A Stochastic Game-Theoretic Model for Smart Grid Communication Networks

Xiaobing He(✉) and Hermann de Meer

Department of Informatics and Mathematics, University of Passau,
Innstr. 43, 94032 Passau, Germany
{Xiaobing.He,Hermann.DeMeer}@uni-passau.de

Abstract. The increasing adoption of new information and communication technology assets in smart grids is making smart grids vulnerable to cyber threats, as well as raising numerous concerns about the adequacy of current security approaches. As a single act of penetration is often not sufficient for an attacker to achieve his/her goal, multistage cyber attacks may occur. This paper looks at the stochastic and dynamic nature of multistage cyber attacks in smart grid use cases and develops a stochastic game-theoretic model to capture the interactions between the attacker and the defender in multistage cyber attack scenarios. Due to the information asymmetry of the interactions between the attacker and the defender, neither of both players knows the exact current game state. This paper proposes a belief-updating mechanism for both players to form a common belief about the current game state. In order to assess threats of multistage cyber attacks, it further discusses the computation of Nash equilibria for the designed game model.

Keywords: Asymmetric information · Positive stop probability · Stochastic game · Multistage cyber attacks · Smart grid · Threat assessment

1 Introduction

Network security is a critical concern with regard to cyber-physical systems. For a long time, security operators have been interested in knowing what an attacker can do to a cyber-physical system and what can be done to prevent or counteract cyber attacks [3,14]. It is suggested that risk assessment must be integral to the overall life cycle of the smart grid systems. A cyber threat assessment helps the system administrator to better understand the effectiveness of the current network security solution and determine the best approach to secure the system against a particular threat, or a class of threats. By offering a deep analysis of existing or potential threats, system administrators are given a clear assessment of the risks to their systems, while possessing a clear vision about the kind of security countermeasures that the respective utility should invest in.

Attack scenarios are dynamically changing in smart grid communication networks, for example, because of existing of legacy and new systems in smart

S. Rass et al. (Eds.): GameSec 2017, LNCS 10575, pp. 295–314, 2017.
DOI: 10.1007/978-3-319-68711-7_16

grid communication networks. Multistage cyber attacks, as important threats in smart grid communication networks, make use of a variety of different exploits, propagation methods, and payloads, resulting in the emergence of many more sophisticated cyber attacks. Current protection mechanisms, which rely on isolation techniques, such as firewalls, data diodes, and zoning concepts, are not sufficiently applicable in cyber-physical systems. For more than a decade, game-theoretic approaches have been recognized as useful tools to handle network attacks [2,7,13,15]. Significant results from game theory concerning cyber situation awareness and network security risk assessment in conventional information and communication technology (ICT) systems have been reported [14,30]. But the application of game theory for the assessment of threats from multistage cyber attacks and the prediction of an attacker's actions in smart grid communication networks are still in their infancy nowadays.

Threat assessment for multistage cyber attacks is not straightforward, given that, at any stage of a cyber attack, the attacker may decide not to proceed or change his/her attack actions. Since the attacker has motivations (costs versus benefits) and finite resources to launch a further attack at any stage, the stage at which the multistage attack stops is not necessary predetermined (stochastic). This paper accounts for this by adding a stopping time to the stochastic model. It is to be noted that an attacker who doesnot have any resource limitations (from an economic point of view) is beyond the scope of this paper. The stop of the attack or the change of attack actions at any stage makes a threat assessment extremely challenging, as it is difficult to know what the attacker will do or to assess possible cyber or physical impacts resulting from his/her attack actions in the next stage.

Cyber attacks on smart grid communication networks can cause physical damage to the power grid. Many existing stochastic game-theoretic threat assessment methods assume symmetric information among the players, which implies that all the players share the same information, i.e., the same signal observed and the same knowledge about states/payoffs in a game. However, in many situations, this assumption is unrealistic. There are many games arising out of communication networks, electronic commerce systems, and society's critical infrastructures involving players with different kinds of information about the game state and action processes over time [11,23,29]. For instance, in cyber-security systems, the attacker knows his/her own skill set, while the defender knows the current and planned resource characteristics of the system. In short, the attacker and the defender do not share their available information with each other.

This paper attempts to design a stochastic game-theoretic model with asymmetric information and positive stop probabilities in order to assess the threat of multistage cyber attacks in smart grid communication networks. The positive stop probability means that the probability of the game to end at any state is positive. Unlike random failures, attackers have motivations and capabilities to launch further attacks. Both the attacker and the defender will act in consideration of the consequences of their corresponding actions, with such consequences including satisfactions, risk versus effort, and effectiveness. In each state of the game, if launching a further attack would have limited benefits, and take months of time and huge amount of computers and memory, the attacker will most probably stop his/her attack. Once the defender observed these phenomena regarding the attacker, he/she will not deploy any corresponding countermeasures. Therefore, this situation will be accounted for by adding a stop probability to the stochastic model; and such a stop probability is positive. The designed stochastic game-theoretic model extends an existing stochastic game-theoretic model with specific characteristics of attacker-defender interactions in smart grid communication networks. The objectives of this attacker-defender stochastic game-theoretic model is to assess cyber attack scenarios at an early stage of the attack, where the defender makes correct optimal proactive defence decisions. Therefore, a defence system can be prewarned, security resources can be better allocated to defeat or mitigate future attacks, and security incidents can be avoided. This paper considers the worst-case scenario where the attacker has complete knowledge of the architecture/infrastructure of the system and hosts' vulnerabilities in the system, and the attacker has full knowledge of the target smart grid defense configurations. Section 2 provides a non-exhaustive overview of existing game-theoretic approaches for cyber attacks, while Sect. 3 presents an attacker-defender stochastic game-theoretic model to represent the attacker-defender interactions. Section 4 analyses the belief-updating mechanisms and presents the feasible computation of Nash equilibria. Finally, Sect. 5 concludes the paper ans discusses future works.

2 Related Work

A game consists of players (in this paper, the attacker and the defender), strategies (i.e., actions of players) available to each player, and utilities depending on the joint decisions of all players. Game theory depicts dynamic interactions between players, involving a complementary methodology of attack trees and/or attack graphs in face of changing attack patterns.

Ismail et al. [10] modelled the problem of optimizing the distribution of defence resources on communication equipment as a one-shot game [22] between the attacker and the defender. That game took into account the interdependency between the cyber and physical components in the power grid. It was assumed that the initial risk, the immediate risk on a node before any incidents or failure propagations is a positive real number and evaluated using other risk assessment methods. The immediate risk and the future cascading risk from interdependent electrical and communication infrastructures were balanced in [10]. The interdependency between the electrical and communication infrastructures were modelled as a weighted directed interdependency graph. Each communication equipment was associated with a load. The worst-case scenario, where both the attacker and the defender have complete knowledge of the architecture of the system, was considered in [10]. The utility functions of both players are composed of three parts: the reward for an attack, the cost of attacking/defending, and the impact of redundant communication equipment. The impact of attacks in the electric and communication infrastructures was evaluated by solving power flow equations and using attack graphs, in conjunction with other risk assessment methods. The dataset of the Polish electric transmission system, provided in the MATPOWER computational packages, was taken as a case study to validate the proposed game-theoretic model, while Nash equilibria for the attacker and the defender for each type of communication equipment in the case study were presented.

