Fault Tolerance in Large Games

[Extended Abstract]

Ronen Gradwohl
Department of Computer Science and Applied Mathematics
The Weizmann Institute of Science
Rehovot, 76100 Israel
ronen.gradwohl@weizmann.ac.il

Omer Reingold*
Department of Computer Science and Applied Mathematics
The Weizmann Institute of Science
Rehovot, 76100 Israel
omer.reingold@weizmann.ac.il

ABSTRACT

A Nash equilibrium is an optimal strategy for each player under the assumption that others play according to their respective Nash strategies. In the presence of irrational players or coalitions of colluding players, however, it provides no guarantees. Some recent literature has focused on measuring the potential damage caused by the presence of faulty behavior, as well as designing mechanisms that are resilient against such faults. In this paper we show that large games are naturally fault tolerant. We first quantify the ways in which two subclasses of large games – λ-continuous games and anonymous games – are resilient against Byzantine faults (i.e. irrational behavior), coalitions, and asynchronous play. We then show that general large games also have some non-trivial resilience against faults.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Fault Tolerance;
J.4 [**Social and Behavioral Sciences**]: Economics

General Terms

Economics, Theory

Keywords

Nash equilibrium, Byzantine faults, Large games

1. INTRODUCTION

Game theory provides a means of studying an interaction between many agents by modelling it as a game. A solution concept such as a Nash equilibrium is then hypothesized to explain or predict the behavior of these agents. Nowadays, with the advent of the Internet, many of the interactions we wish to model are taking place online, in a distributed setting. Such interactions are characterized by unpredictably asynchronous communication, the occasional presence of faults, and perhaps some exogenous communication between the players. However, the concept of a Nash equilibrium provides no guarantees about the optimality of strategies in such a fault-ridden environment, and hence does not suffice.

One approach to correcting this deficiency is to incorporate the characteristics of the distributed setting into the model. This, however, requires precise knowledge of the setting to be modelled, and may make the model much more complicated. A second approach that is taken in some recent literature is to quantify the damage caused by the presence of faulty players (see the section Previous Work below).

In this paper we take a different approach, and show that sometimes the concept of a Nash equilibrium does suffice. That is, sometimes the strategies of a Nash equilibrium remain close to best-responses for each player even if the game is modified to permit asynchronous communication, faulty behavior, and the possibility of collusion.

We study large games – games that involve many agents – because in such games it is particularly relevant to examine the robustness of equilibria against faults. Additionally, in large games it seems plausible that the sheer size of the game renders it robust, whereas in small games such an approach is easily seen to fail. Of course, a large game could be comprised of a small game embedded in a setting with many players, in which case our approach would not work. But then again, such a game is not really a "large" game. Thus, we consider various properties of the game and the equilibrium that we believe are reasonable characterizations of large games, and show how these properties relate to the fault tolerance of the game.

Faulty Behavior.

The guarantee provided by a Nash equilibrium is that each player's strategy is optimal, assuming all others play their designated strategies. One type of faulty behavior consists of Byzantine faults, wherein some of the players do not play according to their equilibrium strategies. Perhaps the actions of these faulty players are altered due to an error in communication, perhaps the players are irrational, or perhaps they have some unknown utility. In any case, their actions can be arbitrary or even adversarial.

Two types of resilience we may wish to have against Byzan-

*This research was supported by grant 1300/05 from the Israel Science Foundation.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
EC'08, July 8–12, 2008, Chicago, Illinois, USA.
Copyright 2008 ACM 978-1-60558-169-9/08/07 ...$5.00.

tine faults are called immunity and tolerance. Immunity means that even if some players fault, the utility of the non-faulty players is not affected by much. Tolerance means that even if some players fault, the original strategies of non-faulty players are still optimal (even conditioned on the actions of the faulty players), although their payoffs may be different from the case in which no players fault.

Another type of faulty behavior is asynchrony. The Nash strategies are optimal if all players play simultaneously, but suppose some player was delayed, and was then exposed to the chosen actions of some of the other players. This player may then use the information he obtained from the exposed actions to get a better payoff for himself. In this case, his original Nash strategy may no longer be optimal. Resilience to asynchrony means that this does not occur – that a player can **not** improve his payoff by too much, even after he is exposed to the actions of some other players.

