Virtual Plots, Real Revolution

Roelof TEMMINGH[a] and Kenneth GEERS[b]
[a] *CEO, Paterva, Pretoria, South Africa*
[b] *Scientist, Cooperative Cyber Defence Centre of Excellence, Naval Criminal Investigative Service (NCIS), Tallinn, Estonia*

Abstract. It is increasingly difficult to separate 'cyberspace' from what we think of as the 'real world'. Human beings respond to stimuli from both. Threats to persons, organizations, and governments require timely and accurate evaluation, but cyber attackers can exploit the imperfect and maze-like architecture of the Internet to make threat evaluation difficult. In cyberspace, it is possible to create fraudulent online identities – potentially millions of them – that could programmatically support any personal, political, or military agenda. In the future, computer botnets may evolve from spam and Distributed Denial of Service (DDoS) generators to semantic creatures that can post opinions, arguments and threats on the Internet. Counterfeit identities on the World Wide Web (WWW), complete with randomized or stolen biographies, pictures, and multi-year histories of Internet activity, will be difficult to separate from real human beings because there is no quick way to determine whether a virtual person really exists.

Keywords. Artificial Intelligence (AI), botnet, cyber threat, cyber warfare, identity theft, Information Operations (IO)

Introduction: a 'semantic botnet'

Cyber attackers exploit the relative anonymity of Internet communications to send unwanted data, including spam, malicious code and Denial of Service (DoS) attacks around the world with near impunity. In the future, computer 'botnets' – networks of compromised and organized computers within a common Command and Control (C2) infrastructure [1] – will evolve to encompass virtual populations of randomly-generated and/or stolen identities, which could be used to support any personal, political, or military agenda. Each fabricated virtual identity will have a 'life' of its own, whose credibility grows over time as the number and variety of its Web postings proliferate.

In 1950, Alan Turing wrote that even the "dullest" human could outperform a computer in a conversation with another human. Turing believed it was inconceivable that a machine could provide a "meaningful answer" to a truly wide variety of questions [2]. In 2009, that may still be the case, but there is a big difference between the formal test that Turing proposed and posting a comment to a blog. Even with adequate time for analysis, there is simply not enough content to evaluate whether it was posted by a man or a machine.

On the Internet, the best way to separate real people from artificial people – without time-consuming, in-depth and unlikely cross examination – is with statistics.

However, for every mathematical defense strategy, there seems to be an effective response for the attacker.

1. The search for intelligent life on the Web

Until recently, most Internet conversations were conducted via email, in relatively interactive, one-on-one correspondence (Internet Relay Chat (IRC) does not count as it was never mainstream). Today, new technologies such as YouTube, Facebook and Twitter allow each of us to be a prolific producer of digital information. The current model is frequently not one-to-one, but one-to-many or many-to-one.

1:n/n:1 communications are popular because they cover a lot of ground very quickly; for example, Sophie can now update her whole family at once instead of each member individually. However, the trade-off here is a loss of interactivity. To pass or fail the Turing Test, some level of cross-examination is required. In short, banal, or low-interaction conversations, a cyber attacker (in the form of an artificial, 'proxy' personality) will find it easier to push information to the world and have it accepted as is, i.e. with no follow-up question and answer time.

Further, it is reasonable to assume that most Internet traffic consists of trivial, everyday information. Serious political and philosophical discussions normally do not take place in the 1:n/n:1 environment. Even when it does, Natural Language Processing (computer analysis of human languages) is unproven technology, and requires significant human oversight to be effective.

Thus, this paper assumes that most of the information found on the Internet is not unique, and can be stolen and repackaged for nefarious purposes. Even among unique photos of a common sight such as the Eifel Tower, the photographers themselves might be hard-pressed to find their own picture among others. Effective authentication technologies such as digital signatures exist, but they are rarely used for common communications, which remain open to theft and malicious manipulation.

