Adversarial Leakage in Games

Noga Alon[*] Yuval Emek[†] Michal Feldman[‡] Moshe Tennenholtz[§]

Abstract

While the maximin strategy has become the standard, and most agreed-upon solution for decision-making in adversarial settings, as discussed in game theory, computer science and other disciplines, its power arises from the use of mixed strategies, a.k.a. probabilistic algorithms. Nevertheless, in adversarial settings we face the risk of information leakage about the actual strategy instantiation. Hence, real robust algorithms should take information leakage into account. To address this fundamental issue, we introduce the study of *adversarial leakage* in games. We consider two models of leakage. In both of them the adversary is able to learn the value of b binary predicates about the strategy instantiation. In one of the models these predicates are selected after the decision-maker announces its probabilistic algorithm and in the other one they are decided in advance. We give tight results about the effects of adversarial leakage in general zero-sum games with binary payoffs as a function of the level of leakage captured by b in both models. We also compare the power of adversarial leakage in the two models and the robustness of the original maximin strategies of games to adversarial leakage. Finally, we study the computation of optimal strategies for adversarial leakage models. Together, our study introduces a new framework for robust decision-making, and provides rigorous fundamental understanding of its properties.

[*] Tel-Aviv University, and Microsoft Israel R&D Center.

[†] Microsoft Israel R&D Center.

[‡] School of Business Administration, The Hebrew University of Jerusalem, and Microsoft Israel R&D Center.

[§] Microsoft Israel R&D Center, and Technion-Israel Institute of Technology.

1 Introduction

Decision-Making lies in the foundations of fields such as Economics, Operations Research, and Artificial Intelligence. The question of what should be the action to be taken by a decision-maker when facing an uncertain environment, potentially consisting of other decision makers, is a fundamental problem which led to a wide variety of models and solutions. The only type of situations for which this question got an agreed-upon answer is in the context of two-player zero-sum games. This setting can model any situation in which a decision-maker aims at maximizing his guaranteed payoff. When mixed strategies are allowed, such desired behavior, termed an agent's maximin (or safety level) strategy, leads to a well defined expected payoff (known as the *value* of the game). Moreover, when presented explicitly in a matrix form, the computation of a maximin strategy is polynomial (by solving a linear program). Various equilibrium concepts have been considered in the game-theoretic literature, but none of them provides a prescriptive advice to a decision-maker which will be as acceptable as the maximin strategy solution in adversarial settings. Since the introduction of the study of two-person zero-sum games [13], maximin strategies have received very little criticism (see [5] for an exception). Moreover, the safety level strategy has been advocated for some non zero-sum settings as well (see [11], following observations by [4]).

Much of the power of a maximin strategy is associated with the use of mixed strategies, a.k.a. randomized algorithms. In such algorithms the randomization phase is assumed to be done in a private manner by the decision-maker, and no information about the instantiation selected in that phase is assumed to be revealed. In reality, however, nothing is really private; for example, competitors will always strive to obtain the private actions of a business, possibly by means of industrial espionage [10]; hence, information leakage should be considered. As a result, it may be of interest to study the effects of adversarial leakage, where a limited amount of information on an agent's instantiation of its mixed strategy may leak in an adversarial manner. We believe that only by considering this situation, it will be possible to construct robust strategies when acting in an adversarial setting. Information leakage appeared in game theory in the context of conditioning a player's strategy about the other player's strategy [12, 8]; however that work did not consider the leakage of mixed strategy instantiations nor its effects on designing robust algorithms in adversarial settings taking information leakage into account.

Our model of adversarial leakage is general. We consider a two-player zero-sum game in strategic form (a.k.a. matrix form), where the MAX player is our decision-maker and the MIN player is the adversary. Both MAX and MIN have a set of (pure) strategies they can choose from. MAX chooses a mixed strategy, that is, a distribution vector over its pure strategies. MIN may base its action on the value of b binary predicates defined on MAX' pure strategies; each such predicate is a Boolean formula on the set of strategies whose value is determined according to the actual instantiation of MAX' mixed strategy. The parameter b can be thought of as the amount of information leakage (or number of leaking bits) regarding the instance of MAX' mixed strategy; MAX would like to maximize his guaranteed expected payoff against any choice of such b binary predicates.

We consider two settings, distinguished by the information structure assumed in them. In the *Strong Model* the MAX player chooses a mixed strategy, which is observable by the MIN player, who can then act upon it in determining the b predicates. In the *Weak Model*, on the other hand, the MIN player chooses the b predicates first, and MAX can observe it and act upon it in choosing his mixed strategy.

Compared with the Strong Model, the information structure of the Weak Model provides a potential advantage to MAX, but is it effectively advantageous? Understanding this issue is particularly interesting in light of the minimax theorem, which essentially states that in a classic two-player zero-sum game, if mixed strategies are allowed, then gaining information about the opponent's mixed strategy prior to playing does not give the agent possessing it any strategic advantage.

Other intriguing questions arise in this setting of adversarial leakage. What would be the best mixed strategy for the MAX player? How well will the original maximin strategy of the game perform? What is the computational complexity of finding the optimal strategy under information leakage? We address all these questions, focusing our attention on general two-person games, where the decision-maker has m strategies to choose from, the adversary has n strategies to choose form, and the payoffs to the decision maker are either 1 or 0. This is known to be a highly applicable model, as it captures games in which a goal is either achieved or not.

