Future Casting Influence Capability in Online Social Networks

Fake Accounts and the Evolution of the Shadow Economy

Matthew Duncan
DRDC Toronto Research Centre

Defence Research and Development Canada

Scientific Report
DRDC-RDDC-2015-R148
August 2015

Abstract

Online Social Networks (OSN) have risen to the point of dominating the daily activities of many Internet users. In response to this success, various groups have sought to exploit these apparently safe and well-populated walled gardens for their own gain. A particularly worrisome target of exploitation is the operation and control of actual accounts. It is now possible to create large numbers of fraudulent accounts with credentials of reasonable face validity. These fake accounts, also called sock-puppets or Sybils, can then be used to generate various influence effects through the apparent collective action of many “individuals”. Because the behaviour of accounts affects the perception of trust and legitimacy of the OSN’s user base, sock-puppet accounts are a serious threat to an OSN’s credibility and economic success. Due in part to the efforts of OSN to combat account fakery, a digital arms race has ensued contributing to the evolution of a technologically sophisticated economic service supply chain involving many players within the cyber-criminal and shadow economy ecosystem. Some of these include malware distributors, Botnet operators, and Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) solving services for account creation and authentication, as well as individuals with expertise in machine learning and artificial intelligence for automated credential generation and control of account behaviour. In its current form, the shadow economy ecosystem represents an influence capability that is open to anyone with money to purchase the accounts. These could range from government agencies to non-state actors to Public Relations (PR) firms to any individual with an axe to grind. In this paper I will describe the complex economic ecosystems that support the generation and sale of fake accounts, review a sample of influence cases involving fake accounts, and examine possible future applications of fake accounts to perpetrate operations for achieving influence effects.

Significance to Defence and Security

This research report is an analysis of current influence capability of Online Social Networks (OSN) with respect to the economic and technological infrastructure that supports it. The current use of OSN for influence is something that is being exploited now and consequently should not be ignored. The use of OSN for interaction and communication is only increasing. As such, it is not a fad that will simply go away in a few years. More importantly, this report also describes how various disruptive technologies and pressures may drive the evolution of the economic and technological infrastructure into a more robust and threatening influence capability in the future. Consequently, the future potential of OSN as an influence capability is something that cannot be ignored. There is greater and greater potential for influence within OSN, and those that are first to establish themselves and leverage this capability will be the ones that most benefit.

Résumé

Les réseaux sociaux ont gagné en popularité au point de dominer les activités quotidiennes de maints internautes. C’est pourquoi divers groupes cherchent à exploiter à leur avantage ces jardins clos fort peuplés en apparence inoffensifs. Un aspect particulièrement inquiétant de cette situation réside en la gestion et le contrôle de comptes réels. En effet, il est désormais possible de créer de nombreux comptes frauduleux à l’aide de justificatifs d’identité qui semblent valides. Ces comptes bidon (également appelés faux-nez) peuvent ensuite être exploités à diverses fins d’influence, sous l’action collective apparente d’un grand nombre de « particuliers ». Ces comptes faux-nez constituent une grande menace à la crédibilité et au succès économique d’un réseau social en ligne, car le comportement des comptes influence la perception de confiance et de légitimité de la base d’utilisateurs dudit réseau. La course aux armements numériques qui s’est ensuivie, en partie attribuable aux efforts déployés par les réseaux sociaux en ligne pour combattre la menace, a contribué à l’essor d’une chaîne d’approvisionnement de services économiques avancés sur le plan technologique, chaîne à laquelle prennent part un grand nombre de joueurs de l’écosystème cybercriminel et d’économie parallèle. Parmi ces joueurs, on compte des distributeurs de maliciels, des opérateurs de réseaux zombies et des services de résolution du Test de Turing complètement automatisé afin de distinguer les ordinateurs des humains (Completely Automated Public Turing test to tell Computers and Humans Apart – CAPTCHA) requis pour la création et l’authentification de comptes, ainsi que des particuliers possédant l’expertise en apprentissage machine et en intelligence artificielle nécessaire à la génération automatisée de justificatifs d’identité et au contrôle de comportement de comptes. Dans sa forme actuelle, l’écosystème d’économie parallèle représente une capacité d’influence accessible à quiconque possédant suffisamment d’argent pour acheter des comptes, y compris des organismes gouvernementaux, des entreprises de relations publiques (RP), des acteurs non étatiques ou toute personne qui y trouve un trouve un quelconque intérêt. Dans le présent rapport, je décrirai les écosystèmes économiques complexes qui soutiennent la production et la vente de comptes faux-nez, j’étudierai un exemple d’influence obtenue à l’aide de faux comptes et j’examinerai les applications futures possibles de ce type de compte pour exercer un trafic d’influence.

Importance pour la défense et la sécurité

Le présent document constitue une analyse de la capacité actuelle d’influence des réseaux sociaux en ligne par rapport à l’infrastructure technologique et économique qui appuie cette capacité. L’exploitation actuelle des réseaux sociaux en ligne à des fins d’influence est une réalité dont on ne doit pas faire abstraction. On utilise de plus en plus ces réseaux pour interagir et communiquer; il ne s’agit pas d’une tendance à la mode qui s’éteindra d’elle-même dans quelques années. Qui plus est, le rapport traite également de la façon dont diverses technologies et pressions perturbatrices peuvent orienter l’évolution de l’infrastructure technologique et économique et en faire une capacité d’influence plus robuste et menaçante dans le futur. Il est donc primordial de tenir compte du potentiel des réseaux sociaux en ligne en tant que capacité d’influence. Cette dernière ne cesse de croître au sein des réseaux sociaux en ligne; les premiers à en faire l’exploitation en tireront le plus grand profit.

Table of Contents

Abstract . . . i

Significance to Defence and Security . . . i

Résumé . . . ii

Importance pour la défense et la sécurité . . . ii

Table of Contents . . . iii

List of Tables . . . iv

1 Introduction . . . 1

1.1 Emergence of the Shadow Economy . . . 2

1.2 OSN as an Influence Capability . . . 3

1.3 Summary . . . 4

2 The Shadow Economy and Influence in OSN . . . 6

2.1 Fake Accounts and Influence Campaigns . . . 8

2.1.1 For-profit Influence Campaigns . . . 8

2.1.2 Not For-profit Influence Campaigns . . . 8

2.1.3 Effectiveness of Influence Campaigns . . . 11

2.2 Casting the Future of Influence Campaigns in OSN . . . 11

3 Conclusions . . . 14

References . . . 15

List of Symbols/Abbreviations/Acronyms/Initialisms . . . 21

Glossary . . . 23

List of Tables

Table 1: Account Creation Barriers and the Shadow Economy Component Providing Services to Circumvent It. 7

1 Introduction

Online Social Networks (OSN) such as Facebook, Twitter, and Google+ have risen in popularity to the point of dominating the daily activities of many Internet users. Given the number of people worldwide with internet access, the scope of their collective success is vast indeed. Active accounts for Facebook, Twitter, and Google+ globally number about one in three people at over 2.5 billion (Facebook, 2013; Google, 2013; Twitter, 2014).[1] That number represents a user base two to three times the size of the largest nations on Earth. Although media technologies such as television can globally claim a user base of around 2.6 billion, this user base is not monolithic and not interactive (IDATE Consulting & research, 2010).[2] On the contrary, the technology behind OSN, like other internet technologies such as E-mail, is specifically designed to enable users to interact directly unfettered by the direction or control of a central agency. From the perspective of social interaction, given the sheer scale of the OSN user base there is potential for inter-personal interaction and influence that is unprecedented in all of human history.

Because OSN were developed to foster information sharing and inter-personal communication, the technology can be seen as an open sandbox influence capability any single user can potentially exploit.[3] Although OSN are economically supported by an advertising based revenue model similar to television, albeit with different distribution and payment mechanics, the key difference is that the creation of influential content is in the hands of users rather than just the corporate entities which own the networks. In fact, OSN make a point of rewarding users for the content they generate and actively promoting those users who are popular. Because popularizing user created content is so central to the function of OSN, anyone wishing to raise the popularity of a public profile will ensure they have a presence in OSN. For example, it is now common practice for politicians, celebrities, and corporate entities to include among their advertising their Facebook page or Twitter account.

