Proceedings of ISIC: the information behaviour conference, Pretoria, South Africa, 28th September to 1st October, 2020.

Digital deception in cybersecurity: an information behaviour lens

M.T. Dlamini, H.S. Venter, J.H.P. Eloff, and M.M. Eloff.

Introduction. Digital deception is a double-edged sword used by both blackhats and whitehats in cybersecurity. A status quo review of the reintroduction of digital deception can reveal challenges and initiatives and show how information behaviour expertise might inform cybersecurity research and vice versa.
Aim. To use a status quo review of digital deception to reveal links between cybersecurity and information behaviour and to stimulate further research.
Method. Critical review of digital deception in cybersecurity regarding whitehats and blackhats using an information behaviour lens.
Findings. There is a need for research that tackles digital deception from both information behaviour and cybersecurity. There is also a need to bridge the gap between the two research fields and link cybersecurity concepts with information behaviour theories.
Conclusions. The reintroduction of digital deception in cybersecurity highlights the challenges for the unreliability of defence-based detection systems. Although many solutions are available from cybersecurity, information behaviour might contribute to multidisciplinary research on digital deception and the future of defence technologies. Understanding the interplay between whitehats and blackhats in cybersecurity can help information behaviour practitioners to design models or frameworks for predicting changes in information-seeking behaviour.

DOI: **https://doi.org/10.47989/irisic2018**

Introduction

In a digital world of ubiquitous information sources (Wilson and Macevičiūtė, 2013), there is a growing trend of falsified online content in the form of *fake news*(Berghel, 2017; Alemanno, 2018), *fake social media profiles* (Malin, et al., 2017) and *fake digital identities* (Hancock, 2009; van der Walt, et al., 2018; Tsikerdekis, et al., 2019). These can be grouped in one term: digital deception (Hancock, 2009). Digital deception is an old concept that is resurfacing in modern information behaviour and cybersecurity in a grander scale and with far reaching consequences (Banerjee, et al., 2019). Despite the plethora of inaccurate and misleading information in the media and on digital platforms, traditional models of information behaviour seem to suggest a normative conception of information as consistently accurate, true, complete, and current, and they neglect to consider whether information might be misinformation (inaccurate information) or disinformation (deceptive information) (Chen, et al., 2015; Karlova and Fisher, 2013; Buschman, 2019). Although fake news and misinformation are addressed in the information science literature, the focus is on the role of librarians, student perceptions and information literacy training and interpretations of the concept of digital deception (Jeffries, et al., 2017; Zakharov, et al., 2019; Buschman, 2019). There is limited evidence of turning to the discipline of cybersecurity for solutions or to find a link between cybersecurity, digital deception and information behaviour. Exceptions are Zhang and Ghorbani (2020) and Gervasio (2019).

Digital deception needs to be addressed as a cybersecurity challenge and informed by cybersecurity as field of study. It, however, also impacts on information behaviour and understanding might be deepened by drawing on insights from information behaviour (a sub-discipline of information science) and cybersecurity in much the same way. Information behaviour, a term that is officially used by the Information Seeking in Context conference (also referred to as the Behaviour Conference) is defined as an umbrella term for all kinds of human interactions with information (Greifeneder, 2014; Julien and Williamson, 2011; Meyer, 2016). For example, it changes the manner in which information seekers and users relate to online or digital content (Nicholas, et al., 2003). This has a direct impact on the human interactions with computers and other digital devices that are used to access digitised information.

Case and Given (2016) have noted that with the internet, information availability is no longer a problem. Hidden in the abundance of online information is a plethora of inaccurate and misleading information in the media and in digital platforms. It is no longer enough to search and find information using search engines; an information seeker needs to go further and verify the integrity of the sources and online information that is presented prior to information use or exchange. Given the rise in big data and predictive data analytics, the validity of data (Greifeneder, 2014) is taking centre stage in both fields of information behaviour and cybersecurity. This is potentially the point where the two fields could come together to address the digital deception problem. For example, Ford (2015) cites some of the widely acknowledged information seeking behaviour models that includes the verification of the accuracy of information e.g., Ellis's model. Though not directly linked to digital deception, Robson and Robinson (2013) provide a good illustration of the complexity of information behaviour studies and the place for evaluation and verification; they make specific reference to assessing the validity of information through utility and credibility. Godbold (2006) refers to '*disputing of the information*' and the work by Wilson and Macevičiūtė (2013) writing on information seeking, one of the information activities falling under the umbrella term of information behaviour, also notes the importance of evaluation and verification. Another information behaviour model that might be considered in the field of cybersecurity is that of Mansourian where he refers to information invisibility and uncertainty. From a cybersecurity perspective Mansourian's concept of information invisibility can be related to privacy and a steganography technique of information hiding.

The above are some of the works that touches on the integrity of information and sources thereof from an information behaviour perspective. The implication is that the field of information behaviour has also noted that not all information or sources thereof are to be trusted. Hence, the need for information verification systems. This need is also emphasised in Maungwa and Fourie (2018) who acknowledge the lack of data validation when discussing information behaviour in competitive intelligence. The need for verification becomes amplified in the advent of purposely planted misinformation or disinformation in the context of digital deception as noted in the first paragraph of this paper. The field of cybersecurity can contribute in providing the information or source verification systems or digital deception detection systems. The authors of this paper have searched widely for existing literature that takes a multidisciplinary approach to address the problem of digital deception from both the perspectives of information behaviour and cybersecurity. Only a few references could be found that does so. For example, the work of Hancock (2009), was found to mention the issue of cybersecurity. However, this was to say that it is a relevant issue but not to be further discussed. The work of Kopp, et al. (2018) which tackles *fake news* using information-theoretic models of digital deception is the closest we could find. Therefore, according to the authors, this paper is among the first research efforts to take a transdisciplinary approach to frame and analyse digital deception with respect to cybersecurity and information behaviour. Although this paper is strongly embedded in cybersecurity based on the authors' background, it makes an attempt to bridge the gap between the two research fields. The goal is to link cybersecurity concepts with information behaviour theories.

Some researchers claim that the greatest impact of digital deception and lack of systems to verify the integrity of online information and their sources that information users consume or exchange was felt in the previous United States of America presidential elections (Berghel, 2017; Lancelot, 2018; Silbey and Hartzog, 2019). Therefore, digital deception in the form of *fake news* is alleged to have interfered with the United States of America electoral system and disrupted the democratic process to sway the votes in favour of a particular candidate (Berghel, 2017; Lancelot, 2018). For this reason, online sources and channels of information related to the United States of America elections came under a spotlight. This has had cascading effects to governments all over the world who are also realising the far-reaching impact of digital deception. The authors believe that this is the main reason why digital deception is now resurfacing in information behaviour and modern cybersecurity in a grander scale.

In the context of cybersecurity, digital deception is defined as a deliberate falsification or manipulation of online content or identity with an intention to purposely create a wrong perception to deceive or misinform unsuspecting information seekers and users (Pawlick, et al., 2019). Hancock (2009) defines digital deception in the context of information behaviour as an intentional control of information in a technologically mediated message to create a false belief in the receiver of the message. In both contexts, the malicious act of digital deception has a deliberate intent to mislead a target information user and the act is carried out via a technological means of communication. It is also referred to as *online deception* (Caspi and Gorsky, 2006) or *cyber deception* (Changwook-Park and Kim, 2019) with the aim to deceive information users that relies on digital content. For the purposes of this paper, the authors adopt and use the term *digital deception* from here onwards. Digital deception has its roots in the military domain (Malin, et al., 2017; Shi, et al., 2019) and it thrives on humans' poor ability to distinguish between falsified and real information (Al-Nemrat, 2018 Frederiksen, 2017; Krause, *et al*., 2019; Pawlick, et al., 2019; Su, 2017). Studies have shown that on average humans have a 55% accuracy rate in differentiating between real and fabricated information (Su, 2017; Asubiaro and Rubin, 2018).

Rapid technology advancements make falsified online content look real (Hao, 2019). Therefore, it would not be a surprise to discover that the 55% figure reported by Su (2017) is even lower. Rapid technology advancements also come with an explosion of big data generated from billions of internet-enabled devices in unprecedented rates. It is reported that by 2016; a total of 44 billion gigabytes (GB) of data were generated per day, globally and that by 2026 the figure will grow exponentially to reach an estimated 463 billion GB per day (Mossburg and Katyal, 2019). The sheer volume of a variety of data that is generated in such high speeds requires systems to attest to its veracity before organisations can derive the value thereof (Younas, 2019; Nguyen, 2018; Gill and Buyya, 2019; van der Walt and Eloff, 2015). Data of all types are being created by all sorts of devices and this creates a fertile ground for digital deception to thrive in the cyberspace.

