

# ‘It’s Not Paranoia If They’re Really After You’<sup>1</sup>: When Announcing Deception Technology Can Change Attacker Decisions

Andrew Reeves  
Defence and Security Institute  
University of Adelaide  
[andrew.reeves@adelaide.edu.au](mailto:andrew.reeves@adelaide.edu.au)

Debi Ashenden  
Institute for Cybersecurity (IFCyber)  
University of New South Wales  
[d.ashenden@unsw.edu.au](mailto:d.ashenden@unsw.edu.au)

## Abstract

*As organisations continue to adopt deception technology, adversaries are becoming aware of this technology. Little is known, however, about how this awareness changes the attacker’s behaviour as they navigate a victim’s network. Concurrently, work is being done to build algorithms that predict attacker paths to recommend where to place deceptive assets, but it is not clear whether attacker awareness of deception alters their behaviour sufficiently to render these algorithms ineffective. We present an ongoing mixed method study to better understand how attackers move through a network when they are aware of the presence of deception. Thematic analysis of think-aloud sessions revealed three key decision-making themes. Themes suggest that several industry heuristics for the use of decoys may be inaccurate and impact the efficacy of decoy placement strategies. In addition, effect sizes indicate that awareness of deception leads attackers to take longer paths through the network, although no more decoys were required to detect them.*

**Keywords:** cyber deception, adversary engagement, active defence, human factors, mixed methods research.

## 1. Introduction

The rapidly evolving cyber landscape necessitates a paradigm shift from traditional perimeter defence towards more sophisticated engagement with adversaries. It is important to distinguish deceptive cyber operations from other types of cyber defence. A recent review by Mohan et al. (2022) distinguished cyber deception from other cyber security measures saying:

“traditional security measures are employed in response to the actions of an attacker [...] whereas deception-based measures are used in anticipation of such actions, manipulating attackers’ perceptions and thus inducing adversaries” (Mohan et al., 2022, p.2).

While theoretically plausible, the utility of cyber deception to successfully shape attacker sensemaking and behaviour remains largely unobserved. Indeed, it is plausible that the mere presence of deception technology within networks, and the attackers’ suspicion of this, will be sufficient to induce behaviour change, however further evidence is needed to establish this with any confidence. These altered attacker behaviours may be valuable in themselves (for example, if they take more time or more resources to achieve) or they may offer exploitable opportunities for defensive purposes (for example revealing tactics, techniques and procedures (TTPs)). Furthermore, while researchers and industry vendors continue to advance the art of building convincing deceptive assets (e.g., fake profiles, documents, and simulated network traffic; Nisrine, 2016) less attention is directed to *where* on a network these deceptive assets should be placed to maximise the effect. Vendors of packaged deception products often resort to heuristics for recommending placement of their deceptive assets (Reeves & Ashenden, 2023). A faux user persona (e.g., Virvilis et al., 2014) may be convincing, but if not deployed adeptly, a prospective attacker may never cross its path. This becomes particularly apparent when considering the broad diversity of technology within, and across, organisational networks and more so when the attacker is vigilant to signs of deception.

While research on cyber deception continues to progress, albeit at a moderate rate (Zhang & Thing, 2021), there are signs that attackers are becoming increasingly vigilant and adept at evading honeypots and other traps (Lu et al., 2020). Therefore, vendors will soon need to ensure that their products are equally able to deceive an attacker who is aware of the presence of their technology as one that is not.

Part of the difficulty in addressing this problem is the lack of existing empirical knowledge regarding attacker decision making, which makes it challenging to predict how attacker paths will change by the presence of deception technology. While game-theory based approaches have helped to elucidate rational decision strategies and to identify interdependencies

<sup>1</sup>quote taken from the 1998 movie *Enemy of the State*.

between attacker and defender (e.g., Do et al., 2017) such approaches can lack ecological validity (Ho et al., 2022). Naturalistic decision making studies on the other hand emphasise ecological validity and tend to differ markedly from rational choice models (Klein, 2008). We suggest that research is needed that examines the efficacy of decoy placement algorithms that use game theory in comparison to real-world naturalistic decision making when attackers are aware of deception. To respond to this challenge, we developed a mixed-methods experiment.

The following section reviews a selection of literature that is relevant to our research questions and builds the case for a think-aloud experiment to observe the effect of decoys on attacker movement and to explore the ability of optimisation models to place decoys within a network.

## 2. Literature Review

Deception is increasingly leveraged as a tool for network defenders to detect and disrupt attackers (Underbrink, 2016). In the simplest form, decoys, such as honeypots or honeytokens, can serve as tripwires that alert to the presence of an active threat, distract an attacker and cause them to waste resources, and allow time for network defenders to learn their TTPs. Despite the intuitive utility of tripwires within networks however, industry uptake remains limited, partly due to an industry culture that prioritises perimeter management, and also because of a lack of knowledge of how best to place decoys for maximum effect (Reeves & Ashenden, 2023).

A properly developed decoy placement strategy needs to consider the entire set of nodes within a network (including users, computers and security groups for example) as well as the dynamic nature of the connections between those nodes (such as temporary user sessions) (Durkota et al., 2019). The complexity and dynamic nature of networks means that knowing where to place decoys is a challenging proposition (Milani et al., 2020). Some researchers suggest that the challenge can best be addressed using optimisation algorithms that decide where is best to place decoys (Ngo et al., 2023) rather than human decision making. In a real-world context, however, such an algorithm needs to be informed by an understanding of attacker decision making about likely paths through a network (lateral movement for example).