2. Internet-enabled intelligence

With an unrestricted Internet connection, the average Web user now has access to more information than her head-of-state did just five years ago. All of Wikipedia, for example, can fit on one hard drive. The data points are now all there; the best strategy is to choose one's sources carefully and to discriminate between good and bad analysis. This is the art of Open Source Intelligence (OSINT).

Computer hackers conduct OSINT just like everyone else. In fact, they also begin their search for information at a target's homepage. Good OSINT can quickly lead an attacker from a name to a date-of-birth, address, education, medical records, and much more. Via social networking sites, the attacker may soon discover intimate details of a person's life, from the clubs she frequents to where she might physically be at any given moment. Eventually, a web of connections to other people, places and things can be constructed.

Good OSINT researchers – and hackers – master both the semantic side of the Web (e.g. the content of a webpage), and the technical side, like the Domain Name Service (DNS), or the 'phone book' of the Internet. The DNS registry catalogues who owns a given website, and often provides a point-of-contact for them in the real world.

Hackers 'enumerate', or conduct in-depth technical reconnaissance, against cyber targets. Technical information, including barely-hidden 'metadata' such as an Internet Protocol (IP) address or a timestamp, is analyzed for anything that can be exploited in the real world. Common applications like webmail are frequently targeted. Sooner or later, hackers normally find an open, misconfigured, or vulnerable Internet access point, which is analogous to a thief finding or forcing open a door or window in the physical world.

The real magic of an effective cyber attack lies in combining technical data with real-world information. Likewise, threat actors can be divided into those who have 'reach' into the real world, and those who do not.

3. It's good to be the king

The combination of Internet monotony and hacker creativity described above can make for a volatile mixture. The average computer programmer could never pass the Turing Test, but she can write a program that updates the world via Twitter on how a bogus Web user is spending his day, or what a bogus Web user thinks about how you are spending yours. And if it is theoretically possible to create one false Web identity, perhaps millions of them already exist.

A large virtual population, scattered all over the world and encompassing different socioeconomic backgrounds, could be programmed to support any personal, social, business, political, military, or terrorist agenda. The nature of an attack could be limited only by the attacker's imagination. For example, in the week before an election, what if both left and right-wing blogs were seeded with false but credible information about one of the candidates? It could tip the balance in a close race to determine the winner.

Via Internet-enabled OSINT, targets can be meticulously profiled by an attacker to learn personal, organizational, or national sensitivities and vulnerabilities. For example, if the target were a multinational corporation (MNC) engaged in oil exploration, OSINT might reveal a wide range of attack vectors: disgruntled employees, friction with indigenous populations, whistleblowers, and/or ongoing lawsuits. A zombie army could be used to target any or all of the above – including judges and jury – by manipulating industry blogs, commenting on news articles, sending targeted email, etc. The MNC, of course, could have its own botnet army pushing its side of the story.

In the impersonal world of cyberspace, who can say for sure whether a message was sent by a real person? Even highly idiomatic language can be stolen by a robot and used (perhaps incorrectly) in another context. It is beside the point to say that one could *eventually* authenticate the information. Propagandists seek first and foremost to bring attention to their cause; ethical considerations are secondary. And the attacker may simply need for the effect to be temporary. If a certain momentum toward the desired goal is achieved – that is, if real people begin to follow the robots – then the attacker can begin to 'plug out' the artificial intelligence. The robots could then be reprogrammed for their next assignment.

Over time, if fake users cannot be distinguished from human users on the Internet, the latter will be forced into a situation not unlike Harrison Ford in *Blade Runner*. The difference will be that there is insufficient interactivity with the robot to spot the fake.

4. The technical details

Today, botnets spam the world, perform DDoS attacks, and hack other computers. Tomorrow, they could be used by ideologues to sway public opinion.

Programmatically, a complex, copy-and-paste algorithm can steal biographical information from web pages, news reports, blogs, and other Internet resources. These in turn can be reconstructed to form the skeleton of an artificial personality. Details from popular news and current events will put meat on the bones. Once created, these artificial 'people' will be instructed to begin interacting with the Web in multiple ways. In due course, they will assume a 'life' of their own, and might even make a few human friends in the process.

