Cyber Deception: Overview and The Road Ahead

Cliff Wang, *Fellow, IEEE* and Zhuo Lu, *Member, IEEE,*

***Abstract*—Since the concept of deception for cyber security was introduced decades ago, several primitive systems, such as honeypots, have been attempted. Until more recently, research on adaptive cyber defense techniques has gained momentum. The new research interests in this area motivate us to provide a high-level overview of cyber deception. We analyze potential strategies of cyber deception and its unique aspects. We discuss the research challenges of creating effective cyber deception based techniques and identify future research directions.**

Keywords: Cyber attack and defense; Cyber deception; Adversarial mind manipulation; Proactive Strategies; Moving target defense;

I. Introduction

The notion of cyber kill chain [1] was introduced to outline the chronicle stages of cyber adversaries, from early reconnaissance to the actual attack. Instead of engaging with adversaries in the early steps of cyber kill chain (the reconnaissance phase), current cyber defense practices mostly focus on reactive response after attacks have happened. This gives adversaries a significantly asymmetric advantage such that they have sufficient time to probe and learn our systems, and then prepare and launch attacks decisively to achieve their objectives within a short period of time, leaving little opportunity for defenders to defeat the attack actions.

In order to reverse defenders' disadvantages, the key change required is to engage with adversaries in the early stage of their cyber kill chain in order to disrupt and disable potential attacks. Recent research on proactive defense concept has led to two major mechanisms: i) moving target defense (MTD, e.g., [2], [3]) that increases complexity, diversity and randomness of the cyber systems in order to disrupt adversaries' reconnaissance and attack planning, and ii) cyber deception (e.g., [4]–[6]), which provides plausible-looking yet misleading information to deceive attackers. Both schemes offer complementary approaches to defeat attack actions.

Although deception has been used since early ages of human history, its adoption into cyber space has been mostly recent. In late 1980s, Stoll [4] first discussed how to use deceptive techniques to trace intruders for computer security. The concept of deception based honeypot followed. To attract potential attackers, honeypots masquerade themselves as service hosts that could be potentially exploited. By collecting and recording information detailing the methods used in attackers' attempts to compromise honeypots, defenders can use the learned knowledge to enhance system security. More recently, a number of deception-based security designs (e.g., [6], [7]) were introduced to confuse or mislead attackers. These cyber deception approaches have shown great promises of disrupting the cyber kill chain at its early stage.

In this article, we introduce cyber deception as an emerging proactive cyber defense technology. Since the research on cyber deception is still nascent, our objective is to provide a high-level overview of its life cycle, analyze unique aspects of cyber deception in comparison to non-deceptive approaches, and lastly outline research challenges to stimulate more research going forward.

Cliff Wang is with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh NC 27695, USA. Email: cliffwang@ncsu.edu.

Zhuo Lu is with the Department of Electrical Engineering and Florida Center for Cybersecurity at University of South Florida, Tampa FL 33620, USA. E-mail: zhuolu@usf.edu.

II. Formal View of Deception

A. Life Cycle of Cyber Deception

One of the most widely adopted definitions of cyber deception proposed by Yuill [5] was "planned actions taken to mislead and/or confuse attackers and to thereby cause them to take (or not take) specific actions that aid computer-security defenses." The essential parts in cyber deception include crafted information by the defender (that will be used to mislead) and wrong actions taken by the adversary as a result of the deception.

Figure 1 shows the life cycle of cyber deception at the conceptual level. Like a traditional deception in physical world, cyber deception is a revolving two-step action.

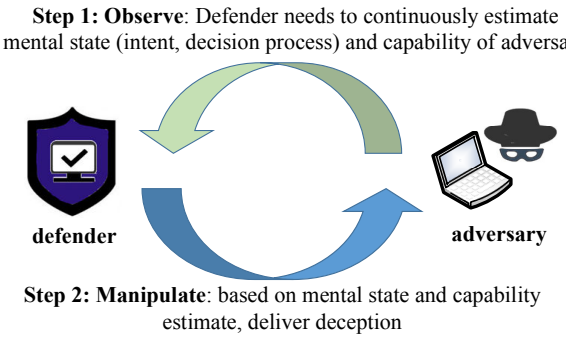

Fig. 1. Life cycle of cyber deception.

