contribute to a loss of inhibition and therefore the moral cost is lower for deceivers.

3.2 Social Media

Social media require us to expand our perspective on how interactions are perceived between a receiver and a sender during deception. For instance, the Interpersonal Deception Theory (IDT) states that the interaction between a sender and a receiver is a game of iterative scanning and adjustments to ensure deception success[2].

Donath[8] has suggested that if deception is prevalent in a system (e.g., Facebook community) then the likelihood of success is reduced. It makes sense that the *prevalence of deception* in an online community is a factor that also determines difficulty in achieving deception. Social media services which encounter high volumes of deception will lead to communities that are more suspicious. This will increase the number of failed deception attempts. Furthermore, by increasing the target's suspicion the difficulty will increase deterring deceivers from the community. Eventually some equilibrium may be reached. This rational however, suggests that communities with low prevalence of deception will likely be more vulnerable to attacks since suspicion will remain low for potential victims. Determining the prevalence of deception in a community remains a challenging task.

Similarly the underlying *software design* of social media can also affect the degree of suspicion: the *level of perceived security* for victims increases the chances of success for the deceiver[11]. Software design can cause several assumptions to be made by users about the level of security it provides. Some aspects of the design can make users more relaxed and less aware of potential risks of being deceived. For example, individuals may falsely assume that profile information on a social networking site is difficult to fake due to additional verification methods such as email confirmation. Moreover, *assurance* and *trust mechanisms* for a system will determine the level of trust between the sender and receiver[11]. Assurance mechanisms can either reduce the probability of a successful deception or increase the penalties for deceivers[11]. High penalties will increase the difficulty for deceivers especially when the chances of being caught are high. Assurance mechanisms are considered to be effective in certain contexts where it is argued that the need for trust can be completely diminished. In social media, assurance mechanisms are much harder to implement and as such penalties and the chances of being caught may be or seem to be lower than those in offline settings and as such the cost of deception is much lower. Finally, *media richness* is also a factor that determines the difficulty in achieving deception. In this context, Galanxhi & Nah[10] in their study of deception in cyberspace found that deceivers experienced more stress when communicating with their victims through text rather than an avatar-supported chat.

3.3 Deceptive Action

Time constraints and *the number of targets* of an attack are factors that also determine the difficulty in achieving online deception. The time available and the time required for a successful attack are important especially in social media services where asynchronous communication takes place. Moreover, the time required for the deception to be detected also determines the effectiveness of the deception method used. For cases where deception must never be discovered, the cost of implementation of the deception method may outweigh the benefits especially when the penalties are high. The social space that deception is applied to and the number of online user targets who are required to be deceived affect the level of difficulty in implementing the deception method. A politician needing to deceive in his/her profile all of their voters will face a more difficult challenge compared to a deceiver deceiving just one individual. The *type of deceptive act* is also another important factor. Complex types of deceptive acts that are guided by multiple objectives (e.g., identity and instrumental) are more difficult to achieve.

3.4 The Potential Victim

In traditional settings the target's *ability to detect* deception may be a factor that determines the difficulty in achieving deception. Online deception seems to be much harder to detect by users. For example, in a study of Internet fraud using page-jacking techniques even experienced users failed to detect inconsistencies present except for a select few who did detect deception showing that it is not impossible[11]. Therefore, in social media a target's ability to detect deception also depends to some extent on his or her Information Communication Technology (ICT) literacy. Deceivers will have to evaluate their potential victim's ICT literacy. Individuals with a high ICT literacy can have a significant advantage over casual Internet users and therefore a cost and benefit analysis for a social engineering attack may be higher in this case.

4. Deception Techniques in the Social Media Environment

There are various techniques reported in the literature that can be used to deceive others in social media environments and they include: bluffs, mimicry (e.g., mimicking a website), fakery (e.g., forging a fake website), white lies, evasions, exaggerations, web page re-directions (e.g., misleading someone to a false profile page) and concealments (e.g., hiding information from one's profile)[21]. We use the communication model proposed by Madhusudan[20] to classify deception techniques for social media and evaluate their effectiveness in achieving deception.

