Security from the Adversary's Inertia–Controlling Convergence Speed When Playing Mixed Strategy Equilibria

Jasmin Wachter [1,*] [ID], **Stefan Rass** [1,*] [ID] **and Sandra König** [2] [ID]

[1] System Security Group, Institute of Applied Informatics, Alpen-Adria-Universität Klagenfurt, Universitätsstrasse 65-67, 9020 Klagenfurt, Austria

[2] Center for Digital Safety & Security, Austrian Institute of Technology, Giefinggasse 4, 1210 Vienna, Austria; sandra.koenig@ait.ac.at

[*] Correspondence: jasmin.wachter@aau.at (J.W.); Stefan.Rass@aau.at (S.R.)

Abstract: Game-theoretic models are a convenient tool to systematically analyze competitive situations. This makes them particularly handy in the field of security where a company or a critical infrastructure wants to defend against an attacker. When the optimal solution of the security game involves several pure strategies (i.e., the equilibrium is mixed), this may induce additional costs. Minimizing these costs can be done simultaneously with the original goal of minimizing the damage due to the attack. Existing models assume that the attacker instantly knows the action chosen by the defender (i.e., the pure strategy he is playing in the *i*-th round) but in real situations this may take some time. Such adversarial inertia can be exploited to gain security and save cost. To this end, we introduce the concept of *information delay*, which is defined as the time it takes an attacker to mount an attack. In this period it is assumed that the adversary has no information about the present state of the system, but only knows the last state before commencing the attack. Based on a Markov chain model we construct strategy policies that are cheaper in terms of maintenance (switching costs) when compared to classical approaches. The proposed approach yields slightly larger security risk but overall ensures a better performance. Furthermore, by reinvesting the saved costs in additional security measures it is possible to obtain even more security at the same overall cost.

Keywords: game theory; Stochastic Control; Mixed Strategy Equilibrium; Control of Expenses; switching costs; incomplete information; Bounded Rationality; information delay; Perron-Frobenius

1. Introduction and Motivation

1.1. Playing a Mixed Strategy Causes Costs

Implementing a pure strategy equilibrium of a game is straightforward and the installation cost of the strategy occur only once at the beginning of the game since the optimal strategy profile is pure and will never be altered. When playing repeated games, however, it may occur that the optimal strategy is mixed, i.e., the optimal strategy is obtained by assigning a positive probability to two or more pure strategies. A mixed strategy is an assignment of probabilities, which declares a law for randomly selecting the individual pure strategies in each round of game to ensure an optimal result regarding the expected utility. In standard models it is assumed that players can switch strategies as frequently as they want. Yet, in real life switching strategies will incur additional costs. For example if we consider game-theoretic models in cybersecurity, strategies may include different configurations of servers, firewalls or other system components. If switching strategies means changing configurations, the change may be costly in terms of time or money (e.g., downtime of servers, hourly rates of staff, etc.).

One possibility to consider switching costs is to compute multiple Nash equilibria and choose the one with the smallest (Shannon) entropy. This approach yields the "purest" of all equilibria. Another possibility is to model the problem in terms of dynamic games: the aim is to find the optimal Markov chain, i.e., to find the best (mixed) strategy based on the current state and the cost of switching to new states. Rass, König and Schauer have discussed these approaches in [1]. They point out that solely considering "more pure" strategies (an thus reducing the frequency of action changes) or minimizing the costs for the next choice is not sufficient: the implementation of a defense strategy needs to be done in a way, such that the defender's moves should not be predictable for an attacker, as this facilitates security breaches. In other words, when employing a security strategy it should not be possible to get a better forecast on the defender's next action when considering the current state of the system and the costs incurred by switching to another strategy. Thus, instead of calling for a dynamic optimization [1], suggest a static framework, where all actions are taken stochastically independent of the current state while still minimizing the switching costs. Despite their strong focus on security principles, there exist even more efficient solutions if we take an "information delay" for the adversary into account, i.e., the time it takes the attacker to recognize a changed situation and adapt to it.

The concept introduced in this work incorporates the time it takes for an attacker to mount an attack. It may happen that an attacker does not have complete information about the present state of the attacked system (such as the current strategy of a defender), but only knows the state of the system some rounds ago. This may happen, for example, if it takes the attacker some time to carry out the attack, i.e., the adversary has some *inertia*. During this period, the attacker may not be able to keep track of the system, and will not detect if the state of the system changes. Thus, his attack is performed after some *delay*, during which no new information can be processed. As a vivid example, consider an intruder who tries to gain unauthorized access to some critical infrastructure, that is surrounded by a wall. Before he starts his attack, he knows the current position of a guard, as he can see him through a window or compromised camera, but while he is entering the premises, the position of the guard may change, without the attacker noticing.

In the following, we will refer to this scenario as an *information delay*. By taking into account the average time the system is unobservable for an attacker prior to his attack, we can construct strategy policies that are cheaper in terms of switching cost. This saving is traded for a slightly larger risk in the primary security goal, but ultimately yielding a better performance overall. Security is never only an economic matte of cost-benefit balance, and impractical security solutions are practically worthless (say, if the optimal security strategy prescribes frequent changes in server configurations, such a strategy would simply not be doable in practice). Taking into account the cost for "running" an optimal defense as such is, in our opinion, an equally important aspect of defense as the security precaution itself. This work aims at providing means to keep the running costs of a defense under control and in balance to the security benefits therefrom.

1.2. Related Work

This work essentially deals with convergence to a Nash equilibrium, which is a well studied matter in the literature, but usually with a totally different goal as ours. Some work [2] indeed assumes a certain "speed" of the attacker, and adapts the defense to it. However, this prior work (and related follow-ups) disregard the potential of moving slightly faster than the adversary to gain an explicit profit from this. Most studies of convergence relate to the speed at which behavior can be adapted to become optimal in the long run [3–7], with some consideration spent on specific settings such as congestion or load balancing. The cost borne in switching between configurations has been considered in [8], where the authors use entropy as a measure to prefer certain strategies with less cost in the change. In the context of password policy choice, [9] considered games about choosing passwords that are (i) easy to remember, (ii) hard to guess, and (iii) easy to change (for the owner). The latter aspect is a well known cause of passwords to follow certain patterns like having counters attached to them or similar. Taking the password change (switching) cost into account can aid looking for a

best password policy and prevent the issue to some extent. Related on different grounds is also [10,11], where convergence to an (ϵ-approximate) equilibrium is studied using Markov chains. Our work relates to this in the sense that we also design Markov chains to play a desired equilibrium, but use an ϵ-approximation to the equilibrium as an area of trade-off to avoid costs from switching. In that sense, we provide a novel use of ϵ-approximations to equilibria for the sake of security economics [12].

1.3. Contribution and Structure of the Article

This contribution aims at generalizing the switching cost model [1] for games where the attacker has incomplete information that can be described by an information delay. By taking into account information delay, the implementation costs can be reduced while still ensuring the security principle that the opponent cannot forecast the next move more precisely. The resulting policy can be described using a Markov chain model. We will show in fact that the switching cost model is a special case of our information delay model.

The structure of the article is as follows: first, we introduce some preliminary concepts and notations required to describe the game setup as well as the costs for switching strategies in Section 2. In Sections 3 and 4, the theoretical framework is explained and all mathematical derivations stated. A numerical example completes the Section 3. Finally, Section 5 summarizes the findings and highlights some open questions that might be relevant for future research.

2. Preliminaries

In the following, we will use uppercase letters to define random variables and sets. Vectors are printed in bold-face. We will write $X \sim F$ if a random variable X is distributed according to a probability distribution F. Distributions on finite ordered sets are described using probability vectors $\mathbf{x} = (x_1, \ldots, x_n)$, $\sum_{i=1}^{n} x_1 = 1$ which represent the probability mass function of the underlying random variable. It is assumed that the random variable follows a discrete distribution, hence it has a density w.r.t. the counting measure. We will use the notation $d \overset{\mathbf{x}}{\leftarrow} PS$ to express that an element d was sampled from the set PS with distribution $\mathbf{x}$; i.e., $Pr(X = d) = x_i$ if d is the i-th element in the ordered set PS.