Jiang et al. [30] proposed a two-player non-cooperative, zero-sum, and finite stochastic game for the attacker and the defender in computer networks. A Markov chain for a privilege model and a privilege-escalating attack taxonomy were presented. By making use of the developed stochastic game model, a Markov chain for the privilege model, and a cost-sensitive model, the attacker's behaviour and the optimal defence strategy for the defender were predicted. He et al. [8] studied a network security risk assessment-oriented game-theoretic attack-defence model to quantify the probability of threats. The payoff matrix was formulated from a cost-benefit analysis, where the cost to the defender when taking actions was made up of the operational cost, the response cost, and the response negative cost. Combined with the vulnerability associated with the nodes, risks of the system were computed as the sum of the threat value of all nodes.

Guillarme et al. [6] presented an attack stochastic game model for adversarial intention recognition for situations featuring strategic interactions between an attacker and a defender. The attack stochastic game model is a coupling of discounted stochastic games and probabilistic attack graphs, although it suffers from zero-sum constraints. In the attack stochastic game model, it was assumed that both the attacker's action and the defender's action, as well as the states experienced by players, were fully observable to both players. This model was inverted to infer the intention of an attacker from observations of his/her (sub-)optimal actions. However, this model does not have the ability to detect intention changes, while the scalability is the principal limitation of this attack stochastic game model.

Nguyen et al. [21] studied a two-player zero-sum stochastic game-theoretic approach to provide the defender with guidelines to allocate his/her resources to secure his/her communication and computer networks. Linear influence networks [19] were used to present the interdependency of nodes in terms of security assets and vulnerabilities. He et al. [9] investigated game-theoretic risk assessment in smart grid communication networks and noticed that the data acquisition and data interpretation for risk assessment and prediction had not been intensively explored. Therefore, [9] established a surveillance architecture to monitor message transactions in communication networks, while surveillance observations were further interpreted as Dirichlet-distributed security events with certain probabilities. By taking the interactions between possible suspicious nodes and the security operators as a repeated zero-sum transmitting-monitoring game, a game-theoretic risk assessment framework was established to compute and forecast the risk of network security impairment. Rass and Zhu [25] presented a sequence of nested finite two-player zero-sum games for developing effective protective layers and designing defence-in-depth strategies against advanced persistent attacks (APTs). In the game-theoretical model, nodes in an infrastructure were equidistantly separated into different levels according to their layers in the infrastructure. Within each level, the game structure was determined by the nodes' vulnerabilities and their distances from the target node. The authors of [25] discussed some closed form solutions for their APTs games and analytically formulated infrastructure design problems to optimize the quality of security across several layers. Under the framework of the HyRiM project, Rass et al. [24] investigated an extensive form game as a risk mitigation tool for defending against APTs. An APT was modelled as a zero-sum one-shot game with complete information, but uncertainty was observed in the game payoffs. Based on a topological vulnerability analysis and an established attack graph, all the attack vectors covered in enumerated attack paths (from the root node to the target node in the attack graph) made up the attacker's action space. By defining players' payoffs as probability-distributed values, instead of real numbers, [24] provided a relative new approach to tackling ambiguous and inconsistent expert opinions in risk management.

The proposed game-theoretic model in this paper differs from the aforementioned approaches in the sense that the model captures the key characteristics (e.g., information asymmetry) of the interactions between the attacker and the defender in smart grid communication networks. None of theses precursor works has looked at the stochastic and dynamic nature of attacks in smart grid use cases (modelled as stochastic games). Both decision makers, the attacker and the defender, have asymmetric information about the underlying system state, while they both maintain a belief (i.e., a probability distribution) about the current system state. This paper provides a common belief-updating mechanism for the attacker and the defender to refresh such a belief. The objectives of this research include contributing towards safety improvements for relevant stakeholders (e.g., smart grid equipment manufacturers, utility companies) in power distributed grids and making recommendations about allocating security resources to reduce cyber security incidents or even safety-related events.

3 Attacker-Defender Stochastic Game-Theoretic Model

To assess threats of multistage attacks, the strategic interactions between the attacker and the defender are modelled as a stochastic game (which covers the step occurrence dependency in multistage attacks). In such a game, the possible actions of the players are restricted, such that there exists an equilibrium point in which the attacker has no chance to successfully obtain his/her ultimate goal. This section introduces action spaces and state transition probabilities of the game between the attacker and the defender. This work designs the attacker-defender stochastic game-theoretic model by a description of an existing stochastic game model and an extension of this model according to the characteristics of the interactions of the attacker and the defender in smart grid communication networks.

3.1 Players

An attacker and a defender are the key "players" in the designed stochastic game-theoretic model. There could be many attackers who are trying to launching attacks and many defenders in the network to protect the system, but this work abstracts those attackers and defenders as one attacker and one defender, respectively. The attacker attains his/her ultimate target via multiple stages. The concept of the defender denotes the security operator (security operator and system administrator are used interchangeably in this paper) who has the task of deploying available defence countermeasures to protect the underlying system, while the attacker attempts to reach the target or the most critical assets located at the centre of the smart grid. This model considers that each of the players has some finite resources to perform actions at each stage of the game. The attacker is considered to be a resource-constrained, determined and rational player. In this way, the attacker will give up when he/she finds it is out of his/her capability to launch any further attacks. Furthermore, it is assumed that once an attack is initiated, the attacker him/herself will never revert the system to any of the previous state (for example, to recover the system from a malfunctioning state to a normal operational state). In this work, the attacker is only able to perform a single action in his/her turn. It is also assumed that the defender does not know whether or not there is an attacker, as that in real systems. Furthermore, the attacker is assumed to be always aware of the active defence mechanisms. Moreover, the defender does not know the objectives and strategies of an attacker. A successful attack may or may not be observable to the defender. The attacker strategically and dynamically chooses his/her targets and attack methods in order to achieve his/her goals, while the defender defines security policies and implements security measures (including email filtering, detection software, patches to prevent and detect attacks, and repairing the system after disruption).

3.2 State Transition Probabilities

A multistage attack, by exploiting vulnerabilities, makes the network system transition from one state to another. However, such a transition also depends on the active defence mechanisms. Therefore, the probability that the state will transition from one to another depends on the joint actions of both players. Unlike accidental failures, an attacker will consider the consequences of his/her actions and compare the reward versus the cost of each elementary attack action [27]. Therefore, the transition probabilities from one state to another depend not only on the decisions of both players to take action, but also the success probability of an attacker going through with his/her action. The probability of success for the attacker at state s is denoted as $p_{suc}(y_{s,b})$ (this work assumes the second player to be the attacker and $y_{s,b}$ (which will be defined later in Eq. (4)) to be the probability that his/her action b is taken at state $s \in S$, S is the state space and $k = |S|$ is the number of game states). Obviously, whether an action by an attacker succeeds depends on the available exploitable vulnerabilities of an asset in the smart grid communication network. For example, attacking an asset with no exploitable vulnerability has zero probability of success. In the attacker-defender stochastic game-theoretic model, success probabilities of an attacker's actions are assigned, based on the intuition and experience (e.g., case studies, common vulnerability scoring system (CVSS), knowledge engineering). Principally, the action of the defender also involves a success probability (e.g., IDSs have detection rates); to simplify the underlying problem, however, such a success probability of the defender with his/her actions is always assumed to be one.