4.1 Deception Model

The model (depicted on Figure 2) consists of a sender (S), the content or message (I), the channel through which communication takes place (C), and the receiver (R). When a receiver's expected model (the S, I, C triangle) is different from the received model then deception has occurred. This is also in line with Ekman's[9] definition of deception who argues that a receiver

must not anticipate deception. By manipulating any of the S, I, C elements or combinations of these deception is achieved. We present an overview of social media and identify factors (S, I, C) and social media types where deception can be achieved with minimal efforts (i.e., low cost) and at the same time results in a fairly high deception success rate (shown in Table 3). These were identified from closely related literature on the topic and we present more information in the following sections.

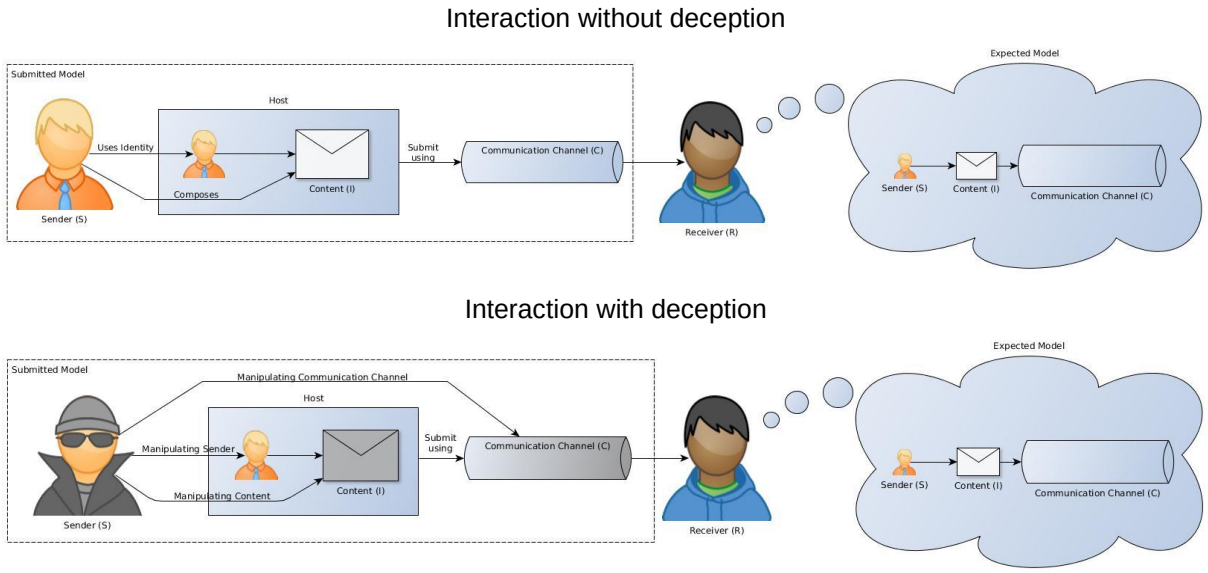


Figure 2: Interaction without/with deception.

Social Media Type	Low difficulty	High deception success
Blogs	S, I	S, I
Collaborative projects	I	-
Microblogging	S, I	S, I
Social news sites	S, I	S, I
Social networking sites	S, I, C	S, I, C
Content communities	I	I
Virtual social worlds	S, I, C	S, I, C
Virtual game worlds	I, C	C

Table 3: Manipulation of Sender's identity information (S), Content (I), and Communication channel (C) with low difficulty and high deception success.

4.2 Content Deception

Manipulating content is presumably the most common way of deceiving others. In social media this can be achieved by falsifying information. Social media that focus primarily on content such as blogs, content communities, social news sites, and microblogging are highly susceptible to such deception. Technology today allows us to manipulate multimedia files to an extraordinary degree. Tampering with images[23] is an efficient way to fake content such as representing that an individual traveled all around the world in one's photos by altering them and broadcast these images using social media. Such strategy may help a deceiver to elevate their social status and gain trust of a victim in order to obtain further information. In addition to videos and images, the ease of manipulating content, that at times, is heavily based on just text allows for a low cost for deception and a high probability of success because of various factors such as a low information literacy of receivers (e.g., critically evaluating content), the lack of expectation for verifiability and even accountability. In addition, social media that offer profile management for their users such as social networking sites and virtual social worlds are also susceptible especially in cases where the initiation of new relationships is advertised. A competent deceiver in affective writing may have a substantial advantage on these types of social media.

