Unlikely Bedfellows? Visualizing Integration of Whaley's Expanded Deception Framework and Soviet *Reflexive Control* Models to Collect Unique Attacker Behaviors

Tim Pappa[1] and Aadam Dirie[2]
[1] Capitol Technology University, Laurel, Maryland, USA
[2]Independent Researcher, Reston, Virginia, USA

tpappa@captechu.edu
aadamdirieofficial@gmail.com

Abstract: This industry cyber deception practitioners' short working paper visualizes the integration of an expanded Bell-Whaley deception framework and Soviet reflexive control modelling to design cyber deception approaches that can collect unique attacker behaviours. While we recognize the application of a deception framework and a cognitive model is unorthodox for collecting cyber threat information, integrating these approaches prompts alternative designs that both disrupt and influence attackers, which can yield rich behaviours as cyber threat information. We will feature unpublished Whaley notes on deception in this expanded Bell-Whaley framework. This practitioners' short working paper will also introduce the application in cyber threat contexts of reflexive control methods for influencing decision-making and categories of "reflexive interactions". We will visualize this integrated approach by modelling initial access by a cybercriminal along a network perimeter, who then starts to pivot within a small non-profit organization's network, demonstrating how a small organization with limited resources can use reflexive control and deception to mimic and dazzle network packet flow to misdirect the attacker to a high-interaction honeypot. This visualized cyber deception design reflects what the attacker observes and likely processes. We will theorize in this visualization how an attacker might respond to this reflexive control and what cyber threat information it could collect.

Keywords: Cyber deception, Barton Whaley, Cyber deception design, Reflexive control

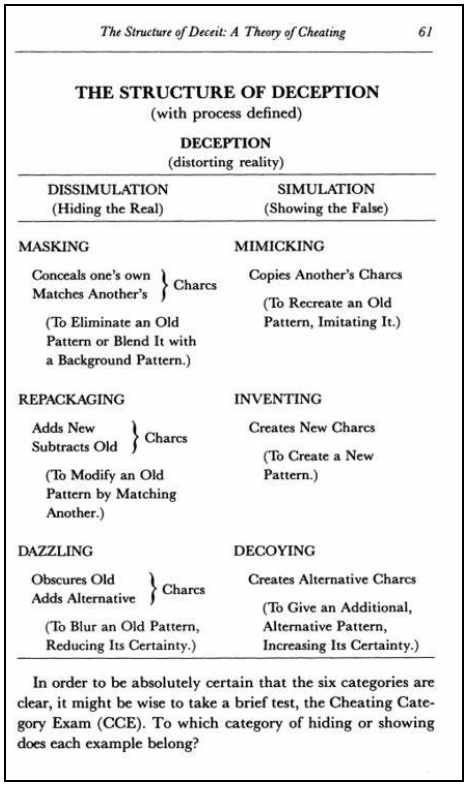
1. Introduction

The information security industry's typical approach to cyber deception has largely been functional. There is a modestly growing market for commercial off-the- shelf software for deception and packaged honeypots. Those honey- prefix builds have evolved significantly in the past twenty years, but these 'turnkey' applications are generally limited and instrumental to a function on a network, rather than a function behaviourally responsive to observed attackers (Javadpour et al., 2024).

Strand's industry handbook on cyber deception, "The Art of Active Defense", represented almost a decade of advocating for more offensive or proactive responses to attackers. Strand (2017) wrote that network defenders should engage attackers whenever possible, essentially 'annoying' them by wasting their time and forcing them to make additional moves, which likely increases collection on an attacker toward some attribution. Strand often characterized cyber deception as "annoyance". Strand's handbook included a considerable number of techniques that can be applied to attackers, but these engagements would usually occur during some phase of incident response or threat hunting. These techniques were generally reactive to attackers, designed to frustrate attackers at the moment of interaction. There was limited knowledge in those moments of an attacker's behaviours or capabilities. This approach largely continues to be the standard, except when augmented with increasingly automated deception software. Otherwise, we have not found much behaviourally based strategy found in cyber deception approaches or practice.