Our results. For the Strong Model, if the value of the game is $q = 1 - \epsilon$ (for small positive ϵ) and 2^b is much smaller than $1/\epsilon$, then MAX can ensure value close to 1 (at least $1 - 2^b\epsilon$), and this is tight. To do so, she simply uses the maximin strategy (that is, the optimal mixed strategy for the original game with no predicates). On the other hand, if 2^b is much bigger than $1/\epsilon$, then for every mixed strategy of MAX, the MIN player can ensure value close to zero (at most $e^{-2^b\epsilon}$). Therefore, for EVERY such game with value $1 - \epsilon$, which is close to 1, a sharp transition occurs at b which is about $\log(1/\epsilon)$: if b is slightly smaller, the value stays close to 1; if it is slightly larger, the value drops to nearly zero.

For games with value q bounded away from 1, even one bit enables MIN to square the value and drop it to at most q^2, and every additional bit squares the value again. There are also examples showing that this is essentially tight. Finally, for any fixed value $q < 1$, $\log\log m + O_q(1)$ bits suffice to enable MIN to drop the value to precisely 0.

For the Weak Model, the situation is different. Clearly, here MAX is in a better shape, hence if the value of the game is $q = 1 - \epsilon$ (for small positive ϵ), MAX can still ensure a value close to 1 if the number of bits is much smaller than $\log(1/\epsilon)$ as in the Strong Model. For games with value q bounded away from 1, however, there are examples in which she can do much better than in the Strong Model, and in fact can ensure no essential drop in the value as long as the number of leaking bits is somewhat smaller than $\log\log m$. More precisely, for any fixed value $0 < q < 1$ and for every large polynomially related m, n, there are examples of games represented by a binary m by n matrix with value $q + o(1)$, so that even if $b = \log\log m - O(1)$, MAX can ensure that the value will stay roughly q. This should be contrasted with the Strong Model, where every additional bit squares MAX' value.

Somewhat surprisingly, once the number of leaking bits is slightly larger, that is, $b = \log\log m + O(1)$, the MIN player can already ensure value 0 in any game with a fixed value $q < 1$. Thus, in the examples above a sharp transition occurs at nearly $b = \log\log m$ under the Weak Model: nearly $\log\log m$ bits have essentially no effect on the value, while slightly more bits already suffice to drop the value to 0.

Note that, in contrast to leakage-free settings, where no advantage is gained by observing the opponent's mixed strategy (due to the minimax theorem), in settings of adversarial leakage, such information can contribute a great deal to the informed player, reflected by the advantage obtained by MAX in the Weak Model compared with the Strong Model.

With respect to computation complexity, computing the optimal strategy in the Strong Model (for the MAX player) against b leaking bits is poly-time for any fixed b, while this problem becomes NP-hard to compute, or even to approximate within any factor, for a general b. In the Weak Model, the optimal strategy of MAX can be computed in polynomial time for every b. As for the MIN player, computing the optimal predicates is polynomial for a fixed number of bits, but is NP-hard in general.

2 Model

We consider two-player zero-sum games defined by an m by n matrix M with $\{0,1\}$ entries, where the rows correspond to MAX' pure strategies and the columns correspond to MIN's pure strategies: $M_{i,j}$ is the payoff of MAX if MAX and MIN play row i and column j, respectively (the payoff of MIN is then $-M_{i,j}$).[1] The matrix M is known to both players.

Given a matrix M and an integer $b \geq 0$, we describe a precise setting of adversarial leakage, as follows:
(1) MAX chooses a distribution vector $\mathbf{p} = (p_1, \ldots, p_m)$ on $[m]$ and MIN chooses a b-bit leakage function $f : [m] \to \{0,1\}^b$.
(2) MAX realizes $i \in [m]$ according to $\mathbf{p}$ (i.e., chooses row i with probability p_i).
(3) MIN observes $f(i)$ (for i realized by MAX) and chooses a strategy $j \in [n]$.
(4) MAX and MIN receive payoffs $M_{i,j}$ and $-M_{i,j}$, respectively.
The two leakage models we consider, referred to as the Strong Model and the Weak Model differ in the order in which the choices in step (1) are made. In the Strong Model MAX first chooses a mixed strategy $\mathbf{p}$ and MIN may base its choice of f on the knowledge of $\mathbf{p}$. In the Weak Model MIN first choose a leakage function f and MAX may base its choice of $\mathbf{p}$ on the knowledge of f.

It will be convenient to formalize the choice of (pure) strategy made by MIN in step (3) as a function $g : \{0,1\}^b \to [n]$. Note that MIN decides on g when it already knows the mixed strategy $\mathbf{p}$ of MAX. This is less important under the Strong Model, where it can be assumed that MIN chooses g simultaneously with its choice of f. However, under the Weak Model, the choice of g must be made at a later stage (when MIN already knows $\mathbf{p}$).

Given a matrix M, a non-negative integer b, a distribution vector $\mathbf{p}$ on $[m]$, a function $f : [m] \to \{0,1\}^b$, and a function $g : \{0,1\}^b \to [n]$, let

$$u(M, b, \mathbf{p}, f, g) = \sum_{w \in \{0,1\}^b} \sum_{i : f(i) = w} p_i M_{i,g(w)}$$

denote the expected payoff of MAX (with respect to these parameters). Denote

$$u_{\mathbf{p}}(M, b) = \min_{f:[m]\to\{0,1\}^b \text{ and } g:\{0,1\}^b \to [n]} u(M, b, \mathbf{p}, f, g) \ .$$

The *value* of M against b leaking bits under the Strong Model is defined as

$$v^{\text{strong}}(M, b) = \max_{\mathbf{p} \in \Delta(m)} u_{\mathbf{p}}(M, b) \ ,$$

where $\Delta(m)$ is the set of all distribution vectors on $[m]$. We denote by $\mathbf{p}^*_b$ a distribution vector that realizes $v^{\text{strong}}(M, b)$, i.e., $u_{\mathbf{p}^*_b}(M, b) = v^{\text{strong}}(M, b)$. The *value* of M against b leaking bits under the Weak Model is defined as

$$v^{\text{weak}}(M, b) = \min_{f:[m]\to\{0,1\}^b} \max_{\mathbf{p} \in \Delta(m)} \min_{g:\{0,1\}^b \to [n]} u(M, b, \mathbf{p}, f, g) \ .$$

[1] While we focus on the natural binary case, some of our results hold for any matrix with entries in $[0,1]$ as well, while some become non-interesting or easily seen to be false.