In contrast, collaborative projects such as Wikipedia are less likely to be affected by this kind of deception (i.e., manipulating I). The difficulty in achieving deception may seem low but the chances of success (at least over the long-term) are also low. This is because the software design of these types of social media where many-to-many communication is supported enables many people to review the content. Examples of content deception can be seen in Wikipedia where not only vandals (people who alter content with intent to deceive others) are eventually detected but there are people who assume a role to fight them[25]. Furthermore, assurance mechanisms such as the requirement for content validity (tracing content back to its source) are built into the system to ensure that content deception becomes more visible. Another example of content deception in these types of social media is with open source software managed by multiple users where it is much harder to add malicious content and succeed in deception because multiple individuals evaluate the code before it is being released. Virtual game worlds also have a low probability for deception success because of the strongly narrated elements such as having a specific role that forces a player to a specific line of actions.

4.3 Sender Deception

Sender deception is achieved by manipulating the sender's identity information (S). Impersonation is a common example which often results in identity deception. This deception falls under the category of identity theft in identity deception[30]. In this case, the deceiver may gain access to an identity and use it to obtain additional information from their peers such as house address, date of birth, and cell number. The failure to authenticate the sender's credentials will lead to a successful deception. Social media's designs that have in-built high self-presentation and self-disclosure enable sender deception at a low cost. Blogs and microblogging can lead to stolen identities because there are no control mechanisms to verify

new users or their associated names. However, the damage caused by deception with these types of social media is also likely to remain low and long-term success is probably not guaranteed. Owners of the identity may become aware of the theft or other individuals familiar with that identity may start identifying behavioral cues that do not match with that identity. In the case of social networking sites and virtual social worlds, the cost of deception increases because, cognitively, an individual will have to satisfy behaviors that are appropriate to the identity he or she impersonates. The benefits however seem to be much higher in a social medium context because access to an individual's social network can lead to an enhanced ability to gain people's trust within the network and obtain information from them. The target in these cases may not necessarily be the individual whose identity is stolen but others within his or her social network. With no control mechanisms in place for identifying a source, arguably, unregistered individuals which do not have an account on a service may be more exposed than registered users for the social media services described above.

Social media (such as collaborative projects or virtual game worlds) with low self-presentation and self-disclosure are likely to be more protected in terms of identity theft. This can be partially attributed to their intended function. Collaborative projects, content communities and virtual game worlds are heavily task-based. A user, who wants to obtain access from the impersonated identity's social network will have to perform just as well as the identity being impersonated in tasks and “act the part.” The cost is likely to be high and the success of the deception low and short-term.

A middle ground between content deception and sender deception involves manipulating information associated with an identity. These deception attacks can be categorized as identity concealment where part of the information for an original identity is concealed or altered and identity forgery where a new identity is formed[30]. For example, people may attempt to fake some of the information in their profiles in order to gain trust or represent themselves in a different way. In customer social networking sites, people may conceal information in order to gain advantages of different offers [5].

4.4 Communication Channel Deception

Manipulating the communication channel requires a higher technical skill level which increases the cost of deception. Tampering with the communication channel includes modifying in-transit messages, re-routing of traffic, eavesdropping, etc. Jamming communications have been used in virtual game worlds. Podhradsky et al.[22] argued that multiplayer games in consoles can be hacked in order to provide access to a user's Internet Protocol address. Once the intruder has access to the host, he/she can kick the player out and proceeds with an identity theft deception. In this case the goal of a deceiver may not be to obtain information but to damage a victim's reputation. It is worth pointing out that there is a fine line between an unintentional disconnection and an intentional departure of a player in a video game. This line becomes blurred when the player is on the losing side and suddenly leaves. As a result, the reliability and reputation of this player are damaged by an invisible deceiver. An advantage that communication channel deception has is the implicit assumption that people make that digital

technology is imperfect and things may not work as good as in real world. Non-verbal behavior[14] such as body movement or patterns of speech can expose deceivers, however, through social media a deceiver may introduce jitter or delays in their video or audio in order to conceal their deception, effectively increasing the chances of success. Victims at the other end of the connection will have a difficult time in differentiating between an unreliable/slow connection and the deceptive act being performed.