- In Step 1, the defender will collect as much intelligence on the adversary as possible, in order to derive an estimate of the adversary's intent, capability and decision process. This is fundamental to the success of cyber deception since without a solid situational understanding of potential attackers, an ad-hoc deception scheme is no better than a random shot into the dark.
- In Step 2, based on the understanding of the adversary, an actual deception scheme will be carefully crafted in order to manipulate and mislead the adversary. Key compositions of a cyber deception scheme can contain

a combination of true and fabricated information and involve various deception techniques.

Through multiple round engagements, the defenders' knowledge on adversarial state (intent, capability and decision process) may improve over time. In addition, deception actions by the defenders can be progressive as well. The defenders may choose to project additional deception information to adjust or reinforce their earlier deception schemes. In practice, a successful cyber deception will not likely be a one-time play, but instead a continuous multi-round process that ties in both defenders and attackers.

B. Deception Formulation: Understand the Adversaries

Perceiving information learned from the adversary leads to the establishment of a good mental state model [8] that can help defenders to identify the intent of adversaries, estimate their capabilities, and reason about potential attack actions. A good mental state model presents a good picture of the adversary that can potentially answer why, what and how questions associated with observed attacker actions, which is a key requirement of any successful deception design. Early work by Daniel Dennett [9] on the Intentional Systems Theory attempted to explain an entity's behavior based on its beliefs and desires. He identified three levels of abstraction when we understand, explain and predict behaviors:

- Physical stance: It is based on the explicit knowledge of the physical constitution and the physical laws that govern systems or the world. For example, if we know that an attacker may learn the physical location, hardware type, and power consumption profile of a server, we can infer that the attacker may be able to estimate the processing capability of the server.
- Design stance: Without the need of knowing physical laws, design stance is based on the knowledge of the system's design. A system can be predicted to work as it is initially designed for. For example, if we know that an attacker can observe the traffic flows in and out of the target server, then the attacker may be able to infer the type of service being offered, clients that may be using the system and the server's potential peers.
- Intentional stance: The behavior or action of an entity is governed by its intentional states (mental states) which are driven by beliefs, desires, intent, and motivation. In the above example, based on the knowledge on what the adversary may already know, if we believe that the adversary's intent is to steal information, we can potentially plant misleading or fake information on the server or in the server's traffic flows that the adversary has been targeting.

If we adopt the Intentional Systems Theory to model an attacker's mental state, we assume that the attacker's plan of action follows his/her physical, design and intentional stances. The task to understand the adversarial mental model relies keenly on identifying the intentional stances, along with the design and physical stances of adversaries. Based on such understandings, the defender can attempt to manipulate adversaries' mindset through the introduction of biases. A cognitive bias [10] represents the deviation from common sense (normal) or rational judgment and decision process. Cognitive biases usually arise from information bias, where misinformation is intentionally crafted and used to cause cognitive biases. Thus, the essence of deception is to understand an adversary and then to introduce cognitive biases in order to mislead. For the aforementioned server attack example, decoy servers with similar hardware profile (physical stance) and fabricated traffic profile (design stance) could be introduced to divert attentions. In addition, different types of service traffic flows could be fabricated at both real and decoy servers to introduce cognitive biases to the attackers such that the real server could be potentially camouflaged.

