A Game-theoretic Taxonomy and Survey of Defensive Deception for Cybersecurity and Privacy

JEFFREY PAWLICK, NYU Tandon School of Engineering and U.S. Army Research Laboratory
EDWARD COLBERT, Virginia Tech Intelligent Systems Lab, Hume Center for National Security and Technology, and U.S. Army Research Laboratory
QUANYAN ZHU, New York University Tandon School of Engineering

Cyberattacks on both databases and critical infrastructure have threatened public and private sectors. Ubiquitous tracking and wearable computing have infringed upon privacy. Advocates and engineers have recently proposed using defensive deception as a means to leverage the information asymmetry typically enjoyed by attackers as a tool for defenders. The term deception, however, has been employed broadly and with a variety of meanings. In this article, we survey 24 articles from 2008 to 2018 that use game theory to model defensive deception for cybersecurity and privacy. Then, we propose a taxonomy that defines six types of deception: perturbation, moving target defense, obfuscation, mixing, honey-x, and attacker engagement. These types are delineated by their information structures, agents, actions, and duration: precisely concepts captured by game theory. Our aims are to rigorously define types of defensive deception, to capture a snapshot of the state of the literature, to provide a menu of models that can be used for applied research, and to identify promising areas for future work. Our taxonomy provides a systematic foundation for understanding different types of defensive deception commonly encountered in cybersecurity and privacy.

CCS Concepts: • **Security and privacy** → *Economics of security and privacy*; *Network security*;

Additional Key Words and Phrases: Cybersecurity, privacy, game theory, deception, taxonomy, survey, moving target defense, perturbation, mix network, obfuscation, honeypot, attacker engagement

1 INTRODUCTION

> "All warfare is based on deception. Hence, when we are able to attack, we must seem unable; when using our forces, we must appear inactive; when we are near, we must make the enemy believe we are far away; when far away, we must make him believe we are near."—Sun Tzu, *The Art of War*

Deception has played an important role in the history of military combat. More generally, deception is commonplace in adversarial or strategic interactions in which one party possesses information unknown to the other. In this article, we are motivated by recent deception research in psychology, criminology, economics, and behavioral sciences.

1.1 Deception Across Disciplines

1.1.1 Military Applications. Deception and secrecy demanded significant attention during World War II and the Cold War (Whaley 2016; Bell and Whaley 2017). But increasing globalization and the proliferation of communication technologies have recently created further challenges for mitigating deception. Increasing availability of information has led not only to more knowledge but also to worse confusion (Godson and Wirtz 2011). National defense requires detailed study of military-relevant deceptions such as advanced persistent threats carried out by state actors (Bodmer et al. 2012).

1.1.2 Psychology and Criminology. Research in psychology and criminology suggests that humans have poor abilities to detect deception (Bond Jr and DePaulo 2008; Vrij et al. 2008). One approach to address this shortcoming focuses on interview techniques. It has been shown that detection rates can be improved by tools that increase cognitive load by asking suspects to recall events in reverse order, to maintain eye contact, or to answer unexpected questions. Some of these have been incorporated into the investigative protocol known as the *Cognitive Interview for Suspects* (CIS) (Geiselman 2012). A second approach uses physiological indicators. For instance, the *Guilty Knowledge Test* (GKT) prompts a suspect with a list of items—for example, a set of articles found at the scene of a crime—and measures the suspect's physiological responses to each item (Lykken 1959). Signs of arousal in a suspect suggest that the suspect possesses guilty knowledge, because the articles are irrelevant to an innocent person.

1.1.3 Cybersecurity. Deception is used in cyberspace interactions for both defense and attack. Examples of malicious deception include email phishing, man-in-the-middle attacks, and deployment of sybil nodes (nodes with forged identities) in a social network. Cybersecurity professionals have developed various defenses against these deceptions. For instance, the field of *adversarial machine learning* protects data analysis algorithms against injection or manipulation of data by attackers (Zhang and Zhu 2015, 2017), and the field of *trust management* evaluates whether distributed nodes should trust possibly malicious signals (Pawlick et al. 2015; Pawlick and Zhu 2017b). Examples of defensive deception include moving target defense (Zhu and Başar 2013), honeypot deployment (Carroll and Grosu 2011), and the use of mix networks (Zhang et al. 2010).

1.1.4 Privacy Advocacy. Recently, privacy advocates have designed technologies for Internet users to *obfuscate* the trails of their digital activity against ubiquitous tracking. Privacy advocates argue that developments such as third-party tracking and persistent cookies have not been sufficiently regulated by law. Therefore, there is a need for user-side technologies to provide proactive privacy protection. One example is *TrackMeNot*, a browser extension that periodically issues random search queries to undermine search engine tracking algorithms (Howe and Nissenbaum 2009). Another example is *CacheCloak*, which protects location privacy by retrieving location-based services on multiple possible paths of a user. An adversary tracking the requests is not able to infer

the actual user location (Meyerowitz and Roy Choudhury 2009). These are instances of deception that are designed for benign purposes.

1.1.5 Behavioral Economics. In economics, the area of *strategic communication* quantifies the amount of information that can be transmitted between two parties when communication is unverifiable (Crawford and Sobel 1982; Kartik 2009). Communication can be evaluated both strategically and physiologically. One recent paper analyzes patterns of eye movement and pupil dilation during strategic deception (Astyk et al. 2010). At the same time, research in behavioral economics finds that sometimes economic agents choose not to deceive, even when it is incentive-compatible (Gneezy 2005; Hurkens and Kartik 2009; Fischbacher and Föllmi-Heusi 2013). Subjects that exhibit so-called *lying aversion* choose to maximize their payoffs less frequently when it requires a lie than when it requires a simple choice. This points toward the influences of morality and ethics on deception.

1.1.6 Economic Markets. In broader economics literature, Akerlof and Shiller (2015) argue that many markets admit equilibria in which deceivers exploit those who are vulnerable. Akerlof and Shiller describe these interactions in politics, pharmaceuticals, finance, and advertising. They use email phishing as an analogy for any kind of deception in which an informed "phisherman" exploits the lack of knowledge or the psychological vulnerabilities of a group of "phools." The essential insight is that opportunities for deception will be exploited in equilibrium. Across all six disciplines, deception involves interaction, conflict, rationality, and uncertainty.

1.2 Cybersecurity and Privacy

Major cybersecurity incidents in the past 10 years include breaches of Home Depot in 2014, insurance company Anthem Inc., in February of 2015, and the U.S. Office of Personnel Management later that year. The ten years from 2005 to 2015 featured over 4,000 (publicized) data breaches (Edwards et al. 2016). During the same 10 years, cyber attacks also affected physical space. The power grid in Ukraine, Iranian nuclear centrifuges, and an American water dam 20 miles north of New York were all infiltrated. Since then, new attack vectors such as the Mirai botnet have highlighted the potential to turn Internet of things (IoT) devices into domestic cyber weapons. At the same time, IoT devices have raised new privacy concerns. Smartphones and wearable electronics combine to collect sensitive data that can predict "a user's mood; stress levels; personality type; bipolar disorder; demographics" (Peppet 2014). Advocates have responded by designing innovative privacy enhancing technologies and trying to formalize and defend appropriate definitions of privacy (Nissenbaum 2004).

1.3 Defensive Deception

Firewalls, cryptography, and role-based access control, although essential components of any security strategy, are unable to fully address these new cybersecurity and privacy threats. Adversaries often gain undetected, insider access to network systems. They obtain information about networks through reconnaissance, while defenders lack an understanding of the threats that they face.

Deception is crucial to counteract this information asymmetry. Consider the following definition (Mahon 2016):

> To deceive $\overset{\text{def}}{=}$ to intentionally cause another person to acquire or continue to have a false belief, or to be prevented from acquiring or cease to have a true belief.

This definition is broad, which matches the diversity of applications of defensive deception to cybersecurity and privacy. However, the breadth of the term deception limits its depth. Finer-resolution definitions are necessary to design deception techniques for specific purposes.

While affirming the breadth and variety of types of deception in our area, we aim to tease out the distinctions between these types. To some extent, this is already present in the literature; authors use specific terms such as moving target defense, perturbation, and obfuscation to refer to different types of deception. Nevertheless, the meaning of these terms is not completely clear. For instance, what is the difference between perturbation and obfuscation? *We attempt to answer this and other questions by proposing a taxonomy of defensive deception for cybersecurity and privacy.*

1.4 Game-Theoretic Taxonomy

We are interested in understanding the question *"What are the various types of deception?"* from the viewpoint of quantitative science. Since deceptive interactions are strategic confrontations between rational agents, an appropriate quantitative science is *game theory*. Game theory "can be defined as the study of mathematical models of conflict and cooperation between intelligent rational decision-makers" (Myerson 1991). Each player in a game makes decisions that may influence the welfare of the other players. Game theory applies naturally to strategic and adversarial interactions in cybersecurity and privacy.[1] It has been used in a variety of cybersecurity contexts. A few application areas include intrusion detection systems (Alpcan and Basar 2003), adversarial machine learning (Zhang and Zhu 2015), and communications jamming (Basar 1983). Applications in physical security include the ARMOR system for Los Angeles airport security (Pita et al. 2008) and the PROTECT system for patrol routes in the port of Boston (Shieh et al. 2012).

Game theory offers a quantitative framework that can model the information structure, actors, actions, and duration of the various types of defensive deception. As a systems science, it models the essential, transferable, and universal aspects of defensive deception. In doing so, it attempts to build up a science of security.

Of course, game-theoretic models are useful only insofar as they capture real features of cyberspace phenomena. Therefore, we need to apply appropriate models to each type of deception. *This is precisely the motivation to develop a game-theoretic taxonomy of deception*: to identify the features of the various types of deception that must be captured by game-theoretic models.

1.5 Contributions and Related Work

Figure 1 gives a conceptual outline of the article. We present the following principle contributions:

(1) We review the game-theoretic models most commonly used to study cybersecurity and privacy (Section 2).

(2) We survey the contributions of 24 recent (2008–2018) articles to the field of game-theoretic models of defensive deception (Section 3).

(3) We develop a taxonomy that defines specific types of deception, including moving target defense, perturbation, mixing, obfuscation, honey-x, and attacker engagement. We also show how the game-theoretic concepts of private information, actors, actions, and duration capture the essential differences between types of deception (Section 4).

(4) We discuss promising areas for future research. These include mimetic deception, theoretical advances, practical implementations, and interdisciplinary security (Section 5).