Since channel deception often involves technology all social media services may be susceptible to an attack, especially those that use a similar set of technologies or architectures. Services that have a higher reliance on their client applications will be more prone to attacks while those that rely on server applications will probably be safer. Services with high media richness tend to rely a lot on client software as is the case with virtual social worlds and virtual game worlds. Deception, by exploiting communication channels, is common in such services[13]. Server-side applications such as social networking sites or content communities are less prone to channel deception because exploits rely on vulnerabilities of web browsers and web servers which are generally hardened and made more secure.

The cost of this deception is quite high. However, the likelihood of success is also high especially for well-orchestrated attacks.

4.5 Hybrid Deception Techniques

Hybrid deception techniques involve the manipulation of multiple elements (S, I, C) in the SIC model described earlier and can be more effective in launching deception attacks. The relationships between S, I, and C, as described by Madhusudan[20], produce a consistent view for a receiver. If one element of the SIC model shows a slightly different behavior, this may give clues about an inconsistent relationship between two elements (e.g., S and I). For example, a message received and signed by one's relative may lose its credibility if the source information of the message does not match with that of the relative.

Various hybrid deception techniques that manipulate the content and the sender's information have been reported in the literature. These include examples such as forgery[20], phishing, identity forgery, web forgery[11], email fraud. These techniques are highly effective in social media such as social networking sites, virtual social worlds, microblogging and blogs, which highlight user identities and provide a one-to-one or one-to-many communications. These online deception attacks have not only been demonstrated to be effective but their consequences can lead to disastrous consequences including the loss of life. In reality, the boy was actually the mother of a former friend who used deception to gain her trust and later sent cruel and hurtful messages to the girl. A service initially designed for people who want to initiate new relationships and the lack of verifying both parties led to a devastating outcome. Online deception can also have financial consequences as is the case with web forgery (e.g., creating a website that represents a fake business). Web forgery involves manipulating the sender's information and the content. Web forgery becomes relevant for social media services due to the increasing trend of including user-developed applications or widgets in many of these services.

Even after an internal review mechanism that is effective in detecting malicious software, vulnerabilities may still be unexpectedly present in these social media applications.

5. Challenges and Opportunities

The cost associated with successful deception in social media environments open up several challenges which include: a) the lack of a standard, unified theory and methods in deception detection for online contexts, b) the lack of universal or context-specific and computationally efficient methods for deception detection for large online environments, and, c) the lack of effort in deception prevention by social media developers.

5.1 Lack of a Standard Unified Theory and Methods for Online Deception

Currently, several theories for both online (e.g., phishing emails) and offline environments (e.g., employment interviews) have been used and proposed for detecting deception including Management Obfuscation Hypothesis (MOH), Information Manipulation Theory (IMT), Interpersonal Deception Theory (IDT), Four Factor Theory (FFT) and Leakage Theory (LT)[14]. These theories focus on detecting leakage cues that a deceiver gives away or strategic decisions made by the deceiver which will reveal deceptive intentions. The main drawback with these techniques is that they rely on a set of verbal and non-verbal cues that may not all apply to the online world. For example, the presence of non-verbals in some social media requires us to rethink what indicators can be used to measure them because they are not likely to exist online in the forms that they exist in the physical world. A shift in focus is required for online deception. Steps in that direction have been made with video blob analysis of hands and movements for detecting movements that are too fast for the eye to detect (100% multiple state classification accuracy but with a limited sample of only 5 interviews)[19], detection of image manipulation by detecting inconsistencies in compression artifacts (30% - 100% depending on type of image, compression and tampering method)[23], machine learning detection using audio and transcribed text to identify patterns that signal deception because of deviations from a baseline (66.4% accuracy when baseline is at 60.2%)[12], and computerized voice stress analysis to identify variations in an individual's speech patterns (56.8% - 92.8% accuracy depending on context)[6]. Furthermore, a promising aspect in social media is the fact that most of the verbal cues are text-based. Methods of using verbal deception detection have been used to successfully identify identity deception using techniques such as similarity analysis of profile information (80.4% - 98.6% accuracy)[30], similarity analysis along with natural language processing to identify identity deception through writing patterns (68.8% accuracy)[25], cross-referencing information between a social network and anonymized social networks that contain the nodes present in the first network to evaluate the trustworthiness of social network profile attributes (40% - 80% recall depending on metric and technique when baseline recall at 20%)[5], and natural language processing to identify text features that betray deceptive emails (75.4% accuracy)[27]. These techniques show that there are options available for addressing issues of deception online. However, these aforementioned techniques cannot be directly applied to