The theoretical promise of algorithmic models for cyber security is increasingly well documented (Apruzzese et al., 2023; Dasgupta et al., 2022; Fraley & Cannady, 2017; Handa et al., 2019), however, there is a conspicuous absence of empirical evidence regarding their effectiveness in real-world scenarios, particularly in the context of decoy placement. Some

authors have begun to address this gap. For example, Ferguson-Walter, Major, et al. (2021) highlight the necessity of human participant studies to ascertain the practical effectiveness of deception techniques and report promising initial findings from a red/blue team simulation exercise. Their Tularosa study supports the intuitive view that cyber attackers are susceptible to cognitive biases in their decision making and that these biases present strategic opportunities for defenders. Specifically, Ferguson-Walter (2024) and Johnson et al. (2021) observed instances of attacker's experiencing confirmation bias, a tendency to favour information that confirms one's existing beliefs while ignoring or dismissing evidence that contradicts them. When aware of deception on the network, attackers' confirmation bias led them to assume unexpected assets were deceptive assets. Consequently, there are signs that announcing the presence of deception on the network may change the path chosen by attackers, and this will be the focus of the current paper.

One reason for the relatively sparse research attention is the difficulty in direct observation of cyber attacker behaviour for research purposes, leading researchers to leverage offensive cyber professionals in simulated exercises as proxies for attackers. For example, Johnson et al. (2021) used red-teamers as a proxy for attackers to identify specific decision-making biases that influence attacker behaviour. The absence of empirical evidence regarding the practical effectiveness of the strategic placement of decoys is a significant research gap and this paper presents an important step in addressing this gap.

## 3. The Current Study

Multiple opinion papers argue that the rapid advancement of optimisation algorithms offers new possibilities for cyber defenders to arm themselves with AI/ML-assisted tools (AlShaikh et al., 2024; Elsadig, 2024; Sarker et al., 2023). However, fewer papers provide empirical demonstrations of the usefulness of such assistance. Recent advancements by Ngo et al. (2023) however have produced algorithms which recommend how to place deception technology within an organisations' existing network to maximise the benefits. The study presented here uses professional populations (red-teamers, pentesters and others working in offensive cyber roles) to examine the efficacy of these models, and to test the ability of decoys placed in this manner to effectively detect the presence of the attacker and to disrupt attacker sensemaking. In an ideal world, our participant population would have been network attackers but due to the difficulties of recruiting such participants we have used a range of offensive cybersecurity professionals. We recognise the limitations of this, but believe that we have still

increased the ecological validity of the study quite significantly from a game theoretic approach.

We adopted a mixed-methods research approach using a qualitative think-aloud technique along with quantitative, experiment-based observation and measurement. We adopted this research design to maximise the depth of the data gathered from each participant, given the predicted difficulty in recruiting individuals with the required skillset. To adequately address our identified research gap, we carefully selected only participants who have practical experience in offensive cyber operations in the form of vulnerability exploitation and lateral movement. We targeted this group following case studies and studies based on game theory which argue that actions performed by attackers in the lateral movement phase are best placed to be detected by deception assets (including decoys) over infiltration or exfiltration (Basak et al., 2019). In addition, it is increasingly being reported that attackers are leveraging the Active Directory (AD) enumeration tool known as Bloodhound (Robbins, 2023) to plan their lateral movements, and yet research attention on this tool is scant. Therefore, we recruited solely individuals with niche experience using Bloodhound for lateral movement and vulnerability exploitation activities.

Explorative Research Questions:

- 1) How does the presence of decoys qualitatively affect attacker sensemaking from their own perspective?
- 2) How many decoys are required in a medium sized AD network (total nodes=972) to detect all attacker paths in our sample, when placed by an optimization algorithm? Does this change when they are aware of the decoys?

Hypotheses:

- 1) Paths chosen by attackers after the reveal of decoys will be longer (i.e., more nodes & edges) than paths chosen before the reveal of decoys.
- 2) A greater number of decoys will be required to detect attacker movement after the reveal than before the reveal of decoys.

## 4. Method

### 4.1 Participants

Participation required previous knowledge of offensive cyber operations and experience with Bloodhound. This niche skillset required the researchers to adopt an expansive approach to advertise the study. Ten individuals participated in the study during November 2023-May 2024. The ten consisted of six cybersecurity consultants (predominantly offensive security, red team, or threat-hunter roles), two dedicated internal red-teamers, one

product tester for offensive cyber technology, and one defence cybersecurity researcher. No reimbursement was offered although participants were welcomed to request access to the final paper and findings. All ten participants identified as men, and their age ranged from 25 to 55. The study was approved by the University of Adelaide Human Research Ethics Committee (HREC; H-2023-215).

### 4.2 Materials

**4.2.1 Active Directory and Bloodhound** Active Directory (AD) is a directory service by Microsoft for identity and access management. Due to being deployed at most enterprises, AD systems are a major target for attackers. In 2021, Microsoft reported 25.6 billion brute force attacks on their AD accounts (Weston, 2022). In these attacks, the attacker first builds an attack graph of the targeted AD system that shows network connections and vulnerabilities. Tools for generating the AD attack graphs are widely available, and tools including Bloodhound developed by Robbins (2023) are frequently used. After the attacker has gained initial entry into the victim network, Bloodhound can be used to scan Active Directory for user accounts, computers, security groups, and other points of interest. Bloodhound provides visual representations of available accesses and exploits that the attacker can use to move within the network. It produces a graphical view of the AD network which includes nodes for users, machines, and access groups. Attackers, pentesters, and red-teamers use bloodhound to plan their lateral movement. In this way, attackers can use Bloodhound to escalate themselves from their likely low-privilege initial entry point to a desired higher privilege target. An example Bloodhound map is provided in Figure 1.