The probability for player 1 (player 1 is the defender) to take action $a \in AS_1$ at state s is denotes as $x_{s,a}$ (which will be defined later in Eq. (3)), while the probability for player 2 (player 2 is the attacker) to take action $b \in AS_2$ at state s is denoted as $y_{s,a}$. Both players take actions simultaneously, meaning that both players take action independently of one another. Thus, when actions $a \in AS_1$ and $b \in AS_2$ are taken from both players, the state transition probability from game state $s \in S$ to state $s' \in S$ can be calculated as

$$q(s'|s,a,b) = x_{s,a} \cdot y_{s,b} \cdot p_{suc}(y_{s,b}).$$

For example, if the probability for player 1 to take action "IDS deployment" is 0.5, the probability for player 2 to take action "Exploit" is 0.4, and the probability that the attacker will successful obtain his/her (sub)goal is 0.2, the game will move from state "normal" to state "malfunctioning" with a state transition probability of

$$q(\text{malfunctioning}|\text{normal}, \text{IDS deployment}, \text{Exploit}) = 0.5 \cdot 0.4 \cdot 0.2 = 0.04.$$

Depending on the exploitable vulnerabilities, it may be that there is no transition between certain game states. For example, it may not be possible for the network to transition from a normal functioning state to a totally failed state without going through any intermediate states. In this work, infeasible state transitions are assigned with a transition probability of zero and hence ignored. Both players make their moves simultaneously, with state transition probabilities being common knowledge to them.

3.3 Game Formalization

In the previous subsections, this paper elaborates players in a game play. At each stage of the game for multistage attacks, the play is in a given state, with every player choosing an action from his/her available action space. With a state transition probability (which is jointly controlled by both players), the current state of the game, and the collection of actions that the players choose, the game will go to another state with an immediate payoff received by each player. Each player has his/her own costs of executing actions, thus the payoff of the game cannot only be described by rewards. Although there may be a dependence of rewards and losses among player's payoffs, because of players' own action execution costs, the payoffs of the attacker and the defender do not sum up to zero. Therefore, the interaction between the attacker and the defender is non-zero-sum. The game is also played with positive stop probabilities in each game state, since the game will end when the attacker decides to stop his/her attacks (completely inactive) and the defender keeps his/her defence countermeasures unchanged. Besides, this paper notices that none of the players knows the exact state of the system, while both players have different kinds of private information about the state and action processes over time. Therefore, in order to apply game theory to assess multistage attacks in smart grid communication networks, the asymmetric information, non-zero-sum, and positive stop probability characteristics of the interaction between the attacker and the defender should be taken into account.

The next concern is on the game type that appropriately captures the players' interactions in the case of multistage cyber attacks. Both players do not know the exact state of the game, but maintain a belief about the current state of the game (where a belief is a probability distribution over the possible states of the game). Taking a two-player non-zero-sum two-stage game for instance, suppose the game has two states and both players do not know the current state of the game (either in state s_1 or state s_2), but they have a belief $\rho_1 = (\rho_1(s_1), \rho_1(s_2)) = (0.8, 0.2)$ about the current state, that is, there is a 80% likelihood that the current game at stage 1 is in state s_1, while there is a 20% likelihood that the current game at stage 1 is in state s_2. The most relevant existing game model that can partially solve this problem is the stochastic game with lack of information on one side (SGLIOS) with positive stop probabilities. Thus, this paper considers SGLIOS with positive stop probabilities as a basic game model and extends it to include the non-zero-sum and information asymmetry of the interactions between the attacker and the defender in smart grid communication networks.

This work starts with the definition of SGLIOS with positive stop probabilities described in [18]. The model of SGLIOS with positive stop probabilities is a two-person zero-sum game and states are a finite set $S = \{s_1, s_2, \cdots, s_\ell, \cdots, s_k\}$ ($k = |S|$ denotes the number of states). Associated with each state s_ℓ ($\ell \in \{1, 2, \cdots, k\}$) is a matrix game $\mathbf{G}_{\{s_\ell\}}$ of size $m_1 \times m_2$, where $m_1 = |AS_1|$ (the number of actions of player 1), $m_2 = |AS_2|$ (the number of actions of player 2), and $\mathbf{G}_{\{s_\ell\}} = \{g_{\{s_\ell\}}(a, b) : AS_1 \times AS_2 \to \mathbb{R} | a = 1, 2, \cdots, m_1; b = 1, 2, \cdots, m_2; \ell = 1, 2, \cdots, k\}$. Additionally, $\emptyset$ is adjoined to S, where $\emptyset$ represents the end of the game. In SGLIOS with positive stop probabilities, at any stage N, there is a probability distribution over states in S. throughout this paper, N takes values from $\mathbb{N}$ and $\mathbb{N}$ is the set of natural number. Player 1 is informed about such a probability distribution at every game stage, but player 2 is never informed about that. There is a probability $\rho_1 \in \Delta(S)$ about the initial state, where $\Delta(S)$ is the set of all probability distributions on S. State transition probabilities are denoted as $q(\cdot|s, a, b)$, which depends on the current state s and actions a and b taken by the defender and the attacker, respectively. Because of the positive stop probability assumption, the sum of transition probabilities from state s to all possible next game state s' is less than one, i.e., $\sum_{s' \in \{S - \emptyset\}} q(s'|s, a, b) < 1$, $\forall a \in AS_1$, $b \in AS_2$. Both players make their moves simultaneously and both of them are informed of their choices (a, b). The game will either end with a probability of $q(\emptyset|s, a, b) > 0$ or transition to a new state s' with a probability of $q(s'|s, a, b) > 0$. Although both players remember actions taken by them, player 2 is not informed of the received immediate payoff $g_{\{s\}}(a, b)$ (which only player 1 knows) of the game. SGLIOS with positive stop probabilities is played with perfect recall (i.e., at each stage each player remembers all past actions chosen by all players and player 1 knows all *past* states that have occurred). There is a common knowledge among both players before they move at stage N and such a common knowledge is a sequence of the form $h_N = \big\{(a_1, b_1), (a_2, b_2), \cdots, (a_{N-1}, b_{N-1})\big\}$ (where $a_\ell \in AS_1$ is the action chosen from player 1 at the ℓ stage, $b_\ell \in AS_2$ is the action chosen from player 2 at the ℓ stage, and $\ell \in \{1, 2, \cdots, N-1\}$). The common knowledge h_N is also called *history* and it represents the choices of actions (i.e., pure strategies) of the two players up to (and excluding) stage N. SGLIOS with positive stop probabilities restricts its attention to behavioural strategies [12].

When the game is in state s at stage N, the action chosen by the players can be deterministic or randomized. A mixed strategy corresponds to a distribution over actions (i.e., pure strategies). Let $\mathbf{x}_s$ $(s \in S)$ denote the mixed strategy of player 1 in state s and $\mathbf{y}_s$ $(s \in S)$ denote the mixed strategy of player 2 at state s. The strategies $\mathbf{x}_s$ and $\mathbf{y}_s$ in state s are used to assign probabilities over the action set AS_1 and AS_2 with cardinality m_1 and m_2, respectively. And the mixed strategies $\mathbf{x}_s$ and $\mathbf{y}_s$ are defined as

$$\mathbf{x}_s := \{(x_{s,1}, \cdots, x_{s,a}, \cdots, x_{s,m_1}) \in \mathbb{R}_+^{m_1} | \sum_{a=1}^{m_1} x_{s,a} = 1, 0 \leq x_{s,a} \leq 1\}, \quad (1)$$

$$\mathbf{y}_s := \{(y_{s,1}, \cdots, y_{s,b}, \cdots, y_{s,m_2}) \in \mathbb{R}_+^{m_2} | \sum_{b=1}^{m_2} y_{s,b} = 1, 0 \leq y_{s,b} \leq 1\}, \quad (2)$$

where

$$x_{s,a} := \mathbb{P}(a|s, h_N), \quad (3)$$
$$y_{s,b} := \mathbb{P}(b|s, h_N), \quad (4)$$

