Ethical and Social Challenges with developing Automated Methods to Detect and Warn potential victims of Mass-marketing Fraud (MMF)

Extended Abstract

Monica Whitty
Cyber Security Centre
WMG, University of Warwick
UK
mw@warwick.ac.uk

Matthew Edwards
Security Lancaster
University of Lancaster
UK
m.edwards7@lancaster.ac.uk

Michael Levi
School of Social Sciences
Cardiff University
UK
levi@cardiff.ac.uk

Claudia Peersman
Security Lancaster
University of Lancaster
UK
c.peersman2@lancaster.ac.uk

Awais Rashid
Security Lancaster
University of Lancaster
UK
marash@comp.lancs.ac.uk

Angela Sasse
Information Security Group
University College London
UK
a.sasse@ucl.ac.uk

Tom Sorell
Politics & International Studies
University of Warwick
UK
t.e.sorell@warwick.ac.uk

Gianluca Stringhini
Security & Crime Science
University College London
UK
g.stringhini@ucl.ac.uk

ABSTRACT

Mass-marketing frauds (MMFs) are on the increase. Given the amount of monies lost and the psychological impact of MMFs there is an urgent need to develop new and effective methods to prevent more of these crimes. This paper reports the early planning of automated methods our interdisciplinary team are developing to prevent and detect MMF. Importantly, the paper presents the ethical and social constraints involved in such a model and suggests concerns others might also consider when developing automated systems.

Keywords

ACM proceedings, mass-marketing fraud, online scams, cybercrime, cybersafety

1. INTRODUCTION

Mass-marketing fraud (MMF) is a serious, complex, and organized crime. Examples include: foreign lotteries and sweepstakes (in which the victim believes they have won money from a lottery and are told to pay a fee in order to release the funds), '419' scams (advance fee fraud, in which victims believe that for a small amount of money they will make a large fortune), and romance scams (taken in by a fake online dating persona, in which the victim sends the 'fake persona' money) [10, 11]. Some MMFs are low-value one off scams on large numbers of victims, whilst others involve developing a relationship (e.g., romantic, business, friendship) where money is defrauded over time, again with multiple simultaneous or sequential victims.

Victims of MMF suffer both financial losses and psychological impacts, with psychological effects sometimes outweighing the financial impact, even when large sums of money are lost [12]. Psychological harm can include: shame, guilt, embarrassment, depression, feeling suicidal, grief, anxiety, and loss of trust. Catching and prosecuting MMF criminals is a difficult task. This is for three main reasons: (1) the criminals often live in a different country to the victims, (2) the methods the criminals use that make them difficult to trace, and (3) prosecution is very time consuming, owing to the large amounts of online data that need to be analysed to establish evidence against the criminals and gain intelligence about their whereabouts and operating tactics.

Law enforcement and others have attempted to prevent the crime by using disruption tactics. Dating sites, for instance, have been asked to share known fake profiles in order to help reduce the number of criminal profiles. Facebook has attempted to take down known fake profiles. Anti-Money Laundering regulations increase the identifiability of transactions and recipients when money is transferred via money transfer companies, such as Western Union and MoneyGram.

There are also numerous guidelines, campaigns, websites, and phone apps available that attempt to educate users about scams in an attempt to prevent victimisation aris-

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3053891

ing from MMF. They suggest basic rules such as: never click on a link in an email; never respond to an email asking for confirmation of banking details; never send any money to strangers you met online, etc. Warnings about online security often focus exclusively on idealised individual behaviour, and assume that people fall for scams because they lack knowledge. However, researchers have found that many victims who fall for MMFs have heard of these scams, some having detailed knowledge [5]. Lea et al., argue that detailed knowledge of a scam increases vulnerability, as these individuals often develop an "illusion of invulnerability" [5]. It has been found that even when authority figures (e.g., police, law enforcement, bank managers) attempt to alert a person to the fact that they have become a victim of a romance scam, the victim often has difficulty believing them. Moreover, even when the victim questions the criminal about his/her authenticity, the criminal will employ persuasive techniques to convince the victim [11]. The number of repeat victims also suggests that this is a difficult population to help to recognise scams. Given that knowledge about a scam, therefore, may not be enough to prevent individuals from becoming defrauded, other types of interventions are needed.