The success of OSN in achieving such heights of interaction between members and the potential ramifications of that success has not gone unnoticed. A variety of individuals and groups have already been hard at work exploiting for monetary gain the influence opportunities afforded by access to a vast population of users. Because OSN sites are designed around interaction between accounts, the for-profit exploitation in the form of spam and malware distribution that infected E-mail has largely carried over to OSN (see Lumenzanu & Feamster, 2012). Like E-mail, this kind of exploitive activity is principally done through the use of fake accounts, also known in the OSN domain as sock-puppets or Sybils (e.g., Thomas, Grier, Ma, Paxson, & Song, 2011; Yang, Wilson, Wang, Gao, Zhao, & Dai, 2011). Although accounts created for this purpose behave with

[1] This statistic ignores the fact that some individuals will hold accounts with multiple OSN. However, this is largely irrelevant to the purpose of this chapter and comparison to other types of media technologies.

[2] The estimate of individual users was based on figures for number of households. As of 2013, there were approximately 1.3 billion households with television sets globally (see IDATE Consulting & research, 2010). Assuming at least 2 persons on average per household as a conservative lower bound, that translates into 2.6 billion users.

[3] In this context “sandbox capability” (a.k.a. sandbox world) refers to a system designed to allow freedom of open creative expression. Members of earlier generations can think of Lego blocks as a sandbox. For younger generations, the computer game Minecraft offers a more recent digital instantiation.

a certain modicum of face validity, this is done principally to avoid detection and enable their real purpose, which is to deliver spam and malware to unsuspecting OSN users.

Being inundated with spam and malware has an obvious negative effect on the OSN user base and thus is a great concern for OSN companies. To preserve their own exclusive monetization of the user base, OSN companies have responded by trying to defend their walled gardens from these kinds of exploitive campaigns. In return, the spammers and venders of malware have responded by adapting to the new defences resulting in a kind of evolutionary arms race that now wages between spammers and the OSN companies. The adaptive cycle plays out continuously with the OSN companies finding new ways to detect and ban the fake accounts and the spammers and malware venders finding new ways to generate fake accounts and keep them from being detected and banned.

1.1 Emergence of the Shadow Economy

One critically important outcome of this competition has been the creation of an entire shadow economic ecosystem evolved to support a variety of online criminal activities, including the creation and maintenance of fake OSN accounts. This ecosystem has evolved to the point where it now represents a robust and sophisticated capability for influence in its own right, one that is much more powerful than anything achievable by a single user operating legitimately within the requisite terms of service of an OSN user space. In short, the shadow ecosystem delivers an end-to-end service that can provide anyone with access to thousands of usable fake OSN accounts. For those wishing to exert influence in the OSN space, there exists a choice of strategic approach with clear and significant differences in efficacy. On the one hand, they can legitimately operate within the terms of service and try to attract and then coordinate the collective behaviour of many independent people, hoping for the spontaneous emergence of a collective coordinated expression of belief. On the other hand, they could simply purchase fake accounts and control the behaviour of many thousands of “individuals” directly, individuals who will obey their every whim immediately and without question. It’s not hard to imagine all kinds of people with money seeking influence who would be more than happy to select the second option.

Appreciating the scope of this potential threat requires some sense of just how many fake accounts there are in OSN. It is also important to note that many fake accounts on OSN sites are not used for malicious, criminal, or deceptive influence purposes. In a recent assessment by Facebook, about 8% of their active accounts were considered fake (Kelley, 2012; Krebs, 2012). However, the majority of these fake accounts were either duplicates (e.g., a secondary account used to protect identity or authenticate other internet services), accounts for non-humans (e.g., pets), or misclassified corporate accounts. The minority, about 1.5%, were classified as undesirable or criminal in intent. Twitter has shown similar numbers, with approximately 3% of accounts deemed to be malicious (Thomas et al., 2011).

Overall, the percentages of fake accounts seems small, but given the scale of OSN the hard numbers mean that current levels of malicious accounts could range in the tens of millions. To put these numbers into further perspective, detection of fake accounts relies on at-abuse time metrics that typically focus solely on spam and malware abuse. Notably, these detection rates are quite high, with over 90% of Twitter accounts suspended within three days of initiating malicious activity (Thomas, McCoy, Grier, Kolcz, & Paxson, 2013). On the other hand, the detection rate of fake accounts at the time of creation is very low; with studies showing detection rates of dormant or non-spam related fake accounts to be typically less than 10% (e.g., Benevenuto, Magno, Rodrigues, & Almeida, 2010; Gao, Hu, Wilson, Li, Chen, & Zhao, 2010; Lee & Kim, 2012). The fact that OSN companies rely on at-abuse time metrics to suspend fake accounts, and only accounts used for spam and malware campaigns, means that action is taken only after the damage is done and only if that damage is caused by spam and malware advertising.

It is also important to distinguish between the criminal use of a legitimate account and the same use through a fake account. Jihadist groups such as the Islamic State (IS) have used OSN to influence and recruit by advertising horrific human rights abuses and promote their brand of violent jihad. Although this use of OSN is considered by some to be illegal and certainly violates many OSN companies' terms of service, it is also being done by those who want to be very clear about where the message is coming from. Another potentially more threatening use of fake accounts comes in the form of deceptive influence campaigns. This use of fake accounts is for influence in which the identity of the account is used to deceive recipients as to the true source and intentions of a message's origin. The term fake account will be used interchangeably to refer to both kinds of influence campaigns with the specific meaning dependent on context.

1.2 OSN as an Influence Capability

The shadow economy was created to service the needs of individuals seeking to exploit OSN for monetary gain. However, the issue at hand is whether the shadow economy infrastructure can be used to support influence campaigns in OSN other than the for-profit spam and malware campaigns that the infrastructure was originally created for. Of critical importance therefore is the question as to what extent OSN can be used to exert a directed intentional influence campaign. Influence over the internet can occur in many ways and the path from point of manipulation to change in behaviour or belief can be very indirect, tenuous, and ultimately affected by many extraneous factors beyond reasonable intentional control. There is perhaps no better example how establishing cause and effect with this kind of multifaceted connection can generate inconsistent and unequivocal results than the ongoing debate about whether video games cause violent behaviour (c.f. Anderson et al., 2010; Ferguson & Kilburn, 2010; Markey & Markey, 2010). However, rather than focus on a myriad of metrics and measures of direct causal linkages, it may be more informative to focus on factors that are merely sufficient, but not always necessary, for influence to occur. OSN tend to emphasize both inter-personal interaction and information sharing and access. Studies have shown that both of these factors play an important role in influence which would imply OSN are a well-positioned capability for types of influence that depend on these factors.

The research suggests that the link between use of OSN and affecting attitudes and behaviors is influenced by what motivates people to search out information and like-minded individuals. This connection has been shown to be particularly relevant for political attitudes and behaviors; as opposed to, say, entertainment and pop culture (for review see Zhang, Johnson, Seltzer, & Bichard, 2009). Specifically, the kind of socially inter-active discussions afforded by OSN has been shown to influence political activities and cognitions (e.g., Hardy & Scheufele, 2005). In addition, online inter-personal interaction can serve both social and information seeking needs; and can influence political behaviors and attitudes (Kaye & Johnson, 2006). People have also been shown to rely on OSN to gratify informational needs and help them to make political decisions (e.g., to obtain information about a prospective candidate, see Postelnicu & Cozma, 2008).

The prominence of OSN in political movements gained substantial attention starting with the so-called Iranian "Twitter revolution" in 2009 and recently culminating in the "Arab spring" in 2011. Although these revolutions made a substantial presence in OSN, particularly Twitter, the actual impact of OSN on these political revolutions has been debated. One notable criticism is the fact that these revolutions occurred in countries where few people had access to OSN (e.g., <1% of Iranian citizens had a Twitter account, see Wolfsfeld, Segev, & Sheafer, 2013). The implication is that OSN likely had little to do with the actual coordination and organization of protests. Rather, the effect of OSN seems to have been in the dissemination of information; particularly spurred by interactions between political organizers and media outlets or news agencies (see Lotan, Graeff, Ananny, Gaffney, Pearce, & Boyd, 2011). In other words, OSN provided a means for people to interact and disseminate information; in these cases allowing the protesters to confer with like-minded peers, get their message out, and perhaps have some hand in shaping their own narrative of what was happening. Recent use of OSN by groups such as the IS have more or less done same thing to attract followers and report on their activities (see Seaboyer, 2015); although there is some evidence IS is also using OSN to coordinate and manage their organization (Ligon, Harms, Crowe, Lundmark, & Simi, 2014).

