DOI:10.1145/2629612

The unknown and the invisible exploit the unwary and the uninformed for illicit financial gain and reputation damage.

BY MICHAIL TSIKERDEKIS AND SHERALI ZEADALLY

Online Deception in Social Media

PROLIFERATION OF WEB-BASED technologies has revolutionized the way content is generated and exchanged through the Internet, leading to proliferation of social-media applications and services. Social media enable creation and exchange of user-generated content and design of a range of Internet-based applications. This growth is fueled not only by more services but also by the rate of their adoption by users. From 2005 to 2013, users and developers alike saw a 64% increase in the number of people using social media;[1] for instance, Twitter use increased 10% from 2010 to 2013, and 1.2 billion users connected in 2013 through Facebook and Twitter accounts.[24] However, the ease of getting an account also makes it easy for individuals to deceive one another. Previous work on deception found that people in general lie routinely, and several efforts have sought to detect and understand deception.[20] Deception has been used in various contexts throughout human history (such as in World War II and the Trojan War) to enhance attackers' tactics. Social media provide new environments and technologies for potential deceivers. There are many examples of people being deceived through social media, with some suffering devastating consequences to their personal lives.

Here, we consider deception as a deliberate act intended to mislead others, while targets are not aware or do not expect such acts might be taking place and where the deceiver aims to transfer a false belief to the deceived.[2,9] This view is particularly relevant when examining social media services where the boundary between protecting one's privacy and deceiving others is not morally clear. Moreover, such false beliefs are communicated verbally and non-verbally,[14] with deception identifiable through cues, including verbal (such as audio and text), non-verbal (such as body movement), and physiological (such as heartbeat).

Training and raising awareness (such as might be taught to security personnel[17]) could help protect users of social media. However, people trained to detect deception sometimes perform worse in detection accuracy than people who are not trained,[17] and evidence of a "privacy paradox" points to individuals sharing detailed information, even though they are aware of privacy concerns,[26] making them more vulnerable to attack. Making things worse, social media, as a set of Internet-based applications, can be broadly defined as including multiple virtual environments.[15,16]

» key insights

- **In social media, deception can involve content, sender, and communication channel or all three together.**
- **The nature of a social medium can influence the likelihood of deception and its success for each deception technique.**
- **Deception detection and prevention are complicated by lack of standard online deception detection, of a computationally efficient method for detecting deception in large online communities, and of social media developers looking to prevent deception.**

PHOTOGRAPH BY NOMAD SOUL

Exploring deception in social media, we focus on motivations and techniques used and their effect on potential targets, as well as on some of the challenges that need to be addressed to help potential targets detect deception. While detecting and preventing deception are important aspects of social awareness relating to deception, understanding online deception and classifying techniques used in social media is the first step toward sharpening one's defenses.

Online Deception

Nature often favors deception as a mechanism for gaining a strategic advantage in all kinds of biological relationships; for example, viceroy butterflies deceive birds by looking like monarch butterflies (which have a bitter taste), ensuring their survival as long as there are not too many in a particular area.[8] Similarly, humans have long used deception against fellow humans.[3] In warfare, Chinese military strategist and philosopher Sun Tzu[29] famously said, "All warfare is based on deception."

Social media services are generally 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 communication takes place; media richness describes the amount of information that can be transmitted at a given moment. Self-representation determines the control users have representing themselves, whereas self-disclosure defines whether one reveals information, willingly or unwillingly. Using these characteristics, Kaplan and Haenlein[16] developed a table including multiple aspects of social media: blogs, collaborative projects (such as Wikipedia), social networking sites (such as Facebook), content communities (such as YouTube), virtual social worlds (such as Second Life), and virtual game worlds (such as World of Warcraft). Table 1 outlines an expanded classification of social media that also includes microblogging (such as Twitter) and social news sites (such as Reddit). We categorize 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 communicated through the medium and average immediacy as news comes in) and their low self-presentation/self-disclosure due to their nature as content-oriented communities.

Social media that give users freedom to define themselves are in the second row of Table 1, and 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, along with increased media richness and social presence, we note a transition from social media using just text for communication to rich media simulating the real world through verbal and non-verbal signals, as well as greater immediacy in virtual game worlds and virtual social communication. The differences between these types of social media affect how deception is implemented and its usefulness in deceiving fellow users.