Digital deception has been studied in different fields such as psychology, politics, economics, communication, and to a lesser extent in cybersecurity and information behaviour (Hancock, 2009; Malin, et al., 2017). Existing cybersecurity literature shows that both whitehats and blackhats have been passively using this concept for more than 30 years (Joshi and Sardana, 2011; Malin, et al., 2017; De Faveri, et al., 2019). Yet, research that explores digital deception in the context of information behaviour and cybersecurity is still insufficient. Therefore, a research gap exists for studies to provide an understanding of digital deception and how it manifests in modern cybersecurity and information behaviour. This paper contributes towards closing this research gap by looking at digital deception from a cybersecurity's perspective through an information behaviour lens (Cramer, 2018; Fulton, et al., 2013).

The structure of the paper is as follows: the next section outlines the current state of the art of digital deception in cybersecurity with respect to both whitehats and blackhats. This sets the scene and provides a glimpse of the current landscape and how it may influence the convergence of information behaviour and cybersecurity. The goal is to briefly pre-empt the road ahead and equip practitioners in both fields with a sneak preview of a looming crisis or opportunity that is presented by digital deception. This paper provides a point of departure on how the two fields can converge towards a better understanding of the rising problem of digital deception. The last section concludes the paper.

Digital deception in cybersecurity

Digital deception is a double–edged sword used by both whitehats and blackhats in modern cybersecurity (Gartzke and Lindsay, 2015; Walker, 2019; Pawlick, et al., 2019). Both blackhats and whitehats exploit what Nahl (2001) refers to as cognitive and affective abilities of each other's mental structures to intentionally influence information behaviour. However, the motivation of the deceiver (which can be either a whitehat or blackhat) in each case is different. For example, whitehats intentionally use digital deception-based mechanisms for securing their systems, and occasionally to track blackhats; whilst blackhats intentionally use them for launching deceptive malicious attacks that attempt to evade detection systems. This interplay between blackhats and whitehats can be best understood from an information behaviour's perspective. For example, Karlova and Fisher (2013), although they might not use the terms blackhats and whitehats, study this interplay from an information behaviour perspective and posits that cues to credibility is used by deceivers to deceive receivers, and receivers use cues to deception to defend against deception. From a cybersecurity perspective, both blackhats and whitehats can each play the role of both a deceiver and receiver. Karlova and Fisher (2013) argue that the success or failure of deception depends on the degree to which receivers may suspect deception. Wilson (2020) in a book review talks of this interplay between whitehats and blackhats as *hacking the hackers*. This is synonymous to whitehats taking an offensive strategy to attack blackhats. Case and Given (2016) argue that studying the underlying human motivation or intent and behaviour is key and very important, yet very difficult to do. This is even more complex in the context of human motivation or intent and behaviour tackled from both a whitehat and/or blackhat; and deceivers' and/or receivers' perspectives.

In agreement with Hancock (2009), the authors of the current paper argue that blackhats are highly motivated to get away with their deceptive behaviour than whitehats. For blackhats, being caught would mean jail time, yet for a whitehat catching a blackhat in action might mean improving their defence systems with a new cyber-attack signature. Therefore, the authors posit that blackhats are more successful in their ability to deceive whitehats' deception detection systems, than whitehats are in luring blackhats to interact with decoy systems. In the field of cybersecurity, the inter-play between whitehats and blackhats is more in line with Hancock's *motivational enhancement effect* than it is with DePaulo and Kirkendol' *motivational impairment effect* (Woodworth, et al., 2005; Hancock, 2009). The ever increasing number of cybercrimes in comparison to only a few isolated successful arrests is evidence to this imbalance. Though, this may be due to the slow wheels of justice or challenges in dealing with digital evidence, the indication is that more research is still required to understand the interplay between blackhats and whitehats in the context of information behaviour. Studying information deception from an information behaviour and information literacy perspective, Karlova and Fisher (2013) confirm that research in this area is insufficient to draw insightful conclusions.

In cybersecurity, digital deception-based tactics, techniques and procedures are also used in blue vs red team exercises where red teams launch attacks whilst blue teams put in place detection systems on the defence. It must be noted though, that reference to the word *defence* also include *offence*, where whitehats may launch a counter attack that targets a specific blackhat or group of blackhats. This relates to Karlova and Fisher's deceivers and receivers, where whitehats may take both roles depending on the context. Hence, in all reference of the word *defence* from a whitehat perspective, this paper uses it to mean both defensive and offensive approaches. Red vs blue team exercises are normally conducted in a simulated environment to identify system vulnerabilities. In general, simulated environments do not give a complete or true picture of the actual threat landscape and would not be effective to draw conclusions for the confluence of information behaviour and cybersecurity. A true picture requires whitehats to track and monitor a blackhat in action attempting to breach a real secure system or honeypot. Whitehats have monitored and tracked blackhats in a number of ways as illustrated in the next subsection.

Digital deception – a whitehat perspective

Whitehats use digital deception-based tactics, techniques and procedures to deceive and mislead blackhats to unknowingly contribute to defence strategies (Changwook-Park and Kim, 2019). This is more of an art than a science of manipulating and providing misleading or fabricated information to an adversary. Information behaviour can help solve this problem from a different angle. Though, this concept now comes in a grander scale and with serious consequences, it is not new to the field of cybersecurity nor is it new to the field of information behaviour. For nearly 30 years now, both whitehats and blackhats have been passively using the concept of digital deception in different ways like spoofing of internet protocol and media access control addresses (Malin, et al., 2017). In the field of information behaviour, it has been discussed as misinformation or disinformation. For example, the work of Karlova and Fisher (2013) posits that digital deception can be viewed from misinformation (mostly referring to inaccurate information) and disinformation (referring to deceptive information) which are types of information behaviour and has a link to information literacy.

Some of the foundational seminal whitehat efforts on digital deception in cybersecurity include the work of Clifford Stoll's; *Stalking the wily hacker* (Stoll, 1988), *What do you feed a Trojan horse* (Stoll, 1987) and *The cuckoo's egg* (Stoll, 1989), Bill Cheswick's *An evening with Berferd* (Cheswisk, 1992, 1997) and Fred Cohen's deception toolkit of 1997 (Joshi and Sardana, 2011; Nasir and Al-Mousa, 2013; Grimes, 2017). These are among the first research efforts (Malin, et al., 2017; Joshi and Sardana, 2011) that discuss how whitehats intentionally lured, interacted with and studied blackhats by providing them fabricated responses, resources and a collection of emulated vulnerabilities.

More than 30 years later, whitehats still continue to use digital deception-based tactics, techniques and procedures to lure and entice blackhats to unknowingly take actions which are in whitehats' favour and contributes to defence strategies (De Faveri, et al., 2019; Whitham, 2017; Joshi and Sardana, 2011; Yuill, et al., 2004). For example, defensive deception-based mechanisms are used to;

- deny blackhats access to confidential resources,
- misdirect blackhats away from legitimate resources,
- delay in responding to blackhats' commands and
- confuse blackhats with plausible yet deceiving information whilst collecting their tactics, techniques and procedures behind the scene (Gutierrez, et al., 2018).

Scholars have noted that tactics, techniques and procedures usually form unique patterns of activities or methods that are specific to a particular blackhat or group of blackhats (Walker, 2019). It would be interesting to note if there would be specific information behaviour characteristics that would help fine tune the patterns with respect to motives and behaviour of groups of blackhats. Whitehats use such tactics, techniques and procedures to create unique attack signatures to correctly identify a blackhat or a group thereof. This is one attempt for whitehats to try and balance the cybersecurity's arm race. The concept of honey-X is one traditional way that whitehats have been using as deception-based tactics, techniques and procedures. Honey-X comes in different forms, such as honeyusers (Malin, et al., 2017)), honeyfiles (Yuill, *et al*., 2004), honeypots (Wallace and Visger, 2018) and honeynets (Shi et al., 2019; Pawlick, et al., 2019; Sokol, et al., 2017). The next section briefly discusses each of these deception-based honey-X and how they are being used by whitehats.