All of the literature that we review (1) studies cybersecurity or privacy, (2) employs *defensive* deception, and (3) uses game theory. This strictly constrains the scope of the article. Notably, we

---

[1]All of the works surveyed in this article use "non-cooperative game theory." In non-cooperative games, "the players' choices are based only on their perceived self-interest, in contrast to the theory of cooperative games" in which groups of players make decisions together in coalitions. Still, "non-cooperative players, motivated solely by self-interest, can exhibit 'cooperative' behavior in some settings" (Fudenberg and Tirole 1991).
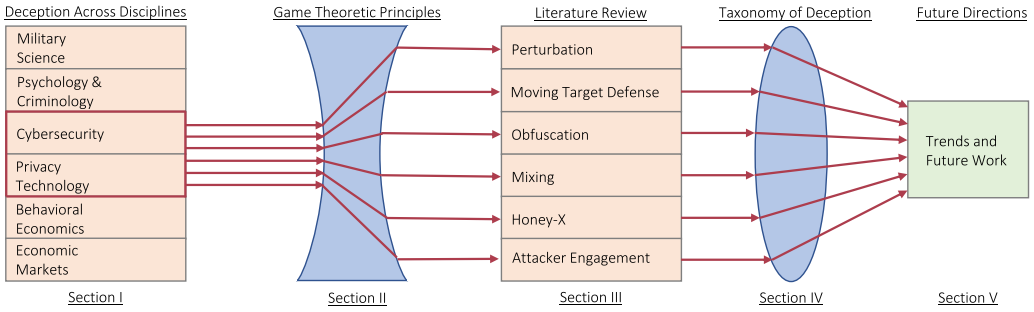
Fig. 1. We begin by discussing deception across several different disciplines, and then we focus on defensive deception in cybersecurity and privacy. Then, we describe a set of game-theoretic principles that are utilized in literature in this area, and we review the literature. Finally, we create a taxonomy that defines each type of deception, and we identify areas for future work.

have excluded papers that use deception for physical security (e.g., (Pita et al. 2008; Shieh et al. 2012), etc.). We have also excluded studies of malicious deception and the defensive methods that aim to mitigate it.

Surveys of game theory and cybersecurity can be found in Roy et al. (2010), Manshaei et al. (2013), and Do et al. (2017), but these do not focus on deception in particular. Deception in the context of military applications has been well studied by Barton Whaley and coauthors (Bell and Whaley 2017; Rothstein and Whaley 2013; Whaley 2016). Some of these works include taxonomies. Specifically, Rothstein and Whaley (2013) make a distinction between "hiding the real" and "showing the false," which we build upon in our taxonomy. Gerwehr and Glenn add that some deceivers aim to inject noise: randomness or intense activity that slows an adversary's ability to act. We find analogous aims in perturbation or obfuscation for the purposes of privacy. They also distinguish between static and dynamic deception. We adopt this distinction, since it motivates modeling efforts using one-shot or multiple-interaction games (Bennett and Waltz 2007).

Our work is also related to Rowe and Rrushi (2016), which categorizes methods of deception according to techniques such as impersonation, delays, fakes, camouflage, false excuses, and social engineering. Finally, the present taxonomy can be compared and contrasted to those developed by Heckman et al. (2015). One of these taxonomies breaks down malicious deception by stages, including design of a cover story, planning, execution, monitoring, and so on. Another distinguishes between defensive deceptions based on whether they enable defensive visibility or restrict attacker visibility.[2] To the best of our knowledge, however, none of the related works develops a taxonomy based on game-theoretic distinctions. *Our taxonomy distinguishes between different types of deception precisely to enable accurate game-theoretic modeling of each type.*

2 REVIEW OF GAME-THEORETIC MODELS

In this section, we introduce some of the most common game-theoretic models used in defensive deception for cybersecurity and privacy: Stackelberg, Nash, and signaling games. Table 1

---

[2]Several other works contribute to the effort to create a taxonomy of deception. Michael Handel makes a distinction between deceptions targeted against capabilities and those against intentions (Bennett and Waltz 2007), which we do not explore in the present work. Oltramari and coauthors define an ontology of cybersecurity based on human factors (2014), but this is more focused on trust than deception. Finally, Neil Rowe (2006) describes a taxonomy of deception in cyberspace based on linguistic theory.

Table 1. Players, Types, Actions, and Utility Functions for Three Games

| Players $\mathcal{P}$	Types Θ	Actions $\mathcal{A}$	Utility $\mathcal{U}$	Duration $\mathcal{T}$
Stackelberg game between leader L and follower F	Typically uniform	$L : a_L \in \mathcal{A}_L$ $F : a_F \in \mathcal{A}_F$	$L : U_L(a_L, a_F)$ $F : U_F(a_L, a_F)$	One-shot leader-follower structure
Nash game between semetric players V and W	Typically uniform	$V : a_V \in \mathcal{A}_V$ $W : a_W \in \mathcal{A}_W$	$V : U_V(a_V, a_W)$ $W : U_W(a_V, a_W)$	Simultaneous move structure
Signaling game between sender S and receiver R	S has multiple types $\theta \in \Theta$	$S : a_S \in \mathcal{A}_S$ $R : a_R \in \mathcal{A}_R$	S of each type $\theta \in \Theta : U_S^\theta(a_S, a_R)$ $R : U_R(\theta, a_S, a_R)$	One-shot sender-receiver structure
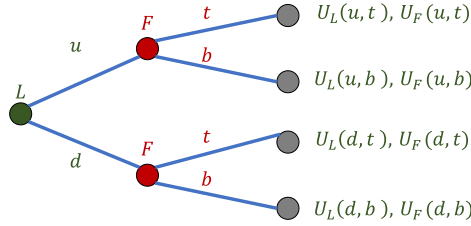
Fig. 2. Stackelberg games consist of a leader L and a follower F. L chooses an action a_L, and F chooses a best response $BR_F(a_L)$. L takes this best response into account when choosing a_L. Stackelberg cybersecurity models often consider the defender to be a leader, and the attacker to be a follower, based on the assumption that the attacker will observe and react to strategies chosen by the defender.

summarizes the components of the games.[3] These components define the structure of the taxonomy that we develop in Section 4.

2.1 Stackelberg Game

Stackelberg games (1934) are perhaps the most fundamental game-theoretic interactions. They are characterized by the following details.

(1) **Players:** $\mathcal{P} = \{L, F\}$, where L is a *leader* and F is a *follower*.
(2) **Actions:** The actions for player L are given by $a_L \in \mathcal{A}_L$. Figure 2 shows a 2×2 game in which $\mathcal{A} = \{u, d\}$, and u denotes moving *up*, while d denotes moving *down*. Player F has actions $a_F \in \mathcal{A}_F = \{t, b\}$, where t denotes *top* and b denotes *bottom*.
(3) **Utilities:** After both players move, L receives utility $U_L(a_L, a_F)$, and F receives utility $U_F(a_L, a_F)$.

In Stackelberg games, the follower moves *after observing the leader's action*. Often, cybersecurity models take the defender as L and the attacker as F, assuming that the attacker will observe and react to defensive strategies.

Stackelberg games are solved backwards in time. Let $\mathcal{P}(\mathcal{S})$ denote the power set of the set $\mathcal{S}$. Then, let $BR_F : \mathcal{A}_L \to \mathcal{P}(\mathcal{A}_F)$ define a *best response function* of the follower to the leader's action.

---

[3]All three games are static games (where we take one-shot interactions to be static, although they are not simultaneous). In addition, while Stackelberg and Nash games typically do not include multiple types, Bayesian Stackelberg and Bayesian Nash games do allow multiple types.
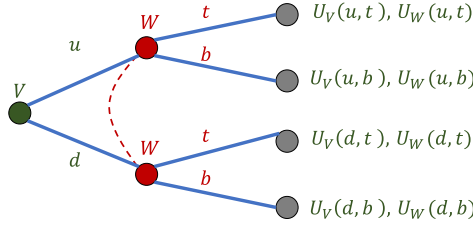
Fig. 3. Nash games are interactions with prior commitment. In this case, the dashed line indicates that W does not know which node describes the game, because she does not know which move V has chosen.

$BR_F(a_L)$ gives the optimal a_F to respond to a_L. The best response function could also include a set of equally good actions. This is the reason for the power set. The best response function is defined by

$$BR_F(a_L) = \arg\max_{a_F \in \mathcal{A}_F} U_F(a_L, a_F). \tag{1}$$

Based on anticipating F's best response, L chooses optimal action a_L^*, which satisfies

$$a_L^* \in \arg\max_{a_L \in \mathcal{A}_L} U_L(a_L, BR_F(a_L)). \tag{2}$$

Then, in equilibrium, the players' actions are (a_L^*, a_F^*), where $a_F^* \in BR_F(a_L^*)$.

2.2 Nash Game

While in Stackelberg games players move at different times, in Nash games (1950) players move simultaneously. More precisely, Nash games are games of *prior commitment*, in which each player commits to his or her strategy before knowing the other player's move. Typically, two-player games of prior commitment are shown in matrix form. Figure 3, however, gives a tree diagram of a two-player game to show the difference between this game and a Stackelberg game. Players V and W act simultaneously, or at least without knowing the other player's action. The dashed line connecting the two nodes for W denotes that W does not know which node the game has reached, because she does not know which move V has chosen.

The concept of *Nash equilibrium* requires each player to choose a strategy that is optimal given the other player's strategy. Let $BR_V : \mathcal{A}_W \to \mathcal{P}(\mathcal{A}_V)$ be defined such that $BR_V(a_W)$ gives the set of actions for V that optimally respond to W's action a_W. Let BR_W be defined similarly. Then a pure strategy Nash equilibrium (1950) is given by a pair (a_V^*, a_W^*) such that

$$a_V^* \in BR_V(a_W^*), \tag{3}$$

$$a_W^* \in BR_W(a_V^*). \tag{4}$$

Nash equilibrium often requires players to choose actions according to probability distributions. These strategies are called *mixed strategies*. Mixed strategies implement the basic idea of randomizing allocations of defense assets to avoid leaving vulnerabilities open to an attacker.

2.3 Signaling Game

Signaling games, like Stackelberg games, are two-player dynamic interactions (Figure 4). Signaling games typically label the players as *sender S* and *receiver R*. The sender has access to some information unknown to the receiver. This is called the sender's *type* $\theta \in \Theta$. The receiver only learns about the type based on the sender's action. For this reason, the sender's action (here a_S) is referred to as a *message*. The message need not correspond to the sender's type.
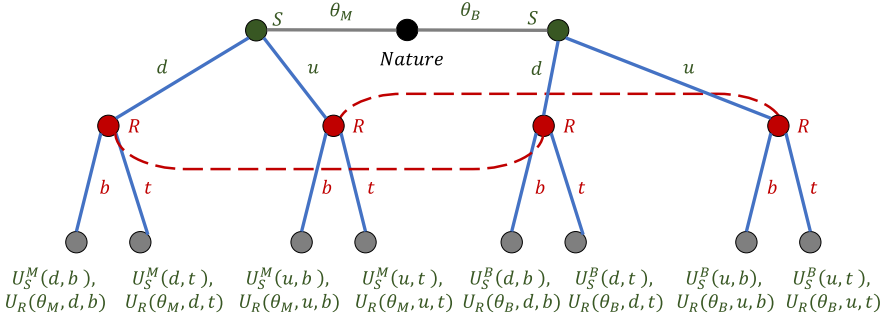
Fig. 4. The figure depicts a signaling game in which a sender S, who has access to private information, transmits a message to a receiver R. The message is not verifiable, so R does not know the underlying information with certainty. In separating and partially separating equilibria, however, it is incentive-compatible for S to transmit a message that at least partially reveals his private information.

In this example, the set of types of S is $\Theta = \{\theta_B, \theta_M\}$, where θ_B represents a *benign* sender, and θ_M represents a *malicious* sender. Let $p(\theta)$ denote the *prior probability* with which S has each type $\theta \in \Theta$. The utility functions depend on type. $U_S^M(a_S, a_R)$ and $U_S^B(a_S, a_R)$ give the utility functions for malicious and benign senders, respectively. $U_R(\theta, a_S, a_R)$ gives the utility function for the receiver when the type of S is θ, the sender message is a_S, and the follower action is a_R. R forms a *belief* $\gamma(\theta \mid a_S)$ that S has type θ given that he sends message $a_S \in \mathcal{A}_S$. To be consistent, this belief should be updated in accord with Bayes' law. A strategy pair in which S and R maximize their own utilities, together with a belief that is consistent, form a *perfect Bayesian Nash equilibrium* (PBNE) (cf. Fudenberg and Tirole (1991)). In some PBNE, S can cause R to form a specific false belief.