At a more general level, OSN provide an influence capability that either provides users with access to a wide inter-connected audience to influence or a large database of information to be influenced (Cha, Haddadi, Benevenuto, & Gummadi, 2010; Zhang et al., 2009). If disseminating information and connecting to people is the principle influence capability offered by OSN, then an effective strategic approach for an agent seeking influence would be to either exploit the capability for their own use and/or stop their opponents from doing the same thing. To do this would require dominating the information and interaction space. The vast numbers of fake accounts offered by the shadow economy would be exactly what is required to do just that. A number of examples discussed in the next section relied on large numbers of fake accounts to employ both the leverage and denial tactical strategies.

1.3 Summary

With the surge in popularity of OSN, fake accounts have become an often unwelcome reality. They are supported by a sophisticated economic supply chain and they are currently being used to influence in a variety of ways, including the use of deception. Although one of the most prevalent uses of fake accounts is for-profit monetary gain through spam and malware distribution, there are already services offering the use of fake accounts for perpetuating not-for-profit influence campaigns (e.g., online reputation management; see Ronson, 2015). Clearly, the great potential of fake OSN accounts represents a strategic capability to exert influence through the collective force of many individuals and the control of the information space. This has important implications as a potentially disruptive technology. Of perhaps more critical importance is the potential of this capability to evolve and become more robust.

A key source of energy for this evolution is the entrepreneurial drive to create more effective for-profit monetization of OSN in the face of the OSN companies' efforts to stop such exploitation by third parties. A second source of energy that will drive the evolution of technology for influence capability are agents, such as state actors, with enough resources to push development forward. As it would appear that both drivers are comfortably in play, the main focus of this paper is to examine the future implications that these efforts will have in driving the evolution of technology as a general but highly effective influence capability. An important consideration is to determine just what aspects of the technology might evolve and how it will come to create future disruptive effects. Ultimately, an estimation of how this capability will manifest as a disruptive technology in the future is clearly needed and is the end goal of this paper. In the following pages I will describe the current state of the shadow economy and the sophistication of its services. To provide some scope of the possibilities for influence, I will then provide some examples of how this service has presently been used to exert influence in a variety of ways. Finally, I will present a hypothetical analysis of key aspects of the ecosystem speculating on how they might evolve as a future influence capability.

2 The Shadow Economy and Influence in OSN

A shadow economic ecosystem was already in place to serve criminal monetization of the internet before the advent of OSN. It is currently a sophisticated end-to-end system serving a number of criminal and grey market activities. These include tools and services for exploiting security vulnerabilities such as the trade in software exploits and system exploit kits (e.g., exploit-as-a-service; see Grier et al., 2012, for a review), tools and services for account creation such as Simple Message Service (SMS) and Phone Verified Accounts (PVA) (e.g., SMS-as-a-service; see also Thomas, 2015, for a review of these services), and tools and services for managing accounts and monetizing the exploits such as spam affiliate programs, uniform resource locator (URL) shorteners, and click through hosting (e.g., spam-as-a-service and the spam value chain; see Levchenko et al., 2011; Thomas, 2013; and Yang, Harkreader, Zhang, Shin, & Gu, 2012, for a review).

Many shadow economy services, especially those that are generally illegal, are only accessible on Blackhat forums or through the Deepweb and Darkweb.[4] One of the more easily accessible public facing (i.e., surface web) components is the account merchants themselves. A Google search for buying OSN accounts produced 98,100,000 and 153,000,000 hits for Twitter and Facebook respectively.[5] Some of the top ranked venders even have Facebook pages offering access to face valid Facebook accounts for conducting influence and public relations (PR) campaigns. According to an analysis by Thomas et al. (2013), account prices vary from 5–10 cents for a Twitter account to as much as $1.50 for a PVA account on Facebook.[6] E-mail accounts are an order of magnitude cheaper with basic Gmail or Hotmail accounts going for 0.5 cents and Gmail PVAs for $0.50. Prices of OSN accounts go up depending on whether accounts are pre-activated, loaded with face valid credentials and social links, or newly created versus aged. Accounts are usually sold in lots ranging from 100–1,000 at a time with some merchants able to fill orders with newly created accounts in as little as 24 hours. Like many shadow economy services, there is little honour among thieves with some merchants selling the same accounts multiple times to unsuspecting buyers.

Initially, services for account creation grew out of attempts to prevent automated account creation in the first place. In the past, account creation required one to input a name and a password. This simple process was of course wide open to automatic account creation abuses so more steps were added to the process to inhibit the use of bots and scripts performing automated creation; especially the creation of bulk accounts. Currently, most OSN sites insert a number of barriers or restrictions in the account creation process with corresponding shadow economy services available for overcoming the barriers to enable bulk account creation. Table 1 lists a number of these barriers and the corresponding shadow economy component that provides a service to

[4] Deepweb refers to Internet accessible sites that cannot be directly reached from a typical web search engine such as Google, Yahoo, Bing, and so on. The Darkweb refers to private networks accessible only between trusted peers such as The Onion Router (TOR) network. The Internet most people are familiar with is called the surface web (Pederson, 2013).

[5] Search terms used were “buy facebook accounts” and “buy twitter accounts” for Facebook and Twitter respectively. These terms were also the top terms auto-suggested by Google’s search engine.

[6] The prices quoted are current as of the publication date of the cited reference; however, they are subject to fluctuations over time.

***Table 1:** Account Creation Barriers and the Shadow Economy Component Providing Services to Circumvent It.*

Bulk Account Creation Barrier or Restriction	Supporting Shadow Economy Service
IP address black listing and throttling: Restricts how many accounts can be created from a single IP address to prevent mass creation from a single point of origin. To counter this, bulk account creation requires access to many different IP addresses, and hence many unique internet connected computers.	Botnets provide access to thousands of compromised computers with unique IP addresses that can act as hosts for account creation or as proxies to hide the Internet Protocol (IP) address of the account creator. Bulk account creators can either purchase these services from Botnet operators or buy the botnet malware to setup and manage their own Botnets.
Unique E-mail for account validation: Each created account requires a validation response through a unique E-mail address that becomes tied to the account. Bulk creation of OSN accounts requires prior bulk creation of corresponding E-mail accounts.	E-mail services from companies such as Google, Hotmail, and Yahoo all provide free accounts for E-mail validation. Although these companies employ similar barriers to prevent bulk E-mail creation, they can be overcome with the same kinds of services as those used for creating bulk OSN accounts.
Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) solving: Requires solving a puzzle designed to be easy for humans but difficult for machine agents that are not human equivalent in IQ for the task. This is essentially the use of a Turing Test to screen out non-human agents.	Despite the intent of these puzzles, machine agents can be used to solve them at low cost by exploiting bugs, brute forcing answers, or Optical Character Recognition (OCR) and pattern matching. However, in practice solution rates for machine agents can vary and are quite low on average (18-30%). A more expensive alternative is to employ human CATPCHA solvers. Human solvers can be hired as cheap labour in countries such as China. A final approach is to trick unsuspecting human users into completing a CAPTCHA. This technique was used by the Koobface Botnet in which CAPTCHAs were presented to users of compromised computers on the pretext of a Microsoft security verification check (Motoyama, Levchenko, Kanich, McCoy, Voelker & Savage, 2010).
Cell phone verification: Similar to E-mail verification, a phone verified account (PVA) requires a SIM card to be tied to the account and a validation response through the Simple Message Service (SMS) text messaging system.	Similar to IP black listing and throttling, this validation requires a unique SIM card phone number. Because most PVAs are handled internally by bulk account creators a PVA is typically 30-100 times as expensive as a non-PVAs. However, the recent advent of SMS-as-a-service and access to bulk SIM cards is helping to automate the PVA process and lower prices for PVAs (Thomas, 2015).

circumvent it. The specialized products and services in Table 1 were either repurposed from existing services (e.g., Botnets) or specifically created to thwart defences against the abuse of account creation procedures (e.g., CAPTCHA solvers). Note that the process of account creation does not require the use of these services. Anyone with basic programming knowledge and a little time can accumulate thousands of fake accounts stocked with reasonably face valid credentials (see Alleyrat, 2010, for some example scripts). However, given the low per account price, an efficient means of bulk creation would be required to generate any monetary benefit.