This approach was also not designed for collection in support of cyber threat intelligence production. These approaches are largely mitigation techniques rather than purposeful design for collection on attacker behaviours. Lansborough et al. (2021) noted that defensive network systems are usually statically configured, generally failing to adapt to an attacker or adapt in ways that an attacker predicts or easily counters. Gonzalez, Aggarwal, Cranford and Lebiere (2020) advocated for engineering dynamic forms of defence with deception integrated into the engineering design, recognizing that most current defence techniques are generic and static, and that like Lansborough et al. noted, attackers often learn and exploit these techniques. Gonzalez et al. wrote that designing effective defence must consider the knowledge of human behaviours, namely the way people make decisions and explore. Incorporating this kind of "dynamic forms of defense" in deception could result in collection of uniquely rich attacker behaviours that demonstrate an attacker's reconnaissance and decision-making.

The foundation of this practitioner's paper is the Bell-Whaley (1982, 2017) deception framework. This framework is the starting point for our designing process



The Structure of Deceit: A Theory of Cheating 61

THE STRUCTURE OF DECEPTION
(with process defined)

DECEPTION
(distorting reality)

DISSIMULATION (Hiding the Real)	SIMULATION (Showing the False)
MASKING Conceals one's own / Matches Another's } Charcs (To Eliminate an Old Pattern or Blend It with a Background Pattern.)	MIMICKING Copies Another's Charcs (To Recreate an Old Pattern, Imitating It.)
REPACKAGING Adds New / Subtracts Old } Charcs (To Modify an Old Pattern by Matching Another.)	INVENTING Creates New Charcs (To Create a New Pattern.)
DAZZLING Obscures Old / Adds Alternative } Charcs (To Blur an Old Pattern, Reducing Its Certainty.)	DECOYING Creates Alternative Charcs (To Give an Additional, Alternative Pattern, Increasing Its Certainty.)

In order to be absolutely certain that the six categories are clear, it might be wise to take a brief test, the Cheating Category Exam (CCE). To which category of hiding or showing does each example belong?

Figure 1: Bell-Whaley's deception framework

This framework has been discussed extensively in literature, primarily military deception literature. Bell-Whaley organized this framework as either *dissimulating* or *simulating*. Dissimulating might include masking, or hiding the real by making it invisible, for example. Simulating might include inventing or showing the false by fabricating something.

This paper will expand this deception framework, introducing some of Whaley's handwritten notes and unpublished manuscripts on that deception typology. This paper contributes to cyber threat alternative modelling by visualizing the application of reflexive control techniques and methods to cyber threat contexts, augmenting cyber threat analysis of cyber threat activity that is increasingly diffuse, in both attribution and demonstration of techniques or attack pathways. Integrating Whaley's expanded deception framework also provides a modelling approach practitioners and researchers designing cyber deception infrastructure and content instrumentally to collect unique information on attackers. While the use of reflexive control as a modelling framework may not be unique to just Russian-speaking attackers, familiarity with reflexive control in cyber threat contexts could be helpful when also considering dynamic nation state attacks embedded in nation state influence campaigns. Understanding these possible contexts can concentrate collection and analysis.

2. Related Work: Expanding Whaley's Deception Modelling Framework

In an early version of Barton Whaley's (1980) unpublished manuscript, *A Typology of Misperception or The Ways We Can Be Wrong*, he wrote that he believed there was a "poverty of theory" about misperception as a foundational deception model.

Whaley wrote in his manuscript that there are two attributes of misperception that are behaviourally or psychologically based: qualitative and quantitative. The qualitative depth of misperception is variety, namely that there are several ways in which someone or a system can be deceived or can misperceive something observable or demonstrated or communicated. Whaley wrote that the quantitative depth or dimension is intensity, namely the degree to which misperception occurs. That degree can be measured in one instance or individual's perception or across several of these possible varieties of misperception.

Whaley (1974) shared an example in another unpublished work describing if he buried or hid a bag of golden coins in his backyard: if he is burying or hiding (dissimulating) the golden coins or his wealth, he is

simultaneously demonstrating or suggesting (simulating) that the golden coins are not at his home but somewhere else, and that he is not wealthy but poor. Applying the Bell-Whaley deception framework then should consider not only simulation or dissimulation, but both.