3 LITERATURE SURVEY

In this section, we survey existing literature in game-theoretic approaches to defensive deception for cybersecurity and privacy. Tables 3 and 4 in the Appendix list the papers, which we now discuss from top to bottom.

3.1 Perturbation

The first set of papers study privacy that is obtained by perturbing sensitive data. Chessa et al. (2015) model the interaction between a set of users who contribute data to identify mechanisms by which the learning agent can improve the quality of its estimation. The users have two strategic variables: whether to contribute at all, and the level of precision to use if they do contribute. They face a trade-off between privacy and the benefit of contributing to data analysis. The paper takes this benefit—although possibly difficult to quantify—to be a non-monetary interest in data as a public good. Chessa et al. study both homogeneous and heterogeneous populations of users. The interactions between these populations of users are modeled as Nash games. After finding that the equilibrium in these games is worse than the population-optimal behavior due to the selfishness of the agents, the authors consider a mechanism for the data collector to improve the equilibrium. They find that an analyst can shift the equilibrium of the game toward more accurate estimation by imposing a minimum level of precision in the strategy space of the users. The paper studies computation of the average of a data set as a foundation for more complex statistics (Chessa et al. 2015).

Shokri (2015) also addresses the issue of data privacy for users who send private information to an untrusted source. Shorki's work concerns two types of privacy, differential privacy and
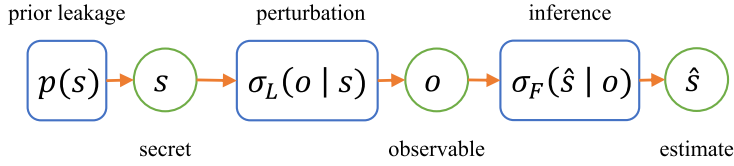
Fig. 5. In the Information Sharing Framework from Shorki (2015), a user hides a secret s by releasing an observable o according to the channel probability $\sigma_L(o|s)$. The adversary uses inference $\sigma_F(\hat{s}|o)$ to obtain estimate $\hat{s}$. This is a Stackelberg game, because the adversary knows the perturbation mechanism when he designs an inference attack.

distortion privacy. Differential privacy obtains a guaranteed limit on the amount of information that users leak to an attacker, but does not tell the user the actual level of privacy achieved. Distortion privacy is a measure of the actual error of the attacker's inferences based on the information that is leaked. The calculation of this metric, however, requires that the attacker have prior knowledge before making an observation. Using a non-zero-sum Stackelberg game, Shokri develops a model for joint distortion-differential privacy. Figure 5 depicts the interaction. The user leads by choosing a protection mechanism. The attacker follows, designing an inference attack and aiming to minimize user privacy. The inference attack is the best response to the user's protection mechanism. The author develops a linear program to obtain the equilibrium (Shokri 2015).

Alvim et al. (2017) study information privacy within the setting of information theory. Specifically, their paper formulates information transmission and leakage using quantitative information flow. A sender possesses a secret, which a defender only knows with some prior probability. The sender transmits the secret through a channel, and the channel leaks some information to the adversary. Using this information, the adversary forms an *a posteriori* belief about the secret. A quantity called the posterior vulnerability of the secret quantifies the degree of the adversary's knowledge. The quantification can be the entropy of the posterior distributions, although a general convex function can also be used. Alvim et al. observe that the expected utility becomes convex, rather than linear, in the mixed-strategy probabilities. For this reason, they call the game an *information leakage game* (Alvim et al. 2017). This formulation makes useful connections to the abundant tools available in information theory.

Theodorakopoulos et al. (2014) address the issue of privacy for location-based services (LBS). The authors emphasize that human locations are not found at discrete spatial-temporal points, but rather follow a trajectory. Theodorakopoulos et al. propose user-centric location privacy preserving mechanisms (LPPMs) that send LBS pseudolocations to protect not only the privacy of past and future locations but also transitions between locations and locations between LBS queries. As a Bayesian Stackelberg game leader, the LPPM strategically sends a perturbed pseudolocation to the LBS to protect its exact real location. Privacy provided by the LPPM is measured by the error with which an adversary deduces the real location of the target. Quality is determined by how much the altered location data from the LPPM affects the functionality of the LBS. Both privacy and quality values are computed for users that prioritize (1) protecting past locations or (2) protecting future locations. The authors' contribution is to improve upon "trajectory-oblivious" LPPMs. They evaluate the LPPM using a set of mobility traces for taxi cabs in the San Francisco Bay area (Theodorakopoulos et al. 2014).

Of course, a large body of literature studies the use of perturbation to protect privacy (especially as quantified by differential privacy). We have only focused on the subset of this literature that uses game theory. We also note that perturbation is very similar to obfuscation, which we survey in Section 3.3. Section 4 defines the difference between the two types and explains why we have categorized some papers as perturbation but others as obfuscation.
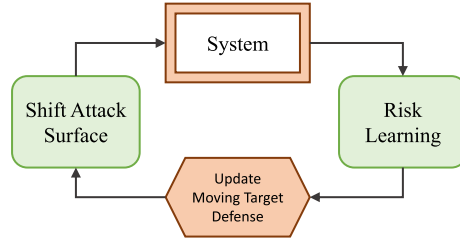
Fig. 6. The moving target defense designed in Zhu and Başar (2013) updates the configuration of a system each round based on learning the risk of attack that the system faces.

3.2 Moving Target Defense

Next, we discuss papers that study moving target defense. Moving target defense invokes concepts of agility, changing attack surfaces, and random configurations. Two different game-theoretic concepts are often used to model moving target defense. They differ in whether time is considered explicitly. One approach is to model randomness using mixed strategies, with the implicit understanding that new realizations from a random variable will be performed over time. A recent paper by Rass et al. notes a shortcoming of this approach: utility functions based on mixed strategies do not inherently capture the cost of switching from one pure strategy to another over time (Rass et al. 2017). The second approach to moving target defense is to use Markov decision processes. In Markov decision processes for moving target defense, the defense configuration is considered a state, and switching costs can be modeled explicitly.

Sengupta et al. (2018) consider the placement of intrusion detection systems (IDS) in a cloud-based architecture. Two cloud servers and a cloud controller are connected to the Internet via a physical router. An attacker who may be located either outside or inside the network attempts to gain access. The network designer is able to place a limited number of IDS. The design problem is where to place them. To limit the knowledge of the attacker, these placements are performed probabilistically. Hence, this is a moving target defense. First the defender chooses a probability distribution for the locations of the IDS. Then the attacker observes this distribution and he chooses which system to attack. The interaction is modeled by a Stackelberg game with mixed strategies. The game can be solved by a mixed-integer linear program, although ultimately the most efficient formulation is as a mixed-integer quadratic program, using a branch-and-cut algorithm. One interesting contribution of this paper is that the authors include a concise taxonomy of moving target defense. They break moving target defense into four categories: shifting the exploration surface, detection surface, attack surface, or prevention surface. This could be useful to add a layer to the taxonomy proposed in the present article.

Zhu and Başar (2013) use Markov decision processes to model moving target defense for network security. An attacker and defender play a series of zero-sum Nash games in which the defender chooses the arrangement of systems with various vulnerabilities, and the attacker selects an attack path that depends for its success on the arrangement of the vulnerabilities. See Figure 6. Tools from control theory are used to analyze the dynamics of the multiple-round interaction. Zhu and Başar show that the steady states of the learning dynamics correspond to equilibrium points in the Nash games. The authors illustrate the game and resulting dynamics with a numerical example. This work points toward modeling the interaction between the attacker and defender as a stochastic game in which the defender would attempt to minimize the total risk over the whole game, rather than optimizing his risk at each individual stage (Zhu and Başar 2013).

Feng et al. (2017) study moving target defense using a combination of a Markov decision process and a Stackelberg game between an attacker and a defender of an abstract resource such as a
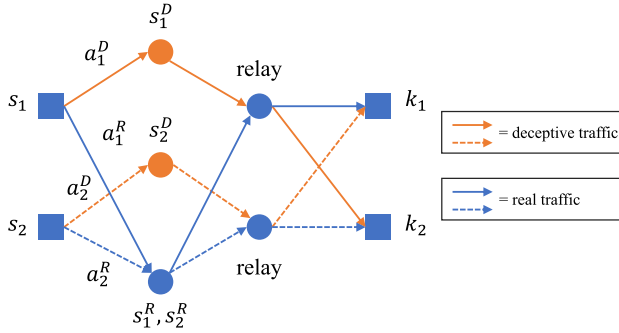
Fig. 7. Deceptive routing (Clark et al. 2012) protects traffic routed from source s_1 to sink k_1 and source s_2 to sink k_2 from jamming attacks. Player 1 sends real traffic with volume a_1^R to legitimate node s_1^R but also sends decoy traffic with volume a_1^D to deceptive node s_1^D. Player 2 has a similar strategy. With some probability, the adversary attacks the deceptive rather than the real traffic.

network asset. The defender begins each turn by claiming a moving target defense strategy, which lasts for $\mathcal{T}$ periods. The defender controls a Markov decision process in which the states represent security configurations of the system. Then the attacker follows by choosing one state to attack. If the attack is successful, then the attacker obtains a reward. The problem can be expressed in min-max form, which the authors show can be reduced to simply a minimization problem. They obtain an algorithm that has $O(|V|^2 \log |V|)$ complexity, where $|V|$ is the number of valid states. One insight from the paper is that the defender's cost can be reduced more easily by increasing the degree of the switching graph than by reducing the costs of switching (Feng et al. 2017).

3.3 Obfuscation

The third set of papers relates to hiding valuable information using external noise. Clark et al. (2012) model a network in which information is sent from a source to a destination over multiple nodes. See Figure 7. The interaction is modeled using a Stackelberg game in which the defender leads by choosing the flow rates of deceptive and real traffic. The attacker observes these flows, and then follows by choosing a path to attack. The authors consider situations in which sources do and do not cooperate to obfuscate their traffic. They model the compounded pathway congestion that comes with the addition of fake data, and the possibility that information sources may take into account the delays being experienced by other nodes in the network when determining their own data pathways and flow rates. A simulation illustrates that altruistic sources improve node functionality. Clark et al. suggest that future work can study other measurements of the efficacy of deception and an incomplete information model in which nodes are only partially informed about the attacker's cost and utility functions (2012).

Zhu et al. (2012) build upon this idea to consider both (1) deceptive routing, and (2) strategic choices of flow rates. To model both of these conflicts, the authors formulate multiple-player Stackelberg games in which the defender leads and the attackers follow. The routing game requires an extension to Stackelberg games to allow $N + 3$ players, where N is the number of attackers. Zhu et al. quantify the effectiveness of the mechanism using a metric called *value of deception*, a ratio of the utility achieved with and without the use of deception. Simulations illustrate that defender utility decreases approximately asymptotically in the attacker's budget, and increases approximately linearly in the number of hops. Future work could include the introduction of learning to similar algorithms, as well as the application of the model to multiple routing pathways (Zhu et al. 2012).