2.1 Fake Accounts and Influence Campaigns

There are numerous examples in the literature showing how control of fake accounts can be used to perpetrate influence campaigns in OSN. However, online influence campaigns that are not restricted to OSN, such as astroturfing (Cho, Martens, Kim, & Rodrigue, 2011; Givel, 2007; Jacobs, 2012; Merchant, 2014;) and fake reviews (Chen, Wu, Srinivasan, & Bharadwaj, 2013; Chen, Wu, Srinivasan, & Zhang, 2013; Mukherjee, Liu, & Glance, 2012; Ott, Choi, Cardie, & Hancock, 2012;) are beyond the scope of this paper as neither are specific to OSN.

2.1.1 For-profit Influence Campaigns

Probably the most widespread use of fake accounts for influence is the distribution of spam advertising and malware in which the exploitation for profit is the unambiguous goal. This list is by no means exhaustive, but it does cover the more typical uses of bulk fake accounts for monetization via spam and malware. One particular campaign on Facebook used 18,000 fake and 40,000 hijacked accounts to send out over 200,000 spam and phishing messages to 3.5 million users. The use of hijacked accounts allowed the spammers to leverage trusted pre-existing social ties, something very time consuming, difficult, and expensive to create from scratch (Gao, et al., 2010). Presumably because it is easier to send unsolicited messages to other users on Twitter in the form of retweets and mentions, many recent campaigns use exclusively fake accounts, although like Facebook, spamming from hijacked accounts seems to be more effective (Thomas et al., 2011).

Several recently analyzed spam and malware campaigns on Twitter sent out over 17 million messages (tweets) from approximately 140,000 fake accounts (Lumezanu & Feamster, 2012; Thomas et al., 2011; Yang et al., 2012). These kinds of spam campaigns can also be highly profitable, with revenue generation estimates for a single campaign ranging from $1–$30 million per month depending on campaign size and products being sold (Akass, 2008; Kanich et al., 2011). It is also worth mentioning that most of the offending accounts advertising spam and malware are banned within 1 to 3 days of activity (Thomas et al., 2013). That these campaigns can be sustained for months is a testament to the fact that the shadow ecosystem is robust enough to provide a continuous supply of new fake accounts.

2.1.2 Not For-profit Influence Campaigns

There are also examples in which fake accounts were used for strictly influence operations where for-profit monetization is not the goal. A report by FireEye Security Reimagined (2015) describes how fake Skype and Facebook accounts were used to induce unsuspecting members of the Syrian opposition forces to download malware on to their phones. The malware was used to steal vital intelligence data on opposition forces and the operations they conduct. The United States (US) Drug Enforcement Agency (DEA) recently used a fake Facebook account to trap suspected drug dealers. This action prompted Facebook to send the US government a letter protesting the use of Facebook in violation of its terms of service (Sullivan, 2014). PR firms will routinely use fake accounts from various OSN sites to establish or repair a person's online reputation and presence (see Ronson, 2015, for an example case study). One high profile campaign, carried out using a hacked account belonging to the Associated Press (AP) rather than a fake, tweeted from the hacked AP account about the health of President Obama causing a temporary but significant drop in the Dow Jones Industrial Average (Memmott, 2013). What these examples all have in common is the small number of accounts used (sometimes only 1), that the campaign was generally effective, and that the accounts were controlled by human operators.

One potential effect of leveraging the shadow economy for influence would be to create the appearance of a lot of people and a lot of available information all extolling a particular position or point of view. In essence, influence can be achieved by controlling the information and inter-personal landscape through sheer volume and availability of information. One rationale for this kind of strategic approach is that if most information sources are saying the same thing people will be more likely to believe and remember it (e.g., Lewandowsky, Stritzke, Oberauer, & Morales, 2005). Alexander (2015) and Chen (2015) have recently reported on troll factories in Russia consisting of scores (perhaps hundreds) of paid operators posting in forums and OSN sites to present or frame a particular message or issue favourable to the Russian government (see also Etling, Alexanyan, Kelly, Farris, Palfrey, & Gasser, 2010; Walker, 2015). Similar types of human controlled fake account troll operations have been discovered in China (see Cho, Caballero, Grier, Paxson, & Song, 2010). Although paid posting is not new, such troll operations provide some of the first evidence that it is being used for political rather than strictly economic motivations (as is seen in fake reviews, c.f. Chen, et al., 2013).

Because human operators are much more expensive than computer controlled bots, there are limits as to the size and scope of campaigns, at least ones that individuals and non-state actors can reasonably afford. Although paying humans to control thousands of accounts is not out of the question for agencies with deep pockets, there have been a number of campaigns utilizing accounts in the range of thousands or tens of thousands which have been found to be controlled by bots (e.g., Lawrence, 2015). Many examples of intentional influence campaigns that leveraged the shadow economy appear to have adopted the other tactic of denying the opponent a voice. In fact, one recurring approach for denying an opponent is what I have termed the Twitter analogue of a Denial of Service (DOS) attack. This particular tactic relies on the fact that Twitter is essentially a stream of posts grouped by arbitrary hashtags; hence, it is possible to flood a discussion thread by injecting large volumes of random or antagonistic posts into the stream. This disrupts the continuity of the discussion drowning out any real content making the thread nearly impossible to follow. Like the example campaigns described above, these campaigns share a number of characteristics; they all occurred on Twitter, they were all directed at censoring discussion of sensitive political issues, and consequently all used the same modus operandi, the OSN DOS attack.

One such event took place over 2 days of heightened political discussion in December 2011 during the aftermath of the Russian general election. An analysis by Thomas, Grier & Paxson (2012) revealed three important characteristics of this campaign. First, was its size compared to other not-for-profit campaigns. Over 25,000 accounts were used to inject 440,000 tweets into various conversations about the election outcome. Moreover, these fake accounts were identified as part of a larger pool of almost 1 million fake accounts which were created up to 8 months prior to their use, remaining dormant in preparation for the campaign. Second, the accounts used in the campaign were controlled by bots. Third, there was clear evidence that the perpetrators of this campaign leveraged the shadow economy spam-as-a-service and exploit-as-a-service marketplace technology. Only 1% of the fake accounts actually logged in from Russia, compared to almost 60% of legitimate accounts that participated in the discussions. In addition, almost 40% of the fake accounts were being controlled from compromised hosts, indicating they were part of a global Botnet. It is of course impossible to tell whether the perpetrators purchased these services or owned them themselves. Regardless, the technology was leveraged to perpetuate the campaign, which allowed the attackers to flood the discussion threads with over twice as many spam tweets as legitimate ones.

Recent analysis by Alexander (2015) showed that the DOS approach for this kind of campaign was not an isolated incident in Russia. There exists an ongoing capability using tens of thousands of bot controlled Twitter accounts to inject spam and pro-Kremlin rhetoric into political discussion about Russia that appear on Twitter. In fact, a subsequent analysis by Thomas et al. (2012) showed that over an 8 month period both before and after the Russian election, approximately 2.4 million spam tweets were sent by the same accounts as those used in the focused December campaign. The December campaign stands out only because it consisted of a relatively large spike in tweets during the 2 day period of heightened discussion activity concerning the election.