C. Deception Schemes and Common Actions

Although the key task for the cyber defender is to protect information from being disclosed, deception-based defense schemes depend on "misinformation" disclosure in order to deceive. To create believable "misinformation", the defender first needs to establish a situation awareness on what exactly the defender knows and does not know, as well on how much the adversaries know. Then, based on this knowledge, the defender will plan strategically on how to selectively combine both true and fictitious information that ultimately can lead to adversaries' cognitive biases. A well-crafted deception scheme may contain disclosure of selected truth in order to convince the adversaries so that the overall cyber deception goal can be achieved.

The success of deception scheme relies on defenders' asymmetric advantage over attackers: defenders know more! Observing, analyzing and understanding the mental state (Step 1 of any deception game, as described in Section II-A) is the key to successful deception action in Step 2. Based on the status of the information known or unknown to both the defender and the attacker, Table I lists the possible actions that the defender may take. While traditional defense focuses on information hiding, deception-based defense may use a combination of true information hiding and fake information disclosure to protect critical information and to mislead the adversary. In addition, in the case of information known to both the defender and the attacker, selected true information is quite often combined with fake information and released together in order to make a deception scheme more convincing.

TABLE I
DEFENDER ACTIONS BASED ON SITUATION OF INFORMATION AVAILABILITY.

		Adversary	
		Known	Unknown
Defender	Known	Undeniable truth, fact; selectively released truth; (**Defender Action**: leverage, selectively release)	Target of reconnaissance and information collection; (**Defender Action**: deceit, protect; denial, selectively release)
	Unknown	Dangerous area. Adversary has the advantage. Defender should strategically minimize this area	Dark space

Any successful deception relies on the asymmetric information advantage over adversaries. Defenders should always try to minimize or eliminate their "blind spots" where they may

not know the information that adversaries know, as listed in row 2 of Table I. When the defender's advantage is reversed, the foundation for any deception scheme is destroyed.

III. Unique Aspects of Cyber Deception

In this section, we present the unique aspects of cyber deception, including key differences of cyber deception from non-cyber deception and a comparison to MTD, another major proactive defense strategy.

A. Key Differences from Non-cyber Deception

Non-cyber space deception includes both physical domain deception and social domain deception (i.e., inter-human deception, including fraud). Cyber deception may share common traits of non-cyber deception and may be related to non-cyber space deception. For example, some attacks, such as advanced persistent threats, can involve physical, social and cyber elements to further the attack effectiveness. Cyber deception also has several unique aspects in terms of time, space and speed constraints, as shown in Table II.

TABLE II
Differences between Cyber Deception and Physical/Social Deception.

	Cyber Deception	Physical Deception	Social Deception
Time	**Past/Current/Future** (if we control an individual system's clock)	**Current**, and always happening in chronical order	**Current**, and always happening in chronical order
Space	**Weaker constraints**, just obey computing and networking practices	**Bounded** by physical-space laws and principles	**Bounded** by normal social interactions
Speed	**Extremely fast** (can be at the speed of light)	**Bounded** by physical-space laws and principles	**Bounded** by human interactions limitations
Sensing	Virtual/Indirect	Physical means	Social interactions

From the time constraint perspective, things happen in chronological order in the physical world. Both physical and inter-personal deception schemes have to follow the chronological order since it is not possible to change the physical world time arbitrarily. In the cyber space, an individual device or a network quite often relies on its own clock for time information. Unless a strict time synchronization with a trusted global source is enforced, it is possible (although not always easy) to manipulate local clock settings such that a deception scheme can exploit "going back into the past". The chronological order of activities could potentially be altered among entities whose clocks were manipulated. This is the first unique aspect of cyber deception: when strict time synchronization is not enforced, time manipulation is quite possible.