Honey-X

Honeyusers, honeyfiles, honeypots, honeybot and honeynets, which when grouped together are also known as honey-X are used by whitehats in defence strategies. According to Pawlick, et al. (2019), a honey-X refers to any defensive digital deception strategy with a prefix word – honey. The goal of any form of honey-X is to make blackhats waste their resources on attacking fake systems instead of real systems whilst their tactics, techniques and procedures are carefully being monitored and tracked by an observer.

Honeyuser. Honeyusers are also known as fake avatars or decoy users or ghost users. This category of honey-X considers deception at the user identity level and are normally used in social networks by both whitehats and blackhats to phish information on their targets. For example, in 2016, Facebook reported that globally, they had 1.7 billion active users and 8.7% of these were honeyusers (Malin, et al., 2017). Honeyusers come in three different categories; i.e. conceal their real identity, use an identity of another real user (i.e. ghost user) and use a fake identity (van der Walt and Eloff, 2015). From an information behaviour perspective, Hancock (2009) refers to these as identity-based digital deception.

Honeyfiles. Honeyfiles are also known as honeytokens, decoy or canary files (Whitham, 2017). Yuill, et al., (2004) posit that they are deceptive documents that emulate real documents which are intended for blackhats to access and interact with. These are purposely injected within a file-system by whitehats to lure unauthorised blackhats in a defence deception operation (Whitham, 2017). Honeyfiles assist whitehats to successfully detect and combat data leakage threats and unauthorised access from blackhats and malicious information seekers.

Honeyfiles can create confusion and uncertainty regarding the value and location of confidential information. This technique relies on the poor ability of blackhats or humans in general to distinguish between real and fake information. This can be best explained by tapping into the field of information behaviour. Well-designed honeyfiles triggers an alert each time a blackhat or malicious insider accesses them. For example, Gutierrez, et al., (2018) uses a password hash honeyfile to prevent blackhats from accessing and using a brute-force attack on the real password hashes. Each time a blackhat makes an attempt to access the password hash honeyfile, the solution sends an alert to a whitehat that is monitoring the system and then misdirect the blackhat into a honeypot for further attack profiling, fingerprinting and analysis (Gutierrez, et al., 2018). An alert might include some digital forensic information for incident response e.g. internet protocol and media access control address of the access device, details of the user making the access attempt.

Honeyfiles are normally accessible to blackhats who have already been lured to gain unauthorised access to a honeypot within a honeynet. Yuill, et al. (2004) argues that honeyfiles can be used to detect and combat unauthorised access in two levels; 1) that is gained through unknown attacks (i.e. targeting zero-day exploits for outsider threat actors) and 2) that which is gained through unintended file-access permissions (i.e. targeting an insider threat actor). Information behaviour studies that explore how information users deal with falsified information can help to improve defence strategies.

Honeypots. A honeypot is a *decoy* physical or virtual computer designed and setup to resemble a real system whose value lies in its illicit use by a blackhat (Wallace and Visger, 2018; Krawetz, 2004; Joshi and Sardana, 2011). Though, deceptive in its tactics, techniques and procedures, a honeypot must have all the qualities of a real resource and appear legitimate even to the best breed of blackhats for them to meaningfully engage with it. Whitehats are also using honeypots as a weapon for counter attack against blackhats (Wallace and Visger, 2018). This can be achieved by setting up a honeypot that contains honeyfiles embedded with malware. Such honeyfiles must closely resemble the organisation's crown jewels like intellectual property and blueprints to be attractive. A whitehat has to make sure that the honeyfiles cannot be opened to have the payload malware executed within their own environment, but ensure that it can only be executed at the blackhat's environment (Wallace and Visger, 2018). This requires careful attention to detail because a lot could go wrong. For example, it can happen that an employee within organisation that owns such a honeypot, stumbles on the honeypot and gets hold of the honeyfiles to open them in their organisation's network infrastructure.

Hence, extra caution is required to appropriately setup honeypots that will be attractive enough to lure blackhats to interact with them, whilst a whitehat gathers valuable information on tactics, techniques and procedures on its target at the background. Honeypots can also be used to collect and reveal obfuscated internet protocol and media access control addresses, origin, identity and intent of the black-hats which can aid the complex task of cybersecurity attribution. Attribution in cybersecurity is defined as a process of identifying an attacker and their geographical location (Goutam, 2015; Egloff, 2020). However, before whitehats can think of using attribution information on a count attack, they need to be mindful of the effects of *stepping stones* (Nicol and Mallapura, 2014; Yang, et al., 2017; Yang, et al., 2018) in the connection chain. *Stepping stones* are discussed in detail later in the paper. Moreover, in using threat intelligence obtained in a honeypot to perform counter attacks, one must exercise extra caution as when detonating a hand grenade. There is a chance that it might explode before it leaves the hand. So is with counter attacks in cybersecurity.

Honeynets. A honeynet extends the concept of a single honeypot to provide a highly controlled network of multiple honeypots (Sokol, et al., 2017). Honeynets are used in larger networks where a single honeypot would not be sufficient to monitor network activities (Shaikhedris, 2018). For example, Honey V is a collection of honeypots which monitors and detects malicious activities at different intensities to provide fine-grained threat intelligence on blackhats (Rashidi, et al., 2018). The challenges of a single honeypot are compounded in a honeynet. For example, a blackhat can compromise and use a network of homogeneous honeypots to create a distributed denial of service to then attack the same network it is supposed to protect at a bigger scale. The end result would be *honeynet weaponisation* – using honeynets as a weapon.

Summary

In summary all forms of honey-X as defence-based deception tactics, techniques and procedures involve an act of planting false information in a *decoy* system or a planned series of actions aimed to confuse or mislead blackhats to go astray or contribute to building effective defence systems (De Faveri, et al., 2019; Changwook-Park and Kim, 2019). Digital deception in defence operations is aimed to aid whitehats to identify, track and study the tactics, techniques and procedures of blackhats to build effective cybersecurity systems. As postulated in Walker, (2019), mounting a good defence requires whitehats to have a good grasp of the threat surface, the types of attacks, their tactics, techniques and procedures and intents. This is particularly effective for newer threats and zero-day exploits that do not yet have a signature.

The use of honey-X must be given careful considerations because of the vulnerabilities that are used to entice a blackhat which may be spinned around to serve as the only entry point for an adversary to escalate their access into a target network (Liska, 2015; De Faveri, et al., 2019). Blackhats can use a honey-X to *step* into a target network and leave with more than can be realised. Hence, whitehats need to practice extra caution in deploying honey-X solutions.

The above are not necessarily an exhaustive list of the issues of deploying honey-X or any deception-based mechanism to defend systems and networks from blackhats, who are always one-step ahead of whitehats in the cybersecurity arm's race. But the authors hope they are at least enough to drive the point home for whitehats to reconsider their approach to defensive-based deception mechanisms. The field of information behaviuor can be looked at for solutions to the problem of digital deception. The question could be asked if honey-X and other deception-based tactics, techniques and procedures still do serve the whitehat community, especially in the advent of anti-deception detection technologies which by-passes detection systems with so much ease.

The next section highlights some of the latest deception-based techniques for offensive operations that are being used by blackhats to evade detection systems.

Digital deception – a blackhat perspective

Whitehats are improving security systems and decreasing the cycle of patching vulnerable systems (Mazurczyk and Caviglione, 2015) and blackhats are fast realising that they can no longer rely on exploiting unpatched vulnerabilities to install and run their malware, and they now prefer to use deception to make their malware become more stealth in order to evade detection. As is the case in the defence, deception is not a new concept even in the offense operations. Kevin Mitnick's *The Art of Deception* (Mitnick and Simon, 2003) is one of the earlier works on deception in offense operations. This seminal work describes how blackhats successfully targeted system users as the weakest link in the cybersecurity chain to use social engineering in order to breach into systems they were not authorised to access. Social engineering is one cybersecurity technique that can be best explained using concepts in information behavioor. This is because its success or failure hinges on how the target users can detect deception from the attacker and such interactions can be best explained outside the field of cybersecurity.

Blackhats are now using sophisticated deception tactics, techniques and procedures to conceal their identity, digital trace and locations (Casey, 2019; Goutam, 2017). For example, *stepping stones*, obfuscated system tools, *stegware*, *deepfakes* and adversarial artificial intelligence and machine learning are some of the tools that are available in a blackhat's arsenal.