and $x_{s,a}$ and $y_{s,b}$ represent the probability that player 1 takes action a and player 2 takes action b, respectively. It is to be noted that actions of players are independently chosen among each other, since both players are playing simultaneously. Let $\mathbf{x} = (\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \cdots, \mathbf{x}_{s_\ell}, \cdots, \mathbf{x}_{s_k})$ be a vector of mixed strategies for player 1 and $\mathbf{x} \in \Omega^{m_1}$ (Ω^{m_1} is the set of all probability vectors of length m_1). Correspondingly, let $\mathbf{y} = (\mathbf{y}_{s_1}, \mathbf{y}_{s_2}, \cdots, \mathbf{y}_{s_\ell}, \cdots, \mathbf{y}_{s_k})$ be a vector of mixed strategies for player 2 and $\mathbf{x} \in \Omega^{m_2}$ (Ω^{m_2} is the set of all probability vectors of length m_2). Let E be a random variable representing the stage the game ends and h_N be the common knowledge among players up to (and excluding) stage N. At each stage N, if player 1 takes action a and player 2 took action b, player 1 receives an immediate payoff $g_{\{s_N\}}(a,b)$, The total payoff function $\mathcal{H}(\cdot)$ (with strategies from both players as parameters) in SGLIOS with positive stop probabilities is given as

$$\begin{aligned}\mathcal{H}(\mathbf{x},\mathbf{y}) &= \sum_{N=1}^{\infty} \mathcal{R}_N(\mathbf{x},\mathbf{y}) \qquad (5)\\ &= \sum_{N=1}^{\infty} \mathbb{E}_{\mathbf{x},\mathbf{y}}\big(\rho_N(s)\mathbf{G}_{\{s\}}|E > N\big)\cdot \mathbb{P}(E > N),\end{aligned}$$

where $\mathbb{P}(E > N)$ means that the game does not end at stage N and the stage E where game ends is longer than N. The expectation operator $\mathbb{E}_{\mathbf{x},\mathbf{y}}\big(\cdot\,|E > N\big)$ is used to mean that player 1 plays strategy $\mathbf{x}$ and player 2 plays strategy $\mathbf{y}$, under the condition that the game does not end at stage N. Equation (5) assumes that the game stage can go to infinite (∞). However, because of the positive stop probability assumption, the game will end after a finite number of stages [28]. Therefore, the game of SGLIOS with positive stop probabilities is a finite game. The fundamental tool in SGLIOS with positive stop probabilities is an updating mechanism which gives at each stage N the belief ρ_N, the posterior distribution on the state space given the history h_N up to stage N. Player 1 is informed about the belief ρ_N but player 2 does not. The updating mechanism for the belief ρ_N is working in this way: initially both players choose strategies $\mathbf{x}$ and $\mathbf{y}$ and give them to chance (chance is a special player, who can be the environment of the system) who then at stage 1 chooses s_1 according to ρ_1. Then the action pair (a_1, b_1) is chosen according to $(\mathbf{x}_{s_1}, \mathbf{y}_{s_1})$ and an immediate payoff $g_{\{s_1\}}(a_1, b_1)$ is received by player 1. Provided that the game does not end, chance chooses another state s_2 according to $\rho_2(s_2) := \mathbb{P}(s_2|a_1, b_1, E > 2)$ or decides to end the game according to $\mathbb{P}(E = 2|a_1, b_1)$. At stage N, chance decides the game to go to state s_N according to $\rho_N(s_N) := \mathbb{P}(s_N|h_N, E > N)$ or ends the game according to $\mathbb{P}(E = N|E > N-1, h_N)$. The value $\rho_N(s)$ represents that the chance believes that the current game state is $s \in S$. It is proved in [18] that the value of the game of SGLIOS with positive stop probabilities exists and is a continuous function on the state space; and there exists also a stationary optimal strategy for the informed player, i.e., player 1. The optimal strategy of player 1 depends only on the updated probability of the current state which he/she independently knows.

Since the interaction between the attacker and the defender in smart grid use cases is non-zero-sum, it is needed to extend SGLIOS with positive stop probabilities (which is zero-sum) to non-zero-sum cases. The game matrices should be first identified. Each player (player 1 or player 2) has his/her own game matrix, which is composed of two parts: his/her reward/loss as the result of an attack and the cost of carrying out his/her action. Essentially, both two players are with contradictory objectives and they are competing with each other. The objective of each player is to maximize his/her own total payoff with strategies $\mathbf{x}$ and $\mathbf{y}$

$$\mathcal{H}_1(\mathbf{x},\mathbf{y}) = \sum_{N=1}^{\infty} \mathcal{R}_{1,N}(\mathbf{x},\mathbf{y}) = \sum_{N=1}^{\infty} \mathbb{E}_{\mathbf{x},\mathbf{y}}\big(\rho_N(s)\mathbf{G}_{\{1,s\}}|E > N\big) \cdot \mathbb{P}(E > N), \quad (6)$$

$$\mathcal{H}_2(\mathbf{x},\mathbf{y}) = \sum_{N=1}^{\infty} \mathcal{R}_{2,N}(\mathbf{x},\mathbf{y}) = \sum_{N=1}^{\infty} \mathbb{E}_{\mathbf{x},\mathbf{y}}\big(\rho_N(s)\mathbf{G}_{\{2,s\}}|E > N\big) \cdot \mathbb{P}(E > N). \quad (7)$$

The reason why both the attacker and the defender share the same belief value $\rho_N(s)$ will be given out in Sect. 4.1. Another characteristic of the interaction between the defender and the attacker is the information asymmetry, where each player has private information about the state of the network system, while such private information among players is asymmetric. The asymmetry stems from the fact that the attacker has knowledge of a particular vulnerability which can be exploited; while the defender knows how to use resources to defend against all possible attacks. In other words, one player either deliberately distorts or does not disclose all the relevant information to another player, during their interaction phases. Since no player completely knows the exact state s of the game, it is assumed that each player (player 1 or player 2) observes a private local state $s_{\{1\}}$ or $s_{\{2\}}$ of the game and the state of the game is composed of both private local states $s = \{s_{\{1\}}, s_{\{2\}}\}$. Each player has to form a belief about the exact state s up to stage N. It is assumed that each player knows all *past* states that have occurred, which means when the game goes to next state, the previous one state will be publicly known to all players. Provided that the game has not ended, the history h_N is common information available to both players whereas private information is only available to that specific player.

According to [18], players can forget the sequence of previous states. So without loss of generality, it is assumed that the state of the two-player game at $N+1$ stage (assuming that the game does not end at N stage) evolves according to the current state s_N and all previous strategies from both players. Similarly, the private local state of each player is evolving according to the current local state $s_{\{1,N\}}$ for player 1 or $s_{\{2,N\}}$ for player 2 and all previous strategies from both players. It is obviously that, at any stage N, the local state $s_{\{1,N\}}$ for player 1 is independent of the local state $s_{\{2,N\}}$ for player 2. Therefore, when both players have taken actions $a \in AS_1$ and $b \in AS_2$, the state transition probability in the case of information asymmetry among players is defined as

$$\begin{aligned} q(s_N|s_{N-1}.a,b) &:= \mathbb{P}(s_N|s_{N-1},a,b) \\ &= \mathbb{P}(s_{\{1,N\}}|s_{\{1,N-1\}},a,b) \cdot \mathbb{P}(s_{\{2,N\}}|s_{\{2,N-1\}},a,b). \end{aligned} \quad (8)$$

The choice of actions for each player at stage N may depend on all past strategies from both players and the player's current local state (the local state is one part of the game state $s_N = \{s_{\{1,N\}}, s_{\{2,N\}}\}$), which is consistent with Eqs. (3) and (4). Given the fact that no player can observe the current game state s_N ($s_N \in S$) at stage N and each player observes only a private local current game state $s_{\{1,N\}}$ or $s_{\{2,N\}}$, the probability for player 1 to choose action a and the probability for player 2 to choose action b at stage N are defined as

$$x_{s_{\{1,N\}},a} := \mathbb{P}(a|s_{\{1,N\}}, h_N) \tag{9}$$

and

$$y_{s_{\{2,N\}},b} := \mathbb{P}(b|s_{\{2,N\}}, h_N), \tag{10}$$

respectively.