Four other DOS Twitter campaigns very similar to the one in Russia were identified and analysed by Verkamp and Gupta (2013). Two appeared in China, and one each in Syria and Mexico. One major difference between these campaigns and the one in Russia is the number of accounts involved. Only 2,000–3,000 accounts were used in each campaign compared to Russia's 25,000. Despite the lower volumes of fake accounts, three of the four campaigns had a greater volume of spam messages compared to legitimate messages; and in the case of China that ratio was almost 3:1. The China and Mexico campaigns also showed a sustained output of message spiking lasting from 2–3 weeks. However, unlike the Russian campaign in which fake accounts were stockpiled months in advance, the fake accounts used in these four campaigns were created only days before they were used.

The results of the analysis by Thomas et al. (2012) and Verkamp and Gupta (2013) of all five campaigns showed clear evidence that the attackers leveraged the capabilities of the shadow economy and its technology. All fake accounts used in the campaigns were purchased in large blocks before the attacks started. The credentials of many of the fake accounts were also clearly bulk generated using simple algorithms as blocks of them showed very distinct patterns of naming, demographics, and profile picture. Although Verkamp and Gupta were unable to get data on the IP addresses of the fake accounts, Thomas et al. showed that many of the fake accounts were being run from compromised computers in a global Botnet. One additional important piece of evidence also emerged. Many of the fake accounts displayed more sophisticated behaviour than is typically seen with bot controlled accounts, such as mimicking the diurnal behaviour typical of human activity. In addition, those fake accounts that were created most recently (i.e., in mid-2012) showed more complex credentials than older accounts created for the campaigns in 2011 (Verkamp & Gupta, 2013).

2.1.3 Effectiveness of Influence Campaigns

Influence campaigns like those described above where run presumably because those who directed them believed they would have a particular effect. This begs the obvious question: Do any of these campaigns actually work? This question may appear loaded in the sense that these campaigns may have produces some kind of influence effect but perhaps not the exact one intended. Nonetheless, it seems that the people who were running these campaigns did so because they believed they would work. So, do they?

Recall that previous research had identified the main locus of influence in OSN as the interaction between people and the information available to them. The most straight forward strategies were to either leverage this capability or prevent an opponent from doing the same. One way of achieving this kind of superiority in the information space would be to employ vast numbers of accounts to dominate both information and interaction. Indeed, many of the campaigns described above leveraged the shadow economy for precisely this purpose. There still remains the question of whether this kind of strategic approach produces the intended effect. There is some evidence to suggest it does.

Reuter and Szakonyi (2011) showed that participation in OSN increased perception of fraud in the Russian 2011 election, but only for those who were active on OSN sites dominated by opposition activists promoting this point of view. The OSN sites dominated by pro-government activists showed no effect on perception of fraud. A report on the use of OSN by ISIS by Seaboyer (2015) showed that controlling the information space by spamming its message has raised the profile of ISIS, diluted or detracted criticism, and created an inflated perception of their strength and power. Other analysts have concluded that the purpose of the Russian troll houses is to interfere with access to information. A number of Russian activists interviewed by Chen (2015) stated that the goal of the troll houses is to pollute and dilute the information space so that legitimate information is lost and people either will not believe what they read of will not bother to search for any truth. These kinds of effects would seem to be the basis for the importance of OSN in influence and the corresponding motivation to exert influence in that space.

Although the information and collaborative space can be affected and some modicum of influence created, it is unclear whether hearts and minds can actually be won in the OSN space (e.g., Zhang et al., 2009). It does seem clear that at the very least one can exert control by simply determining what kinds of information people have access to and the conditions under which they might interact. If the user base is unmotivated or unsophisticated enough to mobilize an effective counter, this seems enough to make it worthwhile for various people to engage in influence campaigns, ones that leverage the power of the shadow economy.

2.2 Casting the Future of Influence Campaigns in OSN

There are many more examples of influence campaigns being run in OSN, certainly more than can be covered here. However, the limited scope that was covered is sufficient to warrant a few conclusions about where things are at present and where they may be going in the future. There have been a number of large not-for-profit influence campaigns, each using thousands to tens of thousands of fake accounts. Out of necessity to control such large numbers, these fake accounts were controlled by bots. There was also evidence that the behaviour of the bots and the face valid credentials were of increasing sophistication. In addition to the influence campaigns, researchers have uncovered fake accounts being run by human operators. Smaller in scope likely due to the increased cost and overhead of managing human operators, these were employed in influence campaigns that required the use of human intelligence. Human intelligence was required here not merely for behavioural mimicry, but to do what easily accessible artificial intelligence (AI) cannot currently do well: reading, interpreting, and generating context appropriate content. This is principally why the Russian and Chinese troll houses exist. It is also why the DOS campaigns were used on Twitter in deference to some other strategy.

The technology required to make a bot perform the requisite behaviours in any OSN space (e.g., posting, liking, inviting users, accepting user invites; see Boshmaf, Muslukhov, Beznosov, & Ripeanu, 2011) is trivial to produce, requiring no more than at most a few tens or a hundred lines of code (see xnite, 2010, for examples of basic Twitter bot code). Of course all of this requires that the bot does not have to come up with any of the content it posts, or decide if a user is a good bet to form a relationship with, or choose which discussions to post into. These competencies are all still the domain of humans. The DOS Twitter campaigns were carried out because they are exactly the kind of thing that current tech-level bots are good at: doing what they are told to do and lots of it. In fact this characteristic of OSN bots has also been the scourge of the OSN space but for benign reasons. It was assumed that a good use for bots on Twitter was to retweet posts to help spread the content farther throughout the network. However, because bots are mindless things they cannot appreciate the context of a tweet or be able check the veracity of the content they are retweeting. This has allowed for the unintended spread of false information inadvertently causing damage to the credibility of the OSN space (see Ferrara, Varol, Davis, Menczer, & Flammini, 2015).

But imagine what would happen if there were OSN bots with the same level of OSN competency as the humans working in the Russian and Chinese troll houses. One could imagine having at one’s disposal a group of workers that do not succumb to human frailties of fatigue, disinterest, or disease. They would not need to draw a salary or require any consideration for the application of labour law; and, they would provide a work force numbering in the thousands or even millions. There would still be a need to manage all these new workers, but the kinds of instructions given to the troll house employees would not be beyond the scope of any AI capability sufficient to build the workers themselves (see Alexander, 2015). These kinds of bots would also require generation of more sophisticated credentials, but even that is not beyond current technology to provide. In fact the AI themselves can help to generate that content. About the only thing that cannot be faked using bots is the account’s age. But even that can be taken care of with foresight and sufficient prior investment.

Current generation bots perform simple tasks well but require much oversight because they need direction and content to perform the tasks they are given. As the scale and sophistication of the influence campaigns increase and the OSN companies get better at detecting fake accounts prior to their use, bots built with current technology will require too much oversight or simply be too easy to spot to be worthwhile. The solution to this pressure is to make them appear (a) more human to avoid detection and (b) more autonomous to make management easier. The evidence to date indicates this kind of evolutionary development is already starting to happen: Bots are evolving towards a point where they can pass the Turing Test and be more autonomous like real OSN users.

But bot AI is not the only shadow economy technology to evolve. Automated credential generation will also become better. However, this technology is fairly well advanced and will likely be passed on to the bot AI to take care of. The real novel challenge for the shadow ecosystem will be to supply the growing need for computational power and unique IP address locations to house the new bot AIs and the command and control servers that manage them. The only sources for this at present are the compromised hosts that make up the world's botnets. Presently, the most common uses for the world's botnets are to send spam (e.g., MegaD) or commit DOS attacks (see Cho et al., 2010). As the shadow ecosystem evolves, this may very well change. Rising demand for these new services may increase the value of botnets leading to greater pressure to develop more sophisticated malware and an improved capability to remain undetected. If the value rises high enough, botnets could become a turf that is fought over by criminal agents seeking exploitation for profit. This could translate into attacks on hosting companies, data centres, and internet service providers (ISP). If competition becomes intense enough that rivals seek to disable each other's capability, the technological infrastructure that runs the internet itself could be at risk (e.g., global Domain Name Servers (DNS), core gateway routers). Given the weak security protecting such infrastructure, this is a seriously troubling possibility (see Jones, 2014; Lewis, 2014; Zhou, 2010).