**4.2.2 Optimal Deployment Strategy Algorithm (ODSA)** Ngo et al. (2023) developed an Optimal Deployment Strategy Algorithm (ODSA) using a Stackelberg game on a directed AD attack graph. Stackelberg games are born from game theory and are used in a security context to involve an ‘attacker’ and a ‘defender’ (Brückner & Scheffer, 2011). In the approach by Ngo et al. (2023), the attacker tries to reach a destination node called Domain Admin via shortest paths only. The defender’s task is to choose nodes in which to allocate honeypots to intercept as many of the attacker’s shortest attack paths as possible. The attacker cannot differentiate a normal node from a honeypot. Furthermore, if the attacker stumbles into a honeypot, the attack campaign fails. This method developed an algorithm that approximated an optimal placement solution. We used this algorithm to place decoys within our simulated

network. Further details on the development of the algorithm are available in Ngo et al. (2023).

#### 4.2.3 Questionnaire

Before commencing the experiment, participants completed an online questionnaire. Measures included a series of standardised measures of constructs of relevance, including decision making preferences, thinking styles, and risk taking propensity. Note this paper does not report on the results of these metrics, which will be published in future work.

#### 4.3 Procedure

The study consisted of two phases: 1) the familiarisation phase, and 2) the think-aloud/experiment phase.

In the familiarisation phase, which occurred one week before the date organised for the experimentation phase, the participants were sent the required file set to allow them to review the Bloodhound map of the simulated network. They were provided with a predefined target to reach (the Domain Admin group) and 10 already compromised nodes (9 users and admin rights to 1 computer) to act as initial entry points. They were asked to role play an attacker and imagine how they would attempt to reach the target by exploiting vulnerabilities.

In the think-aloud phase, participants were asked to explain their chosen path to the researcher. We asked participants to talk through their chosen path in a ‘think aloud’ manner. Participants vocalised their internal dialogue, including thoughts, feelings, and decisions (Van Someren et al., 1994). Think aloud designs have shown utility in cyber security research including to trace analyst decision making following a critical cyber security event (Zhong et al., 2015) and identifying misconceptions in cybersecurity student populations (Thompson et al., 2018). This usually took between 20 to 30 minutes. After the participants had explained their chosen path, they were informed of the presence of decoys within the network with the following script:

*Within the Bloodhound AD map, there are certain nodes that have been set up as decoys. If you access any of these decoy nodes, it will trigger an alert that will be sent to the Security Operations Center (SOC). The objective remains the same: to reach the Domain Admin target. Please take some time to think and plan a path through the network. This may be the same or different to your previously identified paths.*

The think-aloud study was conducted online at a time organised with each participant. Each session was an average of 48 minutes (min: 39, max: 58). The sessions were recorded and transcribed.

After the session ended, the researcher compared each participant’s path to the locations of decoys

placed in the network by the ODSA. Paths that overlapped with ODSA decoys were marked as a ‘hit’ while paths that did not overlap with ODSA decoys were marked as a ‘miss’. Note that this means that none of the nodes were truly decoys at the time of the think-aloud session. This was deliberate to examine the effect of decoy cognisance on attacker behaviour regardless of true deployment. This is similar to the absent-informed condition in Ferguson-Walter (2024).

#### 4.4 Data Analysis

This section details our approach to analysing the qualitative data after the reveal of decoys on the network. The objective of the data analysis was to explore and understand the thematic structure within the transcription data. To achieve this, we employed an innovative approach by integrating OpenAI's ChatGPT 4 with prompt engineering techniques, similar to Zhang et al. (2023).

Prompt engineering is crucial in leveraging the capabilities of language models like ChatGPT (Zhang, et al., 2023). The prompts were structured to serve two primary functions: first, to guide the model to identify and categorise themes within the data, and second, to encourage the generation of insights about the relationships and hierarchies between these themes. The prompt engineering process involved iterative refinement, ensuring that each prompt was clear, contextually relevant, and capable of steering the model towards a meaningful thematic analysis. The prompts used were:

Prompt 1). Read this transcription of an interview to learn if the presence of decoys or deception inside a network influences attacker decision making and behaviour. Focusing on the comments from [ParticipantID] note the recurring topics, cluster into broad themes and present each theme with representative quotes from the data and a brief summary.

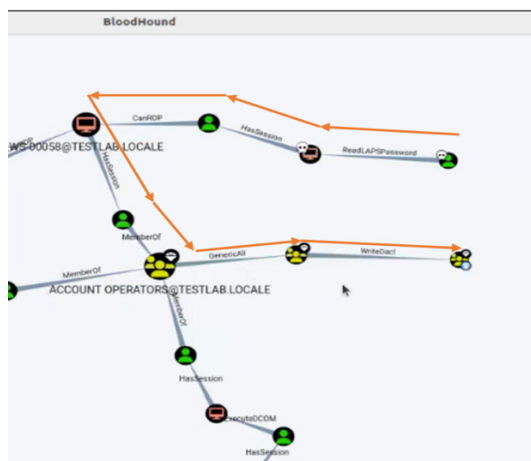
Prompt 2). The following text is a series of themes from a qualitative interview. Help me merge the themes into no more than 10 key themes retaining the quotes and the summary.

Prompt 3). Role play a qualitative researcher, compare and contrast these themes taken from qualitative interviews. Summarise the themes highlighting the similarities and differences of opinion in each theme.