In most social media platforms, communication is generally text-based and asynchronous, giving deceivers an advantage for altering content—an inexpensive way to deceive others. Zahavi[31] identified the difference between assessment signals that are reliable and difficult to fake and conventional signals that are easier to fake; for example, in the real world, if older people want to pass as younger, they might dress differently or dye their hair to produce conventional signals. However, it would be much more difficult to fake a driver's license or other authentic documentation. But social media provide an environment in which assessment signals are neither required nor the norm, making deception easy; for instance, gender switching online may require only a name change.

Social media provide an environment in which assessment signals are neither required nor the norm, making deception easy; for instance, gender switching online may require only a name change.

Difficulty Perpetrating Online Deception

The level of difficulty perpetrating online deception is determined by several factors associated with the deceiver, the social media service, the deceptive act, and the potential victim. Significant difficulty could deter potential deceivers, and lack of difficulty may be seen as an opportunity to deceive others (see Figure 1).

The deceiver. Several factors associated with deceivers determine the difficulty of trying to perpetrate online deception, including expectations, goals, motivations, relationship with the target, and the target's degree of suspicion.[2] Expectation is a factor that determines the likelihood of success in deception. More complex messages have a greater likelihood of being communicated.[20] Goals and motivations also determine the difficulty of perpetrating a deception. Goals are broader and longer term, and motivations consist of specific short-term objectives that directly influence the choice and type of deception. A taxonomy developed by Buller and Burgoon[2] described three motivators for deception: "instrumental," where the would-be deceiver can identify goal-oriented deception (such as lying about one's résumé on a social medium to increase the likelihood of more job offers); "relational," or social capital (such as aiming to preserve social relationships typical in online social networks);[26] and "identity" (such as preserving one's reputation from shameful events in an online profile). These motivators in turn determine the cost or level of difficulty to deceivers in trying to deceive; for example, deceivers motivated to fake their identity must exert more effort offline due to the presence of signals much more difficult to fake than online where many identity-based clues (such as gender and age) may take the form of conventional signals (such as adding information to one's profile page without verification). Difficulty perpetrating a deception is also determined by the deceiver's relationship to a target. Familiarity with a target and the target's close social network make it easier to gain trust and reduce the difficulty of perpetrating deception. Many users assume enhanced security comes with technology so are more likely to trust others online.[4] Moreover, the level of trust individuals afford a deceiver also reduces their suspicion toward the deceiver, thereby increasing the likelihood of being deceived.

Moral cost also increases the difficulty of perpetrating deception.[26] Moral values and feelings can influence what deceivers view as immoral in withholding information or even lying. In the real world, the immediacy of interaction may make it much more difficult to deceive for some individuals. In contrast, in the online world, distance and anonymity[28] contribute to a loss of inhibition; the moral cost is thus lower for deceivers.

Social media. Social media require potential targets and would-be deceivers alike to expand their perspective on how interactions are viewed between receiver and sender during deception; for instance, "interpersonal deception theory"[2] says the interaction between a sender and a receiver is a game of iterative scanning and adjustment to ensure deception success.

Donath[8] suggested that if deception is prevalent in a system (such as Facebook) then the likelihood of successful deception is reduced. It makes sense that the prevalence of deception in an online community is a factor that also determines difficulty perpetrating deception. Social media services that encounter too much deception will inevitably yield communities that are more suspicious. Such community suspicion will increase the number of failed attempts at deception. Moreover, increasing a potential target's suspicion will likewise increase the difficulty, thereby deterring deceivers from entering the community in the first place, though some equilibrium may eventually be reached. However, this rationale suggests communities without much deception are likely more vulnerable to attacks since suspicion by potential victims is low. Determining the prevalence of deception in a community is a challenge.

Similarly, the underlying software design of social media can also affect the degree of suspicion; the level of perceived security by potential victims increases the likelihood of success for would-be deceivers.[11] Software design can cause users to make several assumptions about the level of security being provided. Some aspects of the design can make them more relaxed and less aware of the potential signs of being deceived; for example, potential

Table 1. Social media classifications.

		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

Figure 1. Entities and participants involved in online deception.