Whaley wrote in this manuscript rather candidly that Bell believed the most effective deception techniques in their shared framework were masking and mimicking. Whaley agreed with Bell, adding that he believed dazzling and decoying were the least effective, explaining that these two techniques in their framework only produce "a mere razzle dazzle effect" in most instances. But the effectiveness of these techniques generally depends on the qualitative and quantitative foundations of misperception that Whaley wrote about, in each application of these techniques

3. Related Work on *Reflexive Control* in Cyber Threat Contexts

A Soviet psychosocial framework known as reflexive control was operationalized in the 1960s to manage and influence the behaviours and choices of target decision makers. Influencing someone's decision making process then makes uncertain situations more certain. Chotikul (1986) noted that while the United States had struggled to develop a "sophisticated and discriminating" understanding of Soviet motivations and Soviet decision-making processes, reflexive control largely reflected the Soviets commitment to understanding American social and cognitive processes and cultures with the development of this behavioural influence model.

Chotikul further characterized reflexive control as reflecting a "cognitive map" of a decision maker in their own social and organizational process and context to indirectly influence their choices. Chotikul likened reflexive control to how a chess player might try to gain an advantage over an opponent not only by observing him or her but by signalling intention to predetermine the opponent's anticipated moves in response to his or her signalling or brief moves. Chotikul found that one of the primary goals of reflexive control was to lower risk and increase the predictability of situations so that Soviet decision makers had greater confidence in their own decisions when determining how to try to manage or influence the behaviours of an adversary.

Reflexive control does not require any "chain of feedback" however because there should be confidence already in how the targeted decision maker will process the manipulated information. That management of information directed at the decision maker becomes a "reflex" if that influence results in a decision or behaviour advantageous to operators of reflexive control. Thomas (2004) likened reflexive control to perception management, but this framework is much more of a methodology of cognitive modelling.

Perception management is generally defined from a Western perspective as influencing a target's emotions and motivating choice or decisions, whereas the Soviet or Russian approach in reflexive control is to model how that target collects information and how they process information and then plan to influence that process with controlled or manipulated information. Reflexive control resembles the approach of cyber behavioural analysis, where the context of someone's lived environment and relationships and who they are and who they believe they are shapes much of the design of narratives or approaches to influence that target (Giles, Sherr and House, 2018).

Thomas cited a Soviet general military theorist who promoted the use of reflexive control in the 1970s. This theorist identified four methods for gaining influence over the decision maker. The suggestion is that influencing the decision maker can influence the whole system. The following comparative table characterizes those methods and our suggestions of application of those methods in cyber threat contexts:

Four methods for gaining influence over the decision maker with reflexive control:	*Application of those methods in cyber threat contexts:*
Power pressure Examples include the use of superior force, ultimatums, threats of sanctions, threats of risk, psychological attacks, provocative maneuvers, "exploiting and playing up victory", and showing mercy to an enemy ally that has stopped fighting, for example.	This could include the projection of real and imagined government-industry cyber threat collection tools and information sharing, public pronouncements of disrupting an attack attempt, threatening to release embarrassing information about the attacker, threatening to release information publicly highlighting the attacker's poor tradecraft
Measures to present false information about the situation Examples include concealment or displaying what appears to be weakness in a strong location, abandoning one position to reinforce another, creating false units, weapons bluffing, leaving a route open for an enemy to withdraw from encirclement, and striking an enemy base when the enemy is not there, for example.	This could include two-sided deception of artifacts or servers, *honeypatches* that are patched but appear vulnerable, projecting information security organizational charts displaying additional security teams, maintaining a previously exploited pathway in a network but with canaries along that pathway, taking down boxes and servers that are attributed to an attacker's infrastructure but are dormant
Influencing the enemy's decision-making algorithm Examples include publishing a "deliberately distorted doctrine", striking key leaders, transmitting false background data, and operating in a standby mode, for example.	This could include false or embellished references in public communications and news media to information security programs and teams, dissemination or 'leaking' recent or historical reports on security audits that reveal false vulnerabilities
Altering the decision-making time Examples include unexpectedly starting combat actions and making what appear to be "hasty" decisions that change the mode and character of an operation, for example.	This could include unexpected disclosure of enterprise software removal, delayed implementation of software protections, accelerated adoption of software tools and services throughout the enterprise network