With the prompts engineered, ChatGPT was engaged to conduct the thematic analysis. The model was presented with the prepared prompts and the corresponding data segments. The responses from ChatGPT were recorded and compiled. Each identified theme was documented along with its definition, examples from the data illustrating the theme, and ChatGPT's interpretation of the theme's significance and its connections to other themes.

The final phase involved a thorough review and interpretation of the themes generated by ChatGPT. The research team analysed the relevance of each

identified theme, ensuring that the themes accurately represented the underlying dataset. This process uncovered that Prompt 3 had not been useful. Therefore, the researcher performed this step and manually compared and contrasted the themes generated by Prompt 2. The researcher identified similarities in ChatGPT’s coding of the data and merged similar themes, leading to a final set of three overarching themes. The researcher compared these themes to the original transcription dataset to ensure they presented a sufficient and useful summation of the data.



**Figure 1. A bloodhound map and preferred path (arrows) from initial access nodes (skull symbol) to the Domain Admin target group as provided by Participant 4 pre reveal.**

## 5. Results

Each participant identified at least one path from the initial access nodes to the target. Four participants expressed that they were unable to provide a specific path once the presence of decoys was revealed. This resulted in a usable dataset of 12 attack paths in 6 pre-post pairs. Figure 1 presents an example map produced in Bloodhound by Participant 4 and indicates their preferred path pre reveal.

### 5.1 Qualitative Analysis of the Think-Aloud Study

Thematic analysis of the transcripts after the reveal of decoys produced three key themes.

**Theme 1. Affective and Cognitive Impact of Decoys on Attackers** Following the reveal of the presence of decoys in the network, we observed a considerable affective response from participants. Awareness of the decoys appeared sufficient to induce anxiety and to disrupt their mission planning with one participant stating “honeytokens scare the \*\*\*\* out of me” (Participant 3) and another admitting that “Everything

I normally do is likely to trip up canaries [honeytokens]” (Participant 4)

This anxiety had strategically relevant impacts on attackers by forcing them to be more orchestrated and cautious in their actions.

*“You’d want to be much more deliberate about your persistence. [] Because now anything you try has the risk of getting caught, even more so than usual, so you’re going to expect that you might get kicked out of the network and you’ll need another path back in.”* (Participant 2)

For the right attacker, the overt presence of decoys may be sufficient to induce alternative attack approaches or evasion tactics because *‘it would completely break all of my workflows* (Participant 3). Accordingly, the psychological impact and operational challenges that decoys create for attackers may divert them away from their initial targets, as one participant suggested *‘I might give up and find a softer target’*. (Participant 3)

After the presence of decoys was revealed to participants, some expressed a notable reversal of their affectivity towards certain stimuli. For example, before the reveal, one participant was excited to notice the presence of a privileged account within reach.

*“There’s a session [on this machine] with another user, called ‘TO DEFAULT ADMIN USER’, which sounds quite promising!”* (Participant 1)

However, when later in the session the researcher revealed the presence of decoys, the participant began to distrust the apparent value of this same node.

*“If I’m thinking about decoys, I would imagine that something that says ‘DEFAULT ADMIN’ might be awfully tempting to a pen-tester. So, I might be tempted to avoid it.”* (Participant 1)

Consequently, the presence of decoys may act both as a lure and a deterrent depending on the attacker’s cognisance of deception technology. This may be particularly effective if the decoys are more obvious in nature (i.e., *“If it looks too good to be true”*, Participant 4).

One participant underscored this concurrent ability of decoys to both lure and deter:

*“If something looks really juicy, like something really dumb like a domain admin account that has been logged into a computer that every user in the domain has admin on – that’s attractive – and it’s a clear misconfiguration, and that does happen [in the real*

world]. But, if you see that and you think that there might be canaries, you might immediately think 'well, I am not going to touch that'''. (Participant 3)

**Theme 2. Reconnaissance and Realism in Deception** Participants raised the concept of the realism of decoys and other deceptive assets deployed as part of active defence. Close attention to asset details may reveal the presence of deception.

*"We don't know what might be a decoy [in this experiment], but in a real environment there always might be something. The user account in AD might look different, or the description might be a little bit too well written"* (Participant 5)

This greater attentiveness likely involves a time commitment on the part of the attacker, including added time to validate user profiles via external means (e.g., social media). The concept of Open Source Intelligence (OSINT) wasn't introduced by the researcher, but participants commonly mentioned it regardless. LinkedIn was frequently mentioned due to its relevance both for confirming identifies and investigating job advertisements.

*"Looking at LinkedIn profiles for SysAdmins is something that a lot of red-teamers do before they even send their first phishing email. [] If you see a user account who looks like they're a nice path to domain admin, but that user doesn't exist on LinkedIn, that's an alarm bell"* (Participant 3).

Despite this, creating fake social media profiles to misdirect attacker reconnaissance was considered unrealistic.

*"I wouldn't necessarily go to the point of creating fake LinkedIn profiles for admins, That seems a bit too involved."* (Participant 3)

Consequently, attacker reconnaissance through social media and other methods may remain an ongoing limitation and vulnerability to the use of fake user accounts as decoys. In the words of one participant, *"It's just one of the limitations [of decoys]. It's a hard problem."* (Participant 4). That is not to say, however, that attacker reconnaissance is solely a negative. This phase of attacker activity can also act as a vector for influence by revealing the presence of deception technology within the victim's network, either unwittingly or intentionally.