One of the problems with recognising a scam is that fraudsters are very proficient at taking up assumed identities as well as developing trusting relationships with potential victims. Moreover, some individuals might be more vulnerable to trusting a scam, compared with others [1, 6]. Criminals employ a range of persuasive techniques and for some scams 'groom' individuals, waiting for the right moment to start making requests for money.

It has also been argued that "falling for a scam comes down to errors in decision-making", and "scammers create situations (with their scam offers) that increase the likelihood of poor decision- making" [5]. Cognitive (e.g., overconfidence in a specific topic) and motivational (e.g., the scam triggers positive emotions) processes also explain the psychological reasons for responding to scams. The most consistent finding with regards to reasons why people are scammed include: 'appeals to trust and authority' (i.e., the use of people or institutions of authority to make the scam appear legitimate) and 'visceral triggers' (triggers employed to make potential victims focus on huge prizes and imagined positive future emotional states).

Given the problems users have with identifying scams, in our research we are exploring the use of automated processes to identify communication with a potential scammer and hoping to do so prior to the 'sting' taking place (i.e., prior to any loss of monies). Such automated mechanisms will need to make decisions about the probability of a victim communicating with a scammer by drawing upon their personal and potentially others' personal communication. This paper briefly outlines the research being undertaken, while specifically highlighting the potential ethical and social concerns related to the autonomous agent we are developing. We also point out how these concerns might relate to the development of other types of automatic detection systems.

2. DETECTING COMMUNICATION WITH SCAMMERS

In recent years, scholars in the field of computing science have studied extensively the problem of large-scale cybercrime. In particular, research has focused on operations that involve botnets and networks of compromised computers that act under the control of the same cybercriminal [2]. A key aspect of these illicit operations is that they are run in an automated fashion, and that the malicious activity that leads to the monetization of the operation (e.g., the spread of malware, spam, or phishing) is carried out by computer programs (i.e., malware). Detection is possible due to the spread of similar content and scripts sent across the Internet or at clear signs of automated activity (e.g., synchronization across multiple email senders).

MMFs, however, are more difficult to detect compared with phishing or even spear phishing. MMFs are especially a challenge to detect because a) they typically involve communication with another person, rather than a bot – this means that crime narratives can vary and are more complex; b) often the criminal is developing a relationship that appears authentic to the victims (romantic, friendship, working relationship) over a long-period of time prior to asking for money; c) they can vary the communication when a user demonstrates a lack of trust; and d) they use multiple media channels to communicate with the user (e.g., dating site, instant messaging, email).

The research being undertaken in our project: 'DAPM: Detecting and Preventing Mass-Marketing Fraud' is drawing from psychology, media and communications, criminology and linguistics to help identify deception and persuasive communication and evidence of grooming often evident in MMFs. We are also interested in identifying: the online identities, other communication and online behaviors typical of scammers as well as victims. We are also examining whether the psychological characteristics that are more typical of victims (e.g., romantic beliefs, impulsivity) [1] might be detected in victims' online communications.

We are also examining socio-technical features to identify MMF. In particular, we are building on Huang et al.'s work [4] to investigate characteristics of scammer profiles, such as the utilization of same profile photographs, descriptions across multiple profiles and patters of interaction and contact with other users (e.g., login times). We are also examining socio-technical characteristics typical of users, such as replying to any message they receive or immediately giving away their phone number. We will utilize supervised machine learning techniques such as random forests and clustering of profiles to this end.

In addition to identifying technical features we hope to uncover specific linguistic features: typical of a victim; indicative of scammers' communication; and indicative of the interactions that take place between victims and scammers. For example, previous work has found that victims of romance scam are more likely to hold 'idealized' romantic beliefs – that is, believing that they can find a relationship with someone who will be their 'true' love and this person will be nearly perfect [1]. We might hypothesize, therefore, that someone who is more susceptible to the romance scam will write a profile specifying they are seeking out the perfect partner. We might also hypothesize that scammers will create profiles in a similar way and will communication emphasizing how ‘perfect’ they perceive the victim and that they believe that the victim is their ‘true love’.

We also hope to uncover deceptive communication. For example, previous work has found that liars tend to use more words, and these words are more informal and expressive, compared with people telling the truth. They also make more typographic errors [13]. It has also been found that word count is significantly higher in deceptive communication on Instant Messaging compared with when people speak the trust. Moreover, individuals who were lying were more likely to ask questions that they who were being honest [3]. Linguistic indicators of deception, therefore, might help us to automatically detect mass-marketing fraudsters.