When the leakage model is clear from the context, we may omit the superscripts and write simply $v(M,b)$. Observe that under this notation, $v(M,0)$ is the classical value of (the game defined by) M.

Unless otherwise specified, all logarithms are in base 2.

3 Adversarial leakage in the Strong Model

We first show that for any m by n matrix with $\{0,1\}$ entries of value $q = 1-\epsilon$, the MAX player can guarantee herself at least a payoff of $1-2^b\epsilon$. This can be done, in particular, by playing the maximin strategy.

Proposition 3.1. *Let M be an m by n matrix with $\{0,1\}$ entries. Let $q = 1-\epsilon$ be the value of the game defined by M, that is, $v(M,0) = 1-\epsilon$. Then, for every $b \geq 0$, $u_{\mathbf{p}_0^*}(M,b) \geq 1-2^b\epsilon$.*

Proof. Let $\mathbf{p} = (p_1, \ldots, p_m)$. For every $w \in \{0,1\}^b$, let $S^w = \{i | f(i) = w\}$, and let $p^w = \sum_{i \in S^w} p_i$. Fix some column j. Since $1-\epsilon$ is the value of the game, it holds that for every w, $\sum_{i \in S^w} p_i M_{i,j} + \sum_{i \notin S^w} p_i M_{i,j} \geq 1-\epsilon$. As $M_{i,j} \leq 1$ for every i,j, we have $\sum_{i \in S^w} p_i M_{i,j} + \sum_{i \in [m] \setminus S^w} p_i \geq 1-\epsilon$. Substituting $\sum_{i \in S^w} p_i = 1-p^w$ and rearranging the last inequality yields

$$\sum_{i \in S^w} p_i M_{i,j} \geq p^w - \epsilon. \tag{1}$$

The expected payoff of MAX is given by the expression $\sum_{w \in \{0,1\}^b} \sum_{i:f(i)=w} p_i \cdot M_{i,g(w)}$ and the expected payoff of MAX conditioned on the event that some row $i \in S^w$ is played is given by the expression $\sum_{i:f(i)=w} \frac{p_i}{p^w} \cdot M_{i,g(w)}$, which is at least $\frac{1}{p^w}(p^w - \epsilon)$, by Equation (1). Therefore the expected payoff of MAX is at least $\sum_{w \in \{0,1\}^b} p^w \frac{1}{p^w}(p^w - \epsilon) = 1 - 2^b\epsilon$. □

The above bound is tight, as established in the following proposition.

Proposition 3.2. *For every $\epsilon > 0$ and every $b \geq 0$, there exists a matrix M with $\{0,1\}$ entries so that (1) $v(M,0) = 1-\epsilon$; and (2) $u_{\mathbf{p}_0^*}(M,b) = u_{\mathbf{p}_b^*}(M,b) = 1-2^b\epsilon$.*

Proof. Let $n = 1/\epsilon$ and consider the n by n matrix M in which $M_{i,i} = 0$ for every i, and $M_{i,j} = 1$ for every $i \neq j$. From symmetry considerations, both the maximin strategy and the optimal strategy against b leaking bits is the uniform distribution over the rows. Let f be a function which imposes the following partition on the rows: each one of the first $2^b - 1$ rows constitutes its own subset, and the remaining rows constitute the last subset. In this case, if one of the first 2^b-1 rows is chosen (each with probability ϵ), then MAX' payoff is 0, while if one of the remaining rows is chosen (with a total probability of $1-(2^b-1)\epsilon$), then the payoff obtained by the MAX player is $\frac{\frac{1}{\epsilon}-2^b}{\frac{1}{\epsilon}-(2^b-1)}$. The expected payoff of the MAX player is therefore $(1-(2^b-1)\epsilon) \cdot \frac{\frac{1}{\epsilon}-2^b}{\frac{1}{\epsilon}-(2^b-1)} = 1-2^b\epsilon$. □

The above two propositions essentially say that for games with value $q = 1-\epsilon$ and b such that $2^b\epsilon = o(1)$, MAX can guarantee a payoff of about q^{2^b} by playing the maximin strategy, and this is optimal. The case of general q and b, however, requires more work, and this is the focus of the following statement.

Theorem 3.3. *Let M be an m by n matrix with $\{0,1\}$ entries. Let q be the value of the game defined by M, that is, $q = v(M,0)$. Then, for every $b \geq 0$ and every distribution vector $\mathbf{p}$ of the MAX player, $u_{\mathbf{p}}(M,b) \leq q^{2^b}$.*

Proof. Put $\mathbf{p}^{(1)} = \mathbf{p}$, and let $j_1 \in [n]$ be a pure strategy of MIN (a column of M) ensuring a value of $q_1 \leq q$ against the mixed strategy $\mathbf{p}^{(1)}$ of MAX. Such a pure strategy must exist since q is the value of the game. Define $S_1 = \{i \in [m] \mid M_{i,j_1} = 0\}$. It holds that $\sum_{i \in S_1} \mathbf{p}_i^{(1)} M_{i,j_1} + \sum_{i \in [m]-S_1} \mathbf{p}_i^{(1)} M_{i,j_1} = q_1$, hence $\sum_{i \in [m]-S_1} \mathbf{p}_i^{(1)} = q_1$.