*"A lot of organisations put out a job on LinkedIn or SEEK, for a security engineer for example, and they'll usually say what tools they use or what tools they are*

*expected to be familiar with. So, they may say, 'oh yes, we use CrowdStrike'"* (Participant 3)

**Theme 3. Factors of the Attacker Perspective** Participants discussed the situational and contextual variables that may influence how they approach the task. For example, an attacker that has a strong need to target a particular organisation may take a more cautious approach to avoiding tripwires than an attacker who is content to move on to another target, if required.

*"Assuming you're an attacker that really wants this organisation, you're just going to have to be slower. More methodical."* (Participant 1)

Presuming this high level of motivation to avoid detection, attackers may take a proactive approach to mitigating the threat posed by decoys.

*"If you're really sophisticated, you might even go to the point of buying the products that they are using for yourself and see exactly what the canaries look like"* (Participant 3).

This is a time-consuming approach, however, that may not be considered necessary to some attackers. One participant highlighted how attackers make decisions based on the perceived maturity of a network, and for many networks attackers may disregard the possibility of decoys.

*"They're probably going to just barrel forward on the assumption that canaries don't exist, because they're a pain and it's too hard to get around"*. (Participant 3)

This assumption that decoys don't exist in many organisational networks may be bolstered by a parallel perception that decoys lack appeal to organisations and are rare. Relatedly, some attackers may hold the belief that deception tools offer a poor return on investment and can be burdensome for administrators, leading to low adoption by their victims.

*"A lot of deception technology is really poor return on investment. It makes life miserable for your admins and it doesn't really trip up your attackers that much. The juice isn't worth the squeeze"*. (Participant 1)

Another salient factor of the attackers' perspective is the apparent the sophistication of the cyber security team within the victim organisation. This may be informed by the perceived maturity of the network itself and the broader organisation, with many attackers likely having a limited view of the organisation's ability to defend itself.

“A lot of the SOC analysts’ job is just ‘ctrl-A, delete’ on all of the alerts they get every day, because I used to. For all I knew I was muting real positives because there’s so many alerts that you can’t possibly look at them all” (Participant 4)

Suspecting this to be the case, attackers may be unconcerned regarding the presence of tripwire decoys.

“I always think that if the alert goes to the SOC, I’m done for. But that isn’t necessarily the case, because they might just mark it as a false positive, and that has absolutely happened in many real-world breaches” (Participant 5)

Finally, the attackers’ unique goals and context were highlighted as important components to their decision making.

*If you're an attacker, you're probably going to make sure you tunnel in and get persistence and make sure they can't kick you out. But if you're a pen tester, then you're not going to do that. You're going to do what you need to do to get the rest of your objectives and then write up as a report.* (Participant 3)

## 5.2 Quantitative Hypotheses

While we have gathered a sizeable qualitative data set for this study we recognise that we need a larger sample size to carry out a full quantitative analysis. Consequently this study is ongoing. The quantitative data consists of the length of the chosen path (in number of nodes) pre and post the reveal of deception. An initial examination of the six data points (i.e., only those participants that provided both a pre and post-reveal path) produced promising trends that warrant further discussion in light of the qualitative findings. Assumption checking indicated that it did not meet the assumptions of normality nor equivalence of variance required for a dependent samples t-test. Therefore, a Wilcoxon signed-rank test was conducted to explore the first two hypotheses that (1) the length of attacker paths would increase after the reveal of decoys on the network, and that (2) a greater number of decoys would be required to detect attacker movement post reveal. Given the small  $N$  at time of writing, we interpret the results considering both the  $p$  value and the observed effect size. The effect size at a small  $N$  should be interpreted with caution and does not necessarily indicate the effect size at greater sample sizes (Zhang et al., 2019). Effect sizes (rank-biserial correlations) were calculated using the formula  $r = \frac{Z}{\sqrt{N}}$ , where  $N$  is the number of valid pairs and  $Z$  is the Wilcoxon test statistic. The mean path length before the reveal was 6.7 nodes while the mean path length

after the reveal was 7.7 nodes. Table 1 presents the results of this analysis.

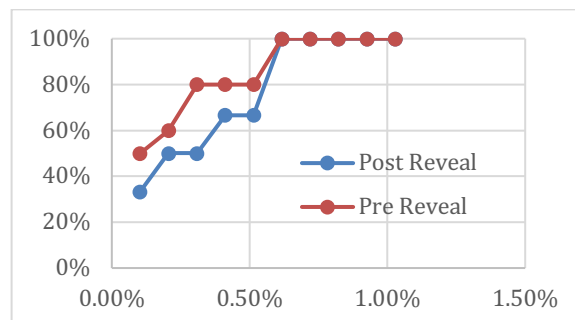
**Table 1. Wilcoxon signed-rank test of mean difference in path length and number of decoys required pre and post the reveal.**

Statistic	Path Length Post-Pre ( $N=6$ )	No. Decoys Required Post-Pre
Z	-1.60	-0.18
$p$	.109	.854
$r$ (ES)	-0.65	-.08
ES	Large	Small
interpretation <sup>^</sup>		

<sup>^</sup>small:0.1, medium:0.3, large:0.5, (Cohen, 1998)

While these results did not meet the usual alpha level required for significance, we believe they warrant further discussion and exploration for two reasons. Firstly, the usable data set size was limited by the inability of some participants to provide a second path. This meant that from an original sample of 10, only 6 participants produced a valid pre-post pair of paths. Given such a small  $N$  and the observed large effect size for path length, we believe further attention is appropriate. Secondly, given the context of our research questions, we believe that the inability of 4 participants to provide a secondary path indicates a level of decision paralysis or other such effect on sense making that is relevant to our aims.