The speculation that the shadow ecosystem may evolve technologically to support ever more sophisticated OSN attacks is only one locus of potential development. State actors, and non-state actors with sufficient economic resources, can certainly develop a capability that is equal to what might evolve in the shadow ecosystem, and there is evidence that this is precisely what is happening (see Greenwald, 2014). Governments are putting resources into creating a future influence capability similar to what was described above and they are actively engaged in perpetrating influence campaigns of their own, campaigns that appear equal or greater in sophistication to those already being conducted (see Fielding, & Cobain, 2011; Jarvis, 2011).

At present, the use of OSN for influence is something that should not be ignored. The future potential of OSN as an influence capability is something that cannot be ignored. There is great potential for influence within OSN and those that are first to establish themselves and leverage this capability will be the ones that most benefit. The use of OSN for interaction, information, and communication is only growing, and like the internet itself, is not a fad that will simply go away in time. It is a new and fledgling ecosystem that will continue to dominate our lives and evolve with them.

3 Conclusions

There is no shortage of ways to exploit OSN for profit or influence. It appears that the evolution of technology to make influence campaigns of various sorts more effective will continue to drive forward to ever increasing heights. As AI grows more sophisticated, the face validity of bots will increase. Users of OSN will not be able to tell the difference between online behaviour generated by a bot vs. that of a real person. Consequently, the interaction between people will become more and more the locus of influence as this space becomes formalized and made routine.

However, the anticipated success of spoofing accounts in OSN may also be the key factor to cause the eventual downfall of OSN as an influence capability. Consider that as fake accounts become more human like, the Reverse Turing Test for detecting and blocking automated exploitation at scale will become impossible. There will eventually be no way to tell the difference between human and bot anymore. This will likely lead to OSN being discredited and ultimately abandoned as spaces for humans to engage in trusted interactions. Hence, the drive to exploit the OSN space for influence may ultimately lead to its destruction.

References

Akass, C. (2008). Storm worm ‘making millions a day’. [Online]. Available: http://www.computeractive.co.uk/pcw/news/1923144/storm-worm-millions-day.

Alexander, L. (2015). Social Network Analysis Reveals Full Scale of Kremlin's Twitter Bot Campaign. [Online]. Available: http://globalvoicesonline.org/2015/04/02/analyzing-kremlin-twitter-bots/.

Alleyrat. (2010). How to create mass hysteria on a college campus using Facebook. 2600: The Hacker Quarterly, 27(2), 18–18.

Anderson, C., Shibuya, A., Ihori, N., Swing, E., Bushman, B., Sakamoto, A., Rothstein, H., & Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries. Psychological Bulletin, 136, 151–173.

Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010). Detecting Spammers on Twitter. Proceedings of the conference on collaboration, electronic messaging, anti-abuse and spam (CEAS), 6, (pp. 12).

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2011). The Socialbot Network: When Bots Socialize for Fame and Money. Proceedings of the 27th Annual Computer Security Applications Conference. ACM, 93–102.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. Proceedings of the Fourth International Association for the Advancement of Artificial Intelligence (AAAI) Conference on Weblogs and Social Media.

Chen, A. (2015). The Agency. New York Times. [Online]. Available: http://www.nytimes.com/2015/06/07/magazine/the-agency.html?_r=0.

Chen, C., Wu, K., Srinivasan, V., & Bharadwaj, R. (2013). The best answers? Think twice: Online detection of commercial campaigns in the CQA forums. Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ACM. 458–465.

Chen, C., Wu, K., Srinivasan, V., Zhang, X. (2013). Battling the internet water army: Detection of hidden paid posters. Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Niagara, Ontario, Canada, ACM.

Cho, C., Caballero, J., Grier, C., Paxson, V., & Song, D. (2010).Insights from the Inside: A View of Botnet Management from Infiltration. USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET).

Cho, C., Martens, M., Kim, H., & Rodrigue, M. (2011). Astroturfing global warming: It isn’t always greener on the other side of the fence. Journal of Business Ethics, 104, 571–587.

Etling, B., Alexanyan, K., Kelly, J., Farris, R., Palfrey, J., & Gasser, U. (2010). Public Discourse in the Russian Blogosphere: Mapping RuNet Politics and Mobilization. (Research Publication No. 2010-11October 2010). Cambridge, Massachusetts: Harvard University, Berkman Center for Internet & Society.

Facebook, Inc. (2014, January). Form 10-K: Annual Report (Commission File Number: 001-35551). Menlo Park, CA: Author.

Fielding, N., & Cobain, I. (2011). Revealed: US spy operation that manipulates social media. The Guardian. [Online]. Available: http://www.theguardian.com/technology/2011/mar/17/us-spy-operation-social-networks.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2015). The Rise of Social Bots: arXiv:1407.5225 [cs.SI]. Unpublished manuscript, Indiana University.

FireEye Security Reimagined. (2015). Behind the Syrian conflict's digital front lines. Regalado, D., Villeneuve, N., & Railton, J: Authors.

Ferguson C. & Kilburn J. (2010). Much ado about nothing: the mis-estimation and overinterpretation of violent video game effects in eastern and western nations. A comment on Anderson et al. (2010). Psychological Bulletin, 136, 174–178.

Gao, H., Hu, J., Wilson, C., Li, Z., Chen, Y., & Zhao, B. (2010). Detecting and characterizing social spam campaigns. Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, 35–47, ACM.

Givel, M. (2007). Consent and Counter-Mobilization: The Case of The National Smokers Alliance. Journal of Health Communication, 12(4), 339–357.

Google, Inc. (2014, January). Form 10-K Annual Report (Commission File Number: 000-50726). Mountain View, CA: Author.

Greenwald, G. (2014). How Covert Agents Infiltrate the Internet to Manipulate, Deceive, and Destroy Reputations. The Intercept. [Online]. Available: https://firstlook.org/theintercept/2014/02/24/jtrig-manipulation/.

Grier, C., Ballard, L., Caballero, J., Chachra, N., Dietrich, C., Levchenko, K., Mavrommatis, P., McCoy, D., Nappa, A., Pitsillidis, A., Provos, N., Rafique, Z., Rajab, M., Rossow, C., Thomas, K., Paxson, V., Savage, S., & Voelker, G. (2012). Manufacturing Compromise: The Emergence of Exploit-as-a-Service. Proceedings of the 2012 ACM Conference on Computer Communications Security.

Hardy, B. W., & Scheufele, D. A. (2005). Examining differential gains of Internet use: Comparing the moderating role of talk and online interactions. Journal of Communication, 55, 71–84.

IDATE Consulting & research. (2010). TV 2010: Markets & Trends Facts & Figures. Montpellier, France: Author.

Jacobs, J. (2012). Faking it : how to kill a business through astroturfing on social media. Keeping Good Companies, 64(9), 567–570.

Jarvis, J. (2011). America's absurd stab at systematising sock puppetry. The Guardian. [Online]. Available: http://www.theguardian.com/commentisfree/cifamerica/2011/mar/17/us-internet-morals-clumsy-spammer.

Jones, S. (2014). Shellshock bug threatens internet’s backbone, analysts warn. Financial Times. [Online]. Available: http://www.ft.com/cms/s/0/2f7d00d0-44a8-11e4-ab0c-00144feabdc0.html.

Kanich, C., Weaver, N., McCoy, D., Halvorson, T., Kreibich, C., Levchenko, K., Paxson, V., Voelker, G., & Savage, S. (2011). Show Me the Money: Characterizing Spam-advertised Revenue. USENIX Security Symposium. 15–15.

Kaye, B. K., & Johnson, T. J. (2006). The age of reasons: Motives for using different components of the Internet for political information. In A. P. Williams, & J. C. Tedesco (Eds.), The Internet election: Perspectives on the Web in campaign 2004. Lanham, MD: Rowman & Littlefield.

Kelley, H. (2012). 83 million facebook accounts are fakes and dupes. [Online]. Available: http://www.cnn.com/2012/08/02/tech/social-media/facebook-fakeaccounts/index.html.

Krebs, B. (2012). Spam volumes: Past & present, global & local. [Online]. Available: http://krebsonsecurity.com/2013/01/spam-volumes-past-present-globallocal.

Lawrence, A. (2015). Social Network Analysis Reveals Full Scale of Kremlin's Twitter Bot Campaign. [Online]. Available: http://globalvoicesonline.org/2015/04/02/analyzing-kremlin-twitter-bots/.