address all types of online deception for all types social media because: a) there is a large variation among social media in terms of design and the type and amount of information that is allowed to be exchanged between users, and, b) it is difficult to determine the context in which optimum accuracy will be achieved for each solution. Put simply, the field lacks a cohesive framework that captures the interdependencies and interactions among different detection methods, types of deception and types of social media.

5.2 Computational Efficiency

The techniques that are currently used in deception detection are highly context-specific and many of them cannot be applied to the online social media environment. Some of the most popular detection deception methods dealing with verbal communication include Content-Based Criteria Analysis (CBCA), Scientific Content Analysis (SCAN) and Reality Monitoring (RM)[14]. The applicability and effectiveness of these detection deception methods to social media is unclear. Methods dealing with verbal cues such as video analysis may be computationally inefficient[19]. Similarly, methods that aim to detect sender deception (identity deception) and use similarity analyses to match identities may be feasible for small datasets but a comparison of all records with one another results in a computational time complexity $O(N^2)$. In some contexts where the profile information is available and text comparison is possible for features on a profile, the time complexity can be reduced to O(w'N) using an adaptive sorted neighborhood method[30]. The method sorts a list of records based on profile features and then moves through the records using a window (w) that compares just the records within that window in order to find duplicates. The adaptive method shortens the window (w') by finding the first (if any) duplicate record in a window and then ignores all further comparisons within that window (hence $w' < w$) which drastically increases the efficiency of the algorithm (1.3 million records parsed in 6.5 minutes). Similarity analyses are the ones that can most likely produce the highest overheads especially in social media where datasets tend to be large. Scalability becomes an issue for large datasets which will require more efficient approaches. For such cases, perhaps techniques such as the Expectancy Violations Theory (EVT) which looks for deviations from a baseline[19] may be an efficient way of filtering suspect cases for further examination. This is a computationally cheaper alternative that can be applied to both cases of sender and content deception. For example, comparing deviations from a normal user baseline will require parsing a database just once leading to a complexity of O(N). Finally, methods used in deception detection in social media need to take into account social context features (such as friends and family of an individual) which have been found to increase the accuracy of detection of deception[18]. The downside to this is that social network analyses (SNA) tend to become dramatically more expensive as networks grow. Simple SNA metrics such as betweeness centrality becomes overwhelmingly difficult to compute as networks grow ($O(N^3)$) where N is the number of nodes and more advanced statistical methods such as exponential random graph models which make use of Markov chain Monte Carlo algorithms become costly to compute. Put simply, the potential for using these newly available social data is there, however, computational efficiency needs to be addressed for large social networks. On a positive note, a

new recent trend online is the formation of small social networking sites[5] and communities for which deception detection methods can become more computationally feasible.