Additionally, the mean number of decoys required to detect all attacker paths was 6 for both before and after the reveal. Interestingly, while the number of decoys required to detect attacker paths did not statistically significantly change post the reveal of decoys, Figure 2 suggests that this relationship may require a deeper examination.



Note: the network size of 972 nodes means that deploying 1 decoy is approximately .1% of the network.

**Figure 2. Percentage of attack paths detected (Y) by the count of decoys as a percentage of network size (X).**

As shown in Figure 2, the same number of decoys detected fewer post-reveal paths where the number of decoys is 1-5 (0.1-0.5% of the network) and

overlapped at a 6 (.6% of the network). It may be that a smaller deployment of decoys will be less able to detect post-reveal paths than pre-reveal paths, although we cannot inferentially test this at our current sample size. A future paper will further explore this hypothesis utilising a larger sample.

## 6. Discussion

The results of the think-aloud sessions with a cohort of offensive cyber professionals uncovered three key themes. The three themes help to explain the complex ways that decoys can have both affective and cognitive implications for attackers, how the presence of decoys can be unveiled through in-network and out-of-network means, and how the influence of decoys on attackers will depend on the attackers' context and perspective. These findings have considerable implications for extant cyber security theory and practice.

Firstly, there is a promising opportunity for the adoption of deception technology itself to be an influence campaign. Our findings support the claim by Roy et al. (2022) that sophisticated adversaries are likely to perform a level of open source intelligence (OSINT) gathering through which they will learn of the presence of deceptive assets in their victim's network, and the salience of this may be sufficient for a subset of attackers to seek softer targets (as was the case for 40% of our participants). The cognitive and behavioural effect of announcing the adoption of deception technology may cause the intended effect in attacker behaviour change regardless of whether any deceptive asset has yet to be deployed. Two promising avenues to manipulate the attacker's awareness via OSINT gathering are job postings for recruitment into the cyber team (For a salient example, see this LinkedIn post for Walmart by Ogden, 2024) and media releases announcing business partnerships with known deception vendors (for example, the funding announcement for CounterCraft by Buckley, 2022) that each reveal which organisations have adopted deception techniques. For the attackers that do continue with their attack, their awareness of deception may lead them to take more time, expend more resources, and forego easily exploitable vulnerabilities (even if these are true security oversights and not decoys).

Furthermore, our results align with the findings of the Tularosa study which found that cyber attackers are susceptible to cognitive biases in their decision making and that these present strategic opportunities for defenders. Specifically, Ferguson-Walter (2024) observed instances of attacker's experiencing confirmation bias, a tendency to favour information that confirms one's existing beliefs while ignoring or dismissing evidence that contradicts them. When

unaware of deception on the network, attackers' confirmation bias led them to assume all assets were legitimate. When they were made aware of deception on the network, "unusual stimuli in the network environment confirm[ed] the presence of fake assets — even when they were not fake." (p. 23) Agreeing with this, we found evidence that attackers will change their behaviour to avoid unusual stimuli even if the stimulus is, in fact, legitimate. In some instances, this may lead attackers to deliberately avoid vulnerabilities that they would otherwise exploit to great effect. We also find agreement with the Tularosa study that the salience of deception appears to add to the cognitive load of the attacker. Ferguson-Walter (2024) found that announcing deception led attackers to evaluate each interaction with the network to assess its authenticity. In our study, we found that deception-aware attackers took longer and more complex paths to the target. Consequently, our study supports the proposition by some authors that the mere announcement that deception is on the network can act as an oppositional human factor against the attacker (Ferguson-Walter, Gutzwiller, et al., 2021).

Organisations announcing that they are using deception may be building a cognitive defensive asset in the mind of the attacker whereby unaddressed critical network vulnerabilities, once alluring, become foreboding. This can provide network defenders greater time to detect and expel the attacker and provides the system administrator valuable leeway to identify and patch vulnerabilities before they are exploited.

Secondly, the findings of this study highlight several assumptions of attacker decision making, implicit in extant industry heuristics, that may be inaccurate. Firstly, there is an implicit assumption that decoys should always blend into their environment. For example, Mohan et al. (2022) conducted a narrative review of the current state of cyber deception research and concluded that the efficacy of deception as defence is based on the capacity to "exploit key biases to appear realistic" (p. 20). Similarly, many industry vendors wish to sell their technology by its ability of "imitate genuine assets" (Rapid7, 2024, p.1). However, as suggested by Theme 1, decoys that present overt indicators of active deception (such as being "*too well written*", Participant 5; or "*too good to be true*", Participant 4) may be valuable lures, deterrents, or disruptors of attacker workflows. This coheres with research by Ferguson-Walter et al. (2021) which indicates that manipulating attacker awareness of deception tactics can be an effective method of shaping behaviour. That is, overt (or semi-obvious) decoys may achieve valuable strategic outcomes if they can lure unsuspecting attackers towards, or deter sophisticated attackers from, key

assets of strategic relevance. This will likely depend on the attacker's context, including their goals and resources (Cranford et al., 2020).

Another intuitive assumption is that attackers fear detection (Aggarwal et al., 2016). This was a partial finding of this study, however, it is important to note that this study, and many like it, was limited by the need to use an offensive cyber defence cohort (specifically penetration testers and red-teamers) as representatives of cyber attackers. Individuals playing the part of an attacker may be influenced by their own experiences in a red team that is not representative of attacker experience. Reeves and Ashenden (2024) used the critical decision method to interview a series of SOC analysts and concluded that "the preconception that SOC analysts have a strong sense of where the attacker is or what their intent is, is not borne out in our data" (p. 19). Consequently, SOCs may be liable to incorrectly close true positive alerts in their efforts to cope with inordinate false positives (Chamkar et al., 2022). Attackers, if aware of this, may not be as strongly averse to activating tripwires as a red team would be.