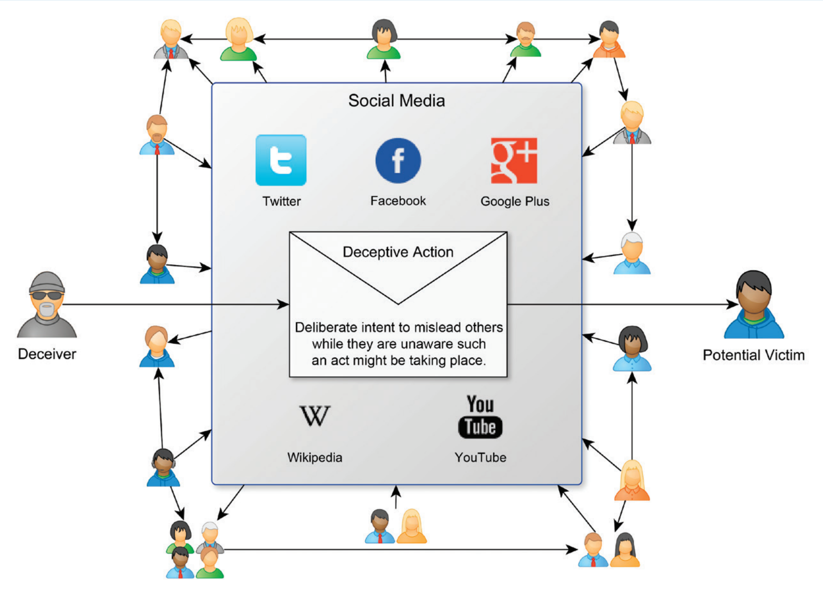
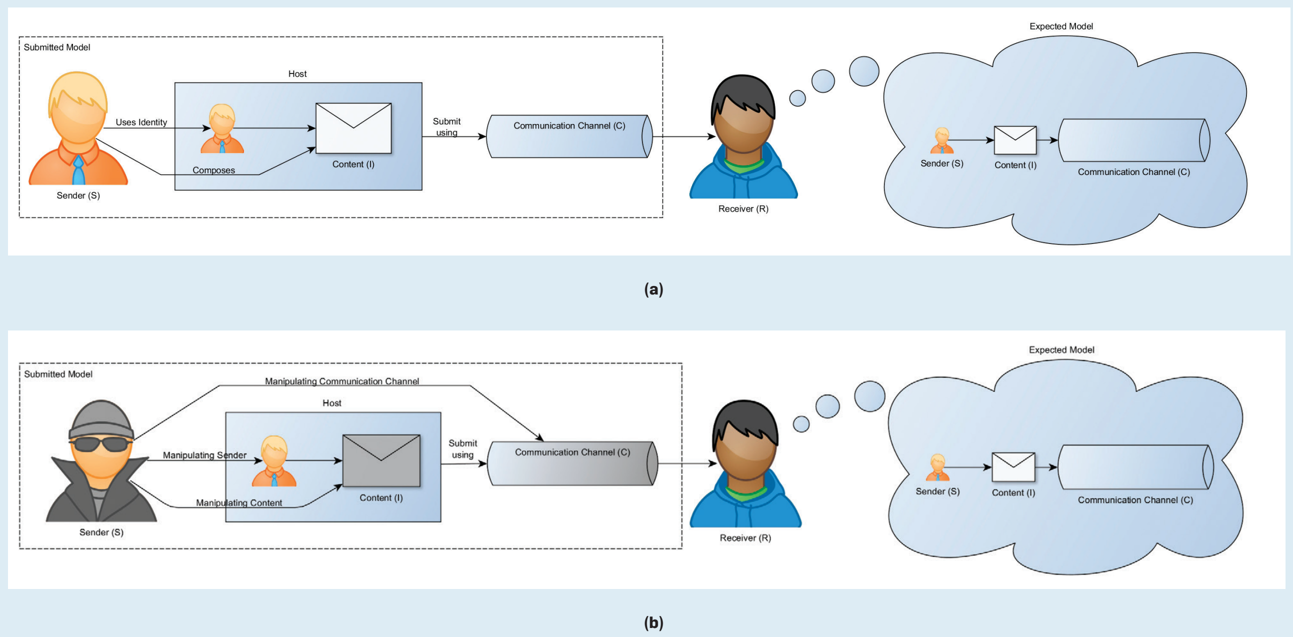


Figure 2. Interaction without and with deception.



targets may falsely assume that faking profile information on a social networking site is difficult due to multiple verification methods (such as email confirmation). Moreover, a system's assurance and trust mechanisms determine the level of trust between sender and receiver.[11] Assurance mechanisms can either reduce the probability of successful deception or increase the penalty for deceivers.[11] A tough penalty means increased difficulty for deceivers, especially when the chances of being caught are high. Assurance mechanisms are considered effective in certain contexts where the need for trust may be completely diminished. In social media, assurance mechanisms are much more difficult to implement, penalties and the chances of being caught may be or seem to be lower than those in offline settings, and the cost of deception is much lower. Media richness is another factor determining difficulty perpetrating deception. In this context, Galanxhi and Nah[10] found deceivers in cyberspace feel more stress when communicating with their victims through text rather than through avatar-supported chat.