5.3 Deception Prevention

Social media application designers need to make an effort to address the issue deception in social media environments in the future. For example, Wikipedia's policy requires information added to articles to be cited back to its source and has ensured that many baseless arguments are exposed to readers. Other social media services need to address the issue of identity verification. For example, paradoxically, individuals who do not have an account today in the popular social networking site, Facebook, are more likely to fall victims of identity theft (for sensitive information) along with their real-life friends. The issue is that friends and other users become wary in the presence of duplicate accounts especially when one has been active by the original owner of an identity. On the other hand, when a deceiver registers an identity that did not exist on a social media service before, users will be more likely to assume that the genuine owner just joined the service. In an attempt to increase their user base, social media services along with their easy registration and access features expose unsuspected individuals world-wide to online deception. An effort to standardize the user registration and verifying users' credentials needs to be investigated in the future.

6. Conclusion

The social media space keeps evolving and continues to be extended with a diverse set of tools and technologies that deceivers can use. While the physical distance that separates the deceiver and the victim may seem large, the damage that can be done is far from negligible. Individuals, organizations and governments are at risk. Understanding how online deception works through social media and future technologies remains a significant challenge. To address this challenge we need to design social media applications with various rules and norms that our traditional physical space does not have. Our desire for innovation has resulted in various online social media designs that we do not yet fully understand and their vulnerabilities are currently being exploited in various ways by various attackers including those involved with deception attacks. It is time we start thinking about how to design social interaction in social media environment to safeguard and protect social media users from the unforeseen consequences of online deception.

Acknowledgements

We would like to express our gratitude to Elisa Bertino and Moshe Vardi for their encouragements and support throughout the preparation of this paper. We thank them for the opportunity they gave us to revise and strengthen this contribution. We also thank the anonymous reviewers for their valuable feedback and comments which help us to improve the quality and presentation of the paper.

References

1. Brenner J, Smith A. *72% of Online Adults are Social Networking Site Users*.; 2013:15. Available at: http://pewinternet.org/Reports/2013/social-networking-sites.aspx .

2. Buller DB, Burgoon JK. Interpersonal Deception Theory. *Commun Theory*. 1996;6(3):203–242.

3. Burgoon J, Adkins M, Jensen JKML, et al. An Approach for Intent Identification by Building on Deception Detection. *Syst Sci 2005 HICSS '05 Proc 38th Annu Hawaii Int Conf*. 2005:21a–21a.

4. Castelfranchi C, Tan Y-H. The role of trust and deception in virtual societies. *Syst Sci 2001 Proc 34th Annu Hawaii Int Conf*. 2001:8 pp.

5. Dai C, Rao F-Y, Truta TM, Bertino E. Privacy-preserving assessment of social network data trustworthiness. *Collab Comput Networking, Appl Work (CollaborateCom), 2012 8th Int Conf*. 2012:97–106.

6. Damphousse KR, Pointon L, Upchurch D, Moore RK. *Assessing the validity of voice stress analysis tools in a jail setting.*; 2007.

7. Dando CJ, Bull R. Maximising Opportunities to Detect Verbal Deception: Training Police Officers to Interview Tactically. *J Investig Psychol Offender Profiling*. 2011;8(2):189–202. Available at: http://dx.doi.org/10.1002/jip.145.

8. Donath JS. Identity and deception in the virtual community. In: Smith MA, Kollock P, eds. *Communities in Cyberspace*. Routledge; 1999.

9. Ekman P. Deception, Lying, and Demeanor. In: Halpern DF, Voiskounsky AE, eds. *States of Mind : American and Post-Soviet Perspectives on Contemporary Issues in Psychology: American and Post-Soviet Perspectives on Contemporary Issues in Psychology*. Oxford University Press; 1997:93–105.

10. Galanxhi H, Nah FF-H. Deception in cyberspace: A comparison of text-only vs. avatar-supported medium. *Int J Hum Comput Stud*. 2007;65(9):770–783.

11. Grazioli S, Jarvenpaa SL. Perils of Internet fraud: an empirical investigation of deception and trust with experienced Internet consumers. *Syst Man Cybern Part A Syst Humans, IEEE Trans*. 2000;30(4):395–410.

12. Hirschberg J, Benus S, Brenier JM, et al. Distinguishing deceptive from non-deceptive speech. In: *Interspeech 2005*. Proceedings of Eurospeech'05; 2005:1833–1836.

13. Hoglund G, McGraw G. *Exploiting Online Games: Cheating Massively Distributed Systems*. First. Addison-Wesley Professional; 2007.