Our preliminary quantitative results indicate that awareness of decoys can lead attackers to take longer paths through the victim's network, although further data collection is required to explore this. It appears that the optimisation algorithms produced by Ngo et al. (2023) are largely as effective at detecting attacker paths when the attacker is aware of the presence of deception technology as when they are not. However, there are indications that when fewer decoys are deployed, they may be less likely to detect aware attackers than unaware attackers.

Should these results hold true in larger samples on a broader array of network types and sizes, it follows that organisations that are unable or unwilling to adopt a decoy deployment of more than 0.5% of their total network size may need to accept an increased risk of decoy-aware attackers successfully infiltrating their networks.

## 6.1 Future work

This research raises a series of open questions that will be explored in future work. The distinction between using decoys to lure and decoys to deter (in Theme 1) indicates a need to determine the extent to which models, such as that by Ngo et al. (2023) are equally efficacious when the goal of the decoy is to either lure or to deter the attacker. Secondly, it remains to be seen how the attacker's perception of the sophistication level of the network effects the optimum placement strategy of decoys (Theme 3).

## 7. Conclusion

Attempts to operationalise cyber deception techniques are hampered by a lack of existing knowledge of

adversary attack paths. Consequently, there is an absence of strategy for how decoys are placed on a network, nor how existing placement method should be updated to capture attackers who are aware of the presence of deception. Thematic analysis of think-aloud sessions with offensive security professionals produced three key decision-making themes which detail the complex ways that decoys can have both affective and cognitive implications, how the presence of decoys can be unveiled through in-network and out-of-network means, and how the influence of decoys on attackers will depend on the attackers' context.

Our findings suggest that announcing the use of deception technology may have valuable outcomes in terms of attacker behaviour, whether any deceptive asset has yet to be deployed. Specifically, awareness of deception may cause some attackers to find a softer target, while others will take more time and forego easily exploitable vulnerabilities.

In addition, this study examined the efficacy of optimisation algorithms to place decoys that detect the presence of the attacker. Awareness of decoys may lead attackers to take longer paths through the network, although no more decoys are required to detect them.

## 8. Acknowledgement

This research was funded by the Australian Government through the Australian Research Council as part of a National Intelligence and Security Discovery Research Grant

## 9. References

- Aggarwal, P., Gonzalez, C., & Dutt, V. (2016). Cyber-security: role of deception in cyber-attack detection. *AHFE 2016 International Conference on Human Factors in Cybersecurity*, July 27-31, 2016, Florida, USA,
- AlShaikh, M., Alsemaih, W., Alamri, S., & Ramadan, Q. (2024). Using Supervised Learning to Detect Command and Control Attacks in IoT. *International Journal of Cloud Applications and Computing (IJCAC)*, 14(1), 1-19.
- Apruzzese, G., Laskov, P., Oca, E. M. d., Mallouli, W., Rapa, L. B., Grammatopoulos, A. V., & Franco, F. D. (2023). The Role of Machine Learning in Cybersecurity. *Digital Threats*, 4(1)
- Basak, A., Kamhoua, C., Venkatesan, S., Gutierrez, M., Anwar, A. H., & Kiekintveld, C. (2019). Identifying stealthy attackers in a game theoretic framework using deception. *GameSec*, Stockholm, Sweden
- Brückner, M., & Scheffer, T. (2011). Stackelberg games for adversarial prediction problems. 17th ACM SIGKDD international conference on Knowledge discovery and data mining,
- Buckley, M. (2022). *U.S. Government Awards CounterCraft \$26MM Ceiling Contract to Support Advanced Cyber Operations*  
<https://www.countercraftsec.com/news/u-s-government-awards-counter-craft-26mm-ceiling-contract-to-support-advanced-cyber-operations/>