In physical space, deception schemes normally have to follow physics principles in order to be plausible. Laws of physics limit what physical world deception can achieve. A deception artifact has to mimic physical attributes of a real entity. For example, fake tanks used in Normandy during the D-day operation need to have the same physical size and color of real tanks in order to fool Nazi air reconnaissance through visual and optical imaging analysis. These artifacts need to be carefully designed such that they will conform to all physical principles governing the operational environment. In addition, a deception design needs to consider the technical capabilities of its recipients. For example, it would be much harder to use simple fake tanks today since new types of sensors (e.g., infrared imaging) will easily identify a real tank from fake ones using advanced sensing capabilities such as a heat profile signature. In contrast, cyber deception may not be constrained by the physical world principles. The requirement to follow physical principles comes into play only when hardware devices are part of the deception scheme, or it is anticipated that the potential deception target (the adversary) may observe the system through hardware side channels.

In the social domain, inter-person deception needs to conform to the norms of human communication and social interactions, and follow common social and cultural practices. Social space deception often relies heavily on inter-human relationship, direct interactions, and in-person verbal and non-verbal communications. Quite often cultural aspects also need to be carefully considered in social space deception. In contrast, cyber deception is largely confined to the cyber space, and has a much weaker social interaction constraint compared to social space deception.

The speed of building and deploying deception schemes in both physical and social spaces is limited by physical world principles and human interaction rules. In contrast, the speed of changes for the cyber space can be extremely fast since normally there are few physical or social constraints unless substantial hardware setup is required. This is another unique characteristic of cyber deception. Generally for cyber deception, the most challenging and time-consuming part of its life cycle is the initial step of observing and estimating the mental model of adversaries. Once deception schemes have been crafted, deploying a scheme or switching between schemes can be made fairly rapidly.

B. Comparison to Moving Target Defense

MTD is another major proactive cyber defense technique that is being actively researched by the community lately. MTD refers to the techniques that continuously change a system's attack surface through adaptation, thereby increasing the uncertainty, complexity, and costs for the attacker. Compared to MTD, the fundamental idea of cyber deception does not focus on transforming our cyber systems continuously, but instead on distracting attackers' attentions away from critical assets by relying on carefully crafted information to create cognitive biases and to mislead adversary's actions so that potential attacks are rendered onto the wrong targets [8]. Table III summarizes the key differences between MTD and cyber deception approaches.

Collectively, MTD and deception are complementary techniques that can be deployed by the defender at the same time, with different focuses but having a common goal to defeat attacks. In the following, we summarize potential advantages of cyber deception over MTD when practically deployed.

- Reducing overhead and saving resources. System adaptation schemes are quite often fairly complex and may require extensive computing and hardware resources to accomplish. Compared to MTD, cyber deception may require no significant system update and can be made simple and effective when correctly set up.

TABLE III
COMPARISONS BETWEEN MTD AND CYBER DECEPTION.

	MTD	Cyber Deception
Technical Approach	Increase system complexity and diversity so that the adversary's observation rate is slower than the change rate. Main focus is to deny information	Not focus on dynamic system changes, but use a combination of information disclosure and hiding that can help project fake information to mislead
Human Engagement	Typically not addressed	Engage to observe and estimate adversary state and to inject misleading information
Social/cultural influence	Typically not addressed	Has to be incorporated as part of "genuine" fake information projection
Information disclosure	Never, focus on denying information	Yes, but only applicable to non-essential, selected truth combined with fake information projection

- Understanding adversaries. Most security mechanisms are designed to create a boundary around cyber systems and aim to stop illicit access attempts. In addition, MTD simply attempts to change the system boundary dynamically without trying to understand where adversaries may attempt to penetrate. Through the engagement with adversaries, cyber deception allows us to gain a better knowledge on adversaries, which may not only help build deception schemes, but also make MTD techniques less ad-hoc and much more relevant and effective.
- Increasing the risk on the adversary's side. Cyber deception relies on passive and proactive probing, observing, and learning on adversaries. The possibility of deploying deception-based approaches may deter attackers who are not willing to take the risk of being exposed, analyzed and further deceived. In contrast, MTD defense is weaker on deterrence from this perspective. Attackers may try all possible methods without worrying about being observed and misled.