2.1. Definitions and Game Setup

We consider a finite two-player game between player 1 and player 2 with pure strategy sets PS_1 and PS_2, respectively. Let $|PS_1| = n$ and $|PS_2| = m$ where $n, m \in \mathbb{N}$. We write $\Delta(PS)$ to denote the simplex over a strategy set PS that contains all probability distributions on PS. The extension to $n > 2$ players will be obvious so we only consider the case with two players. We assume a zero-sum situation, i.e., the attacker (which is player two) has the payoff $u_2 = -u_1$. In our security game scenario let us adopt the defenders perspective, i.e., we act as player 1 in the game. Throughout this work the defender's strategies are determined by the expected damage and the switching costs. We assume that the defender is minimizing two objectives: the primary security goal is minimization of the damage due to a risk and the second goal is reduction of the switching cost.

2.2. Costs for Playing Mixed Strategies

The *damage* that is minimized as the first objective is modeled by a utility function $u_1^{(1)}$: $\Delta(PS_1) \times \Delta(PS_2) \rightarrow \mathbb{R}$,

$$u_1^{(1)}(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{A} \mathbf{y}$$

that descries the expected damage depending on both players actions. For simplicity we assume $\mathbf{A} \in \mathbb{R}^{n \times m}$ is a constant matrix.

The second goal is *switching cost* minimization: by our definition, a switch from strategy $i \in PS_1$ to strategy $j \in PS_1$ will cause cost $s_{ij} \in \mathbb{R}^+$ for player 1. Note that the cost of switching strategies only depends on player 1's actions, i.e., on his past and present strategy which we denote by X_{t-1} and X_t respectively, and $t \in \mathbb{N}$ denotes the t-th gameplay. Thus, we can employ a first order Markov chain to

describe the switching behavior. As the player's switching costs, and therefore his next move, only depend on the present state the switching process is a first order Markov process. As any stochastic process is fully determined by its finite dimensional distribution, we can describe the switching behavior by specifying the joint probability distribution (jpd) of X_{t-1} and X_t, $t \in \mathbb{N}$. As we assume the switching costs are constant over time, the optimal jpd that determines the mode of changing strategies will be constant over time as well. Thus, the resulting, optimal switching policy joint probability distribution of $\Pr(X_{t-1} = i, X_t = j)$ can be modeled as a time-homogeneous process, i.e., it holds $\Pr(X_{t+h-1} = i, X_{t+h} = j) = \Pr(X_{t-1} = i, X_t = j) \; \forall \, h \in \mathbb{N}$. Homogeneity implies that expected switching cost can be described by

$$u_1^{(2)}(X_t, X_{t-1}) = \sum_{i=1}^{n} \sum_{j=1}^{n} s_{ij} \cdot \Pr(X_{t-1} = i, X_t = j)$$

We now model the simultaneous optimization of damage and cost as a multi-objective game (MOG). In a MOG, each player i can have $d_i \geq 2$ utility functions $u_i^{(k)}$, $k \in \{1, \dots, d_i\}$ defined over $\Delta(PS_i) \times \Delta(PS_{-i})$, where $\Delta(PS_{-i})$ denotes the strategy space of the remaining players. In our two-player zero-sum game we have 2 objectives and both players have vector-valued payoffs $\mathbf{u}_1, -\mathbf{u}_1 : \Delta(PS_1) \times \Delta(PS_2) \rightarrow \mathbb{R}$, which yields the two-player zero sum MOG $\Gamma = (\{1,2\}, \{PS_1, PS_2\}, \{\mathbf{u}_1, -\mathbf{u}_1\})$. For this situation the following definition is convenient.

Definition 1 (Pareto-Nash Equilibrium). *In game with a minimizing player 1, a Pareto-Nash equilibrium is a strategy profile* $(\mathbf{x}^*, \mathbf{y}^*) \in \Delta(PS_1) \times \Delta(PS_2)$ *that fulfills*

$$\mathbf{u}_1(\mathbf{x}, \mathbf{y}^*) \geq_1 \mathbf{u}_1(\mathbf{x}^*, \mathbf{y}^*) \geq_1 \mathbf{u}_1(\mathbf{x}^*, \mathbf{y}) \quad \forall \mathbf{x} \in \Delta(PS_1), \; \mathbf{y} \in \Delta(PS_2). \tag{1}$$

where $\mathbf{x} \geq_1 \mathbf{y}$ *means that there exists at least one coordinate i for which* $x_i \geq y_i$ *holds, regardless of the other coordinates.*

Lozovanu, Solomon and Zelikovsky [13] have studied the computation of Pareto-Nash equilibria by scalarizing the utility vector. To this end, each player i defines weights $\boldsymbol{\alpha}_i > \mathbf{0}$, $\|\boldsymbol{\alpha}_i\|_1 = 1$ to scalarize his utilities via $\boldsymbol{\alpha}_i^T \cdot \mathbf{u}_i$. In [13] it was proven that the Nash equilibria of so scalarized games are exactly the Pareto-Nash equilibria in the original multi-objective game.

Letting the defender prioritize a set of two goals by assigning weights α and $1 - \alpha$, the scalarized payoff for the defender is

$$u_1 = (1 - \alpha) \cdot u_1^{(1)} + \alpha \cdot u_1^{(2)}.$$

For readability we will drop the coefficients $(1 - \alpha)$ and α as we can just include them in the constant matrices $\mathbf{A}$ and $\mathbf{S} = (s_{ij})_{i,j=1,\dots,n}$.

2.3. The Switching Cost Model (SCM)

The model introduced in [1] assumes that the switching of strategies is performed independently of the current strategy, i.e., $\Pr(X_{t-1} = i, X_t = j) = \Pr(X_{t-1} = i) \cdot \Pr(X_t = j)$. Thus, any future change in strategy is not predictable with more accuracy when the current system state is known. Hence the utility function $u_1^{(2)}$ can be written as

$$u_1^{(2)}(X_{t-1}, X_t) = \sum_{i=1}^{n} \sum_{j=1}^{n} s_{ij} \cdot \Pr(X_{t-1} = i, X_t = j) = \sum_{i=1}^{n} \sum_{j=1}^{n} s_{ij} \cdot x_i \cdot x_j = \mathbf{x}^T \mathbf{S} \mathbf{x} = u_1^{(2)}(\mathbf{x}).$$

This way, the whole behavior of the system can be described using only the marginal probability vectors $\mathbf{x}$:

$$u_1 = u_1(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{A} \mathbf{y} + \mathbf{x}^T \mathbf{S} \mathbf{x} \tag{2}$$

with constant payoff matrices $\mathbf{A}$ as well as $\mathbf{S} = (s_{ij})_{i,j=1,\ldots,n}$. We stress the fact that $\mathbf{S}$ need not be a symmetric matrix. As a simple example consider a security guard driving to different assets i and j where j is on top of a mountains and i in the valley. The ascend from i to j will certainly take up more resources (e.g., fuel) than the decent from j to i. Thus, $s_{ij} > s_{ji}$ holds indeed. Yet, we assume that $s_{ii} = 0$ (remaining in the current strategy does not incur any switching costs).