Stepping stones

A *stepping stone* is a compromised host that blackhats exploit to reach their target and launch highly targeted attacks. The deceptive nature of *stepping stones* makes it almost impossible to identify the true attacker's source (Goutam, 2017; Gamarra, et al., 2018; Yang, et al., 2017). These *stepping stones* are used to conceal an attack's origin in order to bypass detection systems and eventually avoid attribution. Goutam (2017) posits that *stepping stones* allow blackhats to conceal their identifiable information behind vulnerable digital infrastructure and its shield of anonymity. This relates to Mansourin's model of information invisibility and makes such attacks to be so accurate and precise on their target, yet hard to detect their origin. The uncertainty of the sources of such attacks increases with the stealth of *stepping stones*. Whitehats have not yet figured out how to deal with *stepping stones* and research focused on detecting the origin of attacks is still inadequate and limited (Yang, et al., 2017). Goutam (2017) also posits that *stepping stones* pose a major stumbling block for all efforts of cyberspace attribution. The challenge is amplified when blackhats decide to use multiple *stepping stones* in a single connection chain. Furthermore, researchers (Yang, et al., 2017; Nicol and Mallapura, 2014) have confirmed that more blackhats are now using *stepping stones* to launch their attacks in an effort to avoid detection. *Stepping stones* can be external or internal to the target network. It is better to deal with external *stepping stones* than those that are internal. The challenge with external *stepping stones* is that retaliation may be misguided to a target that knows nothing about the attack it is launching. On the other hand, *stepping stones* inside the target network are most difficult to detect with traditional systems that are often externally facing to detect threats coming from outside than those from inside. The other problem with internal *stepping stones* is that they are highly effective from a blackhat's perspective because they are launched within their target network. Internal *stepping stones*are like threats that are *living off the land*.

Obfuscation system tools - living off the land

Blackhats are using a stealth and deception-based offensive technique called *living off the land* (Banerjee, et al., 2019; Symantec, 2017; Sophos, 2019). This technique requires a blackhat to infiltrate a target network and exploit whatever system tools they find in the compromised environment to launch an attack from within a target environment. The purposes of an *living off the land* attack is in three folds: 1) to first blend in to their victim's network and 2) then hide their activities in a pool of legitimate tools to make it harder to attribute malicious activities, and 3) to evade detection. *Living off the land* exploits vulnerabilities from externally facing perimeter systems which are normally configured to keep external threats out. With *living off the land*, the threats are already inside the perimeter. Instead of looking at attacks coming from external sources, perimeter systems should now also look at internal threats.

For example, blackhats are reported to have used the *living off the land* techniques to hijack a legitimate PsExec remote tool to spread Petya attacks in 2017 (Popli and Girdhar, 2019). In another example, blackhats used a patch and update process within a compromised system as a *stepping stone* to move laterally to distribute malware across an entire network (Symantec, 2017). This is an indication that blackhats continue to shift their evasion techniques to remain undetected. The longer the *dwell* time in a *living off the land* environment, the better the opportunity that blackhats can find, exfiltrate and destroy data. A *dwell* time is the number of days a blackhat is present on a victim network from the first evidence of compromise to detection (FireEye Mandiant, 2019). The approach of using deceptive *living off the land* techniques to hide behind legitimate system tools is a big challenge that is gaining more traction. Moreover, there are at least 35 such legitimate system tools which blackhats are hijacking and exploiting for illegitimate purposes (Symantec, 2017). The field of cybersecurity can tap into the field of information behaviour as it seeks to understand how users deal with hidden information as explained by Wilson (1999).

Stegware – steganographic malware

Blackhats are also reverting back to an old stealth deceptive technique of steganography to provide themselves with a *veil* of information invisibility (Ford, 2015) to evade detection. Deep Secure (2019) argues that steganography presents unlimited opportunities for blackhats to develop a variety of deception-based offensive tools. Steganography is a technique that has always been used by whitehats in covert communications to exchange confidential or secret information. It works by concealing a secret or confidential information within a seemingly innocuous digital media file. A good example of this is found in Dlamini, et al. (2016) where one time passwords (OTPs) are first encrypted and then embedded onto a low fidelity image before they can be transmitted over an unsecured channel to prevent eavesdropping.

Blackhats have taken the same approach to weaponise steganography in what they refer to as stegware (Abarca, 2018; Kay, 2019; Deep Secure, 2019; Mazurczyk and Wendzel, 2018). Stegware conceal malware payload in digital media files to make them bypass in-bound content filtering firewalls and anti-malware solutions to infiltrate their target network. Once inside a target network, blackhats can use the same stegware to bypass out-bound data leakage prevention systems to exfiltrate and siphon stolen confidential information out of unsuspecting companies without being detected (Abarca, 2018). This threat has been used in mal-advertising campaigns on reputable news websites to infect users' devices with malware (Mazurczyk and Wendzel, 2018).

Blackhats are using stegware as a double-edged sword for carrying their malware past in-bound detection systems to infect target networks and also to siphon stolen information past out-bound data leakage prevention systems. Conventional perimeter solutions are not yet able to effectively defend networks against this threat which Deep Secure (2019) argues is an *elephant* in the room that no one wants to talk about. Mazurczyk and Wendzel, (2018) also confirms that there are currently no general effective stegware detection solutions yet. The field of cybersecurity can look to the field of information behaviour for solutions to this problem.

Adversarial artificial intelligence and machine learning - attacks on predictive cybersecurity analytics

Whitehats have proposed predictive cybersecurity analytics (Abraham and Nair, 2015; van der Walt and Eloff, 2015) and cyber threat intelligence systems (Mahlangu, et al., 2019) as new ways of responding to the call of proactive defensive systems. Unfortunately, Blackhats have already realised that some zero-day exploits get detected at first appearance by artificial intelligence and machine learning - inspired malware detectors in predictive cybersecurity analytics systems (Abraham and Nair, 2015; Salem *et al*., 2019). Therefore, blackhats are moving their target to focus on breaching artificial intelligence and machine learning inspired predictive cybersecurity analytics systems using adversarial artificial intelligence and machine learning attacks. Adversarial artificial intelligence and machine learning attacks have recently emerged to deceive and exploit artificial intelligence and machine learning classification models and algorithms into interpreting input training data in a way that is favourable to a blackhat's desired outcomes. Similar to the way whitehats use deception-based tactics, techniques and procedures to lure blackhats to their honey-X whilst they gather threat intelligence information to strengthen defence systems; blackhats also compromise and exploit weaknesses in artificial intelligence and machine learning classification models and algorithms to misinterpret and misclassify input training data to make them work to their advantage (Salem, et al., 2019). For example, blackhats use a carefully selected corpus of adversarial inputs that closely resembles normal inputs which when strategically injected to the normal input will send artificial intelligence and machine learning models on a *goose chase* to make inaccurate predictions that will favour blackhats. This is one way that blackhats use the art of deception to make their malicious threats avoid detection by artificial intelligence and machine learning inspired content filters and perimeter firewalls.

Adversarial artificial intelligence and machine learning threats rely on the blackbox type of complexity of classification models. Data scientists are developing increasingly complex *blackbox-type* of models that use deep learning techniques to capture an unprecedented number of actions to solve complex problems.

The *black-box* models defy human reasoning and often lack human interpretation, even by developers. The complexity that comes from unpredictable model behaviour and the lack of interpretation create a vulnerability for blackhats to exploit (Salem, et al., 2019).

It has been said elsewhere that complexity is an enemy of cybersecurity (Warchal, 2018). It is no longer safe to have a secure hardware and software infrastructure on which to run complex artificial intelligence and machine learning classification models and algorithms. Researchers must also ensure that the algorithms and classification models are secured against malicious use. Therefore, Salem, et al. (2019) asserts that artificial intelligence and machine learning models and algorithms are a new attack surface that whitehats must deal with. The complexity of the models make them susceptible to adversarial artificial intelligence and machine learning. The authors argue that cybersecurity is indeed an arm's race with a constantly moving target – here today and there tomorrow (Dlamini, et al., 2009; Joshi and Sardana, 2011). Adversarial artificial intelligence and machine learning presents a much deeper and harder challenge to digital deception, i.e. *deepfakes*. It is important that artificial intelligence and machine learning models should consider some of the information behaviour models to help address the problem of digital deception. For example, Case and Given (2016) note the emphasis of Robson and Robinson's model on the information sources, communicator and recipient that affects information behaviour. Factoring these into artificial intelligence and machine learning models can help address the problems of misinformation or disinformation in digital deception in artificial intelligence and machine learning defence systems.