Lee, S., & Kim, J. (2012). WarningBird: A Near Real-Time Detection System for Suspicious URLs in Twitter Stream. IEEE Transactions on Dependable & Secure Computing, 10.

Levchenko, K., Pitsillidis, A., Chachra, N., Enright, B., Félegyházi, M., Grier, C., Halvorson, T., Kanich, C., Kreibich, C., Liu, H., McCoy, D., Weaver, N., Paxson, V., Voelker, G., & Savage, S. (2011). Click trajectories: End-to-end analysis of the spam value chain. 2011 IEEE Symposium on Security and Privacy (SP). (pp. 431–446). IEEE.

Lewandowsky S., Stritzke W. G. K., Oberauer K., Morales M. (2005). Memory for fact, fiction, and misinformation: The Iraq War 2003. Psychological Science, 16, 190–195.

Lewis, T. (2014). Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation. John Wiley and Sons: New Jersey, USA.

Ligon, G., Harms, M., Crowe, J., Lundmark, L., & Simi, P. (2014). The Islamic State of Iraq and the Levant: Branding, Leadership Culture and Lethal Attraction. (Final Report prepared for the Department of Homeland Science and Technology Directorate’s Office of University Programs, award number #2012-ST-061-CS0001). Baltimore, MD: University of Maryland, National Consortium for the Study of Terrorism and Responses to Terrorism.

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., & Boyd, D. (2011). The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions. International Journal of Communication, 5, 1375–1405.

Lumezanu, C., & Feamster, N. (2012). Observing common spam in twitter and email. Proceedings of the 2012 ACM conference on Internet measurement conference. 461–466. ACM.

Markey, P., & Markey, C. (2010). Vulnerability to Violent Video Games: A Review and Integration of Personality Research. Review of General Psychology, 14, 82–91.

Memmott, M. (2013). AP Twitter Account Hacked, Tweet About Obama Shakes Market. National Public radio (NPR). [Online]. Available: http://www.npr.org/blogs/thetwo-way/2013/04/23/178620410/ap-twitter-account-hacked-tweet-about-obama-shakes-market.

Merchant, B. (2014). The Program Big Oil's PR Firm Uses to 'Convert Average Citizens'. [Online]. Available: http://motherboard.vice.com/read/a-top-pr-firm-promised-big-oil-software-that-can-convert-average-citizens.

Motoyama, M., Levchenko, K., Kanich, C., McCoy, D., Geoffrey, Voelker, M., & Savage, S. (2010). Re: CAPTCHAs-Understanding CAPTCHA-Solving Services in an Economic Context. Proceedings of the USENIX Security Symposium, 10.

Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting Fake Reviewer Groups in Consumer Reviews. Paper presented at the International World Wide Web Conference Committee (IW3C2), Lyon, France.

Ott, M., Choi, Y., Cardie, C. & Hancock, J. (2011). Finding deceptive opinion spam by any stretch of the imagination. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. 309–319. Association for Computational Linguistics.

Pederson, S. (2013). Understanding the Deep Web in 10 Minutes. White paper report. BrightPlanet deep web intelligence: Sioux Falls, South Dakota. [Online]. Available: http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&cad=rja&uact=8&ved=0CB4QFjAA&url=http%3A%2F%2Fbigdata.brightplanet.com%2FPortals%2F179268%2Fdocs%2Fdeep%2520web%2520whitepaper%2520v3_for%2520approval.pdf&ei=bFEoVdjTKoS1sASV6ICwBg&usg=AFQjCNEjbYnEqii86ZrwMOFBEuFWVfiEWg&bvm=bv.90491159,d.cWc.

Postelnicu, M., & Cozma, R. (2008). Befriending the candidate: Uses and gratifications of candidate profiles on MySpace. Paper presented to the National Communication Association, San Diego, CA.

Reuter, O., & Szakonyi, D. (2012). Online social media and political awareness in authoritarian regimes. (Research Publication No. WP BRP 10/PS/2012). Moscow, Russia: National Research University Higher school of Economics, Basic Research Program.

Ronson, J. (2015). So You've Been Publicly Shamed. Riverhead Books: USA.

Seaboyer, A. (2015). ISIS Social Media Exploitation: Effects On The Contemporary Operating Environment. (Contractor report DRDC-RDDC-2015-C147). Kingston, Ontario: Royal Military College of Canada, Department of Political Science.

Sullivan, J. (2014). [Official letter from Facebook to the United States Drug Enforcement Agency]. Unpublished letter.

Thomas, K. (2013). The Role of the Underground Economy in Social Network Spam and Abuse (Technical Report No. UCB/EECS-2013-201). University of California at Berkeley: Electrical Engineering and Computer Sciences. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-201.html.

Thomas, K. (2015). Black Markets: An Abuse Fighter's Oracle of Delphi. [Online]. Available: http://www.inwyrd.com/blog/.

Thomas, K., Grier, C., Ma, J., Paxson, V., & Song, D. (2011). Design and Evaluation of a Real-time URL Spam Filtering Service. In Proceedings of the 32nd IEEE Symposium on Security and Privacy.

Thomas, K., Grier, C., Paxson, V., & Song, D. (2011). Suspended Accounts In Retrospect: An Analysis of Twitter Spam. In Proceedings of the Internet Measurement Conference.

Thomas, K., Grier, C., & Paxson, V. (2012). Adapting social spam infrastructure for political censorship. In Proceedings of the 5th USENIX conference on Large-Scale Exploits and Emergent Threats. USENIX Association.

Thomas, K., McCoy, D., Grier, C., Kolcz, A., & Paxson, V. (2013). Trafficking fraudulent accounts: the role of the underground market in twitter spam and abuse. USENIX Security Symposium.

Twitter, Inc. (2015). Form 10-K: Annual Report (Commission File Number: 001-36164). San Francisco, CA: Author.

Verkamp, J., & Gupta, M. (2013). Five incidents, one theme: Twitter spam as a weapon to drown voices of protest. Presented as part of the 3rd USENIX Workshop on Free and Open Communications on the Internet. USENIX.

Walker, S. (2015, April). Salutin' Putin: inside a Russian troll house. The guardian. [Online]. Available: http://www.theguardian.com/world/2015/apr/02/putin-kremlin-inside-russian-troll-house.

Wolfsfeld, G., Segev, E., & Sheafer, T. (2013). Social Media and the Arab Spring: Politics Comes First. The International Journal of Press/Politics, 18(2) 115–137.

Xnite. (2010). Twitter for fun and profit. 2600: The Hacker Quarterly, 27(2), 10-10.

Yang, C., Harkreader, R., Zhang, J., Shin, S., & Gu. G. (2012). Analyzing Spammers' Social Networks for Fun and Profit: A Case Study of Cyber Criminal Ecosystem on Twitter. In Proceedings of the 21st International Conference on World Wide Web, 71-80.

Yang, Z., Wilson, C., Wang, X., Gao, T., Zhao, B., & Dai, Y. (2014). Uncovering Social Network Sybils in the Wild. ACM Transactions on Knowledge Discovery from Data (TKDD), 8(2), 2.

Zhang, W., Johnson, T., Seltzer, T., & Bichard, S. (2009). The Revolution Will be Networked: The Influence of Social Networking Sites on Political Attitudes and Behavior. Social Science Computer Review, 12, 1–18.

Zhou, L. (2010). Vulnerability analysis of the physical part of the internet. International Journal of Critical Infrastructures, 6, 402–420.

List of Symbols/Abbreviations/Acronyms/Initialisms

AI	Artificial Intelligence
AP	Associated Press
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans Apart
DEA	Drug Enforcement Agency
DND	Department of National Defence
DNS	Domain Name System
DOS	Denial of Service
DRDC	Defence Research and Development Canada
DSTKIM	Director Science and Technology Knowledge and Information Management
IP	Internet Protocol
IS	Islamic State
ISP	Internet Service Provider
OSN	Online Social Network
PR	Public Relations
PVA	Phone Verified Account
R&D	Research & Development
SMS	Simple Message Service

This page intentionally left blank.