Figure 2: Methods of Reflexive Control and Application in a Cyber Threat Context

Thomas referred to another senior Russian military officer who in the mid-1990s wrote about how computer technology increases the effectiveness of reflexive control, providing opportunities not only to influence people but to influence the collection of data and metadata.

Vasara (2020) highlighted later Russian theorists who emphasized that the "pervasive nature of information technology" in fact facilitates the use of reflexive control. Vasara referred to Kazakov, Kiryshun and Lazukin who compared the use of reflexive content and communication to sending "information packets" much like data packets online. The "sequenced totality" of those information packets can condition the target of that information to believe something and decide based on that influenced belief, but the parallel to network defenders and attackers observing packet exchange and NetFlow is close.

Vasara referred to Dovzhenko and Zavgorodniy who noted that decision makers are increasingly dependent on the development of formalized technical solutions and protocols such as we might find in cyber threat intelligence corporations or intelligence communities and these technical solutions and process infrastructure can be mirror or reflected perhaps more closely than human behaviours.

The same challenge might be experienced by network defenders and attackers, who are both performing behaviours in furtherance of their technical solutions and functions or protocols. There is considerable opportunity for manipulation of information and deception (Jainter and Kantola, 2018).

Clifford (1987) in his work on strategic Soviet deception organized categories of reflexive interactions based only on Soviet theorists developing reflexive control. The following comparative table characterize those categories and our suggestions of application of those methods in cyber threat contexts:

Categories of reflexive interactions:	*Application of those methods in cyber threat contexts:*
Transfer of an image of the situation This involves providing a target with an erroneous or incomplete image of the situation.	This could include brief general references by multiple information security representatives in different conferences and different online platforms to an ongoing network security challenge
Creation of a goal for the opponent This involves putting a target in a position in which they must choose an option that is advantageous to whoever is using reflexive control, such as provoking an adversary with a threat that they must respond to.	This could include giving the impression an organization has collated enterprise functions into what appears to be a single point-of-failure on a network, so that attackers are channeled to focus on that goal of attack given the short window of vulnerability
Form a goal by transferring an image of the situation This involves feigning weakness or displaying what appears to be a false condition.	This could include controlled chatter by company information security engineers on forums suggesting agitation that the company continues to use faulty third-party software
Transfer of an image of someone's own perception of the situation This involves providing a target with false information or some limited truth based on the perspective of someone other than the target.	This could include support of or disagreement with a third-party information security group that published analysis on a company's network vulnerabilities
Transfer of an image of one's own goal An example of this may be a feint by an ice hockey player, where the defender has a changed or uncertain impression of where the other player is headed on the ice.	This could include network defenders mitigating an exfil differently, such as deactivating only a portion of the user or administrator accounts' credentials collected in the exfil
Transfer of an image of one's own doctrine This involves displaying a distorted version of someone's procedures and algorithms for decision-making to the target	This could include publishing early 'working' drafts of network analysis that display some of the protocol and methodology of the information security enterprise
Transfer of one's own image of a situation to make the opponent deduce his own goal This involves suggesting a false version of someone's own perception of a situation to the target.	This could include what appear to be private disclosures to trusted third parties about the actual state of a company's legacy servers, which contrasts with public statements
Control of a bilateral engagement by a third party	This could include simulating network functions using leased services provided by a company the attacker has established access into
Control over an opponent who is using reflexive control	This could include performing functions and anticipated network defense responses to continue luring an attacker
Control over an opponent whose doctrine is game theory	This could include simulating anticipated responses to an attacker enterprise that uses systematic protocols for reconnaissance and attack

Figure 3: Categories of Reflexive Interactions and Application in a Cyber Threat Context

4. Visualizing an Attacker's Attempted Pivot in a Small Non-Profit Network as an Example of this Integrated Approach

In this visualization, we are modelling how an attacker once he or she has established initial access along a network perimeter begins to pivot within the network of a small non-profit organization. The attacker's point of initial access may have been vulnerable because the network defenders anticipated this likely attack pathway. The attacker has begun to observe network packet flow in this space. The following sequence will describe the likely interaction between attacker and the targeted infrastructure.