Deepfakes

There is a looming era of adversarial artificial intelligence and machine learning -inspired *deepfakes* (Citron and Chesney, 2018; Matern, et al., 2019; Day, 2019; Ding, 2019; Hao, 2019; Fletcher, 2018; Knight, 2019) which puts a new twist to make worse the already complex issue of digital deception. *Deepfakes* manipulate digital media files such as video and audio recordings to show people doing or saying things they never did or said (Fletcher, 2018). According to Silbey and Hartzog (2019) *deepfakes* require everyone to question deeply held truths and axioms about the trustworthiness of what people see and hear. For example, an adversarial artificial intelligence and machine learning-inspired Voco - a product of Adobe - takes a text string as input to generate a voice excerpt and within seconds transform the resultant voice excerpt to a different voice excerpt that is totally different from the previous one (Day, 2019). *Deepfakes* take the battle of digital deception between blackhats and whitehats to an unprecedented level. Hao (2019) argues that this is not a time to raise more questions, but to question and scrutinise everything.

Knight (2019) posits that there is a growing concern that *deepfakes* could be used to manipulate and deceive voters in order to influence the next United States presidential elections. Given the unconfirmed rumours that digital deception in the form of *fake news* might have influenced the previous United States presidential elections (Berghel, 2017), the likelihood of *deepfakes* meddling with the next elections is high. Google and Facebook are purportedly reported to be in the process of generating a corpus of *deepfakes* training data to develop artificial intelligence and machine learning inspired deception-based detection solutions (Knight, 2019). These are also susceptible to adversarial artificial intelligence and machine learning and may benefit from incorporating principles from the field of information behaviour.

Summary

In summary, blackhats' offensive-based deception techniques are more stealth than the traditional defensive-based deception approach. Threats that use *stepping stones*, obfuscated system tools, *stegware*, *deepfakes* and adversarial artificial intelligence and machine learning may take a long time to be discovered. This is mainly because they hide under normal activities and cannot easily raise flags. It will take a different type of detection systems that incorporate the principles of other disciplines like information behaviour to effectively detect threats that make use of these techniques.

Evidently, the threat landscape continues to increase in complexity and with far reaching consequences. For example, gathering digital traces that could be used to hold blackhats accountable for their malicious acts is almost impossible. There is a rising need to incorporate concepts and principles from other disciplines to be able to address the increasing complexities.

Another issue which is worth mentioning without going into much details with respect to the tactics, techniques and procedures of blackhats is anti-forensic solutions (Losavio, et al., 2019). Anti-forensic solutions are used by blackhats to make it impossible for whitehats and digital forensic investigators to collect their digital traces for analysis. For example, blackhats use anti-forensic tactics, techniques and procedures to encrypt stolen data or their communication channels to conceal malicious activities and prevent detection. Current defence systems are not able to deal with encrypted data. In future, whitehats need to develop solutions that will deal with anti-forensic tactics, techniques and procedures. The next section concludes the paper.

Conclusion

Though digital deception has been used as a double-edged sword by both blackhats and whitehats, it has recently taken a new twist and is resurfacing in a grander scale with huge ramifications to both cybersecurity and information behaviour. The concept of digital deception is resurfacing at a crucial time for both fields. This is coming when cybersecurity is under public scrutiny for the increasing number of incidents, yet only a few of these incidents can be attributed to specific individuals or groups of blackhats. In the case of information behaviour, there are serious concerns on misinformation or disinformation which has led to growing research interest on the validity of digital information with the rise of big data and predictive data analytics. The reintroduction of digital deception in the field of cybersecurity bring a crisis in terms of making worse an already complex issue with regards to the reliability of defence-based detection systems.

Information behaviour researchers should be stimulated by this paper to enhance research to deepen understanding of human information behaviour to deal with the challenges of deceptive information and digital deceptions. Many opportunities can open up to work with specialists in cybersecurity to find technological solutions that might supplement the work that is currently coming from information literacy training. Researchers in cybersecurity might draw on theories that have been used in information behaviour research such as information theory, social-cognitive theory and sense-making theory (Chen, et al., 2015). At this stage it is still too early to decide on how each of the two disciplines can inform the development of systems that draw on findings from both fields. This might be a good time to strengthen research ties. The authors speculate that in addressing the challenges of digital deception, the field of cybersecurity will eventually have to look at the problem from a multidisciplinary perspective for it to reach its next level of maturity. This paper is meant to open and stimulate more research discussions around the interplay of the fields of information behaviour and cybersecurity around addressing digital deception. The authors also aim to understand the implications of digital deception from an information behaviour lens to the future of cybersecurity and probable shape the future of research in the two disciplines. Although this paper gravitates towards cybersecurity based on the authors' background, it is the first attempt to bridge the gap between the two research fields with a goal to link cybersecurity concepts with information behaviour theories.

Acknowledgements

The support of the University of Pretoria, University of South Africa and the Council of Scientific and Industrial Research is hereby acknowledged. Opinions expressed and conclusions arrived at are solely those of the authors and should not necessarily be attributed to above institutions.

About the author

Moses Dlamini is a Senior Cybersecurity Researcher at Council of Scientific and Industrial Research, Pretoria Campus, Pretoria, South Africa. He is also a part-time lecturer at the University of Pretoria, in the Department of Computer Science. He can be contacted at: TDlamini1@csir.co.za
Hein Venter is a Professor at the University of Pretoria, Hatfield Campus, Department of Computer Science, Pretoria, South Africa. He is also a research group leader for the Information and Computer Security Architectures research group at the University of Pretoria. He can be contacted at: hein.venter@up.ac.za
Jan Eloff is a Professor at the University of Pretoria, Hatfield Campus, Department of Computer Science, Pretoria, South Africa. He is also a Deputy Dean: Research and Postgraduate Studies in the Faculty of Engineering, Built Environment and Information Technology at the University of Pretoria, Pretoria, South Africa. He can be contacted at: jan.eloff@up.ac.za

Mariki Eloff is a Professor at the University of South Africa, College of Economic and Management Sciences, Institute of Corporate Citizenship, Pretoria, South Africa. She is also the chair of the Institute of Corporate Citizenship at the University of South Africa. She can be contacted at: Eloffmm@unisa.ac.za