To summarize, while MTD focuses on changing our system for defense, cyber deception aims at changing the adversary to achieve the same objective. Typically, MTD relies on superior technology to win. Cyber deception, on the other hand, is more on human engineering.

IV. THE PATH AHEAD

In this article, we provided an overview of emerging cyber deception research and identified the life cycle model and key characteristics of cyber deception. Although the use of deception to enhance cyber defense has shown a number of interesting and promising applications, much more new research is needed to drive the area forward.

1) More Accurate Adversarial Model: Successful cyber deception schemes depend on a good understanding of adversaries, which is a challenging task. A well-defined adversarial model that incorporates mental state estimation is critical for both formulating a deception scheme and evaluating its effectiveness.
2) Continuous, Multi-round Engagement: Learning adversarial intent, capability and methods requires continuous direct and indirect engagement. It is usually hard and sometimes impossible to collect information and learn about a potential attacker whom we are not engaged with. Honeypot or honeynet-like systems may attract adversaries and allow us to engage. The interactions enabled by new honey systems may provide a valuable opportunity to learn the intent, capability and potential goal of adversaries, and help us create an initial deception scheme and make follow-on adjustment.
3) Manipulation of Adversarial Mind: The ultimate goal of cyber deception is to introduce cognitive biases to adversaries in order to manipulate their decision process and to mislead them into wrong decisions. We need to leverage advances in other disciplines (such cognitive science, decision sciences and control theory of systems) that can help us improve effectiveness in manipulating adversaries.
4) Usability Analysis and Quantification: A successful deception scheme needs to ensure that it has the minimum impact on normal users. Deception metrics need to incorporate usability as part of its performance evaluation benchmark.
5) Combining Deception and MTD Approaches: While MTD adapts a cyber system to increase its diversity and complexity in order to make it harder and more costly for adversaries to attack, cyber deception can be combined with MTD as a complementary method by using misinformation to lead adversaries into wrong actions and to drain their resources. More research is essential to understand how to combine both proactive techniques seamlessly and how to maximize proactive protection.

We envision that going forward, a well-defined framework of cyber deception with detailed, specific application domains requires collaborative research involving computer and network security, control theory, as well as human cognitive science and psychology. Advances made in cyber deception research will lead to highly effective techniques for proactive cyber defense.

REFERENCES

[1] C. Croom, "The cyber kill-chain: A foundation for a new cyber-security strategy," *High Frontier*, vol. 6, no. 4, pp. 52–56, 2010.
[2] M. Crouse, B. Prosser, and E. W. Fulp, "Probabilistic performance analysis of moving target and deception reconnaissance defenses," in *Proc. of ACM MTD*, 2015.
[3] S. Jajodia, A. K. Ghosh, V. Swarup, C. Wang, and X. S. Wang, *Moving target defense: Creating asymmetric uncertainty for cyber threats*. Springer Science & Business Media, 2011, vol. 54.
[4] C. P. Stoll, *The cuckoo's egg: Tracking a spy through the maze of computer espionage*. Doubleday, 1989.
[5] J. J. Yuill, *Defensive computer-security deception operations: Processes, principles and techniques*. ProQuest, 2006.
[6] A. Juels and R. L. Rivest, "Honeywords: Making password-cracking detectable," in *Proc. of ACM CCS*, 2013.
[7] N. C. Paxton, D. I. Jang, S. Russell, G. J. Ahn, I. S. Moskowitz, and P. Hyden, "Utilizing network science and honeynets for software induced cyber incident analysis," in *Proc. of HICSS*, 2015.
[8] M. H. Almeshekah and E. H. Spafford, "Planning and integrating deception into computer security defenses," in *Proc. of NSPW*, 2014.
[9] D. Dennett, "Intentional systems theory," *The Oxford handbook of philosophy of mind*, pp. 339–350, 2009.
[10] M. G. Haselton, D. Nettle, and D. R. Murray, "The evolution of cognitive bias," *The Handbook of Evolutionary Psychology*, 2005.