Let $\mathbf{p}^{(2)}$ be the distribution vector defined by restricting $\mathbf{p}^{(1)}$ to the rows in $[m] - S_1$, namely,

$$\mathbf{p}_i^{(2)} = \begin{cases} \mathbf{p}_i^{(1)}/q_1 & \text{if } i \in [m] - S_1; \\ 0 & \text{otherwise.} \end{cases}$$

Let j_2 be a pure strategy of MIN ensuring a value of $q_2 \leq q$ against the mixed strategy $\mathbf{p}^{(2)}$ of MAX. Once again, such a pure strategy must exist since q is the value of the game. Define $S_2 = \{i \in [m] - S_1 \mid M_{i,j_2} = 0\}$. As before, it holds that $\sum_{i \in S_2} \mathbf{p}_i^{(2)} M_{i,j_2} + \sum_{i \in [m]-S_1-S_2} \mathbf{p}_i^{(2)} M_{i,j_2} = q_2$, hence $\sum_{i \in [m]-S_1-S_2} \mathbf{p}_i^{(2)} = q_2$.

Continuing in this manner for 2^b steps, we obtain 2^b pairwise disjoint subsets $S_1, \ldots, S_{2^b}$ of $[m]$ with corresponding columns $j_1, \ldots, j_{2^b}$ such that $M_{i,j_k} = 0$ for every $1 \leq k \leq 2^b$ and $i \in S_k$. For convenience we index the words in $\{0,1\}^b$ by $w_1, \ldots, w_{2^b}$ and fix

$$f(i) = w_k \text{ for every } 1 \leq k < 2^b \text{ and } i \in S_k \ , \ f(i) = w_{2^b} \text{ for } i \notin \cup_{k<2^b} S_k;$$

$$g(w_k) = j_k \text{ for every } 1 \leq k \leq 2^b \ .$$

The above construction guarantees that when MAX plays according to $\mathbf{p}$, and MIN follows f and g, the payoff is 1 with probability at most $q_1 q_2 \cdots q_{2^b} \leq q^{2^b}$. It follows that $u_{\mathbf{p}}(M,b) = q_1 \cdots q_{2^b} \leq q^{2^b}$ as required. □

As a corollary of Theorem 3.3, we get the following.

Corollary 3.4. *Let M be an m by n matrix with $\{0,1\}$ entries, and let q be the value of the game defined by M, as above. If $q^{2^b} < 1/m$, then for every distribution vector $\mathbf{p}$ of the MAX player, $u_{\mathbf{p}}(M,b) = 0$. Therefore, for every fixed q satisfying $0 < q < 1$, $b = \log\log m + O_q(1)$ suffices to ensure value 0 for the MIN player.*

Proof. Let $\mathbf{p}$ be the uniform distribution on $[m]$. Theorem 3.3 guarantees the existence of $f : [m] \to \{0,1\}^b$ and $g : \{0,1\}^b \to [n]$ such that if MAX plays according to $\mathbf{p}$ and MIN follows f and g, then the expected payoff is at most $q^{2^b} < 1/m$. We argue that if MIN follows f and g, then in fact, the expected payoff is 0. Indeed, since $p_i = 1/m$ for every $i \in [m]$, a positive expected payoff is possible only if it is at least $1/m$, which derives a contradiction. It follows that for every $w \in \{0,1\}^b$ and for every $i \in [m]$ such that $f(i) = w$, we must have $M_{i,g(w)} = 0$. But this means that the same f and g guarantees that the expected payoff is 0 regardless[2] of the mixed strategy of MAX. □

[2] Note that as here the same f and g work against any mixed strategy of MAX, this works in the weak model as well, providing a proof of Theorem 4.5.

Remark: The corollary is essentially the known simple fact (proved in [6], [7]) that the ratio between the fractional cover and the integer cover of a hypergraph with m edges is at most $\ln m$.

The following theorem shows that both Theorem 3.3 and Corollary 3.4 are essentially tight.

Theorem 3.5. *For every real $0 < q < 1$, for every integer $b \geq 0$ and for every large polynomially related m and n satisfying $q^{2^b} m > 2^b \log n$, there exists an m by n $\{0,1\}$-matrix M that satisfies*
(i) $v(M, 0) = q \pm o(1)$, where the $o(1)$-term tends to 0 as m and n grow; and
(ii) if $\mathbf{p} = (p_1, p_2, \ldots, p_m)$ is the uniform distribution on the rows, then $u_{\mathbf{p}}(M, b) \geq (1 - o(1))q^{2^b}$ (and thus $u_{\mathbf{p}}(M, b) = (1 \pm o(1))q^{2^b}$, by Theorem 3.3).
In particular, for, say, $m = n^2$ and $b \leq \log\log m - \Theta(1)$, $u_{\mathbf{p}}(M, b) > 0$.

Proof. Let M be a random m by n matrix with $\{0,1\}$-entries obtained by choosing each entry $M_{i,j}$, randomly and independently, to be 1 with probability q and 0 with probability $1 - q$. We show that M satisfies the assertion of the theorem with positive probability.