References

- Abarca, S. (2018). *An analysis of network steganographic malware*. (Utica College Master thesis) http://dx.doi.org/10.13140/RG.2.2.33593.62564 (https://www.researchgate.net/profile/Suzanne_Abarca/publication/332511600_AN_ANALYSIS_OF_NETWORK_STEGANOGRAPHIC_MALWARE) (Archived by the Internet Archive at https://web.archive.org/web/20200729081921/https://www.researchgate.net/publication/332511600_AN_ANALYSIS_OF_NETWORK_STEGANOGRAPHIC_
- Abraham, S. & Nair, S. (2015). Exploitability analysis using predictive cybersecurity framework. In P. Jędrzejowicz, N.T. Nguyen, T-P. Hong & I. Czarnowski (Eds.), *Proceedings of the 2nd International Conference on Cybernetics* (pp. 317-323). IEEE. http://dx.doi.org/10.1109/CYBConf.2015.7175953
- Alemanno, A. (2018). How to counter fake news? A taxonomy of anti-fake news approaches. *European Journal of Risk Regulation, 9*(1), 1-5. http://dx.doi.org/10.1017/err.2018.12
- Al-Nemrat, A. (2018). Identity theft on e-government/e-governance & digital forensics. In *Proceedings of the 13th International Symposium on Programming and Systems (ISPS)* (pp. xi). IEEE. http://dx.doi.org/10.1109/ISPS.2018.8378961
- Asubiaro, T.V. & Rubin, V.L. (2018). Comparing features of fabricated and legitimate political news in digital environments (2016-2017). *Proceedings of theAssociation for Information Science & Technology, 55*(1), 747-750. http://dx.doi.org/10.1002/pra2.2018.14505501100
- Banerjee, A., Campbell, T. & Mote, A. (2019). *A look at deception: how to start playing offense against advanced attackers*. Symanec Whitepaper. https://docs.broadcom.com/doc/a-look-at-deception-how-to-start-playing-offense-en (Archived by the Internet Archive at https://web.archive.org//web/20200729094236/https://docs.broadcom.com/doc/a-look-at-deception-how-to-start-playing-offense-en)
- Berghel, H. (2017). Lies, damn lies, and fake news. *Computer, 50*(2), 80-85. http://dx.doi.org/10.1109/MC.2017.56
- Buschman, J. (2019). Good news, bad news, and fake news. *Journal of Documentation, 76*(1), 213-228. http://dx.doi.org/10.1108/JD-05-2018-0074
- Case, D.O. & Given, L.M. (2016). *Looking for information: a survey of research on information seeking, needs, and behaviour* (4th). Emerald Group Publishing.
- Casey, E. (2019). The chequered past and risky future of digital forensics. *Australian Journal of Forensic Sciences*, *51*(6), 649-664. http://dx.doi.org/10.1080/00450618.2018.1554090
- Caspi, A. & Gorsky, P. (2006). Online deception: prevalence, motivation, and emotion. *CyberPsychology & Behaviour, 9*(1), 54-59. http://dx.doi.org/10.1089/cpb.2006.9.54
- Changwook-Park & Kim, Y. (2019). Deception tree model for cyber operation. In *Proceedings of the 2019 International Conference on Platform Technology and Service (PlatCon)* (pp. 1-4). IEEE. http://dx.doi.org/10.1109/PlatCon.2019.8669410
- Chen, X., Sin, S.C.J., Theng, Y.L. & Lee, C.S. (2015). Deterring the spread of misinformation on social network sites: a social cognitive theory-guided intervention. *Proceedings of the Association for Information Science & Technology, 52*(1), 1-4. http://dx.doi.org/10.1002/pra2.2015.145052010095
- Cheswick, W. (1992). An evening with Berferd in which a cracker is lured, endured, and studied. In *Proceedings of the Winter USENIC Conference* (pp. 163-174).
- Cheswick, W. (1997). An evening with Berferd. In D.E. Denning & P.J. Denning (Eds.), *Internet besieged: countering cyberspace scofflaws* (pp. 103-116). ACM Press/Addison-Wesley Publishing Co.
- Chesney, R. & Citron, D. (2018, February 21). *Deep fakes: a looming crisis for national security, democracy and privacy?* https://www.lawfareblog.com/deepfakes-looming-crisis-national-security-democracy-and-privacy (Archived by the Internet Archive at /web/20200804015809/https://www.lawfareblog.com/deepfakes-looming-crisis-national-security-democracy-and-privacy)
- Cramer, T. (2018). WhiteHat security: using AI to enable better SEO results. *EContent, 41*(4), 35-37.
- Day, C. (2019). The future of misinformation. *Computing in Science & Engineering, 21*(1), 108. http://dx.doi.org/10.1109/MCSE.2018.2874117
- Deep Secure. (2019). *Deep secure information eXchange (iX)*. Deep Secure. https://www.deep-secure.com/
- De Faveri, C., Moreira, A. & Souza, E. (2019). Deception planning models for cyber security. In O. Gervasi, B. Murgante, S. Misra, G. Borruso, C. Torre, A.M. Rocha, D. Taniar, B. Apduhan, E. Stankova & A. Cuzzocrea (Eds.), *Proceedings of the17th International Conference on Computational Science and its Applications (ICCSA)* (pp. 1-8). IEEE. http://dx.doi.org/10.1109/ICCSA.2017.8000014
- Ding, L. (2019). *Deepfake technology: a paradox of economic and political possibilities*. Masters of Media, New Media & Digital Culture, M.A. University of Amsterdam. https://mastersofmedia.hum.uva.nl/blog/2019/09/22/deepfake-technology-a-paradox-of-possibilities/ (Archived by the Internet Archive at /web/20200804020624/https://mastersofmedia.hum.uva.nl/blog/2019/09/22/deepfake-technology-a-paradox-of-possibilities/)
- Dlamini, M.T., Eloff, J.H.P., Eloff, M.M. (2009). Information security: the moving target. *Computers & Security, 28*(3-4), 189-198. http://dx.doi.org/10.1016/j.cose.2008.11.007
- Dlamini, M.T., Eloff, M.M., Eloff, J.H.P., Venter, H.S., Chetty, K. & Blackledge, J.M. (2016). Securing cloud computing's blind-spots using strong and risk-based MFA. In *Proceedings of the CONF-IRM 2016*. AIS eLibrary. https://aisel.aisnet.org/confirm2016/22/ (Archived by the Internet Archive at /web/20200804020805/https://aisel.aisnet.org/confirm2016/22/)
- Egloff, F.J. (2020). Contested public attributions of cyber incidents and the role of academia. *Contemporary Security Policy, 41*(1), 55-81. http://dx.doi.org/10.1080/13523260.2019.1677324
- FireEye Mandiant. (2019). *M-Trends 2019 Report*. FireEye Mandiant Services, Special Report. https://content.fireeye.com/m-trends (Archived by the Internet Archive at /web/20200804021015/https://content.fireeye.com/m-trends)
- Fletcher, J. (2018). Deepfakes, artificial intelligence, and some kind of distopia: the new faces of online post-fact performance. *Theatre Journal, 70*(4), 455-471. https://doi.org/10.1353/tj.2018.0097
- Ford, N. (2015). *Introduction to information behaviour*. Facet Publishing.
- Frederiksen, L. (2017). Fake news. *Public Services Quarterly, 13*(2), 103-107. http://dx.doi.org/10.1080/15228959.2017.1301231
- Fulton, E., Lawrence, C. & Clouse, S. (2013). White hats chasing black hats: careers in IT and the skills required to get there. *Journal of Information Systems Education, 24*(1), 75-80.
- Gamarra, M., Shetty, S., Nicol, D.M., Gonzalez, O., Kamhoua, C.A. & Njilla, L. (2018). Analysis of stepping stone attacks in dynamic vulnerability graphs. In *Proceedings of the International Conference on Communications* (pp. 1-7). IEEE. http://dx.doi.org/10.1109/ICC.2018.8422723
- Gartzke, E. & Lindsay, J.R. (2015). Weaving tangled webs: offense, defense, and deception in cyberspace. *Security Studies, 24*(2), 316-348. http://dx.doi.org/10.1080/09636412.2015.1038188
- Gervasio, D.M. (2019). Apps, AI, and automated fake news detection. *Information Outlook (Online), 23*(2), 9-12.
- Gill, S.S. & Buyya, R. (2019). A taxonomy and future directions for sustainable cloud computing: 360 degree view. *ACM Computing Surveys, 51*(5), article 104. http://dx.doi.org/10.1145/3241038
- Godbold, N. (2006). Beyond information seeking: towards a general model of information behaviour. *Information Research, 11*(4), paper 269. http://Informationr.net/ir/11-4/paper269.html (Archived by the Internet Archive at https://web.archive.org/web/20200722214957/http://informationr.net/ir/11-4/paper269.html)
- Goutam, R.K. (2015). The problem of attribution in cyber security. *International Journal of Computer Applications, 131*(7), 34-36.
- Goutam, R.K. (2017). Design issues in stepping stone detection. *International Journal of Computer Applications, 166*(9), 1-4. http://dx.doi.org/10.5120/ijca2017914108
- Greifeneder, E. (2014). Trends in information behaviour research. *Information Research, 19*(4), isic 13. http://informationr.net/ir/19-4/isic/isic13.html#.XyKeY0l7nIU (Archived by the Internet Archive at https://web.archive.org/web/20200728164545/http://informationr.net/ir//19-4/isic/isic13.html#.XyKfLEl7nIV)
- Grimes, R.A. (2017). *Hacking the hacker: learn from the experts who take down hackers*. John Wiley & Sons Inc.
- Gutierrez, C.N., Kim, T., Corte, R.D., Avery, J., Goldwasser, D., Cinque, M. & Bagchi, S. (2018). Learning from the ones that got away: detecting new forms of phishing attacks. *IEEE Transactions on Dependable and Secure Computing, 15*(6), 988-1001. http://dx.doi.org/10.1109/TDSC.2018.2864993
- Hancock, J.T. (2009). Digital deception: why, when and how people lie online. In A.N. Joinson, K.Y.A. McKenna, T. Postmes & U-D. Reips (Eds.), *Oxford Handbook of Internet Pyschology* (chapter 19). Oxford university press. http://dx.doi.org/10.1093/oxfordhb/9780199561803.013.0019 (https://sml.stanford.edu/ml/2009/02/hancock-ohip-digital-deception.pdf)