It is to noteworthy that by knowing the strategy of the other player, one player can make inference about the other player's private information $s_{\{1,N\}}$ (if this player is player 2) or $s_{\{2,N\}}$ (if this player is player 1) from observing their actions. If a player knows the local private state of the other player, he/she can further predict the action of the other player in next stage. Provided that the game continues, state s_N is chosen according to $\rho_N(s_N) = \mathbb{P}(s_N|h_N, E > N)$, the immediate payoff $g_{\{1,s_N\}}(a_N, b_N)$ is received at player 1(correspondingly, $g_{\{2,s_N\}}(a_N, b_N)$ is received at player 2), and both two players computes his/her belief $\rho_{N+1}(s_{N+1})$ on next game state s_{N+1}.

4 Game Analysis

This section analyses the previously specified game model and finds Nash equilibria to construct an attack scenario in which the adversary cannot succeed in performing multistage cyber attacks and arriving at his/her ultimate target. In the previously specified game model, players have asymmetric information about the current state of the game, therefore, each player has to form a belief about the current state of the game. In SGLIOS with positive stop probabilities, player 1 (who can be assumed to be the defender) is informed about the belief value on the current game state but player 2 (who can be assumed to be the attacker) does not. Under the assumption that the true state of the game is independent of the action taken by player 2, the belief value in SGLIOS with positive stop probabilities is not conditional on the strategy taken by player 2 [18]. However, this assumption is not applicable in attacker-defender games where strategies from both player decide the state and the process of the game. Therefore, new belief system updating mechanisms should be described and belief system updates account for a central technical contribution in this paper. To assist equilibria computation for the designed attacker-defender stochastic game-theoretic model, this section first provides the belief update mechanism and then elaborates an easy-to-follow method for Nash equilibria computation.

4.1 Belief System Updates

Actions taken by both players can be summarized through a belief ρ_N of game states. For example, in SGLIOS with positive stop probabilities, under the assumption that the current state of the game is independent of player 2's actions, the belief ρ_N summarizes actions taken by player 1 [18]. In the game of asymmetric information, at stage N, the current game state is unknown to both players; player 1 privately observes a local state $s_{\{1,N\}}$ and player 2 privately observes another local state $s_{\{2,N\}}$. To consist with [18] and the recent work on stochastic game with asymmetric information [23,29], in this work, belief ρ_N on the current state s_N of the game is defined as $\rho_N(s_N) := \mathbb{P}(s_N|h_N, E > N)$.

Provided that the game does not end at N stage, which means the condition $\mathbb{P}(E > N)$ satisfies, for any history $h_N = \big\{(a_1, b_1), (a_2, b_2), \cdots, (a_{N-1}, b_{N-1})\big\}$, it can be observed that player's belief about the current game state s_N is

$$\begin{aligned}\rho_N(s_N) &:= \mathbb{P}(s_N|h_N)\\ &= \mathbb{P}(s_{\{1,N\}}, s_{\{2,N\}}|h_N).\end{aligned} \tag{11}$$

Because of the independence of private local states $s_{\{1,N\}}$ and $s_{\{2,N\}}$, Eq. (11) can be further written as

$$\begin{aligned}\rho_N(s_N) &= \mathbb{P}(s_{\{1,N\}}, s_{\{2,N\}}|h_N)\\ &= \mathbb{P}(s_{\{1,N\}}|h_N) \cdot \mathbb{P}(s_{\{2,N\}}|h_N).\end{aligned} \tag{12}$$

The probability $\mathbb{P}(s_{\{1,N\}}|h_N)$ can be viewed as the probability that player 2 believes that player 1 will be in state $s_{\{1,N\}}$ based on the history h_N of past actions taken from both players. Player 2 might also derive this probability $\mathbb{P}(s_{\{1,N\}}|h_N)$ at N stage based on his/her private local states, however, since the private local states $s_{\{1,N\}}$ and $s_{\{2,N\}}$ ($N \in \mathbb{N}$) are independent, the probability $\mathbb{P}(s_{\{1,N\}}|h_N, s_{\{2,1\}}, s_{\{2,2\}}, \cdots, s_{\{2,N-1\}})$ would be the same as the probability $\mathbb{P}(s_{\{1,N\}}|h_N)$. Therefore, knowledge of private state information $(s_{\{2,1\}}, s_{\{2,2\}}, \cdots, s_{\{2,N-1\}})$ from player 2 does not affect the probability $\mathbb{P}(s_{\{1,N\}}|h_N)$. For player 2, the probability $\mathbb{P}(s_{\{2,N\}}|h_N)$ can be viewed as the probability that player 2 believes that his/her private local state at stage N is $s_{\{2,N\}}$ based on the history of actions from both players. It is to be noted that player 2 knows his current private local state $s_{\{2,N\}}$. However, this paper assumes that after taking any action and before arriving in state $s_{\{2,N\}}$, player 2 can also has a probability $\mathbb{P}(s_{\{2,N\}}|h_N)$ about his/her private local state $s_{\{2,N\}}$. Based on probabilities that player 1 will in state $s_{\{1,N\}}$ and he/she him/herself will be in state $s_{\{2,N\}}$ at stage N, player 2 can derive the probability $\rho_N(s_N)$ that the current game state is s_N at stage N. Similarly, player 1 can also derive the probability that player 2 will be in state $s_{\{2,N\}}$ at stage N with probability $\mathbb{P}(s_{\{2,N\}}|h_N)$ and the probability that he/she him/herself will be in state $s_{\{1,N\}}$ with probability $\mathbb{P}(s_{\{1,N\}}|h_N)$. Therefore, both players can obtain the same belief value that the game play is in state s_N at stage N.

4.2 Finding Nash Equilibria

When dealing with strategic players with inter-dependent payoffs (for example, the attacker's rewards might somehow be losses of the defender), investigating equilibria, mostly notably Nash equilibria, is a method of predicting players' decisions. If we restrict our attention to pure strategies (i.e., actions), a Nash equilibrium may not exists, this is the reason that this work considers only behaviour strategies and the probability used by both players to choose among pure strategies. The attacker-defender game with asymmetric information has finite states and the action spaces AS_1 and AS_2 are finite. The major differences between this attacker-defender game and the SGLIOS with positive stop probabilities are that this attacker-defender game is a non-zero-sum one and the belief system updates in this attacker-defender game are jointly conditioned on strategies from both players. In the SGLIOS with positive stop probabilities, the belief is conditioned only on the strategy of the informed player; while in the attacker-defender game, the belief is conditioned on strategies of both players. If the probability that taking action b_{N-1} is zero, the history h_N will not be observed, which will not happen under the assumption that the game does not end at $N-1$ stage. It was said that the belief in the SGLIOS with positive stop probabilities is continuous [17]. The same continuity property extends to the belief in the proposed attacker-defender game. In the designed attacker-defender game, both players are informed about the belief of game states. Hence, each player can be taken as the informed player in the SGLIOS with positive stop probabilities. It is proved in [18] that the informed player has a stationary optimal strategy. However, [18] does not provide a systematic way to find such optimal strategies.