Operational Infrastructure

This activity can be visualized in a scenario as seen in Fig. 4. In this scenario, a first layer of deception is employed using low and medium interaction honeypots that are placed both external and internal to the network. These honeypots signify low importance assets that look appealing but mainly serve as a means of enriching intelligence about the attacker.

The deceptive network behind the external firewall consists of a hub for means of communication, network intrusion detection and prevention systems for monitoring of incoming and outgoing traffic, as well as various devices and infrastructure that will mainly be utilized to generate synthetic traffic towards the high interaction honeynet. An additional internal firewall is used to segment out a" mimic production environment", which can consist of additional infrastructure.

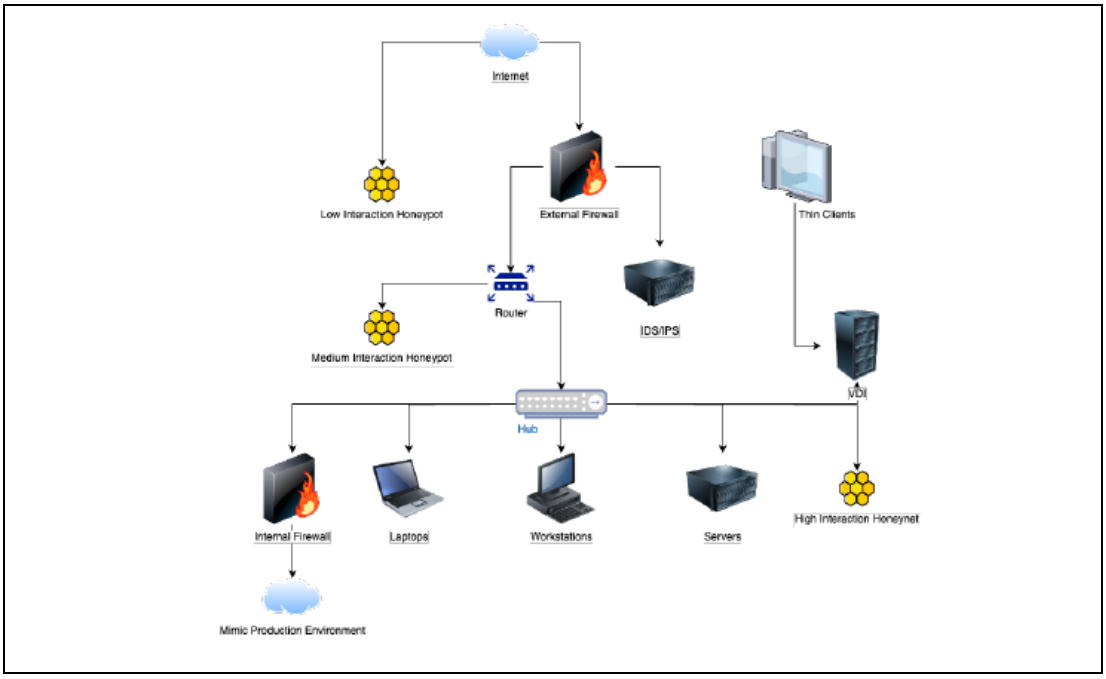


Figure 4: Comprehensive Honeynet Architecture based on the Non-Profit Attack Scenario

Deceptive Application

- *Initial Access & Reconnaissance*: The attacker in this scenario gained initial access inside a network perimeter, allowing them to start gathering information on normal behaviour or ”traffic” within the environment. This is graphically conveyed in Fig. 5, where an attacker can intrude and gain initial access to a device connected to the hub which facilitates monitoring of all traffic through that device. The deceptive network would be designed such that synthetic traffic is crafted by many different devices towards the honeynet, indicating it as a high value target.