Deceptive acts. Time constraints and the number of targets also help determine the difficulty perpetrating online deception. The time available and the time required for a successful attack are important, especially in social media services involving asynchronous communication. Moreover, the time required for deception to be detected also determines the effectiveness of the deception method being used. For instances where deception must never be discovered, the cost of implementing a deception method may outweigh any potential benefit, especially when the penalty is high. The social space in which deception is applied and the number of online user targets who are to be deceived help determine the level of difficulty implementing a deception method; for example, in the case of politicians trying to deceive through their online social media profiles, all potential voters face a more difficult challenge deciding how to vote compared to deceivers targeting just a single voter. Type of deception is another important factor. Complex deceptive acts motivated by multiple objectives (such as faking an identity to manipulate targets into actions that serve the deceiver's goals) are more difficult to perpetrate.

Potential victim. In real-world offline settings, the potential target's ability to detect deception may be a factor determining the difficulty perpetrating deception; for example, in a 2000 study of Internet fraud using page-jacking techniques, even experienced users of social media failed to detect inconsistencies, except for a select few who did detect it, thus showing detection is not impossible.[11] In social media, the potential targets' ability to detect deception also depends to some extent on their literacy in information communication technology. Deceivers must therefore evaluate the technology literacy of their potential victims. Users with high technology literacy have a significant advantage over casual Internet users, so the cost to a deceiver as calculated through a cost-benefit analysis for a social engineering attack may be higher.

Deception Techniques

Various techniques are reported in the literature for deceiving others in social media environments, including bluffs, mimicry (such as mimicking a website), fakery (such as establishing a fake website), white lies, evasions, exaggeration, webpage redirections (such as misleading someone to a false profile page), and concealment (such as withholding 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.

Deception model. The model (see Figure 2) consists of a sender (S), the content or message (I), the channel through which communication takes place (C), and the receiver (R). If a receiver's expected model (the so-called SIC triangle) is different from the received model (any or all SIC elements have been altered) then deception has occurred. This is also in line with Ekman's definition[9] of deception, saying a receiver cannot anticipate deception for deception to be considered deception. Deception is perpetrated by manipulating any of the SIC elements or any combination thereof. We present in the following paragraphs an overview of social media and identify factors and social-media types where deception can be perpetrated with minimal effort at low cost, resulting in a fairly high deception success rate (see Table 2). We identified these factors from the literature.

Content deception. Manipulating content, as in falsifying information, is presumably the most common way to deceive others. Social media that focus primarily on content (such as blogs, microblogs, content communities, and social news sites) are highly susceptible to such deception. Technology allows anyone with access privileges (legitimate and illegitimate) to manipulate multimedia files to an extraordinary degree. Tampering with images[23] is an effective way to fake content (such as representing that one traveled around the world through one's photos, altering them and sharing them through social media). Such a scheme may help deceivers elevate their social status and win a victim's trust to obtain further information. In addition to videos and images, the ease of manipulating content that is at times based on text alone yields low-cost deception and high probability of success due to the targets' low information literacy and lack of expectation for verifiability and even accountability. In addition, social media (such as social network sites and virtual social worlds) offering profile management for users are also susceptible, especially when advertising emphasizes the promise of new relationships. Competent deceivers may thus have a substantial advantage.

Collaborative projects (such as Wikipedia) are less likely to be affected by deception, or manipulating (I). The difficulty in perpetrating deception may seem low, but the likelihood of success (at least over the long term) is also low. This trade-off is due to the software design of these types of social media, where many-to-many communication enables many people to see the content. We see examples of content deception in Wikipedia, where not only vandals (people altering content with intent to deceive others) are eventually detected but other people assume a role in fighting them.[25] Furthermore, assurance mechanisms (such as a requirement for content validity, tracing content back to its source) are built into the system to ensure content deception is more apparent. Another example of content deception in social media involves open source software managed by multiple users where it is much more difficult to add malicious content and perpetrate a deception because multiple individuals evaluate the code before it is released. Virtual game worlds also have low probability for deception due to strongly narrated elements (such as being assigned specific roles that force players to follow a specific course of action).