Since m, n are large and are polynomially related, almost surely (that is, with probability that tends to 1 as m, n tend to infinity) every row of M has $(1 \pm o(1))qn$ 1-entries, and every column of M has $(1 \pm o(1))qm$ 1 entries. This follows easily by the standard known estimates for Binomial distributions, see, for example, [3]. This implies that the value of the game is $(1 \pm o(1))q$: indeed, if MAX (respectively, MIN) plays according to the uniform distribution on the rows (resp., columns), then it guarantees an expected payoff of at least (resp., at most) $(1 \pm o(1))q$. Thus (i) holds almost surely.

We establish the assertion by showing that (ii) holds with high probability as well. For that purpose, we argue that for every choice of a set $J \subset [n]$ of size $|J| = 2^b$, the number of indices $i \in [m]$ so that $M_{i,j} = 1$ for all $j \in J$ is, almost surely, $(1 \pm o(1))q^{2^b} m$. Indeed, for a fixed choice of a set J, the random variable X that counts the number of such indices i is a Binomial random variable with parameters m and q^{2^b}. Therefore the probability that X is not $(1 \pm o(1))q^{2^b} m$ decreases exponentially with $q^{2^b} m > 2^b \log n = \log\left(n^{2^b}\right)$. The assertion is established by applying the union bound over all $\binom{n}{2^b} < n^{2^b}$ choices of the set J. □

Remark: The proof of existence of M as in the last theorem is probabilistic. In Appendix A we give similar explicit examples using either finite geometries, or character sum estimates (Weil's Theorem [14]), or any example of small sample spaces supporting nearly 2^b-wise independent random variables. See [9] or [1] for some such examples.

4 Adversarial leakage in the Weak Model

The following result deals with the Weak Model. It shows that in sharp contrast to the situation with the Strong Model, here there are examples in which $\log\log m - O(1)$ bits of information do not enable the MIN player to gain any significant advantage.

Theorem 4.1. *For every real q, $0 < q < 1$, for every positive δ and for all large polynomially related n, m satisfying*

$$[(\frac{10}{\delta})]^{n2^b} < \delta^{m/\sqrt{n}} \quad \textit{and} \quad [\frac{q(1-q)}{10}]^{2^b} \geq \frac{1}{\sqrt{n}},$$

there is an m by n matrix M with $\{0,1\}$-entries so that the value of the game it determines $v(M,0)$ is $q+o(1)$, and $v^{weak}(M,b) \geq q-\delta$.

In particular, if $m=n^2$ and $b=\log\log m - \Theta(1)$, $v^{weak}(M,b)$ is essentially equal to $v(M,0)$.

The proof of the above theorem is more complicated than the ones in the previous section, and requires several preparations. We need the following known result.

Lemma 4.2 ([2], Lemma 3.2)**.** *Let Y be a random variable with expectation $E[Y]=0$, variance $E[Y^2]$ and fourth moment $E[Y^4] \leq k(E[Y^2])^2$, where k is a positive real. Then $Prob[Y \geq 0] \geq \frac{1}{2^{4/3}k}$.*

Using the above lemma, we prove the following.

Lemma 4.3. *Let q be a real, $0<q<1$, and let $\mathbf{p}=(p_1,p_2,\ldots,p_n)$ be a distribution vector on $[n]$, that is, $p_j \geq 0$ for all j and $\sum_j p_j = 1$. Let $X_1,X_2,\ldots,X_n$ be independent, identically distributed indicator random variables, where each X_j is 1 with probability q (and 0 with probability $1-q$.) Define $X=\sum_{j=1}^n X_j p_j$. Then the probability that X is at least its expectation (which is q) is bigger than $\frac{q(1-q)}{10}$.*

Proof. Define $Y_j = X_j p_j - E[X_j p_j] = X_j p_j - q p_j$, and $Y=\sum_j Y_j$. By linearity of expectation $Y = X - E[X]$, and $E[Y]=0$. In order to apply the previous lemma, we compute the variance of Y, and estimate its forth moment.

$$Var[Y] = \sum_j Var[Y_j] = \sum_j [q(1-q)^2 p_j^2 + (1-q)q^2 p_j^2]$$

$$= q(1-q)\sum_j p_j^2.$$

Similarly

$$E[Y^4] = \sum_j E[Y_j^4] + 6\sum_{i<j} E[Y_i^2]E[Y_j^2]$$

$$= \sum_j [q(1-q)^4 p_j^4 + (1-q)q^4 p_j^4] + 6\sum_{i<j} q^2(1-q)^2 p_i^2 p_j^2$$

$$\leq q(1-q)\sum_j p_j^4 + 6\sum_{i<j} q^2(1-q)^2 p_i^2 p_j^2 \leq \frac{1}{q(1-q)}[q^2(1-q)^2\sum_j p_j^4 + 6\sum_{i<j} q^2(1-q)^2 p_i^2 p_j^2]$$

$$\leq \frac{3}{q(1-q)}(Var[Y])^2.$$

The desired result now follows from Lemma 4.2, (using the fact that $2^{4/3}\cdot 3 < 10$). □