Cultural indicators might also be useful for our agent. Given that it is believed that many scammers are West Africans, there may be some benefits into detecting West African communication in our model.

3. PERSONAL DATA USED FOR IDENTIFICATION

In the research for this project we will be drawing on personal data for our analysis. For example, we will be conducting analysis of profiles from dating sites, employment webpages and so forth. We will be conducting analysis of text produced by both the victim and the criminal. We might also wish to compare communication with the victim and non- criminals to examine any significant difference in stylistic features of this communication. Some of this text, we foresee, will be highly personal communication. In all instances, we will be able to anonymize the data – beyond simply taking out ‘real names’ (e.g., we will anonymize any occupational roles leak out identity information). Much of this anonymization we will conduct automatically rather than manually.

The research will, of course, strictly follow ethical guidelines set out by the British Psychological Society and the UK Research Councils – and so we don’t foresee any specific ethical concerns regarding the research. Participants will be giving informed consent and can withdraw from the research without penalty. Participants’ confidentiality will be respected and their personal data will be stored securely.

Given we are conducting our research following ethical guidelines and principles we are not concerned with the ethics or moral issues regarding our research practice. The ethics and social challenges, however, which do concern us, are in the next stages of the project – which will be undertaken should we reveal effective identification of scammers in our research studies.

4. DEVELOPING A SYSTEM TO AUTOMATICALLY DETECT AND WARN POTENTIAL VICTIMS OF MMF

In addition to the identification platform, should our research successfully detect fraudsters, we intend to develop a proof of concept application to warn users about scam accounts. This application will be composed of a web browser extension to be installed on user computers. In addition, the tool will collect information about the accounts that our backend will detect belong to scammers.

In brief, the system would need to analyze personal data on the end-user’s machine (e.g., online profiles, emails, Instant Messaging, etc.). This could possibly be all data created and sent to their personal computers or there might be a reason to bracket some of these data. In addition to analyzing the data to make decisions about the probability that the user is communicating with a fraudster the system would need to feedback this information to the user – and would need to do so in a convincing and persuasive manner (especially given that victims often do not believe others, including those in authority positions, such as the police, when they are informed that they have been scammed [10].) We will be drawing from expertise in HCI to develop different types of warning messages and test their effectiveness (e.g., visually wording of the messages) [9, 8]. We would therefore need to build a system that end-users might trust and that could win the trust over from criminal to the autonomous system.

5. ETHICAL AND SOCIAL CHALLENGES WITH OUR AUTOMATED SYSTEM

The benefits of an automatic detection system are obvious. Should it be able to detect end-users’ communication with scammers, it has the potential to prevent victims from sending money, downloading harmful malware on their computers used later for identity fraud; prevent victims from forming intimate relationships with a criminal – only later to be traumatized by the end of the relationship; prevent victims from creating compromising material (e.g., naked photographs, cybersex videos) that might be used at a later date as blackmail.

Despite these benefits, there are nonetheless ethical and social challenges associated with such a system. The proposed detection system would need to be making decisions regarding ‘genuine’ people compared with ‘fraudsters’. There is, of course, the possibility of false positives. False positives with such an agent have different ramifications compared with an email that is mistakenly filtered out into the junk box. Imagine the hypothetical scenario described below:

> The end user might be seeking out a partner on a dating site and the person they are communicating with (referred to as *unknown*) might be a good match, with genuine similar interests and outlook in life. The *unknown* exaggerates their attributes (slightly more than users would normally exaggerate). They are also somewhat socially clumsy and are rarely given attention and so flatter the end-user in an exaggerated manner, more akin to a scammer. The end-user is also a General in the American Army. Given this combination of factors, the system warns the end-user communicating that the *unknown* is most likely a scammer and so the end-user decides to end the relationship. In this scenario both parties have found dating difficult and rarely opportunities present themselves to each person. Moreover, the *unknown* was unaware that their data had been analyzed by an automated detection system and warned the end-user that they could be communicating with a criminal.

In the above scenario we are left with the difficult choice of wanting to protect the end-user from potential harm, at the same time as still wanting to ensure possible future happiness. Moreover, the *unknown* has been affected – not only does this person have their personal data used by an automated program, unbeknownst to them, to make decisions about their authenticity, but they have also potentially been prevented from experiencing future happiness with a well-suited partner.