The designed attacker-defender game is non-zero-sum. It is stated in [20] that every non-zero-sum stochastic game has at least one (not necessary unique) Nash equilibrium in stationary strategies and finding these equilibria is non-trivial. The attacker-defender game with uncertainty about current game state for both players makes it extremely challenging. Given the strategies of both players, players continue to accumulate the immediate payoffs. Once the end state of the game is reached, the game is over and no more accumulations are possible. Each player wishes to maximize his/her expected payoff at state s_N. This maximization, in turn, yields player's value of the game. Hence, if the value of the game Γ_N exists, let the vector of values for player 1 be $\mathbf{v}_1$, where $\mathbf{v}_1 = (v_{1,s_1}, v_{1,s_2}, \cdots, v_{1,s_\ell}, \cdots, v_{1,s_k})$ (v_{1,s_ℓ} is player 1's value of the game in state $s_{1,\ell}$ and $v_{1,s_\ell} \in \mathbb{R}$ ($\ell \in \{1, 2, \cdots, k\}$)) and the vector of values for player 2 be $\mathbf{v}_2$, where $\mathbf{v}_2 = (v_{2,s_1}, v_{2,s_2}, \cdots, v_{2,s_\ell}, \cdots, v_{2,s_k})$ (v_{2,s_ℓ} is player 2's value of the game in state s_ℓ and $v_{2,s_\ell} \in \mathbb{R}$ ($\ell \in \{1, 2, \cdots, k\}$)). The value of each player (either the attacker or the defender) includes both short-term (i.e., immediate) payoff and long-term payoff (which is given by the expected value of the sum of state payoffs from the current state) [4]. Taking the value for player 1 for instance, his/her value can be recursively defined as (that for player 2 can be defined in the same way)

$$v_{1,s_N}(\rho_N(s_N)) := \max_{\mathbf{x}_N} \min_{\mathbf{y}_N} \sum_{\mathbf{x}_N,\mathbf{y}_N} \left(\rho_N(s_N)\mathbf{G}_{\{1,s_N\}} + \mathbf{T}_1(s_N,\mathbf{v})\right), \tag{13}$$

where matrix $\mathbf{T}_1(s_N,\mathbf{v})$ is used to represent the long-term payoff (i.e., the future payoff) in a matrix form. The vector $\mathbf{v}$ is a value vector (a sub-vector of the game value vector that is defined above) for player 1 and it depends on the states that the current state s_N can transition to.

A pair of strategy sequence $(\mathbf{x}^*,\mathbf{y}^*)$ forms (Nash) equilibria with strategy pair $(\mathbf{x}^*_{s_N},\mathbf{y}^*_{s_N})$ if

$$\begin{aligned}\mathcal{H}_1(\mathbf{x}^*,\mathbf{y}^*) &\geq \mathcal{H}_1(\mathbf{x},\mathbf{y}^*), \forall \mathbf{x} \in \Omega^{m_1},\\ \mathcal{H}_2(\mathbf{x}^*,\mathbf{y}^*) &\geq \mathcal{H}_2(\mathbf{x}^*,\mathbf{y}), \forall \mathbf{y} \in \Omega^{m_2},\end{aligned}$$

where $\geq$ is used to mean at every stage N, the left-hand-side with strategy profile $(\mathbf{x}^*_{s_N},\mathbf{y}^*_{s_N})$ is greater than the right-hand-side with strategy $(\mathbf{x}_{s_N},\mathbf{y}^*_{s_N})$ or strategy $(\mathbf{x}^*_{s_N},\mathbf{y}_{s_N})$. Therefore, the pair of strategy profile $(\mathbf{x}^*_{s_N},\mathbf{y}^*_{s_N})$ $(N \in \mathbb{N})$ is said to be a Nash equilibrium strategy. At this equilibrium, there is no incentive for either player to deviate from his/her equilibrium strategy $\mathbf{x}^*_{s_N}$ or $\mathbf{y}^*_{s_N}$ at any stage N of the game. In each pair of equilibrium strategies, a strategy for one player is a best-response to the other player and vice versa. A deviation means that one or both of them may have a lower expected payoff, i.e., $\mathcal{H}_1(\mathbf{x},\mathbf{y}^*)$ or $\mathcal{H}_2(\mathbf{x}^*,\mathbf{y})$.

In order to find Nash equilibria for the designed attacker-defender non-zero-sum game in smart grid communication networks, based on the formed work [5,26], this paper studies nonlinear programming (NLP) formulation of the attacker-defender non-zero-sum stochastic game with finite number of strategies and asymmetric information. The theorem and proof of a global minimum to be a (Nash) equilibrium with equilibrium payoff can be found in [1,5], this work is not going to repeat them here again, whereas it provides here an easy-to-follow method to find such (Nash) equilibria in the designed attacker-defender game.

Assuming the game has M stage, where the game ends after the M stage (i.e., $E > M$, $M \geq 1$ and $M \in \mathbb{N}$). The equilibrium solution $(\mathbf{x}^*,\mathbf{y}^*)$ for M-stages games can be obtained by solving the following nonlinear programming problem:

$$\begin{aligned}\text{minimize} \quad & \sum_{N=1}^{M-1} \big(v_{1,s_M} - \mathbf{x}_{s_M} \cdot \rho_M(s_M) \cdot \mathbf{G}_{1,s_M} \cdot \mathbf{y}^T_{s_M} + v_{2,s_M} - \mathbf{x}_{s_M} \cdot \rho_M(s_M) \cdot \\ & \mathbf{G}_{2,s_M} \cdot \mathbf{y}^T_{s_M} + v_{1,s_N} - \mathbf{x}_{s_N} \cdot (\rho_N(s_N)\mathbf{G}_{1,s_N} + \mathbf{T}_1(s_N,\mathbf{v})) \cdot \mathbf{y}^T_{s_N} \\ & + v_{2,s_N} - \mathbf{x}_{s_N} \cdot (\rho_N(s_N)\mathbf{G}_{2,s_N} + \mathbf{T}_2(s_N,\mathbf{v})) \cdot \mathbf{y}^T_{s_N}\big),\end{aligned}$$

subject to

(i) $\rho_M(s_M)\mathbf{G}_{1,s_M}\mathbf{y}_{s_M}^T \leq v_{1,s_M}\mathbf{J}_{m_1}^T$,

(ii) $\rho_M(s_M)\mathbf{G}_{2,s_M}^T\mathbf{x}_{s_M}^T \leq v_{2,s_M}\mathbf{J}_{m_2}^T$,

(iii) $\rho_N(s_N)\mathbf{G}_{1,s_N}\mathbf{y}_{s_N}^T + \mathbf{T}_1(s_N,\mathbf{v})\mathbf{y}_{s_N}^T \leq v_{1,s_N}\mathbf{J}_{m_1}^T, \forall N \in \{1,2,\cdots,M-1\}$,

(iv) $\rho_N(s_N)\mathbf{G}_{2,s_N}^T\mathbf{x}_{s_N}^T + \mathbf{T}_2(s_N,\mathbf{v})^T\mathbf{x}_{s_N}^T \leq v_{2,s_N}\mathbf{J}_{m_2}^T, \forall N \in \{1,2,\cdots,M-1\}$,

(v) $\displaystyle\sum_{a=1}^{m_1} x_{s_N,a} = 1 \quad \forall a \in AS_1, N \in \{1,2,\cdots,M\}$,

(vi) $x_{x_N,a} \geq 0 \quad \forall a \in AS_1, N \in \{1,2,\cdots,M\}$,

(vii) $\displaystyle\sum_{b=1}^{m_2} y_{s_N,b} = 1 \quad \forall b \in AS_2, N \in \{1,2,\cdots,M\}$,

(viii) $y_{s_N,b} \geq 0 \quad \forall b \in AS_2, N \in \{1,2,\cdots,M\}$,

(ix) $\rho_N(s_N) = \mathbb{P}(s_N|h_N, E > M), N \in \{1,2,\cdots,M\}$.