Remark: For $q \leq 1/2$ the estimate in the lemma is tight, up to a constant factor. Indeed, for $\mathbf{p}=(1,0,0,\ldots,0)$ the probability that X is at least q is precisely the probability that $X_1=1$, which is q. For $q=1/k$ with k being an integer there is a simpler argument showing that in this case the probability that X is at least its expectation is at least q (which is precisely tight). The idea is to choose the random vector $(X_1,X_2,\ldots,X_n)$ by first choosing, for each $1 \leq j \leq n$, a random uniform number n_j in $\{1,2,,\ldots,k\}$ with all choices being independent, and then by selecting a uniform random $Z \in \{1,2,\ldots,k\}$, defining X_j to be 1 iff $n_j = Z$. Since the sum $\sum_{Z\in[k]}(\sum_{j:n_j=Z} p_j)=1$, it follows that for each choice of the values n_j, there is at least one Z so that $\sum_{j:n_j=Z} p_j \geq 1/k$, and therefore the probability that the obtained random sum is at least $q = 1/k$ is at least $1/k$, as claimed. Note that for some values of q the probability that X is at least q is strictly smaller than q. Indeed, for example, if $q = 0.501$ and the vector $\mathbf{p}$ is $(0.5, 0.5, 0, 0, \ldots, 0)$, then the probability that X is at least q is the probability that $X_1 = X_2 = 1$, which is q^2, that is, roughly $q/2$.

Corollary 4.4. *Let v be a random vector of length n with $\{0,1\}$ entries obtained by selecting each entry, randomly and independently, to be 1 with probability q, and 0 with probability $1-q$. Let $\mathbf{P}$ be any fixed set of distribution vectors of length n. Then, the probability that the inner product of the vector v with each of the vectors $\mathbf{p} \in \mathbf{P}$ is at least q, is at least $(\frac{q(1-q)}{10})^{|\mathbf{P}|}$.*

Proof. By the FKG Inequality (c.f., e.g., [3], Chapter 6.), the probability that the inner product of v with each of the vectors $\mathbf{p} \in \mathbf{P}$ is at least q is greater or equal to the product of these probabilities. But according to Lemma 4.3, for every $\mathbf{p} \in \mathbf{P}$, the probability that the inner product of v with $\mathbf{p}$ is at least q is greater than $\frac{q(1-q)}{10}$. The desired result follows. □

We are now ready to state the proof of Theorem 4.1:

Proof. (sketch) Take a δ-net $\mathcal{N}$ of distributions of length n with respect to the ℓ_1-norm. Apply the last lemma to every set $\mathbf{P}$ of 2^b of them to conclude, using the union bound, that almost surely for every such set there is a pure strategy of the MAX player that ensures her value at least q with respect to each of these mixed strategies. Here we use the fact that $|\mathcal{N}| \leq [(\frac{10}{\delta})]^n$, and hence there are at most $[(\frac{10}{\delta})]^{n2^b}$ ways to choose a set of 2^b members of $\mathcal{N}$. The desired result follows. The missing details are deferred to the full version. □

Somewhat surprisingly, even in the Weak Model, although there are examples in which the MIN player cannot decrease the value by much using at most $\log\log m - O(1)$ bits of information, if he is allowed to use $\log\log m + O(1)$ bits, he can always decrease the value to 0. This is described in the next (simple) result, which, together with the previous theorem, exhibits an unexpected sharp phase transition at $b = \log\log m$.

Theorem 4.5. *Let M be an m by n matrix with $\{0,1\}$ entries, and let q be the value of the game defined by M. If $q^{2^b} < 1/m$, then $v^{weak}(M,b) = 0$. Therefore, for every fixed q satisfying $0 < q < 1$, $b = \log\log m + O_q(1)$ suffice to ensure value 0 for the MIN player, even in the Weak Model.*

The proof of Theorem 4.5 follows from the proof of Corollary 3.4.

5 Optimal strategy computation

We begin with a simple example of a $\{0,1\}$ matrix M with a maximin strategy $\mathbf{p}_0^*$ that satisfies (i) $u_{\mathbf{p}_0^*}(M,0) = 1/2$; and (ii) $u_{\mathbf{p}_0^*}(M,1) = 0$. On the other hand, there exists another mixed strategy $\mathbf{p}$ of MAX that satisfies (i) $u_{\mathbf{p}}(M,0) = 3/7$; and (ii) $u_{\mathbf{p}_1^*}(M,1) \geq 1/7$. This shows that playing the maximin strategy may be a naive behavior for $b > 0$, and hence motivates the computation of better strategies. The matrix M showing the above is of dimension 9×14 and it is depicted in Table 1. The main ingredient in the construction is a 7 by 7 matrix T_7 with $\{0,1\}$ entries that satisfies the following properties: (1) every row and every column of T_7 contains exactly 3 1-entries; and (2) for every choice of $1 \leq j \leq j' \leq 7$, there exists some $1 \leq i \leq 7$ such that $M_{i,j} = M_{i,j'} = 1$. (Refer to

$1\cdots1$	$0\cdots0$
$0\cdots0$	$1\cdots1$
T_7	T_7

Table 1: $M \in \{0,1\}^{9\times 14}$ of value 1/2, satisfying: (i) $u_{\mathbf{p}_0^*}(M,1) = 0$; and (ii) $u_{\mathbf{p}_1^*}(M,1) \geq 1/7$.

Example 1 in Appendix A.) Playing the first two rows with probability 1/2 each yields an expected payoff of 1/2 for MAX. One can easily verify that this is a unique optimal strategy when $b = 0$, while its expected payoff is clearly 0 when $b = 1$. Yet, by playing the uniform distribution on the bottom 7 rows, MAX ensures an expected payoff of 3/7 when $b = 0$ and an expected payoff of at least 1/7 when $b = 1$.

We now turn to consider the computational complexity of finding the optimal strategy for the MAX player in the Strong Model. The following theorem shows that computing the optimal strategy against b bits is poly-time for any fixed b.