While Clark et al. (2012) and Zhu et al. (2012) apply obfuscation to security, this type of deception also applies directly to privacy. Two recent papers by Pawlick and Zhu (2016, 2017a) model the long-run impact of obfuscation technologies such as *TrackMeNot* (Howe and Nissenbaum 2009) (a browser add-on that issues randomized search engine queries to obfuscate a user's search profile) and *ScareMail* (Grosser 2014) (which appends security-sensitive words to the end of every email in an effort to undermine mass surveillance). These tools add irrelevant data (e.g., randomized browser searches) to relevant data (e.g., real browser searches). The first paper develops a Stackelberg game in which the leader is a machine learning or tracking agent and the follower is a user. The learning agent has the option to promise a level of differential privacy protection. The user reacts to this level and chooses whether to obfuscate the data himself. Under certain conditions, it is incentive-compatible for the learning agent to promise some level of privacy protection to avoid user obfuscation (Pawlick and Zhu 2016). The second paper extends this scenario to multiple users through a mean-field game model. In a mean-field game, each user responds to the average obfuscation of the field of other users. This extended model predicts a threshold of perturbation beyond which users will adopt obfuscation in a cascading manner. The paper also identifies conditions under which obfuscation does not motivate privacy protection from the learner but only results in data pollution (Pawlick and Zhu 2017a). These papers could benefit from future measurements of users' preferences in experimental or empirical settings.

3.4 Mixing

Techniques for both security and privacy use the idea of mixing to prevent linkability. Zhang et al. (2010) study anonymity of the *Tor* network. When information is sent over *Tor*, it passes through three nodes before reaching its destination, in an attempt to make the entry and exit nodes unlinkable. Zhang et al. point out that since connectivity on *Tor* is volunteer-based, attackers could collude to link the entry and exit nodes. They propose *gPath*, which operates on three principles: (1) an exit-eligible node should not be considered for the entry, (2) the middle node should have equal or greater bandwidth than the entry node, and (3) the exit node should be chosen based on the bandwidths of the entry and middle nodes. The authors construct utility functions for the user and attacker. The players use these functions to compute best responses to each other's strategies in four rounds. This is different from Nash equilibrium, since the strategies do not necessarily converge to mutual best responses. Simulations performed on a snapshot of the *Tor* network from 2010 suggest that *gPath* reduces the probability of a successful attack (Zhang et al. 2010).

The idea of mixing can also be applied to cyber-physical systems such as vehicle ad hoc networks (VANETs). Freudiger et al. (2009) use locations called *mix zones* for location privacy in VANETs. Vehicles in these networks communicate using pseudonyms to achieve anonymity. Nevertheless, a constant trace of communications from a single pseudonym creates a link between a vehicle's origin and destination. One technology proposed to overcome this linkability is pseudonym swap in mix zones (Figure 8). At locations such as intersections, multiple vehicles simultaneously exchange their pseudonyms, so that the identities of vehicles that enter the swap zone can only be linked with the identities of vehicles that exit the swap zone with some probability. Freudiger et al. note that this is a strategic interaction, because swapping involves some cost. Selfish actors may decline to swap. Nash equilibrium is used, since pseudonym swapping is a simultaneous interaction. Freudiger et al. consider both complete and incomplete information games. In the incomplete information games, each player is unaware of the privacy loss that the other players have incurred since their last pseudonym swap (and thus their willingness to swap again). Interestingly, VANET nodes tend to swap pseudonyms more often when pseudonym swapping is expensive than when it is cheap (Freudiger et al. 2009). One challenge is that swap locations are presumably those in which communication is most necessary and should not be interrupted (such as at an intersection).
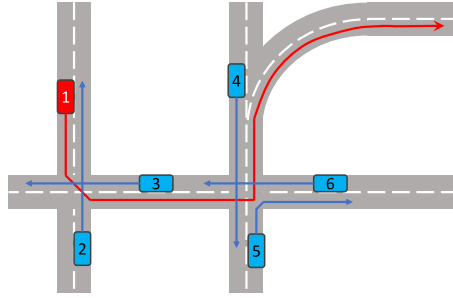
Fig. 8. Freudiger et al. (2009) study pseudonym swapping for location privacy in vehicle ad hoc networks. As Vehicle 1 turns left at its first intersection, it can swap pseudonyms with Vehicles 2 and 3. At its second intersection, it can swap with Vehicles 4, 5, and 6. If these other vehicles have recently swapped, however, then they have little incentive to swap again. Thus, the interactions become multiple-player, simultaneous-move games.

Lu et al. (2012) address this dilemma by proposing pseudonym changes at social spots, where many vehicles often stop for a period of time. Small social spots, e.g., stoplights, are places where vehicles stay briefly. Large social spots, such as parking lots, are places where vehicles remain for longer duration. A privacy mechanism—Key-insulated Pseudonym Self-Delegation (KPSD)—is proposed. Authorized keys (pseudonyms) for each vehicle are kept in a secure location by a Trusted Authority (TA), usually at the car owner's home. The TA loads onto the OnBoard Unit (OBU) a set of temporary keys before each new trip. As in Freudiger et al. (2009), the adversary is global (i.e., has full access to radio network safety messages that provide vehicle time, location, speed, content and pseudonym) and external (only listening in on conversations, not engaging in vehicle sabotage), with access to the vehicle's OBU. This KPSD model is also unique compared to other models because of its ability to prevent theft. Since pseudonyms are changed at social spots and since the keys are not kept with the vehicle but instead with the TA, a thief cannot generate temporary keys needed for vehicle operation. Lu et al. incorporate a game-theoretic argument (based on Nash equilibrium) to discuss the scenarios under which KPSD is incentive-compatible (Lu et al. 2012).

3.5 Honey-X

By honey-x, we refer to deception that uses techniques named with the prefix honey (e.g., honeypot, honeynet, honeybot, etc.). In many implementations of these techniques, defenders have a limited capacity to monitor honeypots, and attackers have constrained time and available technologies to detect honeypots. Therefore, game-theoretic trade-offs arise. Carroll and Grosu (2011) study a network in which a defender can disguise honeypots as real systems (or reveal them as honeypots) and real systems as honeypots (or reveal them as real systems). The goal is to make attackers waste resources attacking honeypots and avoiding real systems. In the proposed signaling game, the defender is the sender and the attacker is the receiver. The equilibrium concept is PBNE. Because the defender never wants to reveal the system type to the attacker, the game does not support any separating equilibria. Carroll and Grosu analyze both pooling and hybrid (partially separating) equilibria. Hybrid equilibria demonstrate cases in which the defender has an incentive to partially reveal the actual system type (Carroll and Grosu 2011).

This work by Carroll and Grosu has inspired several other papers. Çeker et al. (2016) build on this model for honeypot deployment in the setting of distributed denial-of-service attacks. This paper features one major difference: in Çeker et al. (2016), messages are costly, while in Carroll and Grosu
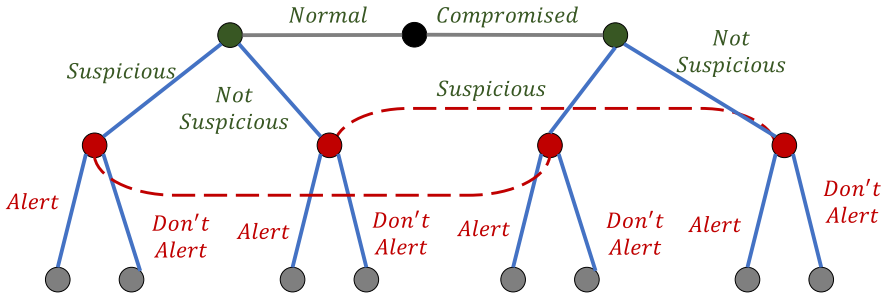
Fig. 9. Mohammadi et al. (2016) use signaling games to model the use of fake avatars as a type of honey-x. The avatars observe suspicious or non-suspicious activity from other user accounts in the network and must form a belief about whether the other users are legitimate accounts or compromised accounts. The goal of the avatars is to correctly report compromised accounts, while avoiding false positives.

(2011) they are costless. In other words, defenders in Çeker et al. (2016) must pay a cost to make normal systems appear to be honeypots or honeypots appear to be normal systems. Because of this cost, separating equilibria emerge in some parameter regimes. Another feature of Çeker et al. (2016) is that the authors propose a parameter valuation method to obtain the utility functions using existing security evaluations. The authors derive a complete set of closed form solutions.

Mohammadi et al. (2016) use signaling games to model the decision processes of a fake avatar that is defending a social network from attack. Figure 9 depicts the model using the signaling game formulated in Figure 4. A user in the network may be normal or compromised. It interacts with an avatar in suspicious or not suspicious ways, and this interaction is taken to be a signaling-game message. The fake avatar observes this interaction, and forms a belief about the real type of the user. The avatar then decides whether to raise an alert. Mohammadi et al. (2016) is not a cheap-talk game, because raising an alert is assumed to be more costly than not raising an alert. But this extra cost of raising an alert applies regardless of the actual user type, so it does not take the form of a "lying cost," as is the case in Çeker et al. (2016). Due to this different structure, the game in Mohammadi et al. (2016) does not support separating equilibria. In the pooling equilibria, the fake avatar alerts the system of an attack if the prior probability that the user is an attacker is sufficiently high (Mohammadi et al. 2016).

Pawlick and Zhu present a final perspective on honeypot deployment using signaling games (2018). They use a cheap-talk game with binary type, message, and action spaces. The unique element, however, is a "detector" that gives off probabilistic evidence when the sender misrepresents his type. They call this model a *signaling game with evidence* (2015, 2018). This model is capable of nesting cheap-talk signaling games (cf. Crawford and Sobel (1982)) and games of verifiable disclosure (cf. Milgrom (1981)) as special cases. The relevant equilibrium concept is an extended version of PBNE. The authors find that the ability to detect honeypots always improves the utility of the attacker, but surprisingly sometimes also improves the utility of the defender. This suggests that in certain parameter regions, it is advantageous for the defender to use imperfect deception. They also find that detectors that prioritize high true-positive rates over low false-positive rates encourage truthful signaling. Finally, they show that receivers prefer equal-error-rate detectors (Pawlick et al. 2018).

Píbil et al. (2012) and Kiekintveld et al. (2015) create a different type of imperfect information model called a honeypot selection game. The honeypot selection game classifies targets into different levels of importance. These are assigned corresponding utility values for the attacker, and the defender plays to minimize the attacker value. The defender disguises the honeypots by choosing

the utility value that they appear to have. A configuration of the network is represented by a vector of the real values of each real machine, and the simulated values of the honeypots. The Bayesian Nash equilibrium can be found using a linear program. The authors examine the defender's optimal strategy and equilibrium utility as a function of the number of honeypots that he deploys. Keintveld et al. also study an extension of the game that allows attackers to probe potential targets before choosing whether to attack. The probe returns accurate or inaccurate responses with some probability. The game is again zero-sum, and the analysis proceeds using a linear program (Píbil et al. 2012; Kiekintveld et al. 2015).

3.6 Attacker Engagement

Most game-theoretic models of deception are either static games (such as simultaneous move Nash games) or single-shot dynamic games (such as Stackelberg games or signaling games). But some initial work has investigated multiple-period games. We use "dynamic" to refer to games having multiple periods, and we call these interactions attacker engagement.