Sender deception. Sender deception is achieved by manipulating the sender's identity information (S). Impersonation is a common example, resulting in identity deception, or identity theft.[30] Deceivers may gain access to an identity and use it to obtain additional information from their peers (such as home address, date of birth, and cellphone number). Failure to authenticate the sender's credentials yields deception. Social media's designs with built-in high self-presentation and self-disclosure enable low-cost sender deception. Blogs and microblogging can lead to stolen identities, as no control mechanisms are in place 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 fairly light, and long-term deceiver success is probably not guaranteed. Authentic-identity owners may become aware of the theft, and other individuals familiar with that identity may start identifying behavioral cues that do not match it. In the case of social network sites and virtual social worlds, the cost of deception increases because users must behave and communicate in ways that are appropriate to the identity they impersonate. The benefits are indeed much greater in a social medium because access to a user's personal social network can lead to enhanced ability to win other 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 that person's social network. With no control mechanisms in place for identifying a source, unregistered individuals without an account may be more exposed than registered users.

Social media (such as collaborative projects and virtual game worlds) with limited self-presentation and self-disclosure are likely to be more protected in terms of identity theft, due, in part, to their intended function. Collaborative projects, content communities, and virtual game worlds are heavily task-based, or contain a fixed narrative from which social behavior is not allowed to

Table 2. Manipulation of sender's identity information (S), content (I), and communication channel (C) with low difficulty and high deception success results.

Social media	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

deviate. Users who want to gain access to the impersonated identity's social network must perform just as well as the identity being impersonated and "act the part." The cost to a deceiver is likely to be great, and the success of the deception low and short term.

Middle ground between content deception and sender deception involves manipulating information associated with an identity. Such 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, would-be deceivers may try to fake some of the information in their profiles to win trust or represent themselves in a different way. In customer social network sites, would-be deceivers may try to conceal information to gain advantage when negotiating to buy or trade something.[5]

Communication-channel deception. Manipulating a communication channel requires greater technical skill, thus increasing the cost of deception. Such manipulation includes modifying in-transit messages, rerouting traffic, and eavesdropping. Jamming communications have been used in virtual game worlds. Podhradsky et al.[22] found multiplayer games in consoles can be hacked to provide access to a user's IP address. Would-be deceivers who gain access to the host can kick the player out and proceed with identity-theft deception. The deceiver's goal may not be to obtain information but to damage the victim's reputation. Worth pointing out is there is a fine line between an unintentional disconnection and an intentional departure of a player in a video game. This line is blurred when the player is on the losing side and leaves suddenly. As a result, the player's reliability and reputation are damaged by the invisible, anonymous deceiver. One advantage of communication-channel deception is the implicit assumption social media users make that digital technology is imperfect and things may not work as well as they do in the real world. However, nonverbal behavior[14] (such as body movement and speech patterns) can expose deceivers through social media by, say, introducing jitter or delays in their video or audio to conceal their deception, effectively increasing the likelihood of success. Victims at the other end of the connection find it difficult to differentiate an unreliable or slow connection from a deceptive act.

Since channel deception generally involves technology, all social media services may be susceptible to attack, especially those using similar technologies or architectures. Services that rely more on their client applications are more prone to attack, while those that rely on server applications are probably safer. Services with high media richness (such as virtual social worlds and virtual game worlds) tend to rely on client software. By exploiting communication channels, deception is common in such services.[13] Server-side applications (such as social networking sites and content communities) are less prone to channel deception because exploits rely on vulnerabilities of Web browsers and Web servers that are generally more secure. The cost of this deception is high, though the likelihood of success is also high, especially for a well-orchestrated attack.

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

Various hybrid deception techniques that manipulate a sender's information have been reported in the literature, including forgery,[20] phishing, identity forgery, Web forgery,[11] and email fraud. They are highly effective in social media (such as social-networking sites, virtual social worlds, microblogging, and blogs) that highlight user identity and provide one-to-one or one-to-many communications. These online deception attacks are not only effective but their consequences can lead to disaster, including loss of life. A service initially designed for people who want to initiate new relationships and the lack of verification can lead to a devastating outcome involving psychological or even physical damage to a victim. Online deception can also have financial consequences, as in Web forgery (such as creating websites representing fake businesses), manipulating the content of the sender's communication. Web forgery is relevant for social-media services due to the popularity of including user-developed applications or widgets. Even after internal review mechanisms that detect malicious software, vulnerabilities may still be present unexpectedly in such applications.