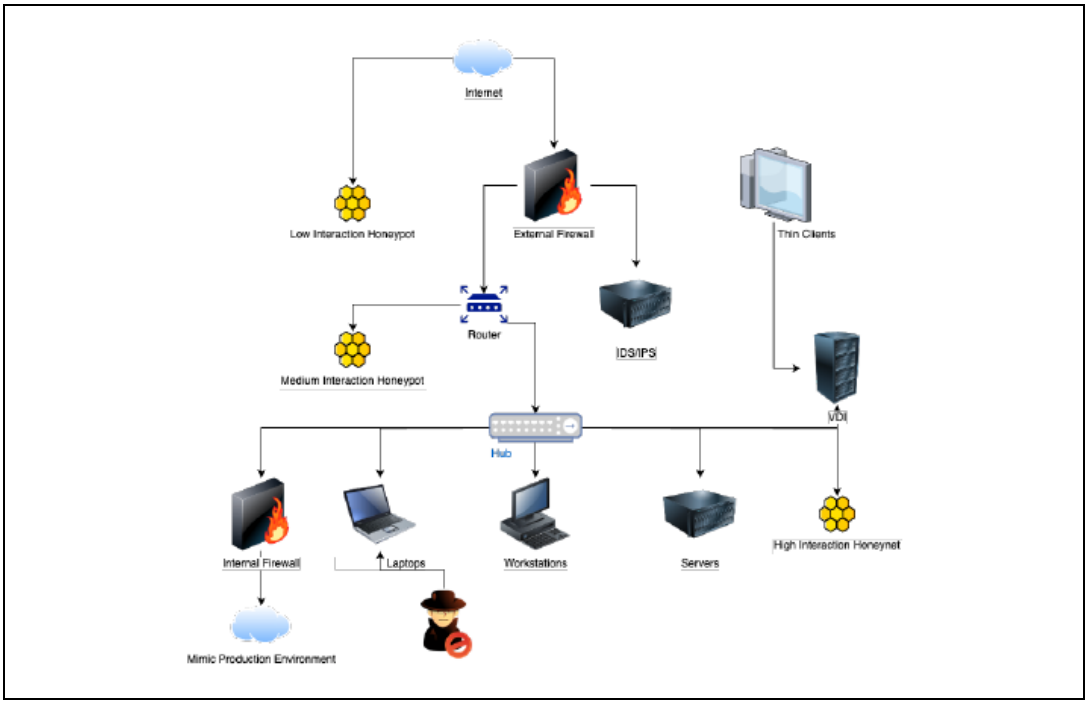


Figure 5: Comprehensive Honeynet Architecture based on Non-Profit Organization’s Network with Attacker Intrusion

- *Interaction with Honeynet*: The attacker in this scenario will notice the network traffic through the hub and identify common behaviour and aim to exploit that in any further actions to gain information or escalate privileges. The synthetic network traffic towards the honeynet can be crafted such that with a very small modification, an attacker can feel confident that they can mask their intentions. An example of this would be embedding exploitable web security vulnerabilities within services hosted within the honeynet infrastructure with those susceptible endpoints being frequented in the synthetic network traffic. This is demonstrated in Fig. 6, where the attacker can send malicious requests to exfiltrate secrets, data, and look for 'crown jewels'. This helps us gain an understanding of the attacker's behaviours.