One way around this problem might be to treat cases like the one in the above scenario as data that are presented to them in a junk box – similar to threats of phishing. Another way might be to give the case a probability score. Unfortunately, however, one of the difficulties with a victim-oriented approach to preventing MMF is that victims tend to prefer to believe the criminal than others who challenge the reality of the situation (e.g., when law enforcement inform citizens that they are a victim romance scam they are often disbelieved) [10].

The above example highlights one of the dilemmas we are faced with in the 'art of the possible'. The solution might be that we include more of the human element – given that it will be challenge to develop a 'perfect' autonomous system that never makes mistakes. In the scenario given, the human element might be an alert system providing a series of steps for the end-user to complete to check the authenticity of the *unknown* – rather a simple warning message that they are about to be scammed. Other solutions might be considering where we place the software – on dating sites to make decisions, to family and friends that help the end user make decisions (especially for the vulnerable whose cognitive abilities are diminished).

6. CONCLUSIONS

In summary, software programs are becoming a part of our everyday lives, with systems built to automate e-commerce, medical decisions and vehicles. Despite the popularity of these systems and the advancement of science to create these systems, there remains the concern as to how to create ethical autonomous systems – or how to implement such systems in an ethical manner. It has been argued that software engineers often "do not cater for the 'messiness' of social life and social research and its continuous impact on design choices" (p.523) [7]. Different ethical, social and psychological concerns might be related to different types of automated programs. Previous work, for example, has raised concerns with drones lowering the thresholds to fire when in warfare. This paper, we believe raises an important concern about needing to balance the need to help and support individuals, whilst still allowing the end user the opportunities they would have otherwise had in their lives. Moreover, like other research on automated programs it raises the concern of privacy. The work we present here will guide our own research as we develop our autonomous agent to prevent MMF and we hope will be considered and guide future development of autonomous agents.

Acknowledgements

The work reported in this paper is supported by award `EP/N028112/1` 'DAPM: Detecting and Preventing Mass-Marketing Fraud (MMF)', from the UK Engineering and Physical Sciences Research Council.

7. REFERENCES

[1] T. Buchanan and M. T. Whitty. The online dating romance scam: causes and consequences of victimhood. *Psychology, Crime & Law*, 20(3):261–283, 2014.

[2] G. Gu, P. A. Porras, V. Yegneswaran, M. W. Fong, and W. Lee. Bothunter: Detecting malware infection through ids-driven dialog correlation. In *Usenix Security*, volume 7, pages 1–16, 2007.

[3] J. T. Hancock, L. Curry, S. Goorha, and M. Woodworth. Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*, pages 22c–22c. IEEE, 2005.

[4] J. Huang, G. Stringhini, and P. Yong. Quit playing games with my heart: Understanding online dating scams. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 216–236. Springer, 2015.

[5] S. Lea, P. Fischer, and K. M. Evans. The economic psychology of scams. *International Association for Research in Economic Psychology and the Society for the Advancement of Behavioral Economics, Nova Scotia, Canada*, 2009.

[6] D. Modic and R. J. Anderson. We will make you like our research: The development of a susceptibility-to-persuasion scale. *Social Science Research Network*, 2014.

[7] A. Rashid, K. Moore, C. May-Chahal, and R. Chitchyan. Managing emergent ethical concerns for software engineering in society. In *Software Engineering (ICSE), 2015 IEEE/ACM 37th IEEE International Conference on*, volume 2, pages 523–526. IEEE, 2015.

[8] A. Sasse. Scaring and bullying people into security won't work. *IEEE Security & Privacy*, 13(3):80–83, 2015.

[9] D. Weirich and M. A. Sasse. Persuasive password security. In *CHI'01 Extended Abstracts on Human Factors in Computing Systems*, pages 139–140. ACM, 2001.

[10] M. T. Whitty. The scammers persuasive techniques model: Development of a stage model to explain the online dating romance scam. *British Journal of Criminology*, 53(4):665–684, 2013.

[11] M. T. Whitty. Mass-marketing fraud: a growing concern. *IEEE Security & Privacy*, 13(4):84–87, 2015.

[12] M. T. Whitty and T. Buchanan. The online dating romance scam: The psychological impact on victims–both financial and non-financial. *Criminology & Criminal Justice*, 16(2):176–194, 2016.

[13] L. Zhou, J. K. Burgoon, J. F. Nunamaker, and D. Twitchell. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group decision and negotiation*, 13(1):81–106, 2004.