Challenges

The costs of deception in social media environments open several technical challenges that developers of social networks, as well as users, must address: lack of a standard, unified theory and methods for online deception detection; lack of a universal or context-specific, computationally efficient method for deception detection in large online communities; and lack of effort by social media developers in deception prevention.

Lack of a standard theory and methods. Several theories concerning online (such as phishing email) and offline environments (such as employment interviews) have been proposed for detecting deception, including "management obfuscation hypothesis," "information manipulation theory," "interpersonal deception theory," "four factor theory," and "leakage theory."[14] All focus on detecting leakage cues deceivers might give away or strategic decisions deceivers make that could reveal deceptive intent. Their main drawback is they rely on a set of verbal and nonverbal cues that may not all apply to the online world; for example, nonverbal cues in some social media communities require deception researchers and site developers to rethink what indicators can be used to recognize them, as they are not likely to exist online in the forms they take in the physical world.

New site-developer focus is required. Steps in that direction are being made with, for example, video blob analysis of hands and movement for detecting movement that is too quick for detection by the human eye (100% multiple state classification accuracy but with a limited sample of only five interviews);[19] detection of image manipulation through 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 due to 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]

The level of trust individuals afford a deceiver also reduces their suspicion toward the deceiver, thereby increasing the likelihood of being deceived.

One notably promising aspect in social media is that most verbal cues are based on text. Verbal deception detection has been used to identify identity deception (such as through similarity analysis of profile information (80.4%–98.6% accuracy);[30] similarity analysis 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 containing the nodes in the first network to evaluate the trustworthiness of social network profile attributes (40%–80% recall, depending on metric and technique when baseline recall is 20%);[5] and natural language processing to identify text features that betray deceptive email messages (75.4% accuracy).[27] These techniques show options are available for addressing online deception.

However, these techniques do not address all types of online deception for all types of social media; for one thing, there is much variation among social media in terms of design and type and amount of information allowed to be exchanged between users, and it is difficult to determine the context in which optimum accuracy will be achieved for each solution. The field lacks a cohesive framework that captures the interdependencies and interactions among different detection methods, types of deception, and types of social media.

Computational efficiency. The techniques being used for deception detection are highly context-specific, and many cannot be applied to the online social media environment. The most popular deception-detection methods dealing with verbal communication include "content-based criteria analysis," "scientific content analysis," and "reality monitoring."[14] Their applicability to social media is unclear. Methods dealing with verbal cues (such as video analysis) may be computationally inefficient.[19] Likewise, 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 results in a computational time complexity $O(N^2)$. In some contexts where profile information is available and text comparison is possible for features in a profile, the time complexity can be reduced to $O(w'N)$ through an adaptive sorted neighborhood method[30] that sorts a list of records based on profile features, then moves through the records using a window (w) comparing 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, then ignores all further comparisons within the window ($w' < w$), drastically increasing the efficiency of the algorithm (1.3 million records parsed in 6.5 minutes).

Similarity analyses are most likely to involve the greatest overhead, especially in social media where datasets tend to be large; scalability is a computational expense for large datasets so require more efficient approaches. For such cases, techniques (such as the "expectancy violations theory," which looks for deviations from a baseline[19]) may be an efficient way to filter suspect cases for further examination. This is a computationally cheaper alternative that can be applied to both sender and content deception; for example, comparing deviations from a normal user baseline requires parsing a database just once, leading to a complexity of $O(N)$.

Finally, methods used in deception detection in social media must account for features of social context (such as friends and family of an individual) that have been found to increase the accuracy of detection of deception.[18] The downside is social network analyses (SNAs) tend to be dramatically more expensive as networks grow. Simple SNA metrics (such as "betweeness centrality") become overwhelmingly difficult to compute as networks grow ($O(N3)$) where N is the number of nodes and more advanced statistical methods (such as exponential random graph models using Markov chain Monte Carlo algorithms) are costly to compute. However, the potential for this newly available social data is apparent, and computational efficiency must be addressed in large social networks. On a positive note, one online trend is formation of small social networking sites[5] and communities for which deception-detection methods may be more computationally feasible.