Zhuang et al. (2010) create a multiple-period signaling game that concerns deception and resource allocation. This paper is about homeland security, in general, and only briefly mentions computer security. Still, it is interesting due to its technical setup and insights. In each period, the defender chooses to invest in either short term "expense" or long term "capital" investments. In addition, the defender chooses how much of this to reveal to the attacker using a signal. The defender may signal truthfully or deceptively, or she may act secretively, i.e., not reveal any information. The attacker observes the defender's signal, updates his belief, and chooses whether to attack. Games are constructed in which the attacker can and cannot observe previous defender investment strategies. The games are solved through backwards induction. In several cases, the defender can deceive the attacker, often by mimicking the strategy of the opposite defender type. For example, when the attacker cannot observe previous defender strategies and has a high cost of attack, the defender can deter an attack by choosing an expensive short term defense rather than long term investment. Zhuang et al. also discuss remaining challenges involved in the area of equilibrium selection (2010).

Durkota et al. (2015) use attack graphs to represent attacker strategies and develop network-hardening methods to increase security. They propose a Stackelberg game in which the defender leads by placing honeypots in the network. The attacker follows and has knowledge of the number of honeypots (not their identities). Attack graphs are used to capture the adversary's multiple-round strategies. Durkota et al. argue that vulnerability databases can be used to construct a menu of possible attack graphs that the adversary can employ, assuming that the defender knows which systems, software, and so on, are present in the network. The defense strategy is obtained using Markov decision processes. One contribution of the paper is an efficient algorithm to solve the Markov decision process using policy search with pruning (Durkota et al. 2015). For the purposes of the classification in Section 4, it is interesting to note that the paper builds upon the foundation of honeypot deployment, but adds the dynamic element of multiple-round attack graphs.

Horák et al. (2017) model the penetration of an attacker into a network using a one-sided partially observable stochastic game. Unlike most approaches to network security, Horák et al. consider the defender rather than the attacker to be the informed player. Instead of ejecting an attacker immediately, the defender decides how long to observe the attacker to gather information. Figure 10 depicts the set of states and transitions. The attacker compromises the network in layers and chooses when to exfiltrate data without knowing for sure whether she has been detected. In equilibrium, the defender's optimal strategy is to keep the attacker inside the network while the attacker is confident that he has not yet been detected and to eject the attacker from the network otherwise. The authors also evaluate the robustness of the solution if the attacker does not know
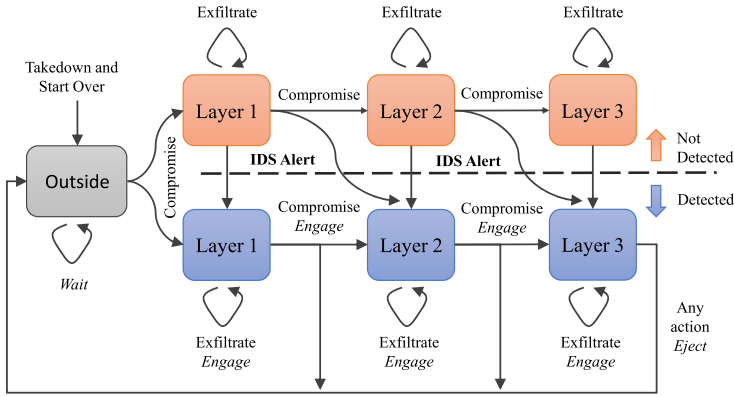
Fig. 10. Horák et al. (2017) model attacker engagement using a one-sided partially observable stochastic game. From left to right, states correspond to layers of a network. In the states at the top of the figure, the attacker is undetected. When the IDS detects the attacker, the state moves to the bottom of the figure. Defender actions are shown in italics. The defender decides when to engage the attacker and gather information rather than eject the attacker.

the capabilities of the IDS. While this paper studies a simplified network structure, it offers an uncommon analysis of the optimal strategies for a powerful and observant defender (Horák et al. 2017).

4 TAXONOMY

The U.S. Department of Defense has highlighted the need for "the construction of a common language and a set of basic concepts about which the security community can develop a shared understanding" (MITRE 2010). More specifically, the term deception has been employed broadly and with a variety of meanings. A "common language and a shared set of basic concepts" would allow deception researchers to better collaborate. It would also further the development of cybersecurity and privacy as a science. Finally, a precise taxonomy would allow game theorists to carefully select models that capture the unique features of each different type of deception.

In Section 3, we sorted deception into six different types: perturbation, moving target defense, obfuscation, mixing, honey-x, and attacker engagement. In this section, we first give general descriptions of the six types. Then, we construct a taxonomy that precisely defines each of the types using game-theoretic principles.

(1) In the area of privacy, deception can limit leakage of sensitive information through the use of noise. This type of deception is often called *perturbation* (Section 3.1).
(2) Deception can limit the effectiveness of attacker reconnaissance through techniques such as randomization and reconfiguration of networks, assets, and defense tools. This is called *moving target defense* (Section 3.2).
(3) Deception can waste effort and resources of attackers by directing them to decoy targets rather than real assets and can protect privacy by revealing useless information aside real information. This is called *obfuscation* (Section 3.3).
(4) Deception can use exchange systems, such as mix networks and mix zones, to prevent linkability. We call this type of deception *mixing* (Section 3.4).
(5) Deception can draw attackers toward specific systems (such as honeypots) by disguising these systems as valuable network assets. To include honeynets, honey-users, and so on, we call this type of deception *honey-x* (Section 3.5).

Fig. 11. We use the term *mimesis* to signify the creation of a specific false belief. In this example, a defender makes an adversary believe that a database exists when, in fact, it does not exist.

Fig. 12. We use the term *crypsis* to refer to preventing an adversary from acquiring a true belief. In this example, a defender hides the true existence of a database from an adversary.

(6) Deception can use feedback to dynamically influence attackers over an extended period of time, to waste their resources and gather intelligence about them. We call this *attacker engagement* (Section 3.6).

4.1 Detailed Definition of Each Type of Deception

A taxonomy is "a collection of controlled vocabulary terms organized into a hierarchical structure" in which each term "is in one or more parent/child (broader/narrower) relationships to other terms in the taxonomy" (NISO 2005). Traditionally, each parent is called a *genus*, and each child is called a *species* (Chisholm 1911). Each species is demarcated from other members of its genus by a *specific difference*. The most fine-grained species in a taxonomy are sometimes called the *infimae species*. In our taxonomy, we can now identify the "types" of deception more formally as infimae species.

Since the purpose of our taxonomy is to identify appropriate game-theoretic models for each infima species of deception, the specific differences that we use are related to game theory (see Table 1). These are the principles of *private information* Θ, *players* $\mathcal{P}$, *actions* $\mathcal{A}$, and *time-horizon* $\mathcal{T}$.

4.1.1 Private Information. One of these principles is the private information. To deceive is to intentionally cause another agent either "to acquire or continue to have a false belief" or "to be prevented from acquiring or cease to have a true belief" (Mahon 2016). These two categories are quite different. Figures 11 and 12 depict both categories. The first category involves instilling in another a specific falsity (Figure 11). In the language of game theory, this requires the creation of a belief. In signaling games or partially observable stochastic games, agents maintain beliefs over the private information Θ of other agents. Deceptive actions manipulate these beliefs to create traps or decoys. However, the second category of deception involves hiding a true belief (Figure 12). This could be employed in cybersecurity to hide a weakness, or it could be deployed in privacy to protect sensitive information. In either case, the defender maximizes her utility function by causing attackers to obtain noisy or uncertain information about the state of reality.

Rothstein and Whaley have distinguished between these two categories as simulation (or "showing the false") and dissimulation (or "hiding the real") (2013). Under simulation, they include subcategories called masking, repackaging, and dazzling. Under dissimulation, they include mimicking, inventing, and decoying. While the distinction between hiding the real and showing the false is appropriate for our purposes, we prefer not to adopt the same subcategories for several reasons. In particular, Rothstein and Whaley (2013) includes decoying as a type of "showing the false." But,
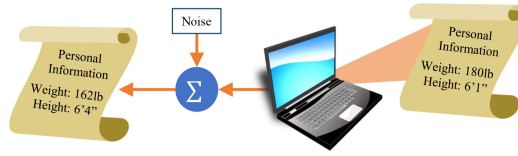
Fig. 13. Intensive deception. The defender alters the same object that is being hidden. In this example, the defender adds noise to private data.

Fig. 14. Extensive deception. The defender hides an object using other objects in the environment. In this example, the defender dynamically changes the location of the private data.

we prefer to label it as "hiding the real." In the literature that we review, decoys are simulated entities, but they are used mostly to hide the presence of the real entity.[4]

Therefore, we prefer a new terminology. Biologists distinguish hiding the real and showing the false using the terms *crypsis* and *mimesis*. "In the former . . . an animal resembles some object which is of no interest to its enemy, and in doing so is concealed; in the latter . . . an animal resembles an object which is well known . . . and in so doing becomes conspicuous" (Cott 1940). We adopt these terms to signify the corresponding categories of cybersecurity deception.

4.1.2   Actors. The next specific difference is the set of actors or players $\mathcal{P}$ involved in deception. Based on the actors involved, cryptic deception can be divided into *intensive* and *extensive* deception (see Figures 13 and 14). Intensive deception modifies an actor (or its own representation) to hide it (Figure 13). For example, some privacy techniques add noise to data about a user before publishing it. The user's own data is modified. By contrast, extensive deception is that category of deception that hides an actor by using outside actors (Figure 14). In cyberspace, examples include mix networks for anonymity. Many messages enter a network, where they are mixed in a chain of proxy servers before leaving the network. Any given message is hidden because of the presence of the other messages.

4.1.3   Actions. Deceivers can act in a number of ways, or actions $\mathcal{A}$. Within cryptic deception, we distinguish between deception that uses information and deception that uses motion. Deception that uses information tends to manipulate the data released about agents' properties, while deception that uses motion either modifies these properties over time or realizes these properties from a random variable. In other words, the first category is associated with creating noise, while the second category is associated with concepts such as agility and randomization.

4.1.4   Duration. Within mimetic deception, we distinguish between static and dynamic scenarios. Dynamic games feature multiple interactions, while static games consist of only one interaction.[5] Currently, the majority of game-theoretic defensive deceptions use static models. The most

---

[4]The use of decoys pertains to obfuscation (Section 3.3). We place technologies such as honeypots (Section 3.5) in a separate category, because they are designed to attract attacker attention, not simply to hide the presence of real systems.
[5]Here, we consider one-shot games to be static even if they are not simultaneous-move games.
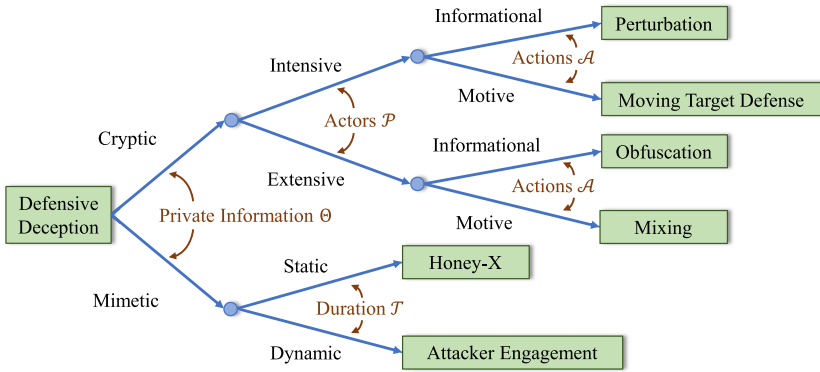
Fig. 15. Tree diagram breakdown of deception into various species. The specific differences correspond to the game-theoretic notions of private information, actors, actions, and duration. We describe a set of infimae species called perturbation, moving target defense, obfuscation, mixing, honey-x, and attacker engagement.