In absence of an accurate adversary model [14], we may strive for a worst-case analysis and assume that the attacker will always try to cause as much damage as possible, i.e., they aim to maximize over u_1:

$$\max_{\mathbf{y} \in \Delta(PS_2)} (\mathbf{x}^T \mathbf{A} \mathbf{y} + \mathbf{x}^T \mathbf{S} \mathbf{x}).$$

Note that for player 2 the expression $\mathbf{x}^T \mathbf{S} \mathbf{x}$ is constant. Thus $\arg\max_{\mathbf{y} \in \Delta(PS_2)} (\mathbf{x}^T \mathbf{A} \mathbf{y} + \mathbf{x}^T \mathbf{S} \mathbf{x}) = \arg\max_{\mathbf{y} \in \Delta(PS_2)} (\mathbf{x}^T \mathbf{A} \mathbf{y}) = \arg\max_i (\mathbf{x}^T \mathbf{A} \mathbf{e}_i)$, where $\mathbf{e}_i \in \mathbb{R}^m$ denotes the i-th coordinate unit vector. By substituting $v := \max_i (\mathbf{x}^T \mathbf{A} \mathbf{e}_i + \mathbf{x}^T \mathbf{S} \mathbf{x})$, the resulting problem can be described through the following optimization problem.

$$v \to \min \qquad \text{subject to} \begin{cases} v \geq \mathbf{x}^T \mathbf{A} \mathbf{e}_i + \mathbf{x}^T \mathbf{S} \mathbf{x} & \textbf{for } i = 1, \ldots, m \\ \sum_{j=1}^n x_j = 1 \\ x_j \geq 0, & \textbf{for } j = 1, \ldots, m \end{cases} \tag{3}$$

2.4. Extension of SCM–Taking into Account Information Delay

In this paper we extend the switching cost model by relaxing the independence assumption, i.e., we let the choice of the next pure strategy X_t depend on the current state X_{t-1}. Thus, we want to model the switching behavior as a Markov process. In order to reduce the switching costs, we may add some *inertia* to player 1 by increasing the conditional probability to remain in the current strategy for each state. Will control the amount of inertia in a way that we can guarantee the distribution of the system after a predetermined amount of gameplays k conditional on the last observed state to be almost the same as the unconditional distribution.
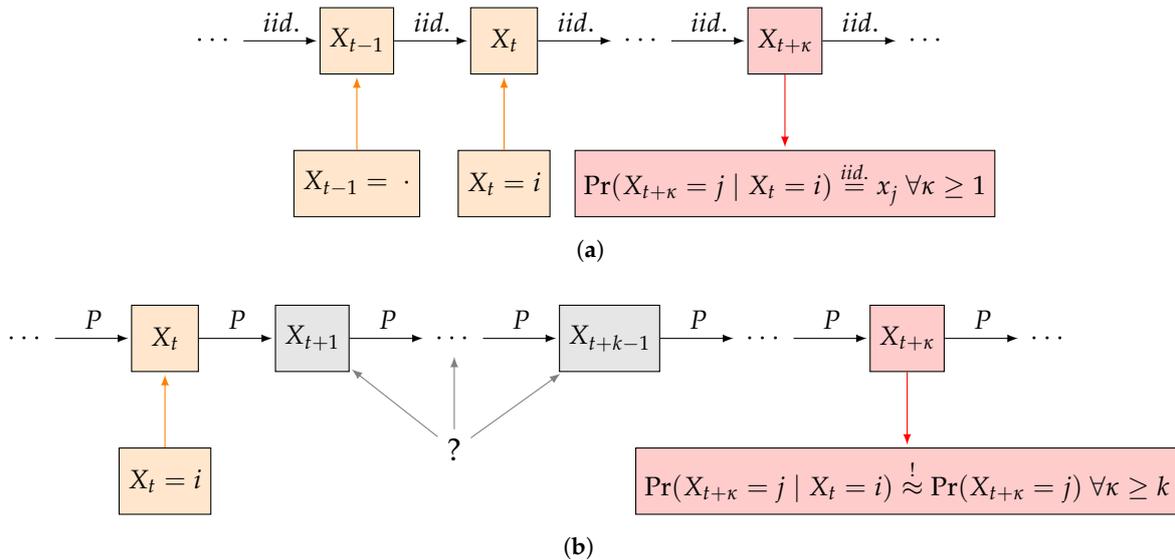

(a)

(b)

Figure 1. Comparison of switching cost optimization in [1] to our approach. (**a**) Model in [1] assumes independent choice of next pure-strategy; (**b**) Our model allows for first-order dependence when choosing next pure-strategy but controls the deviation from the independence assumption after k or more subsequent gameplays.

Hence, we demand the resulting marginal distribution after a fixed number k of consecutive repetitions of the game to be "almost independent" of the initial state, i.e., the conditional and unconditional probabilities after k or more steps need to be almost the same, that is we require

$$|P(X_{t+\kappa} = j \mid X_t = i) - P(X_{t+\kappa} = j)| \leq \epsilon \quad \forall t \in \mathbb{N}, \quad \kappa \geq k, \quad \forall i \in PS_1, \quad \forall j \in PS_1. \qquad (4)$$

where $|\cdot|$ is the sum of absolute deviations of the two probability vectors. We call k the *information delay* that specifies the length of the period an attacker is not able to gain insight into a system prior to attacking (see Figure 1). Furthermore we call ϵ the *maximum deviation of independence*. In a seemingly alternative view, one could propose wrapping up a lot of κ rounds of the game that are interdependent in a single "larger" round, yet such an approach could be flawed for two reasons: first, this would impose an independence assumption between any two batch of round in the game. Second, the timing of the game rounds may be naturally induced by the "periodicity" of the business as such (e.g., work hours per day, shifts, or similar).

In this dynamic framework we need to redefine the objective function u_1. Obviously, there is a conflict in notation and conceptualization here when optimizing over $u_1^{(1)}$ plus $u_1^{(2)}$, as $u_1^{(1)}$ is a function with arguments $\mathbf{x}$ and $\mathbf{y}$, i.e., the arguments are the marginal distributions each players assign to his set of pure strategies, but $u_1^{(2)}$ (in contrast to the formulation in (3)) is a function of the joint probability distribution of player 1's strategies. Yet, there is a direct connection between $\mathbf{x}$ and $\Pr(X_{t-1} = i, X_t = j)$: as we are dealing with mixed strategies, the defender will often switch pure strategies and by law of large numbers the distribution over pure strategies PS_1 will converge to $\mathbf{x}$ after an infinitude of gameplays. Accordingly, the dynamic (i.e., switching) behavior of player 1, which is described using a homogeneous discrete Markov chain (HDMC) needs to have $\mathbf{x}$ as a stationary as well as the unique limiting distribution in order for the two objective goals to be consistent.

Bearing in mind that the limiting behavior of any HDMC can be described using a one-step transition matrix $\mathbf{P}$ of dimension $|PS_1| \times |PS_1|$ and an initial distribution $\boldsymbol{\pi}_0$ that describes the starting state of the process, we will make use of the following theorems for our results:

Theorem 1. *(Limit [15]) Every aperiodic irreducible HDMC with finite state space has a unique limiting state $\boldsymbol{\pi}$.*

So if we are dealing with aperiodic irreducible homogeneous discrete Markov chains with finite state space $E = \{1, \ldots, \ell\}$ we can ensure the existence of a unique limiting state $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_\ell)^T$, which is always a stationary state. Additionally, it can be shown that the limiting distribution $\boldsymbol{\pi}$ of a such a stochastic process is independent of the initial distribution $\boldsymbol{\pi}_0$. Moreover, by the following ergodic theorem, it is possible to specify the speed of convergence to the limit state for an aperiodic irreducible HDMC with finite state space $E = \{1, \ldots, \ell\}$ and transition matrix $\mathbf{P}$. Let $P^k(i, j)$ denote the transition probability from state $i \in E$ to $j \in E$ after k steps. Note that the following theorem is a consequence of the Perron-Frobenius Theorem. [1]

Theorem 2. *(Geometric Ergodicity [16]) Let $\mathbf{P}$ the transition matrix of an irreducible, aperiodic Markov chain with finite state space $E = \{1, \ldots, \ell\}$. Then for all probability vectors $\boldsymbol{\pi}_0$ if holds*

$$\lim_{n \to \infty} \boldsymbol{\pi}_0^T \cdot \mathbf{P}^n = \boldsymbol{\pi}^T$$

$\boldsymbol{\pi} = (\pi_1, \ldots, \pi_\ell)^T$, $\pi_j > 0$ *for all* $j \in E$ *and* $\boldsymbol{\pi}$ *is the only solution to*

---

[1]   Note that by Perron-Frobenius Theorem for aperiodic irreducible HDMCs with finite state space the largest eigenvalue of the transition matrix is always 1 and its eigenvector is the steady state distribution. Further, the second largest eigenvalue that determines the speed of convergence to the steady state.

$$\boldsymbol{\pi}^T \cdot \mathbf{P} = \boldsymbol{\pi}^T, \qquad \sum_{i \in E} \pi_i = 1. \tag{5}$$

Moreover, the speed of convergence to the limiting state π is geometric, i.e., there exists a constant $c > 0$ (that depends on $\mathbf{P}$ only) such that

$$|P^k(i,j) - \pi_j| \le c \cdot |\lambda_2|^k \tag{6}$$

$\forall i,j \in E$ is the row vector with all ones, λ_2 denotes the second largest eigenvalue of $\mathbf{P}$ in terms of absolute values.