Deception prevention. Social media application designers must address deception in social media environments; for example, Wikipedia's editing policy requires information added to articles to be cited back to its source and has exposed many baseless arguments to readers. Other social media services must address identity verification; for example, individuals who do not have a Facebook account are paradoxically more likely to fall victim to identity theft (for sensitive information), along with their real-life friends. Friends and other users become wary in the presence of duplicate accounts, especially when a social media account has been active by the original owner of an identity. On the other hand, when a deceiver registers an identity that did not previously exist in a social media service, users are more likely to assume the genuine owner only recently joined the service. In an attempt to increase their user base, social media services, using easy registration and access features, expose unsuspecting users to online deception. An effort to standardize user registration and credential verification must be investigated by government agencies and technical organizations, as elements of everyday life shift to an all-online presence.

Conclusion

Social media keep being extended through a diverse set of tools and technologies available to deceivers. While the physical distance separating a deceiver and a potential target may seem large, the damage that could be done could be enormous. Individuals, organizations, and governments are at risk. Understanding how online deception works through social media is a challenge. To address it, the social media industry must design applications with rules and norms lacking in traditional physical space. Vast numbers of users' desire for innovation and personal connection, as well as romance, has resulted in online designs not yet fully understood, with vulnerabilities exploited by attackers, including those engaging in deception attacks. Researchers, developers, and communities must address how to design social interaction in social-media environments to safeguard and protect users from the consequences of online deception.

Acknowledgments

We thank Elisa Bertino and Moshe Y. Vardi for their encouragement and support throughout the preparation of this article and for the opportunity they gave us to revise and strengthen it. We also thank the anonymous reviewers for their valuable feedback and comments that helped us improve it.