- Hao, K. (2019). The biggest threat of deepfakes isn't the deepfakes themselves. *MIT Technology review*. https://www.technologyreview.com/s/614526/the-biggest-threat-of-deepfakes-isnt-the-deepfakes-themselves/ (Archived by the Internet Archive at /web/20200804021240/https://www.technologyreview.com/2019/10/10/132667/the-biggest-threat-of-deepfakes-isnt-the-deepfakes-themselves/)
- Jeffries, S., Kroondyk, J., Paolini, F. & Radisauskas, C. (2017). Says who?: librarians tackle fake news. *College & Research Libraries News, 78*(10), 538.
- Joshi, R.C. & Sardana, A. (2011). *Honeypots: a new paradigm to information security* (1st ed.). Science Publishers, CRC Press, Taylor & Francis Group.
- Julien, H. & Williamson, K. (2011). Discourse and practice in information literacy and information seeking: gaps and opportunities. *Information Research, 16*(1), paper 458. http://InformationR.net/ir/16-1/paper458.html (Archived by the Internet Archive at https://web.archive.org/web/20200722215724/http://informationr.net/ir/16-1/paper458.html)
- Karlova, N.A. & Fisher, K.E. (2013). A social diffusion model of misinformation and disinformation for understanding human information behaviour. *Information Research, 18*(1), paper 573. http://InformationR.net/ir/18-1/paper573.html (Archived by the Internet Archive at https://web.archive.org/web/20200722214819/http://informationr.net/ir/18-1/paper573.html#.XyF0GUl7nIU)
- Kay, B. (2019). The rise of stegware. *InfoSecurity-magazine.com*. https://www.infosecurity-magazine.com/opinions/rise-stegware-1/ (Archived by the Internet Archive at /web/20200804021409/https://www.infosecurity-magazine.com/opinions/rise-stegware-1/)
- Knight, W. (2019). Even the AI behind deepfakes can't save us from being duped. *com*. https://www.wired.com/story/ai-deepfakes-cant-save-us-duped(Archived by the Internet Archive at /web/20200804021607/https://www.wired.com/story/ai-deepfakes-cant-save-us-duped/)
- Kopp, C., Korb, K.B. & Mills, B.I. (2018). Information-theoretic models of deception: modelling cooperation and diffusion in populations exposed to "fake news". *PLoS ONE, 13*(11), e0207383. https://doi.org/10.1371/journal.pone.0207383
- Krause, N.M., Wirz, C.D., Scheufele, D.A. & Xenos, M.A. (2019). Fake news: a new obsession with an old phenomenon? In J.E. Katz & K.K. Mays (Eds.), *Journalism and Truth in an Age of Social Media* (pp. 58-78). Oxford university press. http://dx.doi.org/10.1093/oso/9780190900250.003.0005
- Krawetz, N. (2004). Anti-honeypot technology. *IEEE Security & Privacy Magazine, 2*(1), 76-79. http://dx.doi.org/10.1109/MSECP.2004.1264861
- Lancelot, J.F. (2018). Russia today, cyberterrorists tomorrow: US failure to prepare democracy for cyberspace. *Journal of Digital Forensics, Security and Law, 13*(4), 23-31.
- Lindsay, J.R. (2015). Tipping the scales: the attribution problem and the feasibility of deterrence against cyberattack. *Journal of Cybersecurity, 1*(1), 53-67. https://doi.org/10.1093/cybsec/tyv003
- Liska, A. (2015). Advanced intelligence capabilities. In A. Liska (Ed.), *Building an Intelligence-Led Security Program* (pp. 153-169). Syngress Publishing.http://dx.doi.org/10.1016/B978-0-12-802145-3.00010-7
- Losavio, M.M., Pastukov, P., Polyakova, S., Zhang, X., Chow, K.P., Koltay, A., James, J. & Ortiz, M.E. (2019). The juridical spheres for digital forensics and electronic evidence in the insecure electronic world. *Wiley Interdisciplinary Reviews: Forensic Science, 1*(5), 1-13. http://dx.doi.org/10.1002/wfs2.1337
- Mahlangu, T., January, S., Mashiane, T.C., Dlamini, M.T., Ngobeni, S.J. & Ruxwana, N.L. (2019). 'Data poisoning' - achilles heel of cyber threat intelligence systems. In N. van der Waag-Cowling & L. Leenen (Eds.), *Proceedings of the 14th International Conference on Cyber Warfare and Security, (ICCWS 2019).*Academic Conferences Ltd.
- Malin, C.H., Gudaitis, T., Holt, T. & Kilger, M. (2017). *Deception in the digital age: exploiting and defending human targets through computer-mediated communications* (1st ed.). Academic Press Inc.
- Matern, F., Riess, C. & Stamminger, M. (2019). Exploiting visual artifacts to expose deepfakes and face manipulations. In *Proceedings of the Winter Applications of Computer Vision Workshops (WACVW)* (pp. 83- 92). IEEE. http://dx.doi.org/10.1109/WACVW.2019.00020
- Maungwa, T. & Fourie, I. (2018). Exploring and understanding the causes of competitive intelligence failures: an information behaviour lens. *Information Research, 23*(4), paper isic1813. http://www.informationr.net/ir/23-4/isic2018/isic1813.html (Archived by WebCite® at http://www.webcitation.org/74FBb1rq7)
- Mazurczyk, W. & Caviglione, L. (2015). Information hiding as a challenge for malware detection. *IEEE Security & Privacy, 13*(2), 89-93. http://dx.doi.org/10.1109/MSP.2015.33
- Mazurczyk, W. & Wendzel, S. (2018). Information hiding: challenges for forensic experts. *Communications of the ACM, 61*(1), 86-94. http://dx.doi.org/10.1145/3158416
- Meyer, H. (2016). Untangling the building blocks: a generic model to explain information behaviour to novice researchers. *Information Research, 21*(4), paper isic1602. http://InformationR.net/ir/21-4/isic/isic1602.html (Archived by WebCite® at http://www.webcitation.org/6mHhVYy54)
- Mitnic, K.D. & Simon, W.L. (2003). *The art of deception: controlling the human element of security*. John Wiley & Sons.
- Mossburg, W. & Katyal, V. (2019). *Value-based data risk management: what data is worth fighting for?* https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-risk-value-based-data-risk-management.pdf (Archived by the Internet Archive at /web/20200804022048/https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-risk-value-based-data-risk-management.pdf)
- Nahl, D. (2001). A conceptual framework for explaining information behaviour, *SIMILE: Studies in Media & Information Literacy Education, 1*(2), article 6. http://dx.doi.org/10.3138/sim.1.2.001 (http://www2.hawaii.edu/~donnab/lis610/nahl_2001.html)
- Nasir, Q. & Al-Mousa, A. (2013). Honeypots aiding network forensics: challenges and notions. *Journal of Communications, 8*(11), 700-707. http://dx.doi.org/10.12720/jcm.8.11.700-707
- Nguyen, T.L. (2018). A framework for five big v's of big data and organizational culture in firms. In *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data)* (pp. 5411-5413). IEEE. http://dx.doi.org/10.1109/BigData.2018.8622377
- Nicholas, D., Dobrowolski, T., Withey, R., Russell, C., Huntington, P. & Williams, P. (2003). Digital information consumers, players and purchasers: information seeking behaviour in the new digital interactive environment. *Aslib Proceedings, 55*(1/2), 23-31. http://dx.doi.org/10.1108/00012530310462689
- Nicol, D.M. & Mallapura, V. (2014). Modeling and analysis of stepping stone attack. In *Proceedings of the 2014 Winter Simulation Conference* (pp. 3036-3047). IEEE. http://dx.doi.org/10.1109/WSC.2014.7020142
- Pawlick, J., Colbert, E. & Zhu, Q. (2019). A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy. *ACM Computing Surveys, 52*(4), 82. http://dx.doi.org/10.1145/3337772
- Popli, N.K. & Girdhar, A. (2019). Behavioural analysis of recent ransomware and prediction of future attacks by polymorphic and metamorphic ransomware. In N. Verwa & A. Ghosh (Eds.), *Computational intelligence: theories, applications and future direction - Volume II.* (Part of the Advances in Intelligent Systems and Computing book series, volume 799) (pp. 65-80). Springer. http://dx.doi.org/10.1007/978-981-13-1135-2_6
- Rashidi, B., Fung, C., Hamlen, K.W. & Kamisinski, A. (2018). HoneyV: a virtualized honeynet system based on network softwarization. In *NOMS 2018 – 2018 IEEE/IFIP Network Operations and Management Symposium* (pp. 1-5).http://dx.doi.org/10.1109/NOMS.2018.8406205
- Robson, A. & Robinson, L. (2013). Building on models of information behaviour: linking information seeking and communication, *Journal of Documentation,* 69(2), 169-193.https://doi.org/10.1108/00220411311300039
- Salem, A., Zhang, Y., Humbert, M., Berrang, P., Fritz, M. & Backes, M. (2019). ML-leaks: model and data independent membership inference attacks and defences on machine learning models. In *Proceedings of 2019 Network and Distributed System Security Symposium.*http://dx.doi.org/10.14722/ndss.2019.23119
- Shaikhedris, S.S.A. (2018). *Collection and analysis of attacker data using honeynet*. SUST Repository. (Sudan University of Science and Technology MSc Thesis) http://repository.sustech.edu/handle/123456789/22738 (Archived by the Internet Archive at /web/20200804022308/http://repository.sustech.edu/handle/123456789/22738)
- Shi, Z.R., Procaccia, A.D., Chan, K.S., Venkatesan, S., Ben-Asher, N., Leslie, N.O., Kamhoua, C. & Fang, F. (2019). Learning and planning in feature detection games. *org*. https://arxiv.org/abs/1905.04833 (Archived by the Internet Archive at /web/20200804022419/https://arxiv.org/abs/1905.04833)
- Silbey, J. & Hartzog, W. (2019). "The upside of deep fakes". *HeinOnline, Maryland Land Review, 78*(4), 960-966.
- Sokol, P., Misek, J. & Husak, M. (2017). Honeypots and honeynets: issues of privacy. *EURASIP Journal on Information Security*, *2017*(1), article 4. http://dx.doi.org/10.1186/s13635-017-0057-4
- Sophos (2019). Sophos labs 2019 threat report. Sophos. https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/sophoslabs-2019-threat-report.pdf. (Archived by the Internet Archive at /web/20200804022545/https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/sophoslabs-2019-threat-report.pdf)
- Stoll, C. (1987). What do you feed a Trojan Horse?, In *Proceedings of the 10th National Computer Security Conference, Baltimore, USA, 21 -24 September 1987.*
- Stoll, C. (1988). Stalking the wily hacker. *Communications of the ACM, 31*(5), 484-497. https://doi.org/10.1145/42411.42412
- Stoll, C. (1989). *The cuckoo's egg*. http://bayrampasamakina.com/tr/pdf_stoll_4_1.pdf (Archived by the Internet Archive at /web/20200804022644/http://bayrampasamakina.com/tr/pdf_stoll_4_1.pdf)