The following proof is from [17]. We will limit ourself tho the case when $\mathbf{P}$ is diagonalizable, which is the case for our construction of $\mathbf{P}(\theta)$.

Proof. An irreducible aperiodic Markov chain has a positive transition matrix $\mathbf{P}$. Let $\mathbf{h}_i$, $i \in E$, $\mathbf{h}_i \in \mathbb{R}^{\ell \times 1}$ denote the right eigenvectors of $\mathbf{P}$ and $\mathbf{g}_i^T$, $i \in E$, $\mathbf{g}_i \in \mathbb{R}^{\ell \times 1}$ the left eigenvectors of $\mathbf{P}$.

By Perron-Frobenius Theorem the largest eigenvalue λ_1 is unique and possesses a strictly positive left eigenvector. For stochastic matrices like $\mathbf{P}$ it additionally holds that the largest eigenvalue is $\lambda_1 = 1$, the right eigenvector to λ_1 is $\mathbf{1}$ and its left eigenvector is the one that fulfills (5). Thus, $\mathbf{g}_i^T = \boldsymbol{\pi}^T$.

Now we can write $\mathbf{P}$ in its spectral representation $\mathbf{P} = \lambda_1 \mathbf{B}_1 + \cdots + \lambda_\ell \mathbf{B}_\ell$ where $\mathbf{B}_i = \mathbf{h}_i \cdot \mathbf{g}_i^T$. As $\mathbf{B}_i \cdot \mathbf{B}_j = \mathbf{B}_i$ if $i = j$ and $\mathbf{B}_i \cdot \mathbf{B}_j = 0$ if $i \ne j$, we have $\forall k \in \mathbb{N}$

$$\mathbf{P}^k = (\lambda_1 \mathbf{B}_1 + \cdots + \lambda_\ell \mathbf{B}_\ell)^k = \lambda_1^n \mathbf{B}_1 + \cdots + \lambda_\ell^n \mathbf{B}_\ell = \mathbf{B}_1 + \lambda_2^n \mathbf{B}_2 + \cdots + \lambda_\ell^n \mathbf{B}_\ell = \mathbf{1} \cdot \boldsymbol{\pi}^T + \lambda_2^n \mathbf{B}_2 + \cdots + \lambda_\ell^n \mathbf{B}_\ell.$$

As $1 > \lambda_2 \ge \cdots \ge \lambda_\ell$ we have

$$\lim_{k \to \infty} \mathbf{P}^k = \mathbf{1} \cdot \boldsymbol{\pi}^T.$$

Now for all initial states i and resulting states j the absolute difference of the components of $\mathbf{P}^k$ and the corresponding entries in $\mathbf{1}\boldsymbol{\pi}^T$ (i.e., $|P^k(i,j) - \pi_j|$) is bounded by

$$|P^k(i,j) - \pi_j| = |\lambda_2^n B_2(i,j) + \cdots + \lambda_\ell^n B_\ell(i,j)| \le |\lambda_2|^n |B_2(i,j)| + \cdots + |\lambda_\ell|^n |B_\ell(i,j)|,$$

where $B_l(i,j)$ denotes the respective entry of $\mathbf{B}_l$, $l \in \{2, \ldots, \ell\}$.

Finally $\lambda_2 \ge \cdots \ge \lambda_\ell$ yields

$$|P^k(i,j) - \pi_j| \le |\lambda_2|^n \left(|B_2(i,j)| + \cdots + |B_\ell(i,j)| \right).$$

Finally, taking

$$c := \sup_{i,j} \; |B_2(i,j)| + \cdots + |B_\ell(i,j)| \tag{7}$$

we get the Expression (6). $\square$

Geometric ergodicity means that the absolute difference of the steady state to the marginal distribution after k steps given any initial distribution is bounded by $c \cdot |\lambda_2|^k$. Subsequently, λ_2 determines the speed of convergence to the steady state distribution given an arbitrary initial distribution $\boldsymbol{\pi}_0$: The smaller λ_2, the faster the convergence to the steady state. Considering Equation (4) it is obvious, that if we want the distributions of $X_{t+\kappa}$ and $X_{t+\kappa} \mid X_t$ for an arbitrary instantiation of X_t to differ by ϵ at maximum $\forall \kappa \ge k$, we need to control the second largest eigenvalue of $\mathbf{P}$, i.e.,

$$c \cdot |\lambda_2|^k \le \epsilon \quad \implies \quad |\Pr(X_{t+\kappa} = j \mid X_t = i) - \Pr(X_{t+\kappa} = j)| \le \epsilon \tag{8}$$

$\forall i,j \in E, \quad \forall t \in \mathbb{N}, \quad \forall \kappa \ge k$.

Now we want construct an irreducible aperiodic HDMC described by a transition probability matrix of a for which it holds

- ϵ-**Convergence:** the resulting conditional probabilities after k or more repetitions of the game given an initial are approximately the ergodic state (i.e., they satisfy (4))
- **Equilibrium**: the limiting as well as the marginal distribution of the process equal the Nash-Equilibrium-solution from (3)
- **Cost reduction**: the total costs are reduced.

In (8) have seen that ϵ-convergence can be achieved by controlling the second largest eigenvalue of the conditional probability matrix. The following result will help construct the sought transition matrix for the intended convergence control:

Theorem 3. *(Sklar [18]) Every cumulative distribution function $F_{\mathbf{X}}(\mathbf{X})$ of a random vector $\mathbf{X} = (X_1, \dots, X_n)^T$ can be expressed by its marginal distributions $F_{X_1}(x_1), \dots, F_{X_n}(x_n)$ and a copula $C : [0,1]^n \to [0,1]$ such that $F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = C\left(F_{X_1}(x_1), \dots, F_{X_n}(x_n)\right).$*

Note that we are dealing with first order HDMCs and that the whole behaviour of the chain is determined by one single two-dimensional joint distribution function for all (X_{t-1}, X_t), i.e., $\forall t$. For brevity, we will abbreviate $F_{X_{t-1}, X_t}(i,j)$ by F_{ij}. As we require both the marginal probabilities of X_t and X_{t-1} to equal the Nash-Equilibrium-solution from (3), the joint distribution of the random vector (X_{t-1}, X_t) can be constructed using the marginal distribution of X_t only, i.e., $F_{ij} = C\left((F_{X_{t-1}}(i), F_{X_t}(j))\right) = C\left(F_{X_t}(i), F_{X_t}(j)\right).$

Using Sklar's Theorem and the fact that we are only considering absolutely continuous discrete random variables X_t with σ-finite measures F_{X_t}, the first order Markov Process has not only a joint cdf, but also a discrete density $\Pr(X_{t-1} = i, X_t = j)$. Thus, there exists an $f \in [0,1]^{n \times n}$, for which it holds that for all $1 \leq i, j \leq n$: $f_{ij} = \Pr(X_{t-1} = i, X_t = j)$, $\sum_{i=1}^n f_{ij} = x_j$ for all j and $\sum_{j=1}^n f_{ij} = x_i$ for all i. Therefore, the discrete density, which is represented by $(f_{ij})_{i,j=1,\dots,n}$, has marginals prescribed by $\mathbf{x}$ and we hereafter write $f_{ij}(\mathbf{x})$ to denote this dependency. As such $f_{ij}(\mathbf{x})$, $i, j \in PS_1$ is not necessarily a function of $\mathbf{x}$, but rather chosen in a way constrained by $\mathbf{x}$ regarding the marginals.