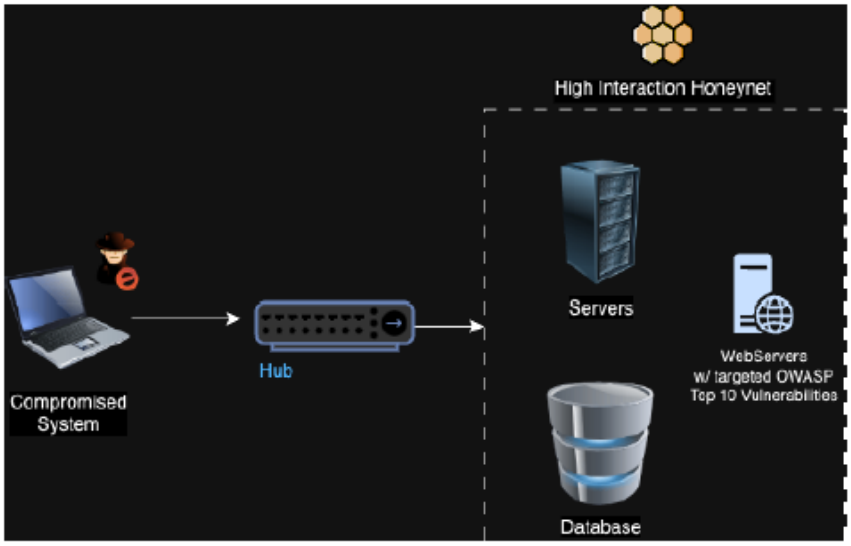


Figure 6: Attacker interaction with High Interaction Honeynet

Reflexive Control Application

- *Reflecting the Attacker's Anticipated Observations and Processing*: The topography of the network the attacker observes throughout the pathway of the attacker's approach and movement inside the network reflects anticipated placement of infrastructure, including external and internal firewalls. The generation of synthetic traffic from multiple network origins toward the high-interaction honeypot is what an attacker might expect to see in this relatively common design, where there is direction toward higher functioning and higher priority programs. Because the attacker may find this topography common, he or she may scrutinize the design of this infrastructure less. Depending on the storyline and plausibility of web security vulnerabilities in this network infrastructure, the attacker may be anticipating the identification and exploitation of these common web security vulnerabilities.
- *Theorizing What Kind of Attacker Behaviours We Might Collect*: The attacker's familiarization with the common topography of this network may ground anticipated behaviours we expect to see, but the attacker's pivot upon observing the synthetic flow of data to the high-interaction honeypot may reveal not only his or her tactics, techniques, and procedures (TTPs) while on a network host, but how dazzling with a common topography and familiar NetFlow processes may lower his or her scrutiny of our deceptive design and direct the attacker toward collection points. Luring the attacker to attempt to exploit some of the web security vulnerabilities we embedded in hosted services could not only demonstrate the attacker's tool selection and execution of exploit functions and techniques, but also the attacker's process in this kind of network context where there is a range of possible movement for the attacker to recon further and search for sensitive data or credentials. This scenario is much more of a naturalistic sandbox where we see not only demonstration, but processing and choice.

5. Conclusion and Future Work

This industry cyber deception practitioner's paper expanded Bell-Whaley's deception framework and then conceptually applied Soviet reflexive control behavioural modelling to cyber threat contexts. This provided a foundation for exploring this alternative approach to cyber deception design and design for custom building deceptive infrastructure that not only influences attackers but collects unique information on attackers. This is an unorthodox approach for operationalizing collection for cyber threat intelligence.

The two tables included in this paper where we applied reflexive control techniques and approaches to cyber threat contexts was designed to suggest some possibilities for cyber threat practitioners and researchers, where they might understand that application in cyber contexts better than some of the research literature on reflexive control. This may just be a starting point for understanding and applying reflexive control, but this approach of reframing reflexive control in cyber threat contexts should make this modelling more viable. There has not been much attention in literature to how reflexive control can be applied in cyber threat contexts, other than more general online information campaigns or cyber-enabled disinformation campaigns.

Our visualized scenario could have been presented in several ways, depending on the attacker's behaviour in the visualization. What is most important is that the infrastructure the attacker observes, and experiences is what he or she expects to see. The infrastructure reflects what is anticipated or normal in terms of network topography. The 'reflex' occurs when we use that anticipated or typical pathway for attacker decision making and movement to provide or demonstrate network data or behaviours the attacker anticipates, to influence that attacker to make a different decision or move in a different direction.

This visualized scenario demonstrated a "reflexive interaction" that was behaviourally responsive to the attacker rather than any static function we put into position, because we did not adapt the infrastructure - we used the infrastructure design most attackers would expect to see but then added simulated and dissimulated network behaviours and data to alter the attacker's collection and processing before he or she made a decision. We would argue that we would influence the attacker's sensemaking, in fact. The information we can collect in this "reflexive interaction" is dynamic, which is potentially much more valuable behavioural information to collect for analysts that could help clarify attribution or undefined analysis.