Theorem 5.1. *Given an m by n matrix M with $\{0,1\}$ entries and a fixed $b \geq 0$, computing the optimal strategy against b leaking bits ($\mathbf{p}_b^*$) under the Strong Model is poly-time.*

Proof. An optimal strategy $\mathbf{p}_b^* = (p_1, \ldots, p_m)$ can be computed by solving the following linear program:

$$\begin{aligned}
&\text{maximize } z \text{ s.t.}\\
&\sum_{w\in\{0,1\}^b} \sum_{i: f(i)=w} p_i M_{i,g(w)} \geq z \quad \forall f : [m] \to \{0,1\}^b, \forall g : \{0,1\}^b \to [n]\\
&\sum_{i\in[m]} p_i = 1\\
&p_i \geq 0 \quad \forall i \in [m] \ .
\end{aligned}$$

Since there are 2^{bm} possible functions f and n^{2^b} possible functions g, this linear program admits a polynomial number of variables, but an exponential number of constraints. However, a closer analysis shows that it can be rewritten with a polynomial number of constraints.

The composition of f and g is essentially a mapping $h : [m] \to [n]$ with image of cardinality at most 2^b. Fixing some subset $J \subseteq [n]$, $|J| \leq 2^b$, it is easy to compute a mapping h_J that minimizes the expected payoff of MAX over all mappings h with image J: h_J simply maps each row $i \in [m]$ to a column $j \in J$ that minimizes $M_{i,j}$. Therefore an optimal strategy $\mathbf{p}_b^*$ can be computed by solving the linear program

$$\begin{aligned}
&\text{maximize } z \text{ s.t.}\\
&\sum_{j\in J} \sum_{i: h_J(i)=j} p_i M_{i,j} \geq z \quad \forall J \subseteq [n], |J| \leq 2^b \qquad (2)\\
&\sum_{i\in[m]} p_i = 1\\
&p_i \geq 0 \quad \forall i \in [m] \ ,
\end{aligned}$$

whose size is polynomial as long as b is a constant. □

We next show that for general b, computing the optimal strategy against b bits is NP-hard. Moreover, we show that it is NP-hard to approximate the optimal value by any factor.

Theorem 5.2. *Given an m by n matrix M with $\{0,1\}$ entries, it is NP-hard to approximate $v(M,b)$ by any factor under the Strong Model.*

Proof. We show that given an m by n matrix M with $\{0,1\}$ entries and some $b \geq 0$, it is NP-hard to decide whether $v(M,b) > 0$. This is done by reduction from set cover (SC). An instance of SC is composed of a finite set of elements $U = \{1,\ldots,m\}$, a collection $\mathcal{C} = \{C_1,\ldots,C_r\}$ of subsets of U and an integer k. The question is whether there is a subcollection $\mathcal{C}' \subseteq \mathcal{C}$, $|\mathcal{C}'| \leq k$, such that every element in U belongs to at least one member of $\mathcal{C}'$.

Given an instance of SC, $\langle U, \mathcal{C}, k\rangle$, we construct the following instance of our problem. Let M be a binary matrix with $m = |U|$ rows and $n = r$ columns such that $M_{i,j} = 0 \Leftrightarrow i \in C_j$. Fix $b = \log k$. We show that there is a set cover of size at most k if and only if $v(M,b) = 0$.

Sufficiency: Suppose the size of the set cover is greater than k. Then, we show that taking the uniform distribution over the whole action set (i.e. setting $p_i = \frac{1}{m}\ \forall i \in [m]$) yields $v(M,b) > 0$.

Consider inequality 2 and let $\mathbf{p}$ be the uniform distribution as described above. For every choice of $J \subseteq [n]$, the left-hand side of the inequality is composed of a finite set of summands. In order to show that the obtained payoff is greater than zero, it is sufficient to show that at least one summand is greater than zero. Indeed, since the set cover is greater than $k = 2^b$, there must exist some row $i \in [m]$, call it i', such that $M_{i',j} = 1$ for every $j \in J$, and also $\mathbf{p}(i') > 0$ (since $\mathbf{p}$ has a full support). Consequently, $v(M,b) > 0$.

Necessity: Suppose there exists a set cover of size at most $k = 2^b$. Then, there is a set of columns S, $|S| \leq 2^b$, s.t. for every $i \in [m]$ there exists $j \in |S|$ for which $M_{i,j} = 0$. Let g be a function that maps every $w \in \{0,1\}^b$ to a different column in S (arbitrarily). By the choice of S, it must hold that for every $i \in [m]$, $M_{i,g(f_g(i))} = 0$. Therefore, for every distribution vector $\mathbf{p}$, every summand in inequality 2 equals zero. Consequently, $v(M,b) = 0$. □

In contrast to the last theorem, under the Weak Model, computing the optimal strategy for a given $f : [m] \to \{0,1\}^b$ is trivially poly-time: For every $w \in \{0,1\}^b$, let $S^w = \{i : f(i) = w\}$, and let M^w denote the sub-matrix $M^w \in \mathbb{R}^{|S^w| \times n}$, induced by S^w. MAX will choose w that maximizes $v(M^w, 0)$ and play the corresponding maximin strategy.

Finally, we consider the computational complexity of finding the optimal f function for the MIN player in the Weak Model. For a general b, the exact same reduction from Set Cover, presented in the proof of Theorem 5.2, shows that computing the optimal f function is NP-hard and that it is NP-hard to find an f function that approximates the optimal value (for the MIN player) within any factor. For a fixed b, however, the computation of an optimal (from the perspective of MIN) $f : [m] \to \{0,1\}^b$ becomes polynomial. Indeed, there are n^{2^b} possible functions $g : \{0,1\}^b \to [n]$. For each such function g the computation of an optimal f is straightforward.