Under the above-mentioned prerequisites, we are able to redefine $u_1^{(2)}$ so that it only depends on $\mathbf{x}$:

$$u_1^{(2)}(X_{t-1}, X_t) = \sum_{i=1}^n \sum_{j=1}^n s_{ij} \cdot \Pr(X_{t-1} = i, X_t = j) = \sum_{i=1}^n \sum_{j=1}^n s_{ij} \cdot f_{ij}(\mathbf{x}) = u_1^{(2)}(\mathbf{x})$$

for the just defined joint probability matrix $f = (f_{ij})_{i,j=1,\dots,n} : PS_1 \times PS_1 \to [0,1]^{n \times n}$, $1 \leq i, j \leq n$.

Note that it is necessary to specify parametric functions $f(\mathbf{x}, \theta)$ to model the jpd, as n^2 parameters are not estimable given the number of constraints. It is not possible to directly optimize the individual f_{ij} over $[0,1]^{n \times n}$. Therefore, we need to constrain f to a parametric family of functions, i.e., $f =: f(\boldsymbol{\theta}, \mathbf{x})$, where the optimization is performed by adjusting the parameter vector $\boldsymbol{\theta} \in \Theta$. Then, given $f(\boldsymbol{\theta}, \mathbf{x})$ which represents, the respective one step transition matrix $\mathbf{P}$ can directly be computed via

$$\mathbf{P} = diag(\mathbf{x})^{-1} \cdot f(\boldsymbol{\theta}, \mathbf{x}) \tag{9}$$

where $diag$ is a diagonal matrix. Unfortunately, even when using parametric families of functions for f in most cases controlling the value of λ_2 from $diag(\mathbf{x})^{-1} f(\boldsymbol{\theta}, \mathbf{x})$ will be difficult. For reversible Markov chains one could compute upper bounds via Cheeger's and Poincare's inequality [19], yet we will work with a direct construction scheme for $f(\boldsymbol{\theta}, \mathbf{x})$ for which it is possible to obtain *exact* control of λ_2.

3. Efficient Switching by Considering Information Delay

In the defined framework it is possible to construct an aperiodic irreducible HDMC with state space $E \subseteq PS_1$ that satisfies ϵ-convergence as well as the equilibrium condition while reducing costs at the same time.

To do so, we first set the parameters

- $k \in \mathbb{N}$: the information delay
- $\epsilon > 0$: the maximum deviation from the steady state distribution after k rounds of game play when an arbitrary initial state X_0 is given.

W.l.o.g. we assume X_0 is the last instantiation of the process which is known to the attacker. In the next step we compute the optimal solution $\mathbf{x}^*$ for $\mathbf{x}$ from (3). Then, using $\mathbf{x}^*$ we will only include those pure strategies $i \in PS_1$ in our framework for which $x_i^* > 0$ holds in Theorem (3). Excluding zero probability states is necessary, as otherwise the transition Matrix that we will construct is not positive, which is a necessary condition in Theorem 2. W.l.o.g. denote the included strategies $E = \{1, \ldots, \ell\}$, $1 \leq \ell \leq n$ and their probability vector $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_\ell)^T$. For the class of functions f we choose the following family that depends on one parameter $\theta \in (0, 1]$ and the probability vector $\tilde{\mathbf{x}}^*$:

$$f : [0, 1) \times (0, 1]^\ell \to [0, 1]^{\ell \times \ell}, \qquad f(\theta, \tilde{\mathbf{x}}^*) = \theta \cdot diag(\tilde{\mathbf{x}}^*) + (1 - \theta) \cdot \tilde{\mathbf{x}}^* \cdot (\tilde{\mathbf{x}}^*)^T \tag{10}$$

W.l.o.g. let the 0 entries of $\mathbf{x}^*$ be $x_{\ell+1}^*, \ldots, x_n^*$. For f we can easily prove the following:

1. $f_{ij}(\tilde{\mathbf{x}}^*, \theta) := \Pr(X_{t-1} = i, X_t = j)$
2. $\sum_{i=1}^{\ell} f_{ij}(\tilde{\mathbf{x}}^*, \theta) = x_j^*$ for all $j \in PS_1$ with $x_j^* \neq 0$, $\theta \in [0, 1)$
3. $\sum_{j=1}^{\ell} f_{ij}(\tilde{\mathbf{x}}^*, \theta) = x_i^*$ for all $i \in PS_1$ with $x_i^* \neq 0$, $\theta \in [0, 1)$

Now observe that by the definition of f in (10) statement (9) is equivalent to

$$\mathbf{P} = \theta \cdot diag(\tilde{\mathbf{x}}^*)^{-1} \cdot diag(\tilde{\mathbf{x}}^*) + (1 - \theta) \cdot diag(\tilde{\mathbf{x}}^*)^{-1} \cdot \tilde{\mathbf{x}}^* \cdot \tilde{\mathbf{x}}^{*T}$$

which yields

$$\mathbf{P} = \mathbf{P}(\theta) = \theta \cdot \mathbf{I} + (1 - \theta) \cdot \mathbf{1} \cdot (\tilde{\mathbf{x}}^*)^T. \tag{11}$$

where $\mathbf{I}$ denotes the identity matrix and $\mathbf{1} \in \mathbb{R}^\ell$ is the vector of all 1s. Note that by strict positivity of $(1 - \theta)$ the constructed HDMC with one step transition matrix $\mathbf{P}(\theta)$ is aperiodic and irreducible. Furthermore, the so constructed Markov chain has the limiting state $\tilde{\mathbf{x}}^*$, i.e., the limiting marginal distribution of the chain is the Nash-equilibrium solution from (3).

In this setting it can easily be verified that the largest eigenvalue of $\mathbf{P}(\theta)$ is 1 and that all remaining eigenvalues are θ:

Theorem 4. *The second largest eigenvalue λ_2 of $\mathbf{P}(\theta) = \theta \cdot \mathbf{I} + (1 - \theta) \cdot \mathbf{1} \cdot (\tilde{\mathbf{x}}^*)^T$ is $-\theta$ and has algebraic multiplicity $\ell - 1$.*

Proof. The characteristic polynomial of $\mathbf{P}(\theta)$ is

$$|\mathbf{P}(\theta) - \lambda\mathbf{I}| = |\theta \cdot \mathbf{I} + (1 - \theta) \cdot \tilde{\mathbf{x}}^* \cdot \mathbf{1}^T - \lambda \cdot \mathbf{I}| = |\underbrace{(1 - \theta) \cdot \tilde{\mathbf{x}}^* \cdot \mathbf{1}^T}_{:=\mathbf{Q}} - \underbrace{(\lambda - \theta)}_{:=\tilde{\lambda}} \cdot \mathbf{I}| = |\mathbf{Q} - \tilde{\lambda} \cdot \mathbf{I}|$$

Thus, in order to determine the eigenvalues λ_i of $\mathbf{P}(\theta)$ we need to determine the eigenvalues $\tilde{\lambda}_i$ of $\mathbf{Q}$ and it holds[2]:

$$\lambda_i = \tilde{\lambda}_i + \theta \qquad \forall i \in \{1, \ldots, \ell\}.$$

As $\mathbf{Q}$ consists of equal rows $(1 - \theta) \cdot (\tilde{x}_1^*, \tilde{x}_2^*, \ldots, \tilde{x}_\ell^*)$ the rank of $\mathbf{Q}$ is one. Thus, there exists only one eigenvalue $\tilde{\lambda}_1$ of $\mathbf{Q}$ which is not 0. As the trace of $\mathbf{Q}$ is the sum of its eigenvalues, and $\tilde{\lambda}_2 = \cdots = \tilde{\lambda}_\ell = 0$, it holds

$$\tilde{\lambda}_1 = (1 - \theta)\tilde{x}_1^* + (1 - \theta)\tilde{x}_2^* + \cdots + (1 - \theta)\tilde{x}_\ell^* = 1 - \theta.$$

---

[2] Note that the eigenvectors of $\mathbf{P}(\theta)$ and $\tilde{\mathbf{x}}^* \cdot \mathbf{1}^T$ are equal, as rescaling and adding a multiple of $\mathbf{I}$ does not alter the eigenvalues.