The following steps have been field-tested with good results:

Her name is Violet:

- Visit the Census database (http://www.census.gov/genealogy/names/names_files.html)
- Select random first name
- Select random last name

She looks real:

- Select random but common first name/lastname combination
- Search Google (Images) for "fname lname @ Facebook" inurl:profile, medium size with face recognition
- Select random image after page 1

She has a real job:

- Mine the LinkedIn Directory (http://www.linkedin.com/pub/dir/fname/lname)
- Mine the ZoomInfo Developer API (http://developer.zoominfo.com)
- Pick random data and combine creatively

She said what?

- Violet opens a social networking account
- She befriends people
- She posts to their site

5. Evaluating the credibility of a cyber threat

In chess, it is often said that a threat is mightier than its execution. The very existence of a threat – regardless of whether it can be realized – tends to have a harmful effect on the victim, which may behave differently or even begin to act in a way that undermines its long-term security.

OSINT can yield enough information about a target to make even an empty threat seem credible. It is always difficult to quickly and accurately evaluate newly-discovered information, but cyber threats are especially complicated due to the power of modern OSINT and the relative anonymity behind which cyber attackers can hide. For example, phishing attacks are successful even though they normally employ only one layer of deceit: the website itself. Intelligent attackers can weave a much more intricate web of deception than that; an entire organization could successfully be faked if the time were taken to invest in enough third-party references.

In cyber terminology, the classic ‘I know where you live’ can be articulated as ‘I know your Oracle server runs on 10.7.0.33, its administrator is Bob, and Bob likes passwords that relate to Manchester United’. OSINT specialists, especially those with some knowledge of computer hacking, could quickly develop the following threat: ‘You have an appointment today with Dr. Livingstone at the Olympic Hotel … if I were you, I would cancel it’. Business leaders, military officers, and even heads-of-state have personal lives that can be targeted.

Botnet armies could be used to amplify a threat or to artificially enhance its credibility. If an attacker threatened a corporation or a government with strikes or civil unrest, a barrage of hard-to-verify complaints on Web fora could augment the threat, especially if the attacker had been seeding the fora for some time. The challenge for the attacker is to make the fabricated ‘evidence’ seem real while making verification a complex and time-consuming challenge.

When evaluating a cyber threat, it is important to remember that what makes a cyber attack easy – the power, ubiquity, vulnerability and anonymity of the Web – can also lessen its credibility. Good OSINT can lead to a significant bluff. In fact, the problem of attribution is the most complicating factor in cyber threat analysis. If the attacker is careless and leaves a large digital footprint (e.g. his home IP address), law enforcement may be able to take quick action. If the cyber attacker is smart, and covers his digital tracks, then deterrence, evidence collection, and prosecution become major challenges.

In almost all cases, computer log files alone do not suffice. Unmasking a cyber attacker requires the fusion of cyber and non-cyber data points. Investigators must enter the real world if they want to arrest a computer hacker. There will always be clues: if the goal is extortion, where is the money to be paid, and is there a point-of-contact? If the threat is Denial of Service, the target could ask for a proof of capability. The point is to generate a level of interactivity with the cyber threat actor that might be used against it. Further, cross-checking suspect information against trusted sources is always one of the best defenses.

From a technical perspective, solutions to the attribution problem exist. They include the increased use of Public Key Infrastructure (PKI), Internet Protocol version 6 (IPv6), and biometrics. Neural networks have also played a considerable role in reducing credit card fraud [3], and their ability to locate suspicious patterns in voluminous network traffic could be helpful outside the financial sector in the future. However, wide-scale deployment and proper implementation of such technologies are still years away. The widespread use of anonymous email services to support criminal activity, for example, has convinced some that an international convention is needed to regulate its use [4].

In the short term, one inexpensive counter to the threat posed by fake online identities is the simple use of a live video feed. As in *Blade Runner*, before you can really trust someone, it may be necessary to look her in the eye.