Glossary

Astroturfing

Astroturfing is the practice of masking the origin of a message or organization to make it appear as though it originates from a genuine and credible source. Astroturfing is most typically used to make it appear as though a message or organization has genuine and honest grass roots origin or support thereby granting credibility to the benefiting agency.

Bot

A computer program used to perform or automate actions typically carried out by a human. In OSN, bots are typically used to control the behaviour of accounts by performing the requisite canonical functions associated with OSN account behaviour.

Botnet

A network of Internet-connected computers communicating with each other to complete repetitive tasks and objectives generally controlled from a command and control server. Illegal botnets consist of computers that have been compromised by malware and are being used without the knowledge and consent of their owners.

CAPTCHA

A puzzle designed to be easy for humans to solve but difficult for machine agents that are not human equivalent in IQ for the particular task. This is essentially the use of a Turing Test (see below) to screen out non-human agents trying to perform a task restricted to humans only (e.g., creating an OSN account).

DNS

Domain Name System is a naming system for Internet-connected computers that translates the text form of an internet computer's name used by humans into the unique numerical IP Address (see below) used by computers to communicate on the Internet. For example, www.google.ca translates into the IP Address 173.194.40.248.

DOS Attack

A DOS attack is a type of network based attack in which a computer is flooded with such a high volume of irrelevant requests that it overwhelms the computer's capacity. This renders the computer unable to process legitimate requests, perhaps even causing the computer to crash. A DOS attack ultimately prevents legitimate users from accessing the computer. For example, a DOS attack against an online bank would prevent customers from using the online banking site.

Core Gateway Router

A router is a network device that forwards network traffic between computer networks. A core gateway router or core router performs this function between Internet networks and is an essential part of the operational infrastructure of the Internet, also known as the Internet backbone. Disabling these routers would cause the part of the Internet serviced by the router to cease functioning.

IP Address

The unique numerical address that provides both an identifier and location to every Internet-connected device. The presently used standard (IPv4) consists of four 8-bit numbers (total 32-bits) often written in decimal notation with each 8-bit number separated by a period (e.g., 173.194.40.248).

Malware

Short for malicious software, malware is any software created for intentionally malicious purposes to distinguish it from unintentionally harmful software. Malware typically includes computer viruses, worms, trojans, ransomware, spyware, scareware, and adware. It may take the form of executable code or as scripts embedded in other applications (e.g., web browsers, E-mail).

Spam and Spamming

Spam and spamming refer both to unsolicited messages sent through electronic messaging systems such as E-mail and SMS, and the act of flooding a system with repeated messages to dilute, cover, or drown out legitimate messages. Spam and spamming generally refer to such message abuse sent through all forms of electronic message systems.

Turing Test

The Turing test is a test of a machine's ability to exhibit intelligent behavior indistinguishable from that of a human. A machine passes a Turing Test when a human evaluator is unable to distinguish between the behaviour produced by a machine and that produced by a human. The reverse Turing Test is one in which the roles of judge are reversed. In the case of CAPTCHA solvers, a computer is used to judge whether a solver is human or machine.

DOCUMENT CONTROL DATA

(Security markings for the title, abstract and indexing annotation must be entered when the document is Classified or Designated)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g., Centre sponsoring a contractor's report, or tasking agency, are entered in Section 8.)

DRDC – Toronto Research Centre
Defence Research and Development Canada
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9
Canada

2a. SECURITY MARKING (Overall security marking of the document including special supplemental markings if applicable.)

UNCLASSIFIED

2b. CONTROLLED GOODS

(NON-CONTROLLED GOODS)
DMC A
REVIEW: GCEC DECEMBER 2012

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Future Casting Influence Capability in Online Social Networks : Fake Accounts and the Evolution of the Shadow Economy

4. AUTHORS (last name, followed by initials – ranks, titles, etc., not to be used)

Duncan, M.

5. DATE OF PUBLICATION (Month and year of publication of document.)

August 2015

6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)

34

6b. NO. OF REFS (Total cited in document.)

67

7. DESCRIPTIVE NOTES (The category of the document, e.g., technical report, technical note or memorandum. If appropriate, enter the type of report, e.g., interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Scientific Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

DRDC – Toronto Research Centre
Defence Research and Development Canada
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9
Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)

02CA

9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)

DRDC-RDDC-2015-R148

10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

Online Social Networks (OSN) have risen to the point of dominating the daily activities of many Internet users. In response to this success, various groups have sought to exploit these apparently safe and well-populated walled gardens for their own gain. A particularly worrisome target of exploitation is the operation and control of actual accounts. It is now possible to create large numbers of fraudulent accounts with credentials of reasonable face validity. These fake accounts, also called sock-puppets or Sybils, can then be used to generate various influence effects through the apparent collective action of many "individuals". Because the behaviour of accounts affects the perception of trust and legitimacy of the OSN's user base, sock-puppet accounts are a serious threat to an OSN's credibility and economic success. Due in part to the efforts of OSN to combat account fakery, a digital arms race has ensued contributing to the evolution of a technologically sophisticated economic service supply chain involving many players within the cyber-criminal and shadow economy ecosystem. Some of these include malware distributors, Botnet operators, and Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) solving services for account creation and authentication, as well as individuals with expertise in machine learning and artificial intelligence for automated credential generation and control of account behaviour. In its current form, the shadow economy ecosystem represents an influence capability that is open to anyone with money to purchase the accounts. These could range from government agencies to non-state actors to Public Relations (PR) firms to any individual with an axe to grind. In this paper I will describe the complex economic ecosystems that support the generation and sale of fake accounts, review a sample of influence cases involving fake accounts, and examine possible future applications of fake accounts to perpetrate operations for achieving influence effects.

Les réseaux sociaux ont gagné en popularité au point de dominer les activités quotidiennes de maints internautes. C'est pourquoi divers groupes cherchent à exploiter à leur avantage ces jardins clos fort peuplés en apparence inoffensifs. Un aspect particulièrement inquiétant de cette situation réside en la gestion et le contrôle de comptes réels. En effet, il est désormais possible de créer de nombreux comptes frauduleux à l'aide de justificatifs d'identité qui semblent valides. Ces comptes bidon (également appelés faux-nez) peuvent ensuite être exploités à diverses fins d'influence, sous l'action collective apparente d'un grand nombre de « particuliers ». Ces comptes faux-nez constituent une grande menace à la crédibilité et au succès économique d'un réseau social en ligne, car le comportement des comptes influence la perception de confiance et de légitimité de la base d'utilisateurs dudit réseau. La course aux armements numériques qui s'est ensuivie, en partie attribuable aux efforts déployés par les réseaux sociaux en ligne pour combattre la menace, a contribué à l'essor d'une chaîne d'approvisionnement de services économiques avancés sur le plan technologique, chaîne à laquelle prennent part un grand nombre de joueurs de l'écosystème cybercriminel et d'économie parallèle. Parmi ces joueurs, on compte des distributeurs de maliciels, des opérateurs de réseaux zombies et des services de résolution du Test de Turing complètement automatisé afin de distinguer les ordinateurs des humains (Completely Automated Public Turing test to tell Computers and Humans Apart – CAPTCHA) requis pour la création et l'authentification de comptes, ainsi que des particuliers possédant l'expertise en apprentissage machine et en intelligence artificielle nécessaire à la génération automatisée de justificatifs d'identité et au contrôle de comportement de comptes. Dans sa forme actuelle, l'écosystème d'économie parallèle représente une capacité d'influence accessible à quiconque possédant suffisamment d'argent pour acheter des comptes, y compris des organismes gouvernementaux, des entreprises de relations publiques (RP), des acteurs non étatiques ou toute personne qui y trouve un trouve un quelconque intérêt. Dans le présent rapport, je décrirai les écosystèmes économiques complexes qui soutiennent la production et la vente de comptes faux-nez, j'étudierai un exemple d'influence obtenue à l'aide de faux comptes et j'examinerai les applications futures possibles de ce type de compte pour exercer un trafic d'influence.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g., Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Sock Puppet; Sybil; Online Social Media; Influence; Shadow Economy; Underground Economy; CAPTCHA; Twitter; Facebook; Goolge+