Table 2. Infimae Species and Definition

Infima Species	Definition
Perturbation	Cryptic, intensive, informational deception
Moving Target Defense	Cryptic, intensive, motive deception
Obfuscation	Cryptic, extensive, informational deception
Mixing	Cryptic, extensive, motive deception
Honey-X	Mimetic, static deception
Attacker Engagement	Mimetic, dynamic deception

popular set of mimetic deceptions that are static are honeypots, honeynets, and honeytokens, to which we refer generically using the term *honey-x*. We use the name *attacker engagement* to denote all forms of mimetic deception that are dynamic.

4.2 Synthesizing the Taxonomy

Figure 15 uses the specific differences of private information, actors, actions, and duration to create a taxonomy that breaks the genus of deception into multiple levels of species, ending with the infimae species of perturbation, moving target defense, obfuscation, and so on. Based on the taxonomy, Table 2 lists the definitions of each of the infimae species. The definitions are unambiguous as long as the specific differences are clearly defined. They are also mutually exclusive, because we have used mutually exclusive specific differences. Finally, note that the order of application of the specific differences does not matter. This implies that the taxonomy could be also represented using a binary lattice in four dimensions, one dimension for each of the specific differences.

4.2.1 Classification of Literature. In the Appendix, Tables 3 and 4 justify the classification of each paper into its infimae species based on the value of its specific differences. Most classifications are straightforward. To evaluate the structure of the taxonomy, however, consider several border cases.

First, Shokri (2015) uses the term "obfuscation," but we have classified the deception in this paper as perturbation. Noise is applied to the valuable information itself, so our taxonomy categorizes it as intrinsic rather than extrinsic. Nevertheless, obfuscation and perturbation are closely related. If

	Motive	
Moving Target Defense Feng. et al. 2017 − Stack. (uses MDP) Sengupta et al. 2018 − Stack. (with leader mixed-strategies)		Mixing Zhang et al. 2010 − Best response in multiple stages (user-adversary) Freudiger et al. 2009 − Nash (user-user) Lu et al. 2012 − Nash (user-user)
Rass et al. 2017 − Nash (mixed-strategies) Zhu and Başar 2013 − Nash (mixed-strategies)		
Perturbation Chessa et al. 2015 − Nash (user-user) Alvim et al. 2017 − Nash (utilizes are a priori)		Obfuscation Clark et al. 2012 − Stack. (user-adversary) Zhu et al. 2012 − Stack (user-adversary) Pawlick and Zhu 2016 − Stack. (user-adversary)
Shorki 2015 − Stack. (user-adversary) Theodorakopoulos et al. 2017 − Stack. (user-adversary)		Pawlick and Zhu 2017a − Stack (user-adversary) and Mean-Field Game (user-user)

Intrinsic   Extrinsic

Fig. 16. The game-theoretic model is listed for each paper in cryptic deception, to compare and contrast the models used for each infimae species. Parenthetical remarks give reasons for the chosen models.

	Motive	
Moving Target Defense Games which include either mixed strategies or Markov decision processes		Mixing Nash Games
Perturbation	and	Obfuscation
Nash games or mean-field games for user-user interactions Stackelberg games for user-adversary interactions		

Intrinsic   Extrinsic

Fig. 17. Several trends are apparent in the mapping between infimae species and game-theoretic models.

we view each specific difference of an infimae species as a bit, then we can say that obfuscation and perturbation have a Hamming distance of one.

Second, consider Zhu et al. (2012). The main idea of the paper is to protect valuable network traffic by sending external decoy traffic. Therefore, the infimae species is obfuscation. But the paper also considers mixed strategies, which implement the agility or motion that is characteristic of moving target defense. Again, obfuscation and moving target defense have a Hamming distance of one.

A third border case is Zhang et al. (2010), which studies protection of traffic in the *Tor* network. This application is similar to the deceptive routing problems in Clark et al. (2012) and Zhu et al. (2012), which we have classified as obfuscation. Still, we have classified (Zhang et al. 2010) as mixing, because *Tor* requires the participation of other users in the mix network; therefore, it is extrinsic.

4.2.2 Observed Relationships Between Models and Infimae Species. Since our taxonomy differentiates species of deception based on game theory, it is interesting to see which games were used to model each species. Figure 16 lists the papers from Section 3 that study crypsis. Most papers use Nash games and Stackelberg games. The papers that use Nash games (or related concepts) are listed on an orange background, and the papers that use Stackelberg games are listed on a blue background. The figure reveals that authors do not concur on a one-to-one mapping between species and games. This is to be expected, since (1) different games capture different aspects of the same interaction, (2) modeling approaches are still evolving in this nascent field, and (3) the cybersecurity landscape itself is evolving.

Nevertheless, some trends are evident. These are depicted in Figure 17. First, all of the papers on mixing use Nash games or related concepts. (Zhang et al. (2010) use several rounds of

Honey-X	Attacker Engagement
Carroll and Grosu 2011 − Signaling Çeker et al. 2016 − Signaling Mohammadi et al. 2016 − Signaling Pawlick and Zhu 2018 − Signaling	Zhuang et al. 2010 − Multi-Period Signaling Durkota et al. 2015 − Stackelberg (with Markov decision process)
Píbil et al. 2012 − Bayesian Nash Kiekintveld et al. 2015 − Bayesian Nash	Horák et al. 2017 − One-sided partially- observable stochastic game
Static	Dynamic

Fig. 18. The game-theoretic model is listed for each paper in mimetic deception, to compare and contrast the models used for each infimae species.

Honey-X	Attacker Engagement
Emphasis on attacker belief: Signaling games Emphasis on defender defense allocation: Bayesian Nash games	Multiple-period games Interaction between games and MDPs One-sided stochastic games
Static	Dynamic

Fig. 19. Trends are indicated in each species of mimesis.

best-response, which is based on the idea of Nash equilibrium.) This is logical, because each user in a mix network or a mix zone simultaneously decides whether to participate. A second trend spans both perturbation and obfuscation. Papers that study user-user interactions are modeled using Nash games (or in one case, a mean-field game), and papers that study user-adversary interactions are generally modeled using Stackelberg games. Finally, moving target defense tends to be modeled in two different ways. One way is to use mixed strategies to model randomness in defensive configurations. The second way is to use Markov decision processes to explicitly model the temporal component of changing defensive characterizations.

Figures 18 and 19 list the papers that study mimesis. Papers on the left-hand side study honey-x (static mimesis), and papers on the right-hand side study attacker engagement (dynamic mimesis). The mapping between deception species and games is not one-to-one. Honey-x is modeled using two different approaches. One approach uses signaling games, to emphasize the formation of the attacker's belief about whether systems are normal systems or honeypots. The other approach uses Bayesian Nash games. This approach follows along the lines of resource allocation problems, and obtains an overall network configuration that is optimal for the defender.

The right-hand sides of Figures 18 and 19 list three approaches to attacker engagement. Zhuang et al. (2010) use a multiple-period game that has a state that reflects information from past periods. The solution is obtained using dynamic programming. Durkota et al. (2015) represent a dynamic network attack using a Markov decision process. The attacker deploys the whole Markov decision process as a Stackelberg follower, and the defender places honeypots as a Stackelberg leader. Finally, Horák et al. (2017) use a one-sided partially observable Markov decision process. This model is perhaps the closest of the three to a general *competitive Markov decision process*, in which a defender and an attacker choose dynamic policies to optimize long-term utility functions (Filar and Vrieze 2012).

4.2.3 Preferred Relationships Between Models and Infimae Species. Within these trends, certain modeling approaches best capture the essential elements of each species of deception. Figure 20
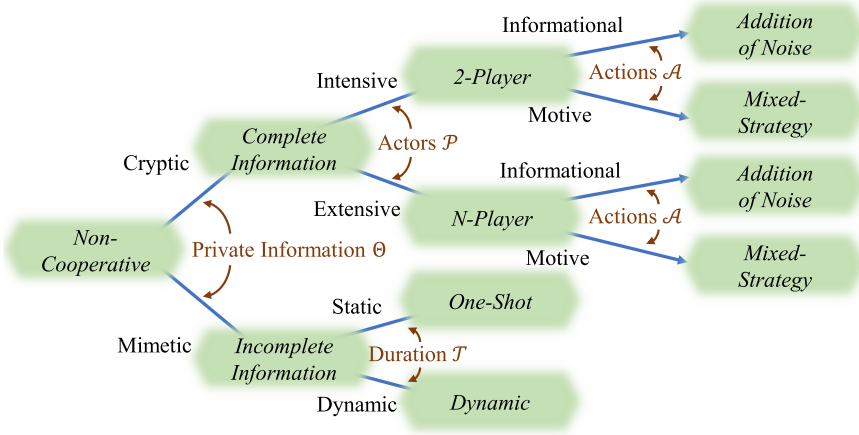
Fig. 20. Taxonomy with promising modeling approaches overlaid in a hierarchical fashion. For instance, cryptic, intensive, and motive deception (i.e., moving target defense) can be modeled by non-cooperative, complete-information, two-player games with mixed strategies. Each of the modeling approaches is shown in a blurred shape to indicate that alternate approaches are possible.

depicts these approaches. The root node is non-cooperative games.[6] Within non-cooperative games, games with incomplete information are well suited for mimetic deception, since they can explicitly model the false attacker beliefs that defenders attempt to inculcate. One-shot models such as signaling games model honey-x, while dynamic models such as partially observable stochastic games are necessary to effectively study attacker engagement.

Within cryptic deception, two players are often sufficient to study intensive deception, in which a single defender attempts to directly deceive an attacker. By contrast, extensive deception involves multiple defenders who compete against each other in addition to competing against an attacker. In perturbation and obfuscation, actions often consist of choosing the variance of noise to add to published data. In mixing and moving target defense, actions often consist of choosing mixed strategies. Through these mixed strategies, defenders probabilistically choose routing paths or network configurations.

5 DISCUSSION AND FUTURE DIRECTIONS

Our taxonomy and literature survey suggest a number of open challenges and promising future directions.

5.1 Mimesis

In the papers that we survey, cryptic deception predominates over mimetic deception. One explanation is that crypsis is the goal of all privacy research. Another explanation is that the straight-forward technique of randomization falls within the category of crypsis. Finally, research in mimesis may face greater ethical concerns than in crypsis. For instance, ought a Chief Information Officer to publish a document that falsely states the number of data servers that the company runs in order mislead an adversary? Clearly, there are some aspects of deception that could infringe on other values such as trustworthiness. These caveats notwithstanding, more opportunity remains in mimetic deception.

---

[6]Future research can consider cooperative games, but most current studies use non-cooperative models.

5.2 Theoretical Advances

The vast majority of the papers that we have surveyed use Stackelberg games and Nash games. Many combine these with rich non-strategic models. For example, Feng et al. (2017) combines a Stackelberg game with a Markov decision process, and Zhu and Başar (2013) combines a Stackelberg game with dynamic systems analysis. Few papers, though, include advanced game-theoretic models. For instance, the literature that we surveyed did not include any cooperative games. In addition, dynamic games were not often considered, partly because general stochastic games are fundamentally difficult to analyze. This may also reflect the relative newness of research in defensive deception for cybersecurity and privacy—and the application of game theory to this area.