This visualization did not include a deceptive narrative or storyline, but we consider narratives or storylines that are communicated or displayed inside and outside of a network to be critical to effective cyber deception. Integrating reflexive control cognitive modelling in a particular cultural context or online environment would arguably enhance a deception storyline or narrative and the related deceptive infrastructure, because it would presumably reflect an attacker's reconnaissance and attack pathway and his or her anticipated environment when he or she reached the perimeter or just inside the target network.

Our planned future work will include a more comprehensive visualization of this integrated modelling of Whaley's expanded deception framework and reflexive control approaches with a storyline or narrative. Reflexive control approaches would naturally include storylines or narratives that reflect the information pathway and environment decision makers are used to, so in these contexts that might include information shared publicly by a corporate retailer about a change in network enterprise software or a contract with a new vendor that may present a vulnerability an attacker is confident they can exploit. Those deception storylines or narratives can also influence targeting decisions, if an attacker has reason to target a corporation or organization because of some relationship affiliation or support for a country or organization a nation state attack group or hacktivist collective may want to attack.

References

Bell, J.B. and Whaley, B., 2017. *Cheating and deception*. Routledge.

Bell, J.B. and Whaley, B., 1982. *Cheating: deception in war & magic, games & sports, sex & religion, business & con games, politics & espionage, art & science*. St Martin's Press.

Chotikul, D., 1986. *The Soviet theory of reflexive control in historical and psychocultural perspective: preliminary study*. Monterey, California: Naval Postgraduate School.

Clifford, R., 1987. Reflexive control in Soviet military planning. *Soviet Strategic Deception*, pp.293-312.

Jaitner, M.L. and Kantola, H., 2016. Applying principles of reflexive control in information and cyber operations. *Journal of Information Warfare*, *15*(4), pp.27-38.

Giles, K., Sherr, J., House, C., 2018. Russian Reflexive Control. Kingston: Royal Military College of Canada.

Gonzalez, C., Aggarwal, P., Cranford, E.A. and Lebiere, C., 1825, January. Design of dynamic and personalized deception: A research framework and new insights for cyberdefense. In *Proceedings of the 53rd hawaii international conference on system sciences* (Vol. 1834).

Javadpour, A., Ja'fari, F., Taleb, T., Shojafar, M. and Benzaïd, C., 2024. A comprehensive survey on cyber deception techniques to improve honeypot performance. *Computers & Security*, p.103792.Kambow, N. and Passi, L.K., 2014. Honeypots: The need of network security. *International Journal of Computer Science and Information Technologies*, *5*(5), pp.6098-6101.

Landsborough, J., Carpenter, L., Coronado, B., Fugate, S., Ferguson-Walter, K. and Van Bruggen, D., 2021, January. Towards Self-Adaptive Cyber Deception for Defense. In *HICSS* (pp. 1-10).

Mokube, I. and Adams, M., 2007, March. Honeypots: concepts, approaches, and challenges. In *Proceedings of the 45th annual southeast regional conference* (pp. 321-326).

Strand, J., Asadoorian, P., Robish, E. and Donnelly, B., 2013. *Offensive Countermeasures: The Art of Active Defense*. CreateSpace Independent Publishing Platform.

Thomas, T., 2004. Russia's reflexive control theory and the military. *Journal of Slavic Military Studies*, *17*(2), pp.237-256.

Vasara, A., 2020. *Theory of reflexive control: origins, evolution and application in the framework of contemporary Russian military strategy*. National Defence University.

Whaley, B., 1982. Toward a general theory of deception. *The Journal of Strategic Studies*, *5*(1), pp.178-192.

Whaley, B., 1980. A Typology of Misperception or The Ways We Can Be Wrong. *Unpublished manuscript draft*.

Whaley, B., 1974. Deception: Its Decline and Revival in International Conflict. *Unpublished manuscript draft*.