Thus, $\lambda_1 = 1$ and $\lambda_2 = \cdots = \lambda_\ell = \theta$. $\square$

Thus, we have proven that by proper choice of f and θ we can control the switching of policies in the way we need it. Furthermore, as we remain in the current position more often, the costs of switching are reduced while it is ensured that the marginal distribution converges after a predetermined number of rounds k at a given accuracy ϵ.

By construction of $\mathbf{P}(\theta)$ and the fact that $s_{ii} = 0$ for all $i \in E$ it furthermore holds

$$
\begin{aligned}
u_1^{(2)}(\tilde{\mathbf{x}}^*, \theta) &= \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \tilde{s}_{ij} \cdot f_{ij}(\tilde{\mathbf{x}}^*) = \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \tilde{s}_{ij} \cdot P(i,j) \cdot \tilde{x}_i^* = \sum_{i \neq j} \tilde{s}_{ij} \cdot (1-\theta) \cdot \tilde{x}_i^* \cdot \tilde{x}_j^* \\
&= (1-\theta) \cdot (\tilde{\mathbf{x}}^*)^T \tilde{\mathbf{S}} \tilde{\mathbf{x}}^* = (1-\theta) \cdot (\mathbf{x}^*)^T \mathbf{S} \mathbf{x}^* = (1-\theta) \cdot u_1^{(2)}(\mathbf{x}^*)
\end{aligned}
\tag{12}
$$

Thus, we can reduce the initial switching costs by a proportion of θ.

This way using a homogeneous discrete Markov chain we can describe a policy that employs lower average switching cost, while still controlling the damage caused by adversaries. Of course (11) is not the only construction scheme for $\mathbf{P}(\theta)$ that allows for a direct control of λ_2. There certainly exist other ones employing different copulae with similar characteristics. We chose the upper construction due to its simplicity and elegant properties regarding its second largest eigenvalue.

We will now prove a sufficient condition, when the general cost can be reduced while obtaining almost the same security. In the following it is again assumed that the scalarization constants are already included in the payoff matrices. Furthermore, assume that the information delay k is given. As mentioned before, in the information delay model we only include those strategies $\in PS_1$, where $x_j^* > 0$, that is we consider $\tilde{\mathbf{x}}^* = (\tilde{x}_1^*, \ldots, \tilde{x}_\ell^*)^T$, $E = \{1, \ldots, \ell\}$, $\ell \leq n$.

Let $\tilde{\mathbf{x}}_j^\kappa$ denote the conditional probability vector after κ steps, i.e., $\tilde{\mathbf{x}}_j^\kappa$ is the jth colum of the positive transition matrix:

$$
\tilde{\mathbf{x}}_j^\kappa = (P^\kappa(1,j), P^\kappa(2,j), \ldots, P^\kappa(\ell,j))^T
$$

Now assume it takes an attacker $\kappa \geq k$ steps to carry out an attack. Then, by the law of total probability and (12), the total utility for player one when considering information delay is given by

$$
\begin{aligned}
u_1(\mathbf{x}^*, \mathbf{y}, \theta, \kappa) &= \sum_{j=1}^{n} \left(\Pr(\mathbf{X}^*_{t+\kappa} | X_t^* = j)^T \mathbf{A} \mathbf{y} \right) \Pr(X_t^* = j) + (1-\theta)(\mathbf{x}^*)^T \mathbf{S} \mathbf{x}^* \tag{13} \\
&= \sum_{j=1}^{\ell} \left((\tilde{\mathbf{x}}_j^\kappa)^T \tilde{\mathbf{A}} \mathbf{y} \right) \tilde{x}_j^* + (1-\theta) \cdot (\tilde{\mathbf{x}}^*)^T \tilde{\mathbf{S}} \tilde{\mathbf{x}}^* \tag{14} \\
&=: u_1^{(1)}(\tilde{\mathbf{x}}^*, \mathbf{y}, \theta, \kappa) + u_1^{(2)}(\tilde{\mathbf{x}}^*, \theta) \tag{15}
\end{aligned}
$$

where $\tilde{\mathbf{A}} \in \mathbb{R}^{\ell \times \ell}$ denotes the payoff matrix $\mathbf{A}$ where all rows of states $j \in PS_1$ with $x_j^* = 0$ were deleted, and $\tilde{\mathbf{S}} \in \mathbb{R}^{\ell \times \ell}$ is defined analogously. Setting $\ell = n$, $\theta = 0$, $k = 1$ yields (2).

Now by loosening the independence assumption to lower the switching costs, we deviate from the optimal (under independent sampling of strategies) solution $\mathbf{x}^*$, which might increase the value of the first objective function $u_1^{(1)}$. The total cost incurred after $\kappa \geq k$ steps is reduced if the reduction of switching costs $u_1^{(2)}$ is higher than the increase in the value of the first objective function $u_1^{(1)}$, i.e., if

$$
u_1(\mathbf{x}^*, \mathbf{y}, 0, 1) - u_1(\mathbf{x}^*, \mathbf{y}, \theta, \kappa) \geq 0
$$

which yields

$$
\mathbf{x}^{*T} \mathbf{A} \mathbf{y} - \sum_{j=1}^{\ell} \left((\tilde{\mathbf{x}}_j^\kappa)^T \tilde{\mathbf{A}} \mathbf{y} \right) \tilde{x}_j^* + \theta \cdot \mathbf{x}^{*T} \mathbf{S} \mathbf{x}^* \geq 0
$$

or by deletion of the zero-entries in $\mathbf{x}^*$ and the correspronding rows in $\mathbf{A}$:

$$\theta \cdot (\mathbf{x}^*)^T \mathbf{S} \mathbf{x}^* \geq \left(\sum_{j=1}^{\ell} (\tilde{\mathbf{x}}_j^\kappa)^T \tilde{x}_j^* - (\tilde{\mathbf{x}}^*)^T \right) \tilde{\mathbf{A}} \mathbf{y} \tag{16}$$

Expression (16) states that the total cost can be guaranteed to be reduced if the switching cost reduction is larger than the costs incurred by deviating from $\mathbf{x}^*$. We will now show, how θ can be chosen with respect to a given maximum deviation of independence $\epsilon > 0$ in order to ensure a total cost reduction. First, assume that (4) holds; a sufficient criterion for this to hold is $c \cdot |\theta|^k \leq \epsilon$. By (7) $|c| \leq \sup_{i,j} |B_2(i,j)| + \cdots + |B_\ell(i,j)|$. Then, for the resulting conditional probability vector with information delay, it holds for all $j \in E$ and for all $\kappa \geq k$:

$$\left(\sum_{j=1}^{\ell} (\tilde{\mathbf{x}}_j^\kappa)^T \tilde{x}_j^* - (\tilde{\mathbf{x}}^*)^T \right) \tilde{\mathbf{A}} \mathbf{y} \leq \max_{j \in \{1,\dots,\ell\}} |(\tilde{\mathbf{x}}^* - \tilde{\mathbf{x}}_j^\kappa)^T \tilde{\mathbf{A}} \mathbf{y}|$$

and as the maximum deviation from independence is ϵ we have

$$|(\tilde{\mathbf{x}}^* - \tilde{\mathbf{x}}_j^\kappa)^T \tilde{\mathbf{A}} \mathbf{y}| \leq \max_i |(\tilde{\mathbf{x}}^* - \tilde{\mathbf{x}}_j^\kappa)^T \tilde{\mathbf{A}} \mathbf{e}_i| = \|\tilde{\mathbf{A}}^T \cdot (\tilde{\mathbf{x}}^* - \tilde{\mathbf{x}}_j^k)\|_\infty \overset{(8)}{\leq} \epsilon \cdot \|\tilde{\mathbf{A}}^T\|_\infty = \epsilon \cdot \max_{j=1,\dots,\ell} \sum_{i=1}^n |\tilde{a}_{ij}| \tag{17}$$

Thus, we have proven the following theorem:

Theorem 5 (Cost reduction).

$$\left(\epsilon \cdot \max_{j=1,\dots,\ell} \sum_{i=1}^n |\tilde{a}_{ij}| \leq \theta \cdot \mathbf{x}^{*T} \mathbf{S} \mathbf{x}^* \right) \wedge (c \cdot |\theta|^k \leq \epsilon) \implies \underbrace{u_1(\mathbf{x}^*, \mathbf{y}, 0, 1) - u_1(\mathbf{x}^*, \mathbf{y}, \theta, \kappa) \geq 0}_{\text{lower cost}} \tag{18}$$

This implies that, if θ and ϵ satisfy the condition (18), then the total costs are reduced. The following example illustrates the results.