6. Attacking zombie armies with mathematics

Cyberspace mirrors the real world, and as such, it is complex and highly dynamic. Nonetheless, security analysts must find signals within the noise, or a targeted attack in a sea of normal network traffic. By way of example, let us examine an attempt to hack a simple, online poll.

The Internet Movie Database (IMDB) ranks Sergio Leone's *Il buono, il brutto, il cattivo* as the top-rated 'Western' film of all time, with an average user-determined score of 8.9 on a scale of 1 to 10 [5]. High IMDB rankings are lucrative in DVD sales, so a rival production company might try to raise the value of its own, low-ranked Western *Five Bloody Graves* by artificially increasing its number of high votes.

The IMDB, and the copyright holders of *Il buono*, must defend their turf. A sound strategy could consist of a two-step process:

1. the discovery of statistics that distinguish humans from computer programs as they vote in an online poll, and
2. using these statistics to support traffic analysis and database integrity.

Is it possible to separate human voters from robotic voters in a given data set? The trick is to keep sorting the data until identifiable fault lines appear. The goal of an attacker is to skew the poll result without being discovered; the goal of an IMDB security analyst is to identify the artificial votes and discard them. In concrete terms, the analyst should try to isolate portions of the data set that look different than those created by humans. While human beings are occasionally irrational, their behavior on the whole can be qualified and quantified as human. For example, when asked to vote on a scale from 1 to 10, human results normally lie within a 'bell curve': some are high, some low, but most votes fall somewhere in the middle.

Statistical analysis should reveal characteristics that distinguish humans from robots throughout the entire voting process. For example, if a computer program were to rate films in a purely random fashion, there would be no qualitative bell curve at all (instead, an equal number of 1s, 2s, 3s, etc). In terms of voting frequency, humans may typically cast their ballots over lunch or before bedtime; computers do not share the same requirements for nourishment and rest, so any serious divergence on vote frequency may be a sign of bot infestation. Humans are also prone to some highly subjective choices: top-ranked *Il buono* has received over 100,000 votes, while fourth-ranked *The Wind* has barely 2,000 to its credit. *The Wind* thus may be a 'hidden gem'; qualitative distinctions such as current popularity and off-beat taste may be difficult to program accurately.

On the technical side, it is possible to analyze the Internet traffic that brought the vote from the remote computer to IMDB in the first place. The 'source' Internet Protocol (IP) address can be geo-located on the Earth with the help of DNS. A good security analyst brings some knowledge of culture and politics to her analysis, and understands that there should not be too large of a discrepancy between what she expects to find and what she does find in the data.

Think of an IP address as a car. Not every parking space should be occupied by a red, 1989 Fiat Uno, just as not every entry in a computer log file should contain the same IP address. At the other extreme, randomizing IP addresses also does not work; one might then see just as many Maseratis in the lot as there are Hondas. To make his cyber attack credible, a hacker needs to make the final distribution of his source IP addresses mirror real Internet traffic patterns, which would require a large and sophisticated botnet.

Internet browser activity also offers computer network defenders valuable data points for analysis. When a human accesses a webpage, she typically waits for images, forms, and advertisements to load in the browser. Computers lack the curiosity and patience of a human. Robotic voters may move mechanically from one data request to the next; all such regimented Web requests should be investigated for other non-human properties.

Finally, cyber defense against virtual army attacks should involve a statistical analysis of the alleged identities themselves. The basic strategy is similar to a game of ‘twenty questions’. Is the user male or female? Young or old? In entertainment or politics? Strange patterns and sudden ratio changes should be investigated. Advanced analysis might consist of an algorithm that combines first name, last name, country of origin, IP address and vote to known or expected baselines. Attackers can never be completely sure of what a security analyst expects to see, so their attack will always require some guesswork and likely entail some miscalculations.

7. Conclusion

In 2009, hackers steal data, send spam, and deny service to other computers. In the future, they may also control virtual armies, in the form of millions of artificial identities that could support any personal, business, political, military, or terrorist agenda. This attack vector exists because humans now communicate via ubiquitous software that is by nature impersonal and non-interactive. Further, given the pure amplification power of the Internet, it is not necessary that every target fall for the scam. And it may not matter if the ruse is eventually discovered, because the attacker may desire to sway public opinion only for a short period of time, such as prior to an election [6], business deal [7], or military operation [8].