- Su, Q. (2017). A multi-dimensional analysis of deception. In *Proceedings of the 2017 International Conference on Asian Language Processing (IALP)* (pp. 160-163). IEEE. http://dx.doi.org/10.1109/IALP.2017.8300569
- Symantec. (2017). Living off the land and fileless attack techniques. *Symantec, Internet Security Threat Report (ISTR) (An ISTR Special Report)*. https://docs.broadcom.com/doc/istr-living-off-the-land-and-fileless-attack-techniques-en. (Archived by the Internet Archive at /web/20200804023325/https://docs.broadcom.com/doc/istr-living-off-the-land-and-fileless-attack-techniques-en)
- Tsikerdekis, M., Morse, T., Dean, C. & Ruffin, J. (2019). A taxonomy of features for preventing identity deception in online communities and their estimated efficacy. *Journal of Information Security and Applications, 47*(2019), 363-370. https://doi.org/10.1016/j.jisa.2019.06.002
- van der Walt, E., Eloff, J.H.P. & Grobler, J. (2018). Cyber-security: identity deception detection on social media platforms. *Computers & Security, 78*, 76-89. https://doi.org/10.1016/j.cose.2018.05.015
- van der Walt, E. & Eloff, J.H.P. (2015). A big data science experiment - identity deception detection. In *Proceedings of the2015 International Conference on Computational Science and Computational Intelligence* (pp. 416-419). http://dx.doi.org/10.1109/CSCI.2015.169
- Walker, R. (2019). *Combating strategic weapons of influence on social media*. Naval Postgraduate School. (Naval Postgraduate School, Monterey, California Thesis) https://apps.dtic.mil/dtic/tr/fulltext/u2/1080481.pdf. (Archived by the Internet Archive at /web/20200804023530/https://apps.dtic.mil/dtic/tr/fulltext/u2/1080481.pdf)
- Wallace, D. & Visger, M. (2018). The use of weaponized “Honeypots” under the customary international law of state responsibility. *The Cyber Defence Review, 3*(2), 33-42.
- Warchal, A. (2018). Dimension of security system complexity and risk detection praxis. In I. Kabashkin, Yatskiv (Jackiva) & O. Prentkovskis (Eds.), *Reliability and statistics in transportation and communication* (Lecture Notes in Network and Systems, vol. 68) (pp. 479-488). Springer. http://dx.doi.org/10.1007/978-3-030-12450-2_46
- Whitham, B. (2017). Automating the generation of enticing text content for high-interaction honeyfiles. In *Proceedings of the 50th Hawaii International Conference on System Science* (pp. 6069-6078). ScholarSpace. http://dx.doi.org/10.24251/HICSS.2017.733
- Wilson, T.D. (1999). Models in information behaviour research. *Journal of Documentation, 55*(3), 249-270. http://dx.doi.org/10.1108/EUM0000000007145
- Wilson, T.D. (2020). Book review of Buchanan, Ben: The hacker and the state. Cyber attacks the new normal of geopolitics. *Information Research, 25*(1), revs 680. http://www.informationr.net/ir/reviews/revs680.html (Archived by the Internet Archive at https://web.archive.org/web/20200713113825/http://informationr.net/ir/reviews/revs680.html)
- Wilson, T.D. & Macevičiūtė, E. (2013). What's newsworthy about 'information seeking'? An analysis of Google's News Alerts. *Information Research, 18*(1), paper 557. http://InformationR.net/ir/18-1/paper557.html (Archived by the Internet Archive at https://web.archive.org/web/20200728170912/http://informationr.net/ir//18-1/paper557.html#.XyKNEkl7nIU)
- Woodworth, M., Hancock, J. & Goorha, S. (2005). The motivational enhancement effect: implications for our chosen modes of communication in the 21st In *Proceedings of the 38th Annual Hawaii International Conference on System Science* (p. 22a). IEEE. http://dx.doi.org/10.1109/HICSS.2005.607
- Yang, J., Zhang, Y., King, R. & Tolbert, T. (2018). Sniffing and chaffing network traffic in stepping-stone intrusion detection. In *Proceedings of the32nd International Conference on Advanced Networking and Applications Workshops (WAINA 2018)* (pp. 515-520). IEEE. http://dx.doi.org/10.1109/WAINA.2018.00137
- Yang, J., Zhang, Y. & Zhao, G. (2017). Integrate stepping-stone intrusion detection technique into cybersecurity curriculum. In *Proceedings of the 31st International Conference on Advanced Information Networking and Applications Workshops* (pp. 1-6). IEEE. http://dx.doi.org/10.1109/WAINA.2017.29
- Younas, M. (2019). Research challenges of big data. *Service Oriented Computing and Applications*, *13*(2), 105-107. https://doi.org/10.1007/s11761-019-00265-x
- Yuill, J., Zappe, M., Denning, D. & Feer, F. (2004). Honeyfiles: deceptive files for intrusion detection. *Proceedings from the 5th Annual IEEE SMC Information Assurance Workshop* (pp. 116-122). IEEE. http://dx.doi.org/10.1109/IAW.2004.1437806
- Zakharov, W., Li, H. & Fosmire, M. (2019). Undergraduates' news consumption and perceptions of fake news in science. *Portal: Libraries and the Academy, 19*(4), 653-665. http://dx.doi.org/10.1353/pla.2019.0040
- Zhang, X. & Ghorbani, A.A. (2020). An overview of online fake news: characterization, detection, and discussion. *Information Processing & Management, 57*(2), 1. http://dx.doi.org/10.1016/j.ipm.2019.03.004

How to cite this paper

Dlamini, M.T., Venter, H.S., Eloff, J.H.P., & Eloff, M.M. (2020). Digital deception in cybersecurity: an information behaviour lens. In *Proceedings of ISIC, the Information Behaviour Conference, Pretoria, South Africa, 28 September - 1 October, 2020. Information Research, 25*(4), paper isic2018. Retrieved from http://InformationR.net/ir/25-4/isic2020/isic2018.html (Archived by the Internet Archive at https://bit.ly/3qRXNIr) https://doi.org/10.47989/irisic2018