References

1. Brenner, J. and Smith, A. *72% of Online Adults are Social Networking Site Users.* Pew Internet & American Life Project, Washington, D.C., Aug. 5, 2013; http://pewinternet.org/Reports/2013/social-networking-sites.aspx
2. Buller, D.B. and Burgoon, J.K. Interpersonal deception theory. *Communication Theory 6*, 3 (Aug. 1996), 203–242.
3. Burgoon, J., Adkins, M., Kruse, J., Jensen, M.L., Meservy, T., Twitchell, D.P., Deokar, A., Nunamaker, J.F., Lu, S., Tsechpenakis, G., Metaxas, D.N., and Younger, R.E. An approach for intent identification by building on deception detection. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (Big Island, HI, Jan. 3–6). IEEE, New York, 2005.
4. Castelfranchi, C. and Tan, Y-H. The role of trust and deception in virtual societies. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences* (Maui, HI, Jan. 3-6). IEEE, New York, 2001.
5. Dai, C., Rao, F.-Y., Truta, T.M., and Bertino, E. Privacy-preserving assessment of social network data trustworthiness. In *Proceedings of the Eighth International Conference on Networking, Applications and Worksharing* (Pittsburgh, PA, Oct. 14-17). IEEE, New York, 2012, 97–106.
6. Damphousse, K.R., Pointon, L., Upchurch, D., and Moore, R.K. *Assessing the Validity of Voice Stress Analysis Tools in a Jail Setting: Final Report to the U.S. Department of Justice.* Washington, D.C., 2007; http://www.ncjrs.gov/pdffiles1/nij/grants/219031.pdf
7. Dando, C.J. and Bull, R. Maximising opportunities to detect verbal deception: Training police officers to interview tactically. *Journal of Investigative Psychology and Offender Profiling 8*, 2 (July 2011), 189–202.
8. Donath, J.S. Identity and deception in the virtual community. In *Communities in Cyberspace*, M.A. Smith and P. Kollock, Eds. Routledge, New York, 1999, 29–59.
9. Ekman P. Deception, lying, and demeanor. In *States of Mind: American and Post-Soviet Perspectives on Contemporary Issues in Psychology*, D.F. Halpern and A.E. Voiskounsky, Eds. Oxford University Press, New York, 1997, 93–105.
10. Galanxhi, H. and Nah, F.F.-H. Deception in cyberspace: A comparison of text-only vs. avatar-supported medium. *International Journal of Human-Computer Studies 65*, 9 (Sept. 2007), 770–783.
11. Grazioli, S. and Jarvenpaa, S.L. Perils of Internet fraud: An empirical investigation of deception and trust with experienced Internet consumers. *IEEE Transactions on Systems, Man and Cybernetics 30*, 4 (July 2000), 395–410.
12. Hirschberg, J., Benus, S., Brenier, J.M. et al. Distinguishing deceptive from non-deceptive speech. In *Proceedings of the Ninth European Conference on Speech Communication and Technology* (Lisbon, Portugal, Sept. 4–8, 2005), 1833–1836.
13. Hoglund, G. and McGraw, G. *Exploiting Online Games: Cheating Massively Distributed Systems.* Addison-Wesley Professional, Boston, 2007.
14. Humpherys, S.L., Moffitt, K.C., Burns, M.B., Burgoon, J.K., and Felix, W.F. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems 50*, 3 (Feb. 2011), 585–594.
15. Kaplan, A.M. and Haenlein, M. The early bird catches the news: Nine things you should know about microblogging. *Business Horizons 54*, 2 (Mar.–Apr. 2011), 105–113.
16. Kaplan, A.M. and Haenlein, M. Users of the world, unite! The challenges and opportunities of social media. *Business Horizons 53*, 1 (Jan.–Feb. 2010), 59–68.
17. Kassin, S. and Fong, C. 'I'm innocent!': Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior 23*, 5 (Oct. 1999), 499–516.
18. Li, J., Wang, G.A., and Chen, H. PRM-based identity matching using social context. In *Proceedings of the IEEE International Conference on Intelligence and Security Informatics* (Taipei, June 17–20). IEEE, New York, 2008, 150–155.
19. Lu, S., Tsechpenakis, G., Metaxas, D.N., Jensen, M.L., and Kruse, J. Blob analysis of the head and hands: A method for deception detection. In *Proceedings of the 38th Annual Hawaii International Conference on Systems Sciences* (Big Island, HI, Jan. 3–6). IEEE, New York, 2005.
20. Madhusudan, T. On a text-processing approach to facilitating autonomous deception detection. In *Proceedings of the 36th Annual Hawaii International Conference on Systems Sciences* (Big Island, HI, Jan. 6–9). IEEE, New York, 2003.
21. Nunamaker Jr., J.F. Detection of deception: Collaboration systems and technology. In *Proceedings of the 37th Annual Hawaii International Conference on Systems Sciences* (Big Island, HI, Jan. 5–8). IEEE, New York, 2004.
22. Podhradsky, A., D'Ovidio, R., Engebretson, P., and Casey, C. Xbox 360 hoaxes, social engineering, and gamertag exploits. In *Proceedings of the 46th Annual Hawaii International Conference on Systems Sciences* (Maui, HI, Jan. 7–10). IEEE, New York, 2013, 3239–3250.
23. Popescu, A.C. and Farid, H. Exposing digital forgeries by detecting traces of resampling. *IEEE Transactions on Signal Processing 53*, 2 (Feb. 2005), 758–767.
24. Shen, X. Security and privacy in mobile social network. *IEEE Network 27*, 5 (Sept.–Oct. 2013), 2–3.
25. Solorio, T., Hasan, R., and Mizan, M. A case study of sockpuppet detection in Wikipedia. In *Proceedings of the Workshop on Language Analysis in Social Media*, A. Farzindar, M. Gamon, M. Nagarajan, D. Inkpen, and C. Danescu-Niculescu-Mizil, Eds. (Atlanta, June 3). Association for Computational Linguistics, Stroudsburg, PA, 2013, 59–68.
26. Squicciarini, A.C. and Griffin, C. An informed model of personal information release in social networking sites. In *Proceedings of the 2012 International Conference on Social Computing* (Amsterdam, Sept. 3–5). IEEE, New York, 2012, 636–645.
27. Stone, A. Natural-language processing for intrusion detection. *Computer 40*, 12 (Dec. 2007), 103–105.
28. Suler, J. The online disinhibition effect. *CyberPsychology & Behavior 7*, 3 (June 2004), 321–326.
29. Tzu, S. *The Art of War* (translated by Samuel B. Griffith). Oxford University Press, New York, 1963.
30. Wang, G.A., Chen, H., Xu, J.J., and Atabakhsh, H. Automatically detecting criminal identity deception: An adaptive detection algorithm. *IEEE Transactions on Systems, Man, and Cybernetics 36*, 5 (Sept. 2006), 988–999.
31. Zahavi, A. The fallacy of conventional signalling. *Philosophical Transactions of the Royal Society of London 340*, 1292 (May 1993), 227–230.

Michail Tsikerdekis (tsikerdekis@uky.edu) is an assistant professor in the College of Communication and Information of the University of Kentucky, Lexington, KY.

Sherali Zeadally (szeadally@uky.edu) is an associate professor in the College of Communication and Information at the University of Kentucky, Lexington, KY.

© 2014 ACM 0001-0782/14/09 $15.00