14. Humpherys SL, Moffitt KC, Burns MB, Burgoon JK, Felix WF. Identification of fraudulent financial statements using linguistic credibility analysis. *Decis Support Syst*. 2011;50(3):585–594.

15. Kaplan AM, Haenlein M. The early bird catches the news: Nine things you should know about micro-blogging. *Bus Horiz*. 2011;54(2):105–113.

16. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. *Bus Horiz*. 2010;53(1):59–68.

17. Kassin S, Fong C. "I'm Innocent!": Effects of Training on Judgments of Truth and Deception in the Interrogation Room. *Law Hum Behav*. 1999;23(5):499–516.

18. Li J, Wang GA, Chen H. PRM-based identity matching using social context. In: *Intelligence and Security Informatics, 2008. ISI 2008. IEEE International Conference on*.; 2008:150–155.

19. Lu S, Tsechpenakis G, Metaxas DN, Jensen ML, Kruse J. Blob Analysis of the Head and Hands: A Method for Deception Detection. *Syst Sci 2005 HICSS '05 Proc 38th Annu Hawaii Int Conf*. 2005:20c–20c.

20. Madhusudan T. On a text-processing approach to facilitating autonomous deception detection. *Syst Sci 2003 Proc 36th Annu Hawaii Int Conf*. 2003:10 pp.

21. Nunamaker Jr. JF. Detection of deception: collaboration systems and technology. *Syst Sci 2004 Proc 37th Annu Hawaii Int Conf*. 2004:1 pp.

22. Podhradsky A, D'Ovidio R, Engebretson P, Casey C. Xbox 360 Hoaxes, Social Engineering, and Gamertag Exploits. In: *System Sciences (HICSS), 2013 46th Hawaii International Conference on*.; 2013:3239–3250.

23. Popescu AC, Farid H. Exposing digital forgeries by detecting traces of resampling. *Signal Process IEEE Trans*. 2005;53(2):758–767.

24. Shen X (Sherman). Security and privacy in mobile social network [Editor's Note]. *IEEE Netw*. 2013;27(5):2–3.

25. Solorio T, Hasan R, Mizan M. A Case Study of Sockpuppet Detection in Wikipedia. In: Farzindar A, Gamon M, Nagarajan M, Inkpen D, Danescu-Niculescu-Mizil C, eds. *Proceedings of the Workshop on Language Analysis in Social Media*. Stroudsburg, PA: The Association for Computational Linguistics; 2013:59–68.

26. Squicciarini AC, Griffin C. An Informed Model of Personal Information Release in Social Networking Sites. In: *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*.; 2012:636–645.

27. Stone A. Natural-Language Processing for Intrusion Detection. *Computer (Long Beach Calif)*. 2007;40(12):103–105.

28. Suler J. The Online Disinhibition Effect. *CyberPsychology Behav*. 2004;7(3):321–327.

29. Tzu S. *The Art of War. Translated by Samuel B. Griffith*. New York: Oxford University; 1963.

30. Wang GA, Chen H, Xu JJ, Atabakhsh H. Automatically detecting criminal identity deception: an adaptive detection algorithm. *Syst Man Cybern Part A Syst Humans, IEEE Trans*. 2006;36(5):988–999.

31. Zahavi A. The Fallacy of Conventional Signalling. *Philos Trans R Soc London Ser B Biol Sci*. 1993;340(1292):227–230.

Michail Tsikerdekis is an Assistant Professor in the College of Communication and Information at the University of Kentucky. His research interests include sociotechnical systems and social interaction design in social media especially in the domain of social networks and collaborative projects.

Sherali Zeadally is an Associate Professor in the College of Communication and Information at the University of Kentucky. He is a Fellow of the British Computer Society and a Fellow of the Institution of Engineering Technology, England. His research interests include computer/system/cyber security, computer networking (wired/wireless), mobile computing, energy-efficient networking, and performance evaluation of systems and networks.

© ACM, 2014. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in *Communications of the ACM*, Vol. 57, No. 9, (September 2014)
http://doi.acm.org/10.1145/2629612