References

[1] N. Alon, O. Goldreich, J. Håstad and R. Peralta, *Simple constructions of almost k-wise independent random variables*, Proc. 31^{st} IEEE FOCS, St. Louis, Missouri, IEEE (1990), pp. 544–553.

[2] N. Alon, G. Gutin and M. Krivelevich, Algorithms with large domination ratio, J. Algorithms 50 (2004), 118–131.

[3] N. Alon and J. H. Spencer, *The Probabilistic Method, Third Edition*, Wiley, 2008, xv+352 pp.

[4] R. J. Aumann, On the non-transferable utility value: a comment on the Roth-Shaper examples, *Econometrica*, 53(3):667–677 (1985).

[5] R. J. Aumann, M. Maschler, Some Thoughts on the Minimax Principle, *J. Management Science*, 18(5):54–63 (1972).

[6] D. S. Johnson, Approximation algorithms for combinatorial problems, J. Comput. System Sci. 9 (1974), 256–278.

[7] L. Lovász, On the ratio of optimal and fractional covers, Discrete Mathematics 13 (1975), 383–390.

[8] A. Matsui, Information leakage forces cooperation, Discussion paper, Northwestern University (1988).

[9] J. Naor and M. Naor, *Small-bias probability spaces: efficient constructions and applications*, Proc. 22^{nd} annual ACM STOC, ACM Press (1990), pp. 213–223.

[10] H. Nasheri, Economic Espionage and Industrial Spying (Cambridge Studies in Criminology), Cambridge University Press (2005).

[11] M. Tennenholtz, Competive safety analysis: robust decision-making in multi-agent systems, *Journal of Artificial Intelligence Research*, 17:363–378 (2002).

[12] M. Tennenholtz, Program Equilibrium, Games and Economic Behavior, vol. 49, no. 2, pages 363–373 (2004).

[13] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton (1944).

[14] A. Weil, Sur les courbes algebriques et les varietes qui s'en deduisent, Actualites Sci. et Ind. No. 1041 (1948), Herman, Paris.

APPENDIX

A Explicit Constructions

In this appendix we describe several explicit constructions of n by n $\{0,1\}$-matrices representing games with value q, such that if the MIN player has b bits and b is smaller than $\log \log n + O(1)$, then the MAX player can guarantee a payoff of at least roughly q^{2^b}. This shows (by explicit examples) that the statements of Theorem 3.3 and Corollary 3.4 are essentially tight.

Example 1:

Let p be a prime power and let r be a positive integer. Fix $n = p^r - 1$. Let $M = (M_{u,v})$ be the following n by n binary matrix whose rows and columns are indexed by the set N of all nonzero vectors of length r over $GF(p)$. For each such u, v, $M_{u,v} = 1$ if and only if the two vectors u and v are orthogonal over $GF(p)$ (namely, their inner product over $GF(p)$ is zero). Note that M is a symmetric matrix, and every row and every column of it contains exactly $p^{r-1} - 1$ 1-entries. Indeed, this is the number of non-zero solutions of a single linear equation in r variables over $GF(p)$. It is easy to check that the maximin strategy of the game determined by M is the uniform distribution over N, yielding a value of $q = \frac{p^{r-1}-1}{p^r-1}$. Note that for large $n = p^r - 1$ this is very close to $1/p$.

We claim that for every set $J \subseteq N$ of at most $\log_p n$ columns, there are at least $p^{r-|J|} - 1$ rows u so that $M_{u,v} = 1$ for every $v \in J$. Note that if $p^{r-|J|}$ is large, then this number is very close to $q^{|J|}n$, implying that by playing the uniform distribution on the rows of M, MAX can ensure a value close to $q^{|J|}$. Note also that if $b \leq \log r - O(1) = \log \log n - O_p(1)$, then 2^b is much smaller than r, and hence $p^{r-|J|}$ is large provided $|J| \leq 2^b$. Fix a subset $J \subseteq N$ of cardinality at most $\log_p n$. By definition, row u satisfies $M_{u,v} = 1$ for every $v \in J$ if and only if the inner product of u and v over $GF(p)$ is zero for every $v \in J$. This is a homogeneous system of $|J|$ linear equations in the r variables representing the coordinates of u. This system clearly admits at least $p^{r-|J|} - 1$ non-trivial solutions; each such non-trivial solution corresponds to a row with the desired properties, proving the claim.

This completes the description of the first set of examples. Note that it works for every value q which is about $1/p$, where p is a prime power.

Example 2 (sketch):

Let p be a prime and let M be a $p \times p$ binary matrix, where $M_{i,j} = 1$ if and only if $i - j$ is a quadratic residue modulo p (where here zero is considered a quadratic residue). The value of the game represented by M is $(p+1)/(2p)$ which, for large p, is roughly $1/2$. Using Weil's Theorem, it is not difficult to show that for every subset S of Z_p of size at most $(0.5 - \delta) \log p$, the number of rows i such that $M_{i,j} = 1$ for all $j \in S$, is $(1 + o(1)) \frac{p}{2^{|S|}}$. A similar example holds for characters of other orders instead of the quadratic character, providing examples with values close to $1/d$ for any desired positive integer $d > 1$ (where here we have to choose a prime p so that d divides $p - 1$- by Dirichlet's Theorem on primes in arithmetic progressions it is known that there are infinitely many such primes for any such d).

More generally, one can use any construction of small sample spaces supporting nearly 2^b-wise independent binary random variables to supply additional examples. We omit the details, which will appear in the full version of the paper.