Constraints (i) and (iv) are the value bounds for the attacker-defender game, which are satisfied for any pair of strategy profile. The mixed strategies $\mathbf{x}_{s_N}$ and $\mathbf{y}_{s_N}$ ($N = \{1,2,\cdots,M\}$) are defined in Eqs. (1) and (2), respectively. Constraints (v)–(viii) are conditions that the probability $x_{s_N,a}$ to select action a for player 1 in state s_N and the probability $y_{s_N,b}$ to select action b for player 2 in state s_N is greater than zero and the sum of all such probabilities for each player is one. Any pair of strategy profile satisfies constraints (v)–(viii). The constraint (ix) is a prior belief constraint and the belief ρ_1 for the first stage, which is presumed to be known to both players, is a probability distribution over state space S, i.e., $\rho_1 \in \Delta(S)$. Because of the recursion definition of belief values of constraint (ix) and the recursive optimization involved in the long-term payoff (i.e., $\mathbf{T}_1(s_N,\mathbf{v})$ or $\mathbf{T}_2(s_N,\mathbf{v})$) of constraints (iii) and (iv), it is non-trivial to find global minima.

In an one-stage game, each player (either the attacker or the defender) would play with the stationary strategy that maximizes his/her expected immediate payoff at the current game stage. Hence $(\mathbf{x}_{s_1}^*, \mathbf{y}_{s_1}^*)$ will be one optimal strategy profile. There can be mutiple stationary Nash equilibria in each game state and hence there will be multiple global minima. For example, for a stochastic game with one stage and the payoff matrix for player 1 (who has three actions: A, B and C) and player 2 (who has two actions: D and E) is $\mathbf{G}_{\{1,s_1\}}$ and $\mathbf{G}_{\{2,s_1\}}$, respectively (to be noted that those values in payoff matrices are artificial numbers for illustration)

$$\mathbf{G}_{\{1,s_1\}} = \begin{array}{c|cc} & \text{D} & \text{E} \\ \hline \text{A} & 6 & 2 \\ \text{B} & 1 & 3 \\ \text{C} & 5 & 4 \end{array}, \text{and} \quad \mathbf{G}_{\{2,s_1\}} = \begin{array}{c|cc} & \text{D} & \text{E} \\ \hline \text{A} & 4 & 3 \\ \text{B} & 1 & 5 \\ \text{C} & 2 & 2 \end{array}.$$

Presuming that each player knows that the probability distribution $\rho_1(s_1)$ is 1, and the game value for player 1 (the row player) and player 2 (the column player) are denoted as v_{1,s_1} and v_{2,s_1}, respectively. Therefore, the nonlinear programming formulation of this one-stage game can be expressed as

$$\text{minimize} \left(v_{1,s_1} - \mathbf{x}_{s_1} \cdot \begin{bmatrix} 6 & 2 \\ 1 & 3 \\ 5 & 4 \end{bmatrix} \cdot \mathbf{y}_{s_1}^T + v_{2,s_1} - \mathbf{x}_{s_1} \cdot \begin{bmatrix} 4 & 3 \\ 1 & 5 \\ 2 & 2 \end{bmatrix} \cdot \mathbf{y}_{s_1}^T \right),$$

subject to

$$\text{(i)} \quad \begin{bmatrix} 6 & 2 \\ 1 & 3 \\ 5 & 4 \end{bmatrix} \mathbf{y}_{s_1}^T \leq v_{1,s_1} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T,$$

$$\text{(ii)} \quad \begin{bmatrix} 4 & 3 \\ 1 & 5 \\ 2 & 2 \end{bmatrix}^T \mathbf{x}_{s_1}^T \leq v_{2,s_1} \begin{bmatrix} 1 & 1 \end{bmatrix}^T,$$

$$\text{(iii)} \quad \sum_{a=1}^{3} x_{s_1,a} = 1 \quad \forall a \in \{A, B, C\},$$

$$\text{(iv)} \quad x_{s_1,a} \geq 0 \quad \forall a \in \{A, B, C\},$$

$$\text{(v)} \quad \sum_{b=1}^{2} x_{s_1,b} = 1 \quad \forall b \in \{D, E\},$$

$$\text{(vi)} \quad x_{s_1,b} \geq 0 \quad \forall b \in \{D, E\}.$$

There are three stationary mixed equilibria available for this one-stage game (by solving a constrained minimization problem), which are shown in Table 1 with their corresponding values for each player. All Nash equilibria and game values in Table 1 are further verified by the Gambit software tool [16]. Suppose that the first player is the defender of a system and the second player is the attacker. For the first Nash equilibrium in Table 1, to obtain maximum payoffs ("6" for the defender and "4" for the attacker, as shown in Table 1), the defender is suggested play the pure strategy "A" with a probability of 1 (i.e., play the action "A" in all game repetitions) and the attacker play the pure strategy "D" with a probability of 1. The same interpretation can be applied to the third Nash equilibrium, i.e., the defender plays the pure strategy "C" with a probability of 1 and the attacker plays the pure strategy "E" with a probability of 1 to maximise their payoffs. Regarding the second Nash equilibrium, the game suggests that the defender play his/her pure strategy "C" with a probability of 1, while it suggests that the the attacker play his/her pure strategy "D" with a probability of approximately 0.67 and his/her pure strategy "E" with a probability of approximately 0.33. If actions (i.e., pure strategies) are continuously and taking daily (24 h), the mixed Nash equilibrium strategy $\left(\frac{2}{3}, \frac{1}{3}\right)$ for the attacker can

Table 1. Nash equilibria and their corresponding game values in the sampled game.

# of Nash equilibrium	Player 1			Player 2		Game value	
	A	B	C	D	E	Player 1	Player 2
1	1	0	0	1	0	6	4
2	0	0	1	2/3	1/3	14/3	2
3	0	0	1	0	1	4	2

also be interpreted that the attacker temporarily runs the pure strategy "D" for approximately 16 h and runs the pure strategy "E" for the remainder of the day. If the actions "D" and "E" are instantaneous actions (which are taken at discrete time instants), the mixed Nash equilibrium strategy $\left(\frac{2}{3}, \frac{1}{3}\right)$ for the attacker can be interpreted as the (asymptotic) frequency with which the strategies "D" and "E" are chosen in the game. After obtaining the mixed Nash equilibrium, the defender and the attacker can subsequently use it in the following way: when the game begins, both players (the defender and the attacker) randomly choose actions (i.e., pure strategies) from their corresponding action spaces, a game payoff from the chosen action pair will be received at each player. When the game is played again, both players again randomly choose actions from their corresponding action spaces in this round. It is to be noted that the actions from both players in this round may be different from that taken in the previous one. A game payoff will again be received at each player. The actions in each round are chosen randomly, however, the player should be aware of that the (asymptotic) frequency of chosen actions must be that suggested from the mixed Nash equilibrium. Therefore, when averaging payoffs in all repetitions of the game, the average payoff is optimal for each player only if the actions are chosen with their frequencies that are prescribed by the equilibrium strategy. For example, for the attacker, in any game round, he/she should always aware of that the (asymptotic) frequency of choosing actions "D" and "E" in all game repetitions should be $\frac{2}{3}$ and $\frac{1}{3}$, respectively.