- Chamkar, S. A., Maleh, Y., & Gherabi, N. (2022). The human factor capabilities in security operation center (SOC). *EDPACS*, 66(1), 1-14.
- Cranford, Gonzalez, Aggarwal, P., Tambe, M., & Lebiere, C. (2020). What Attackers Know and What They Have to Lose *Human Factors and Ergonomics Society Annual Meeting*, 64(1), 456-460.
- Dasgupta, D., Akhtar, Z., & Sen, S. (2022). Machine learning in cybersecurity: a comprehensive survey. *The Journal of Defense Modeling and Simulation*, 19(1), 57-106.
- Do, C. T., Tran, N. H., Hong, C., Kamhoua, C. A., Kwiat, K. A., Blasch, E., Ren, S., Pissinou, N., & Iyengar, S. S. (2017). Game theory for cyber security and privacy. *ACM Computing Surveys (CSUR)*, 50(2), 1-37.
- Durkota, K., Lisý, V., Bošanský, B., Kiekintveld, C., & Pěchouček, M. (2019). Hardening networks against strategic attackers using attack graph games. *Computers & Security*, 87..
- Elsadig, M. A. (2024). ChatGPT and Cybersecurity: Risk Knocking the Door.
- Ferguson-Walter, K. J. (2024). An empirical assessment of the effectiveness of deception for cyber defense. PhD Dissertation, doi: 10.7275/z0rb-ek46
- Ferguson-Walter, K. J., Gutzwiller, R. S., Scott, D. D., & Johnson, C. J. (2021). Oppositional human factors in cybersecurity, 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW),
- Ferguson-Walter, K. J., Major, M. M., Johnson, C. K., & Muhleman, D. H. (2021). Examining the efficacy of decoy-based and psychological cyber deception. 30th USENIX security symposium (USENIX Security 21),
- Fraleay, J. B., & Cannady, J. (2017, 30 March-2 April 2017). The promise of machine learning in cybersecurity. SoutheastCon 2017,
- Handa, A., Sharma, A., & Shukla, S. K. (2019). Machine learning in cybersecurity: A review. *WIRES Data Mining and Knowledge Discovery*, 9(4), e1306.
- Ho, E., Rajagopalan, A., Skvortsov, A., Arulampalam, S., & Piraveenan, M. (2022). Game Theory in Defence Applications: A Review. *Sensors (Basel)*, 22(3).
- Johnson, C. K., Gutzwiller, R. S., Gervais, J., & Ferguson-Walter, K. J. (2021). Decision-Making Biases and Cyber Attackers. 2021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW), 140-144.
- Klein, G. (2008). Naturalistic decision making. *Human factors*, 50(3), 456-460.
- Lu, Z., Wang, C., & Zhao, S. (2020). Cyber deception for computer and network security: Survey and challenges. *arXiv preprint arXiv:2007.14497*.
- Milani, S., Shen, W., Chan, K. S., Venkatesan, S., Leslie, N. O., Kamhoua, C., & Fang, F. (2020). Harnessing the power of deception in attack graph-based security games. *GameSec 2020*, College Park, MD, USA,
- Mohan, P. V., Dixit, S., Gyaneshwar, A., Chadha, U., Srinivasan, K., & Seo, J. T. (2022). Leveraging computational intelligence techniques for defensive deception, *Sensors*, 22(6), 2194.
- Ngo, H. Q., Guo, M., & Nguyen, H. (2023). Near Optimal Strategies for Honeypots Placement in Dynamic and Large Active Directory Networks. *International Conference on Autonomous Agents and Multiagent Systems*,
- Nisrine, M. (2016). A security approach for social networks based on honeypots. *Colloquium on Information Science and Technology (CiSt)*,
- Ogden, H. [in/haroldogden]. (2024, 2024). *The Cyber Deception team at Walmart is growing! [...]*LinkedIn.
- Rapid7. (2024). *Deception Technology Solution: Learn the tricks, traps, and technology to reliably detect intruders earlier in the attack chain.* <https://www.rapid7.com/solutions/deception-technology>
- Reeves, A., & Ashenden, D. (2023). Understanding decision making in security operations centres: building the case for cyber deception technology. *Frontiers in Psychology*, 14, 1165705.
- Robbins, A. (2023). *Bloodhound: Six Degrees of domain admin.* In <https://github.com/BloodHoundAD/BloodHound>
- Roy, S., Sharmin, N., Acosta, J. C., Kiekintveld, C., & Laszka, A. (2022). Survey and taxonomy of adversarial reconnaissance techniques. *ACM Computing Surveys*, 55(6), 1-38.
- Sarker, I. H., Janicke, H., Maglaras, L., & Camtepe, S. (2023). Data-driven intelligence can revolutionize today's cybersecurity world: A position paper. International Conference on Advanced Research in Technologies, Information, Innovation and Sustainability,
- Thompson, J. D., Herman, G. L., Scheponik, T., Oliva, L., Sherman, A., Phatak, D., & Patsourakos, K. (2018). Student misconceptions about cybersecurity concepts *Journal of Cybersecurity Education, Research and Practice*, 2018(1), 5.
- Underbrink, A. (2016). Effective cyber deception. *Cyber Deception: Building the Scientific Foundation*, 115-147.
- Van Someren, M., Barnard, Y. F., & Sandberg, J. (1994). The think aloud method: a practical approach to modelling cognitive. *London: Academic Press*, 11(6).
- Virvilis, N., Vanautgaerden, B., & Serrano, O. S. (2014). Changing the game: The art of deceiving sophisticated attackers. 2014 6th International Conference On Cyber Conflict (CyCon 2014),
- Weston, D. (2022). New security features for Windows 11 will help protect hybrid work. <https://www.microsoft.com/en-us/security/blog/2022/04/05/new-security-features-for-windows-11-will-help-protect-hybrid-work/>
- Zhang, Hedo, R., Rivera, A., Rull, R., Richardson, S., & Tu, X. M. (2019). Post hoc power analysis: is it an informative and meaningful analysis? *Gen Psychiatr*, 32(4),
- Zhang, H., Wu, C., Xie, J., Kim, C., & Carroll, J. M. (2023). QualiGPT: GPT as an easy-to-use tool for qualitative coding. *arXiv preprint arXiv:2310.07061*.
- Zhang, H., Wu, C., Xie, J., Lyu, Y., Cai, J., & Carroll, J. M. (2023). Redefining qualitative analysis in the AI era: Utilizing ChatGPT for efficient thematic analysis. *arXiv preprint arXiv:2309.10771*.
- Zhang, L., & Thing, V. L. (2021). Three decades of deception techniques in active cyber defense-retrospect and outlook. *Computers & Security*, 106, 102288.
- Zhong, C., Yen, J., Liu, P., Erbacher, R., Etoty, R., & Garneau, C. (2015). An integrated computer-aided cognitive task analysis method for tracing cyber-attack analysis processes. 2015 Symposium and Bootcamp on the Science of Security,