5.3 Practical Implementations

Game theory has been successfully deployed for physical security in several applications. For instance, Pita et al. (2008) describes the ARMOR system, which is a game-theoretic protocol deployed at Los Angeles International Airport. This protocol uses Stackelberg games. Similarly, Jain et al. (2010) uses Stackelberg games to optimize the assignment of Federal Air Marshals to U.S. commercial flights. Yet it is difficult to identify successful implementations of game-theoretic concepts to cybersecurity. Of course, it is possible that commercial endeavors use game-theoretic defenses, but prefer not to publish the results. Still, several challenges frustrate the deployment of game theory in cybersecurity.

One obstacle is that some cybersecurity professionals may be wary of so-called *security through obscurity*, i.e., deceptive security mechanisms that rely only on an attacker's lack of information. It is true that deceptive mechanisms should be combined with traditional approaches such as cryptography and access control. In our opinion, though, game theory provides precisely the right set of tools to offer security professionals *provable guarantees* on what can be achieved by deception. Many game theory models assume that the adversary has full access to the strategy of the defender. Hence, they obtain worst-case guarantees and are not undermined if an attacker learns the defender's strategy. Indeed, some government organizations have enthusiastically adopted deception for cybersecurity. For example, researchers at the American National Security Agency (NSA) recently conducted a series of observed red-team exercises involving deception. Data collectors monitored the actions and rationale of the attackers and defenders (Ferguson-Walter et al. 2017). Such exercises are invaluable in making sure that game-theoretic models accurately capture elements of human psychology.

A second obstacle is, paradoxically, the high demand for technical network security professionals. Since industry and government organizations have a high demand for security analysts, and since these analysts need to constantly monitor network traffic, it is challenging to establish collaborations between the analysts and academic researchers with backgrounds in game theory. This collaboration is crucial to ensure that game theorists solve relevant problems. However, work in game theory also offers the possibility to relieve the burden on analysts. For example, game theory can be used to optimally design IDS alerts to limit the number of false positives that must be analyzed.

5.4 Interdisciplinary Security

Our article has analyzed deception as a quantitative science. Deception, however, is interdisciplinary. Economics research offers well-developed game-theoretic models. Often, however, the challenge is to apply these models in accurate, domain-specific ways. Work in experimental economics emphasizes behavioral or sub-rational aspects of deception. These aspects are critical for applications in which the attacker does not play the theoretically optimal strategy. Additionally,

psychology can be employed to analyze attacker preferences and develop accurate threat models. Finally, criminology can be useful to detect signs of deception by humans. As attacks become increasingly automated, though, the effectiveness of both psychology and criminology will have to be reevaluated.

5.5 Conclusion

The next generation of cybersecurity and privacy techniques will leverage tools commonly employed by attackers for the purpose of defense. For applications ranging from protection of civil liberties to network security to defense of the Internet of Battle Things, defensive deception will play a principal role. We have developed a taxonomy of defensive deception for cybersecurity and privacy viewed through the lens of game theory. This taxonomy provides a scientific foundation for future defensive deception research and a common language that can be used to conceptualize defensive deception. This work also provides a menu of game-theoretic models and defensive deception techniques that can be leveraged for future research. Finally, we have summarized many of the contributions of game theory to the various species of deception over the past ten years. We hope that these will provide a conceptual basis for the next stage of research in defensive deception in cybersecurity and privacy.

APPENDIX

Tables 3 and 4 list our classification justifications for papers in cryptic and mimetic deception, respectively.

Table 3. Classification Justification for Papers in Cryptic Deception

Authors and Year	Category	Justification
Chessa et al. (2015)	Cryptic	The objective is to hide private data.
	Intrinsic	The private data itself is modified.
	Informational	Noise is added to a quantity.
Shokri (2015)	Cryptic	The objective is to hide private data.
	Intrinsic	The private data itself is modified.
	Informational	Noise is added to a quantity.
Alvim et al. (2017)	Cryptic	The objective is to hide private data.
	Intrinsic	The private data itself is modified.
	Informational	Noise is added to a quantity.
Theodorakopoulos et al. (2014)	Cryptic	The objective is to hide location.
	Intrinsic	The location itself is modified.
	Informational	Noise is added to the location.
Rass et al. (2017)	Cryptic	The goal is to hide configuration.
	Intrinsic	The configuration is centrally controlled.
	Motive	The configuration is varied over time.
Sengupta et al. (2018)	Cryptic	The goal is to hide IDS locations.
	Intrinsic	The locations are chosen centrally.
	Motive	The locations are is realized randomly.
Zhu and Başar (2013)	Cryptic	The goal is to hide vulnerability locations.
	Intrinsic	The configuration is centrally controlled.
	Motive	The configuration is varied over time.

(Continued)

Table 3. Continued

Authors and Year	Category	Justification
Feng et al. (2017)	Cryptic	The goal is to hide a resource configuration.
	Intrinsic	The configuration is centrally controlled.
	Motive	The configuration is varied over time.
Clark et al. (2012)	Cryptic	The goal is to hide the network traffic path.
	Extrinsic	Traffic is hidden using other useless traffic.
	Informational	This uses uncertainty rather than motion.
Zhu et al. (2012)	Cryptic	The goal is to hide the network traffic path.
	Extrinsic	Traffic is hidden using other useless traffic.
	Informational	This uses uncertainty rather than motion.
Pawlick and Zhu (2016)	Cryptic	The goal is to hide true user behavior.
	Extrinsic	True behavior is hidden in distracting actions.
	Informational	This uses uncertainty rather than motion.
Pawlick and Zhu (2017a)	Cryptic	The goal is to hide true user behavior.
	Extrinsic	True behavior is hidden in distracting actions.
	Informational	This uses uncertainty rather than motion.
Zhang et al. (2010)	Cryptic	The goal is to hide entry-exit node pairings.
	Extrinsic	The mix network relies on other nodes.
	Motive	Exit and entry nodes are realized randomly.
Freudiger et al. (2009)	Cryptic	The goal is to hide user-pseudonym links.
	Extrinsic	Swaps require multiple participants.
	Motive	Hiding links depends on multiple exchanges.
Lu et al. (2012)	Cryptic	The goal is to hide user-pseudonym links.
	Extrinsic	Swaps require multiple participants.
	Motive	Hiding links depends on multiple exchanges.

Table 4. Classification Justification for Papers in Mimetic Deception

Authors and Year	Category	Justification
Carroll and Grosu (2011)	Mimetic	Honeypots appear to be normal systems.
	Static	The interaction is a one-shot game.
Çeker et al. (2016)	Mimetic	Honeypots appear to be normal systems.
	Static	The interaction is a one-shot game.
Mohammadi et al. (2016)	Mimetic	Honeypots appear to be normal systems.
	Static	The interaction is a one-shot game.
Pawlick and Zhu (2018)	Mimetic	Honeypots appear to be normal systems.
	Static	The interaction is a one-shot game.
Píbil et al. (2012)	Mimetic	Honeypots appear to be normal systems.
	Static	The interaction is a one-shot game.
Kiekintveld et al. (2015)	Mimetic	Honeypots appear to be normal systems.
	Static	The interaction is a one-shot game.
Zhuang et al. (2010)	Mimetic	Defender types can signal the opposite type.
	Dynamic	The game occurs over multiple periods.
Durkota et al. (2015)	Mimetic	Honeypots appear to be normal systems.
	Dynamic	Attacks are multi-stage.
Horák et al. (2017)	Mimetic	The attacker believes he is undetected.
	Dynamic	The game is infinite-horizon.

REFERENCES

George A. Akerlof and Robert J. Shiller. 2015. *Phishing for Phools: The Economics of Manipulation and Deception*. Princeton University Press.

Tansu Alpcan and Tamer Basar. 2003. A game theoretic approach to decision and analysis in network intrusion detection. In *Proceedings of the IEEE Conference on Decision and Control*, vol. 3. IEEE, 2595–2600.

Mário S. Alvim, Konstantinos Chatzikokolakis, Yusuke Kawamoto, and Catuscia Palamidessi. 2017. Information leakage games. In *Decision and Game Theory for Security*. Springer, 437–457.

Sharon Astyk, Aaron Newton, and Colin F. Camerer. 2010. Pinocchio's pupil: Using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *Amer. Econ. Rev.* 100, 3 (2010), 984–1007.

Tamer Basar. 1983. The Gaussian test channel with an intelligent jammer. *IEEE Trans. Info. Theory* 29, 1 (1983), 152–157.

J. Bowyer Bell and Barton Whaley. 2017. *Cheating and Deception*. Routledge.

Michael Bennett and Edward Waltz. 2007. *Counterdeception Principles and Applications for National Security*. Artech House.

Sean Bodmer, Max Kilger, Gregory Carpenter, and Jade Jones. 2012. *Reverse Deception: Organized Cyber Threat Counter-exploitation*. McGraw Hill Professional.

Charles F. Bond Jr. and Bella M. DePaulo. 2008. Individual differences in judging deception: Accuracy and bias. *Psychol. Bull.* 134, 4 (2008), 477.

Thomas E. Carroll and Daniel Grosu. 2011. A game theoretic investigation of deception in network security. *Secur. Commun. Nets.* 4, 10 (2011), 1162–1172.

Hayreddin Çeker, Jun Zhuang, Shambhu Upadhyaya, Quang Duy La, and Boon-Hee Soong. 2016. Deception-based game theoretical approach to mitigate DoS attacks. In *Decision and Game Theory for Security*. Springer, 18–38.

Michela Chessa, Jens Grossklags, and Patrick Loiseau. 2015. A game-theoretic study on non-monetary incentives in data analytics projects with privacy implications. In *Proceedings of the IEEE Computer Security Foundations Symposium*. 90–104.

Hugh Chisholm. 1911. Predicables. In *Encyclopedia Britannica* (11th ed.). Cambridge University Press.

Andrew Clark, Quanyan Zhu, Radha Poovendran, and Tamer Başar. 2012. Deceptive routing in relay networks. In *Decision and Game Theory for Security*. Springer, 171–185.

Hugh Cott. 1940. *Adaptive Coloration in Animals*. Methuen.

Vincent P. Crawford and Joel Sobel. 1982. Strategic information transmission. *Econometrica: J of the Econometric Soc.* (1982), 1431–1451.

Cuong T. Do, Nguyen H. Tran, Choongseon Hong, Charles A. Kamhoua, Kevin A. Kwiat, Erik Blasch, Shaolei Ren, Niki Pissinou, and Sundaraja Sitharama Iyengar. 2017. Game theory for cyber security and privacy. *ACM Comput. Surveys* 50, 2 (2017), 30.

Karel Durkota, Viliam Lisỳ, Branislav Bošanský, and Christopher Kiekintveld. 2015. Optimal network security hardening using attack graph games. In *Proceedings of the International Joint Conference on Artificial Intelligence*. 526–532.

Benjamin Edwards, Steven Hofmeyr, and Stephanie Forrest. 2016. Hype and heavy tails: A closer look at data breaches. *J. Cybersecur.* 2, 1 (2016), 3–14.

Xiaotao Feng, Zizhan Zheng, Prasant Mohapatra, and Derya Cansever. 2017. A Stackelberg game and Markov modeling of moving target defense. In *Decision and Game Theory for Security*. Springer, 315–335.

K. J. Ferguson-Walter, D. S. LaFon, and T. B. Shade. 2017. Friend or "Faux": Deception for cyber defense. *J. Info. Warfare* 16, 2 (2017), 28–42. https://www.jstor.org/stable/26502755.