3.1. Example and Sensitivity Analysis

Consider the following two-player zero-sum Matrix-game with switching costs. We define the parameters

$$\mathbf{A} = \begin{pmatrix} 4 & 8 & 6 \\ 9 & 3 & 8 \\ 7 & 6 & 3 \end{pmatrix}, \qquad \mathbf{S} = \begin{pmatrix} 0 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 3 & 0 \end{pmatrix}.$$

For simplicity it is assumed that the scalarization constants are already included in the payoff matrices. Furthermore, assume that the information delay is $k = 3$.

The standard equilibrium (we will denote it as $\hat{\mathbf{x}}$), if we only consider $\mathbf{A}$, is $\hat{\mathbf{x}} = (0.511, 0.311, 0.178)^T$ with a value 6.0889. The cost of switching strategies independently, as it is assumed in standard models, however, would incur an extra cost of 0.88 per round, which yields 6.9689 in total.

Now, computing $\mathbf{x}^*$ from (3) yields $\mathbf{x}^* = (0.2, 0, 0.8)^T$. The switching costs are 0.32 and the maximum damage caused by the adversary is 6.4. This yields average costs of 6.72 in each repetition of the game. Note that this strategy only includes the first and the last strategy $\in PS_1$.

We will lower this average cost by taking into account the adversaries inertia, which is represented by the information delay of $k = 3$ rounds. As mentioned before, in the information delay model we only include those strategies j, where $x_j^* > 0$, hereafter denoted as $\tilde{\mathbf{x}}^* = (0.2, 0.8)^T$, $E = \{1, 3\}$ and

$\ell = 2$. By (7), $|c| \leq \sup_{i,j} |\mathbf{B}_2(i,j)|$. In our case the matrix $\tilde{\mathbf{x}}^* \cdot \mathbf{1}^T$ has left eigenvectors $f_1 = (1,1)^T$, $f_2 = (-4,1)^T$ and right eigenvectors $g_1^T = (0.2, 0.8)$, $g_2^T = (-0.2, 0.2)$. Thus,

$$\mathbf{B}_2 = \begin{pmatrix} -4 \\ 1 \end{pmatrix} \cdot (-0.2, 0.2) = \begin{pmatrix} 0.8 & -0.8 \\ -0.2 & 0.2 \end{pmatrix} \text{ and therefore } |c| \leq 0.8$$

Solving the inequalities $\epsilon \cdot \max_{j=1,\dots,\ell} \sum_{i=1}^{n} |\tilde{a}_{ij}| \leq \theta \cdot \mathbf{x}^{*T} \mathbf{S} \mathbf{x}^*$ and $c \cdot |\theta|^k \leq \epsilon$ from (18) for $\tilde{\mathbf{A}} = \begin{pmatrix} 4 & 8 & 6 \\ 7 & 6 & 3 \end{pmatrix}$ yields $0.8 \cdot \theta^3 \leq \epsilon \leq \frac{4 \cdot \theta}{225}$. Replacing the inequalities by equalities we obtain $\theta = \frac{1}{3\sqrt{5}} \approx 0.149$ and $\epsilon = \frac{4 \cdot \theta}{225} \approx 0.00265$, i.e., if the adversary knows the initial state, the individual conditional probabilities $P^k(i,j)$ after 3 or more steps will differ from each component $\tilde{\mathbf{x}}^*$ by about a quarter of a percent point at maximum.

Inserting $\theta = 0.149$ into $\mathbf{P}^k(\theta) = (\theta \cdot \mathbf{I} + (1-\theta) \cdot \tilde{\mathbf{x}}^* \cdot \mathbf{1}^T)^k$ for $k = 3$ it can be seen that the maximum deviation to the components of $\tilde{\mathbf{x}}^*$ is indeed no more than ϵ:

$$\mathbf{P} = \begin{pmatrix} 0.3192 & 0.6808 \\ 0.1702 & 0.8298 \end{pmatrix}, \qquad \mathbf{P}^3 = \begin{pmatrix} 0.202646 & 0.797354 \\ 0.199338 & 0.800662 \end{pmatrix}$$

For $\theta = 0.149$ the switching costs are $\sum_{i=1}^{\ell} \sum_{j=1}^{\ell} s_{ij} \cdot f(\tilde{\mathbf{x}}^*)_{ij} = (1 - 0.149) \cdot 0.32 = 0.27232$ and the value for $u_1^{(1)}$ is obtained using Expression (13):

$$0.2 \cdot \max_{i \in \{1,2,3\}} (0.202646, 0.797354) \cdot \begin{pmatrix} 4 & 8 & 6 \\ 7 & 6 & 3 \end{pmatrix} \cdot \mathbf{e}_i$$

$$+ \qquad 0.8 \cdot \max_{i \in \{1,2,3\}} (0.199338, 0.800662) \cdot \begin{pmatrix} 4 & 8 & 6 \\ 7 & 6 & 3 \end{pmatrix} \cdot \mathbf{e}_i$$

$$= \qquad 0.2 \cdot \max\{6.39206, 6.40529, 3.60794\} + 0.8 \cdot \max\{6.40199, 6.39868, 3.59801\}$$

$$= \qquad 0.2 \cdot 6.40529 + 0.8 \cdot 6.40199$$

$$= \qquad 6.40265.$$

Thus, the total average cost incurred is 6.67497. The switching costs were reduced by 14.9% in each round, while the maximum damage caused by an adversary was only increased by 0.04140625%. Henceforth, taking into account the adversaries inertia can cause a dramatic cost reduction, while still ensuring almost the same security.

Remark 1. *Another possibility to find admissible θ is to apply the bisection method to θ until the maximum deviation of the entries of $\mathbf{P}^3$ to $\tilde{\mathbf{x}}^*$ is smaller than a predetermined ϵ. As an example we chose $\epsilon = 0.005$ and obtained $\theta = 0.187$, which yields even lower switching costs (0.26016). We obtained*

$$\mathbf{P} = \begin{pmatrix} 0.3496 & 0.6504 \\ 0.1626 & 0.8374 \end{pmatrix}, \qquad \mathbf{P}^3 = \begin{pmatrix} 0.205231 & 0.794769 \\ 0.198692 & 0.801308 \end{pmatrix}.$$

In this case, the average value of $u_1^{(1)}$ is 6.405228 and the total cost per round is 6.665388.

4. Minimizing the Total Cost

If one wishes not only to find a way to efficiently implement the Nash-equilibrium solution $\mathbf{x}^*$ from (3), but also allows for other ergodic states $\pi \neq \tilde{\mathbf{x}}^*$ to minimize the total cost while still ensuring ϵ-convergence after k steps, one can rewrite the utility function (13) using the transition matrix $\mathbf{P}$:

$$u_1(\boldsymbol{\pi}, \mathbf{y}, \theta, \kappa) = \sum_{j=1}^{\ell} \boldsymbol{\pi}^T \mathbf{P}^\kappa \tilde{\mathbf{A}} \mathbf{y} + (1 - \theta) \cdot \boldsymbol{\pi}^T \tilde{\mathbf{S}} \boldsymbol{\pi} \tag{19}$$