Technologies exist, such as PKI, IPv6, and biometrics, to mitigate this threat. Smart system administrators, through network traffic analysis and rigorous database oversight, can also theoretically ensure a high level of data integrity. And if an attacker tried to fly ‘under the radar’ by using an insignificant number of bots for an attack, there would likely be a correspondingly insignificant impact on the target data set to merit the effort.

Unfortunately, the widespread use of good defensive tactics and technologies is not on the horizon. Most system administrators do not have the time, expertise, or staff to undertake a sophisticated analysis of their own data. Furthermore, clever programming can obfuscate many common signatures: if IP addresses and browser settings are scattered within the attack in a realistic way, the bar for cyber defenders is raised considerably.

For the foreseeable future, individual Web users must improve their own ability to evaluate threats emanating from cyberspace [9]. In most cases, the key is credibility. Illustrations from the Turing Test and *Blade Runner* suggest that sufficient interactivity with a computer should reveal that it is not human. But in the 1:n/n:1 computing environment in which we now live, the danger is that adequate dialogue is becoming rarer all the time.

References

[1] Freiling, Felix C., Holz. Thorsten, and Wicherski, Georg. “Botnet Tracking: Exploring a Root-Cause Methodology to Prevent Distributed Denial-of-Service Attacks”. S. De Capitani di Vimercati et al. (Eds.): ESORICS 2005, LNCS 3679, pp. 319–335, 2005.

[2] Oppy, Graham and Dowe, David. “The Turing Test”. *The Stanford Encyclopedia of Philosophy (SEP)*, http://plato.stanford.edu/entries/turing-test/, 2008.

[3] Rowland, Jan B. “The role of automated detection in reducing cyber fraud.” *The Journal of Equipment Lease Financing*; Spring 2002; 20, 1; pg. 2.

[4] Mostyn, Michael M. “The need for regulating anonymous remailers”. *International Review of Law*, Computers & Technology; Mar 2000; 14, 1; pg. 79.

[5] Top Rated “Western” Titles, The Internet Movie Database, www.imdb.com/chart/western.

[6] Consider the enormous impact of the 2004 Madrid train bombings on Spain's national elections three days later: “Europe: An election bombshell; Spain, a week on.” *The Economist*. London: Mar 20, 2004. Vol. 370, Iss. 8367; pg. 41.

[7] Financial institutions often take the loss when their clients are defrauded: Patterson, Aubrey B. “Fighting hackers, fraud and wrong perceptions.” *American Bankers Association. ABA Banking Journal*; Apr 2003; 95, 4; pg. 14. However, the court case of *Ahlo, Inc. vs. Bank of America*, in which malicious code on the company's computer was likely used to steal almost $100,000 from its bank account, demonstrated that coverage is not absolute: Cocheo, Steve. “Privacy rumblings grow louder.” *American Bankers Association. ABA Banking Journal*; Jun 2005; 97, 6; pg. 56.

[8] All political and military conflicts now have a cyber dimension. The conflict between Russia and separatists in Chechnya has clearly demonstrated the power of well-timed Internet propaganda: Geers, Kenneth. "Cyberspace and the changing nature of warfare." *SC Magazine*, http://www.scmagazineus.com/Cyberspace-and-the-changing-nature-of-warfare/article/115929/, August 27, 2008.

[9] In 2006, identity theft was already the fastest-growing crime in the United States, affecting almost 20,000 persons per day: Ramaswamy, Vinita M. “Identity-Theft Toolkit.” *The CPA Journal*; Oct 2006; 76, 10; pg. 66. Nearly a third of all adults in the U.S. reported that security fears had compelled them to shop online less or not at all: Acoca, Brigitte. “Online identity theft.” *Organisation for Economic Cooperation and Development. The OECD Observer;* Jul 2008; 268; pg. 12.