5 Conclusion and Future Work

To assess the threat of multistage cyber attacks in smart grid communication networks, this paper designs a stochastic game-theoretic model according to the characteristics of the interactions between the attacker and the defender in smart grid use cases. Firstly, the majority of the existing game-theoretic threat and risk assessment models are reviewed. Then, this paper elaborates players and state transition probabilities of the designed stochastic game-theoretic model. Since each player has partial knowledge of the game state, a belief-updating mechanism for both players to form a common belief about the current state of the game is proposed. Moreover, this paper discusses the use of nonlinear programming for Nash equilibria computation. One important aim of future work is the application of the proposed stochastic game-theoretic model to evaluate a multistage cyber attack scenario. Additionally, cyber attacks can also introduce disruptive events in power grids. Therefore, further studies of payoff formulation with an understanding of cascading effects of multistage cyber attacks would be of great significance.

Acknowledgments. The research leading to the results presented in this paper was supported by the European Commission's Project No. 608090, HyRiM (Hybrid Risk Management for Utility Networks) under the 7th Framework Programme (FP7-SEC-2013-1). The authors acknowledge Stefan Rass from Alpen-Adria-Universität Klagenfurt for his invaluable discussions and comments.

References

1. Barron, E.N.: Game Theory: An Introduction. Wiley, Hoboken (2007)
2. Chen, L., Leneutre, J.: Fight jamming with jamming – a game theoretic analysis of jamming attack in wireless networks and defense strategy. J. Comput. Netw.: Int. J. Comput. Telecommun. Network. **55**(9), 2259–2270 (2011)
3. European Union Agency for Network and Information Security. ENISA smart grid security recommendations. Technical report, European Union Agency for Network and Information Security (2012)
4. Feinberg, E.A., Shwartz, A. (eds.): Handbook of Markov Decision Processes: Methods and Applications, vol. 40. Springer US, New York (2002). doi:10.1007/978-1-4615-0805-2
5. Filar, J.A., Schultz, T.A., Thuijsman, F., Vrieze, O.J.: Nonlinear programming and stationary equilibria in stochastic games. Math. Program. **50**(1), 227–237 (1991)
6. Le Guillarme, N., Mouaddib, A-I., Gatepaille, S., Bellenger, A.: Adversarial intention recognition as inverse game-theoretic plan for threat assessment. In: IEEE 28th International Conference on Tools with Artificial Intelligence, January 2017
7. Hamman, S.T., Hopkinson, K.M., McCarty, L.A.: Applying Behavioral game theory to cyber-physical systems protection planning. In: Cyber-Physcial Systems: Foundations, Principles and Applications, pp. 251–264. Elsevier (2017)
8. He, W., Xia, C., Wang, H., Zhang, C., Ji, Y.: A game theoretical attack-defense model oriented to network security risk assessment. In: 2008 International Conference on Computer Science and Software Engineering. IEEE (2008)
9. He, X., Sui, Z., de Meer, H.: Game-theoretic risk assessment in communication networks. In: IEEE 16th International Conference on Environment and Electrical Engieering (EEEIC). IEEE, June 2016
10. Ismail, Z., Leneutre, J., Bateman, D., Chen, L.: A methodology to apply a game theoretic model of security risks interdependencies between ICT and electric infrastructures. In: Zhu, Q., Alpcan, T., Panaousis, E., Tambe, M., Casey, W. (eds.) GameSec 2016. LNCS, vol. 9996, pp. 159–171. Springer, Cham (2016). doi:10.1007/978-3-319-47413-7_10
11. Jones, M.G.: Asymmetric information games and cyber security. Ph.D. Thesis, Georgia Institute of Technology (2013)
12. Kuhn, H.W.: Extensive games and the problem of information. Ann. Math. Stud. **28**(28), 193–216 (1953)

13. Liang, X., Xiao, Y.: Game theory for network security. IEEE Commun. Surv. Tutorials **15**(1), 472–486 (2013)
14. Lye, K.-W., Wing, J.M.: Game strategies in network security. Int. J. Inf. Secur. **4**(1–2), 71–86 (2005)
15. Manshaei, M.H., Zhu, Q., Alpcan, T., Basa̧r, T., Hubaux, J.-P.: Game theory meets network security and privacy. J. ACM Comput. Surv. (CSUR) **45**(3), 25:1–25:39 (2013)
16. Mckelvey, R.D., McLennan, A.M., Turocy, T.L.: Gambit: software tools for game theory, Version 14.1.0 (2014). http://www.gambit-project.org. Accessed 04 June 2017
17. Melolidakis, C.: On stochastic games with lack of information on one side. Int. J. Game Theory **18**(1), 1–29 (1989)
18. Melolidakis, C.: Stochastic games with lack of information on one side and positive stop probabilities. In: Raghavan, T.E.S., Ferguson, T.S., Parthasarathy, T., Vrieze, O.J. (eds.) Stochastic Games and Related Topics. TDLC, vol. 7, pp. 113–126. Springer, Netherlands (1991). doi:10.1007/978-94-011-3760-7_10
19. Miura-Ko, R.A., Yolken, B., Bambos, N., Mitchell, J.: Security investment games of interdependent organizations. In: 46th Annual Allerton Conference on Communication, Control, and Computing. IEEE (2009)
20. Nash, J.: Non-cooperative Games. Ph.D. Thesis, Princeton University (1950)
21. Nguyen, K.C., Alpcan, T., Basa̧r, T.: Stochastic games for security in networks with interdependent nodes. In: International Conference on Game Theory for Networks. IEEE, June 2009
22. Osborne, M.J., Rubinstein, A.: A Course in Game Theory. MIT Press, Cambridge (1994)
23. Ouyang, Y.: On the interaction of information and decision in dynamic network systems. Ph.D. Thesis, University of Michigan (2016)
24. Rass, S., König, S., Schauer, S.: Defending against advanced persistent threats using game-theory. PLoS ONE **12**(1), 1–43 (2017)
25. Rass, S., Zhu, Q.: GADAPT: a sequential game-theoretic framework for designing defense-in-depth strategies against advanced persistent threats. In: Zhu, Q., Alpcan, T., Panaousis, E., Tambe, M., Casey, W. (eds.) GameSec 2016. LNCS, vol. 9996, pp. 314–326. Springer, Cham (2016). doi:10.1007/978-3-319-47413-7_18
26. Rothblum, U.G.: Solving stopping stochastic games by maximizing a linear function subject to quadratic constraints. In: Game Theory and Related Topics, pp. 103–105 (1979)
27. Sallhammar, K., Knapskog, S.J.: Using game theory in stochastic models for quantifying security. In: The 9th Nordic Workshop on Secure IT-Systems (2004)
28. Shapley, S.L.: Stochastic games. Proc. Natl. Acad. Sci. U.S.A. **39**(10), 1095–1100 (1953)
29. Vasal, D.: Dynamic decision problems with cooperative and strategic agents and asymmetric information. Ph.D. Thesis, University of Michigan (2016)
30. Zhang, H., Jiang, W., Tian, Z., Song, X.: A stochastic game theoretic approach to attack prediction and optimal active defense strategy decision. In: International Conference on Networking, Sensing and Control (ICNSC), pp. 648–653. IEEE (2008)