Jerzy Filar and Koos Vrieze. 2012. *Competitive Markov Decision Processes*. Springer Science & Business Media, New York.

Urs Fischbacher and Franziska Föllmi-Heusi. 2013. Lies in disguise-an experimental study on cheating. *J. Euro. Econ. Assoc.* 11, 3 (2013), 525–547.

Julien Freudiger, Mohammad Hossein Manshaei, Jean-Pierre Hubaux, and David C. Parkes. 2009. On non-cooperative location privacy: A game-theoretic analysis. In *Proceedings of the ACM Conference on Computer and Community Security*. ACM, 324–337.

D. Fudenberg and J. Tirole. 1991. *Game Theory*. MIT Press.

R. Edward Geiselman. 2012. The cognitive interview for suspects (CIS). *Amer. Coll. Forensic Psychol.* 30, 3 (2012), 1–16.

Uri Gneezy. 2005. Deception: The role of consequences. *Amer. Econ. Rev.* 95, 1 (2005), 384–394.

Roy Godson and James J. Wirtz. 2011. *Strategic Denial and Deception: The Twenty-first Century Challenge*. Transaction Publishers.

Benjamin Grosser. 2014. Privacy through visibility: Disrupting NSA surveillance with algorithmically generated "scary" stories. University of Wisconsin-Milwaukee. Retrieved from https://bengrosser.com/projects/scaremail/.

Kristin E. Heckman, Frank J. Stech, Roshan K. Thomas, Ben Schmoker, and Alexander W. Tsow. 2015. *Cyber Denial, Deception and Counter Deception*. Springer.

Karel Horák, Quanyan Zhu, and Branislav Bošanský. 2017. Manipulating adversary's belief: A dynamic game approach to deception by design in network security. In *Decision and Game Theory for Security*. Springer, 273–294.

Daniel C. Howe and Helen Nissenbaum. 2009. TrackMeNot: Resisting surveillance in web search. *Lessons Ident. Trail: Anon., Priv., Ident. Netw. Soc.* 23 (2009), 417–436. Retrieved from http://www.nyu.edu/pages/projects/nissenbaum/papers/ch23(HoweNissenbaum)Web.pdf.

Sjaak Hurkens and Navin Kartik. 2009. Would I lie to you? On social preferences and lying aversion. *Exper. Econ.* 12, 2 (2009), 180–192.

Manish Jain, Jason Tsai, James Pita, Christopher Kiekintveld, Shyamsunder Rathi, Milind Tambe, and Fernando Ordóñez. 2010. Software assistants for randomized patrol planning for the LAX airport police and the Federal Air Marshal Service. *Interfaces* 40, 4 (2010), 267–290.

Navin Kartik. 2009. Strategic communication with lying costs. *Rev. Econ. Studies* 76, 4 (2009), 1359–1395.

Christopher Kiekintveld, Viliam Lisỳ, and Radek Píbil. 2015. Game-theoretic foundations for the strategic use of honeypots in network security. In *Cyber Warfare.* Springer, 81–101.

Rongxing Lu, Xiaodong Lin, Tom H. Luan, Xiaohui Liang, and Xuemin Shen. 2012. Pseudonym changing at social spots: An effective strategy for location privacy in vanets. *IEEE Trans Vehic. Technol.* 61, 1 (2012), 86–96.

David T. Lykken. 1959. The GSR in the detection of guilt. *J. Appl. Psychol.* 43, 6 (1959), 385.

James Edwin Mahon. 2016. The definition of lying and deception. In *The Stanford Encyclopedia of Philosophy* (winter 2016 ed.), Edward N. Zalta (Ed.).

Mohammad Hossein Manshaei, Quanyan Zhu, Tansu Alpcan, Tamer Bacşar, and Jean-Pierre Hubaux. 2013. Game theory meets network security and privacy. *ACM Comput. Surveys* 45, 3 (2013), 25.

Joseph Meyerowitz and Romit Roy Choudhury. 2009. Hiding stars with fireworks: Location privacy through camouflage. In *Proceedings of the 15th annual International Conference on Mobile Computing and Networking.* ACM, 345–356. Retrieved from http://dl.acm.org/citation.cfm?id=1614358.

Paul R. Milgrom. 1981. Good news and bad news: Representation theorems and applications. *Bell J. Econ.* 12, 2 (1981), 380–391.

MITRE. 2010. Science of cyber-security. https://fas.org/irp/agency/dod/jason/cyber.pdf.

Amin Mohammadi, Mohammad Hossein Manshaei, Monireh Mohebbi Moghaddam, and Quanyan Zhu. 2016. A game-theoretic analysis of deception over social networks using fake avatars. In *Decision and Game Theory for Security.* Springer, 382–394.

Roger B. Myerson. 1991. *Game Theory: Analysis of Conflict.* Harvard University Press.

John F. Nash. 1950. Equilibrium points in n-person games. *Proc. Nat. Acad. Sci. U.S.A.* 36, 1 (1950), 48–49.

NISO. 2005. Guidelines for the construction, format, and management of monolingual controlled vocabularies. https://groups.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf.

Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. Law Rev.* 79 (2004), 119.

Alessandro Oltramari, Lorrie Faith Cranor, Robert J. Walls, and Patrick D. McDaniel. 2014. Building an ontology of cyber security. In *Proceedings of the Conference on Semantic Technology for Defense, Intelligence, and Security (STIDS'14).* 54–61.

Jeffrey Pawlick, Edward Colbert, and Quanyan Zhu. 2018. Modeling and analysis of leaky deception using signaling games with evidence. *IEEE Trans. Inform. Forens. Secur.* 14, 7 (2018), 1871–1886.

Jeffrey Pawlick, Sadegh Farhang, and Quanyan Zhu. 2015. Flip the cloud: Cyber-physical signaling games in the presence of advanced persistent threats. In *Decision and Game Theory for Security.* Springer, 289–308.

Jeffrey Pawlick and Quanyan Zhu. 2015. Deception by design: Evidence-based signaling games for network defense. In *Proceedings of the Workshop on the Economics of Inform. Security and Privacy.* Delft, The Netherlands. http://arxiv.org/abs/1503.05458

Jeffrey Pawlick and Quanyan Zhu. 2016. A Stackelberg game perspective on the conflict between machine learning and data obfuscation. In *Proceedings of the IEEE Workshop on Information Forensics and Security.* https://arxiv.org/abs/1608.02546

Jeffrey Pawlick and Quanyan Zhu. 2017a. A mean-field Stackelberg game approach for obfuscation adoption in empirical risk minimization. In *Proceedings of the Global Signal and Information Processing Workshop on Control and Game Theoretic Approaches to Security and Privacy.*

Jeffrey Pawlick and Quanyan Zhu. 2017b. Strategic trust in cloud-enabled cyber-physical systems with an application to glucose control. *IEEE Trans Inform. Forens. Secur.* 12, 12 (2017), 2906–2919.

Scott R. Peppet. 2014. Regulating the Internet of things: First steps toward managing discrimination, privacy, security and consent. *Tex. L. Rev.* 93 (2014), 85. Retrieved from http://heinonlinebackup.com/hol-cgi-bin/getpdf.cgi?handle=hein.J.s/tlr93&section=5.

Radek Píbil, Viliam Lisỳ, Christopher Kiekintveld, Branislav Bošanskỳ, and Michal Pechoucek. 2012. Game theoretic model of strategic honeypot selection in computer networks. In *Decision and Game Theory for Security.* Springer, 201–220.

James Pita, Manish Jain, Janusz Marecki, Fernando Ordóñez, Christopher Portway, Milind Tambe, Craig Western, Praveen Paruchuri, and Sarit Kraus. 2008. Deployed ARMOR protection: The application of a game theoretic model for security at the los angeles intl. airport. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems: Industrial Track.* International Foundation for Autonomous Agents and Multiagent Systems, 125–132.

Stefan Rass, Sandra König, and Stefan Schauer. 2017. On the cost of game playing: How to control the expenses in mixed strategies. In *Decision and Game Theory for Security*. Springer, 495–505.

Hy Rothstein and Barton Whaley. 2013. *The Art and Science of Military Deception*. Artech House.

Neil C. Rowe. 2006. A taxonomy of deception in cyberspace. In *Proceedings of the International Conference on Information Warfare and Security* (2006).

Neil C. Rowe and Julian Rrushi. 2016. *Introduction to Cyberdeception*. Springer.

Sankardas Roy, Charles Ellis, Sajjan Shiva, Dipankar Dasgupta, Vivek Shandilya, and Qishi Wu. 2010. A survey of game theory as applied to network security. In *Proceedings of the IEEE International Conference on System Sciences*. 1–10.

Sailik Sengupta, Ankur Chowdhary, Dijiang Huang, and Subbarao Kambhampati. 2018. Moving target defense for the placement of intrusion detection systems in the cloud. In *Decision and Game Theory for Security*. Springer, 326–345.

Eric Shieh, Bo An, Rong Yang, Milind Tambe, Craig Baldwin, Joseph DiRenzo, Ben Maule, and Garrett Meyer. 2012. Protect: A deployed game theoretic system to protect the ports of the United States. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems—Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 13–20.

Reza Shokri. 2015. Privacy games: Optimal user-centric data obfuscation. *Proc. Priv. Enhanc. Technol.* 2 (2015), 299–315.

George Theodorakopoulos, Reza Shokri, Carmela Troncoso, Jean-Pierre Hubaux, and Jean-Yves Le Boudec. 2014. Prolonging the hide-and-seek game: Optimal trajectory privacy for location-based services. In *Proceedings of the ACM Workshop on Privacy in the Electronic Society*. 73–82.

Heinrich Von Stackelberg. 1934. *Marktform und Gleichgewicht*. Springer.

Aldert Vrij, Samantha A. Mann, Ronald P. Fisher, Sharon Leal, Rebecca Milne, and Ray Bull. 2008. Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law Hum. Behav.* 32, 3 (2008), 253–265.

Barton Whaley. 2016. *Practise to Deceive: Learning Curves of Military Deception Planners*. Naval Institute Press.

Nan Zhang, Wei Yu, Xinwen Fu, and Sajal K. Das. 2010. gPath: A game-theoretic path selection algorithm to protect tor's anonymity. In *Decision and Game Theory for Security*. Springer, 58–71.

Rui Zhang and Quanyan Zhu. 2015. Secure and resilient distributed machine learning under adversarial environments. In *Proceedings of the 18th International Conference on Information Fusion (Fusion'15)*. IEEE, 644–651.

Rui Zhang and Quanyan Zhu. 2017. A game-theoretic analysis of label flipping attacks on distributed support vector machines. In *Proceedings of the 51st Annual Conference on Information Sciences and Systems (CISS'17)*. IEEE, 1–6.

Quanyan Zhu and Tamer Başar. 2013. Game-theoretic approach to feedback-driven multi-stage moving target defense. In *Decision and Game Theory for Security*. Springer, 246–263.

Quanyan Zhu, Andrew Clark, Radha Poovendran, and Tamer Başar. 2012. Deceptive routing games. In *Proceedings of the IEEE Conference on Decision and Control*. 2704–2711.

J. Zhuang, V. M. Bier, and O. Alagoz. 2010. Modeling secrecy and deception in a multiple-period attacker-defender signaling game. *Eur. J. Operation. Res.* 203, 2 (2010), 409–418.