Here, k is is the information delay (an input parameter), and $\boldsymbol{\pi} \in \mathbb{R}^\ell$ is an $\ell \times 1$ probability vector over $E = \{1, \dots, \ell\} \subseteq PS_1$, that only includes non-0-probability strategies from PS_1. $\tilde{\mathbf{A}}$, $\tilde{\mathbf{S}}$ likewise denote the cleaned from zeros payoff and switching cost matrices only including the strategies from E. For $\boldsymbol{\pi}$, the transition matrix is defined as $\mathbf{P} = \theta \cdot \mathbf{I} + (1 - \theta) \cdot \mathbf{1} \cdot (\boldsymbol{\pi})^T$. The global optimum can then be found by solving the following optimization problem:

$$\min_{\boldsymbol{\pi}, \theta} \max_i \sum_{j=1}^{\ell} \boldsymbol{\pi}^T \mathbf{P}^k \tilde{\mathbf{A}} \mathbf{e}_i + (1 - \theta) \boldsymbol{\pi}^T \tilde{\mathbf{S}} \boldsymbol{\pi} \tag{20}$$

subject to $c \cdot \theta^k \leq \epsilon$, $\theta \in [0, 1)$, $\pi_j > 0 \; \forall j \in E$, $\sum_{j \in E} \pi_j = 1$. The optimization is performed in the following way: first, it is decided which stategies from PS_1 to include in E. i.e., we choose a subset of stategies from PS_1 and w.l.o.g denote them $\{1, \dots, \ell\}$. Then, the matrix $\mathbf{A}$ is reduced to $\tilde{\mathbf{A}} \in \mathbb{R}^{\ell \times m}$, which obtained by deleting all rows of stategies that are not included in E. Analogously, $\tilde{\mathbf{S}} \in \mathbb{R}^{\ell \times \ell}$ is obtained by deleting all columns and rows of strategies $\notin E$. Having obtained $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{S}}$ and using the spectral representation of $\mathbf{P}^k$ we can reformulate (20):

$$\min_{\boldsymbol{\pi}, \theta} \max_i \sum_{j=1}^{\ell} \boldsymbol{\pi}^T \left(\mathbf{1} \cdot \boldsymbol{\pi}^T + \theta^k (\mathbf{B}_2 + \cdots + \mathbf{B}_\ell) \right) \tilde{\mathbf{A}} \mathbf{e}_i + (1 - \theta) \cdot \boldsymbol{\pi}^T \tilde{\mathbf{S}} \boldsymbol{\pi} \tag{21}$$

subject to $c \cdot \theta^k \leq \epsilon$, $\theta \in [0, 1)$, $\pi_j > 0 \; \forall j \in E$, $\sum_{j \in E} \pi_j = 1$. Note that all $\mathbf{B}_i$, $i = 2, \dots, \ell$ depend continuously on $\boldsymbol{\pi}$.

5. Discussion

It is interesting to note that despite optimality-by-design, some short-term deviations from an equilibrium can indeed be rewarding. Extending the concepts put forth in this work to dynamic games (e.g., leader-follower scenarios) is a natural next step. The methods used here lend themselves also to a treatment of perhaps continuous time chains, as limits of sequences of discrete chains with vanishing pauses in the limit. For practical matters, our work can provide a tool to fix implausible or impractical equilibria, by avoiding "hectic" changes if the equilibrium is mixed, while retaining a good security-investment trade-off. In the end, reinvesting the saved cost in additional security measures will yield even more security at the same cost.

At first glance our result seem to contrast earlier findings. For example, Reference [20] states that a defending player may actually benefit from revealing information about the defense strategy to the adversary and Reference [21] suggest that centrally allocating resources and publicly announcing the defensive allocation yields higher success probabilities for a defender. Both approaches deal with publicly announcing defense strategies to influence alleged attackers. This is different from our situation as we do not consider influencing the attacker (neither by providing potentially misleading information nor by hiding information). Rather we investigate how players behave if an information delay is part of the setting of the game, i.e., if it needs to be taken into account due to the situation at hand.

Author Contributions: Conceptualization, J.W.; Formal analysis, J.W.; Funding acquisition, S.R.; Investigation, J.W.; Methodology, J.W.; Project administration, S.R.; Resources, J.W.; Supervision, S.R.; Validation, J.W., S.R. and S.K.; Visualization, J.W.; Writing—original draft, J.W., S.R. and S.K.; Writing—review & editing, S.R. and S.K.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SCM   Switching Cost Model
jpd   Joint Probability Distribution
HDMC   homogeneous discrete Markov chain

References

1. Rass, S.; König, S.; Schauer, S. On the Cost of Game Playing: How to Control the Expenses in Mixed Strategies. In *Decision and Game Theory for Security*; Rass, S.; An, B.; Kiekintveld, C.; Fang, F.; Schauer, S. Eds.; Springer: New York, NY, USA, 2017; pp. 494–505, ISBN 978-3319687100.
2. Dijk, M.; Juels, A.; Oprea, A.; Rivest, R.L. FlipIt: The Game of "Stealthy Takeover". *J. Cryptol.* **2013**, *26*, 655–713. [CrossRef]
3. Fudenberg, D.; Levine, D.K. *The Theory of Learning in Games*; MIT Press: London, UK, 1998.
4. Chien, S.; Sinclair, A. Convergence to approximate Nash equilibria in congestion games. *Games Econ. Behav.* **2011**, *71*, 315–327, doi:10.1016/j.geb.2009.05.004. [CrossRef]
5. Even-Dar, E.; Kesselman, A.; Mansour, Y. Convergence time to Nash equilibrium in load balancing. *ACM Trans. Algorithms* **2007**, *3*, 32. [CrossRef]
6. Even-Dar, E.; Kesselman, A.; Mansour, Y. Convergence Time to Nash Equilibria. *Lect. Notes Comput.Sci.* **2003**, *2719*, 502–513.
7. Pal, S.; La, R.J. Simple learning in weakly acyclic games and convergence to Nash equilibria. In Proceedings of the 53rd Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 29 September–2 October 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 459–466.
8. Zhu, Q.; Başar, T. Game-Theoretic Approach to Feedback-Driven Multi-stage Moving Target Defense. In *4th International Conference on Decision and Game Theory for Security—Volume 8252*; GameSec 2013; Springer-Verlag, Inc.: New York, NY, USA, 2013; pp. 246–263.
9. Rass, S.; König, S. Password Security as a Game of Entropies. *Entropy* **2018**, *20*, 312. [CrossRef]
10. McDonald, S.; Wagner, L. Using Simulated Annealing to Calculate the Trembles of Trembling Hand Perfection. In Proceedings of the 2003 Congress on Evolutionary Computation, Canberra, Australia, 8–12 December 2003.
11. Hespanha, J.P.; Prandini, M. Nash equilibria in partial-information games on Markov chains. In Proceedings of the 40th IEEE Conference on Decision and Control, Orlando, FL, USA, 4–7 December 2001; IEEE: Piscataway, NJ, USA, 2001; pp. 2102–2107.
12. Anderson, R. Why information security is hard—An economic perspective. In Proceedings of the 17th Annual Computer Security Applications Conference (ACSAC 2001), New Orleans, LA, USA, 10–14 December 2001; pp. 358–365.
13. Lozovanu, D.; Solomon, D.; Zelikovsky, A. Multiobjective Games and Determining Pareto-Nash Equilibria. *Bul. Acad. Stiint. Republicii Mold. Mat.* **2005**, *49*, 115–122.
14. Rios Insua, D.; Rios, J.; Banks, D. Adversarial Risk Analysis. *J. Am. Stat. Assoc.* **2009**, *104*, 841–854. [CrossRef]
15. Parzen, E. *Stochastic Processes*; Dover Publications, Inc.: Mineola, NY, USA, 2015.
16. Cinlar, E. *Introduction to Stochastic Processes*; Springer-Varlag: New York, NY, USA, 1975.
17. Lorek, P. Speed of Convergence to Stationarity for Stochastically Monotone Markov Chains. Ph.D. Thesis, University of Wroclaw, Wroclaw, Poland, 2007.
18. Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **1959**, *8*, 229–231.
19. Diaconis, P.; Stroock, D. Geometric bounds for eigenvalues of Markov chains. *Ann. Probab.* **1991**, *1*, 36–61. [CrossRef]
20. Cotton, C.; Li, C. Profiling, screening and criminal recruitment. *J. Public Econ. Theory* **2014**, *17*, 964–985. [CrossRef]
21. Bier, V.; Oliveros, S.; Samuelson, L. Choosing what to protect: Strategic defensive allocation against an unknown attacker. *J. Public Econ. Theory* **2007**, *9*, 563–587. [CrossRef]