A final type of faulty behavior consists of collusion among players. Nash equilibria only guarantee that no player can benefit by a unilateral deviation from his strategy, but says nothing about deviations by a colluding set of players. Resilience against collusion means that coalitions of players can not improve their payoffs even by a coordinated deviation.

Our Results.

Our results are roughly of the following form: Given a large game of type X and a Nash equilibrium of that game, the strategies given by the equilibrium remain nearly best-response (up to, say, a small constant additive factor) even in the presence of at most x faults, asynchronous exposure to at most y other players' actions, and with the possibility of forming a coalition of size at most z.

We begin our study of fault tolerance in large games with a subclass of games that has received much attention in the economics literature (c.f. [16]) and has recently appeared in some computer science literature as well ([9, 10]). This subclass consists of λ-continuous games, and it turns out that it is particularly well-suited to our demands of fault tolerance. Roughly, in such games no player is influenced too much by the other players. More precisely, if some δ-fraction of players changes their actions, then this can only change the utility of a non-member by at most $\lambda \cdot \delta$. For games that are λ-continuous, we observe that x, y, and z can be as large as a constant fraction of all the players (if λ is some arbitrarily large constant).

λ-continuous games have very strong fault tolerance, but perhaps the restriction on the game is too strong. More precisely, in these games a player's utility given **any** of the others' actions changes little if a small fraction of players changes their action. For fault tolerance, however, we are really only interested in what happens when the actions of (non-faulty) players come from specific distributions – namely, the strategies of the Nash equilibrium. Thus, we consider weakening the restriction on the game, but then requiring some condition on the distribution of the equilibrium strategies.

The next natural subclass of games we study is that of anonymous games, in which the utility function of each player is a function of his own action and the distribution of other players' actions (how many players played each possible action). Thus, in anonymous games a player does not care which player performed each action, but only how many performed each action. Anonymous games are widely studied in economics, and have also appeared in recent computer science literature (c.f. [10]). Note that anonymous games can be very far from being λ-continuous: For example, if a player's utility is determined by the majority of the other players' binary actions, then a change of even 1 player's action can flip the utility from 1 to 0.

Anonymous games are fault tolerant if we consider Nash equilibria that are "mixed enough" – this means that players play every action with at least some minimal probability. Section 6.1 contains some discussion about the applicability and naturalness of this assumption, as well examples demonstrating its necessity. In any case, in such games the equilibria are resilient to roughly $\sqrt{n}$ faults, as well as coalitions of the same size. Additionally, the equilibria can tolerate asynchronous exposure to the actions of nearly a constant fraction of the players. Furthermore, if the game is both λ-continuous and anonymous, then the Nash equilibria are resilient to asynchronous exposure to all $n-1$ other players.

Finally, in general large games, we give an example in which there can be no fault tolerance, even to one Byzantine fault (see Section 7). However, not all is lost. We show that if the faulting players are chosen randomly (but their actions remain adversarial), then the strategies provided by Nash equilibria remain nearly best-responses for most players (again assuming that the strategies are mixed enough). We give a reduction from the case of arbitrary games with randomly-chosen faulting players to anonymous games with arbitrarily-chosen faulting players. Using this technique we show that the number of faulting players can be as large as the number of faulting players in any anonymous game – roughly $\sqrt{n}$ for Byzantine faults and up to a constant fraction for asynchronous exposure.

We note that our results are optimal in the following sense: the restrictions placed on the games, the equilibria, and the number of faulting players are pretty much necessary. It is not possible to weaken them by much, and we describe some games that serve as counter-examples.

Previous Work.

The original motivation for this work arose from two recent papers of Kalai [16, 17], in which he shows that in large normal-form games that are anonymous and λ-continuous, the Nash equilibria are extensively robust. This means that even when the game is altered in such a way as to allow sequential play, the possibility of revision, and communication between the players, the original Nash equilibria of the game survive. In particular, the Nash equilibria of such games are ex post Nash: a player's strategy remains best-response even after he sees the chosen actions of all the other players.

In a previous work [12] we attempt to generalize the work of Kalai, and introduce the notion of partial exposure. We show that in **any** large game, players' Nash strategies remain nearly optimal even after they are exposed to the chosen actions of *some* of the other players.

These properties of ex post Nash and partial exposure can be seen as a form of resilience to asynchrony, since players who play later are exposed to the chosen actions of those who played earlier. The resilience holds since the properties guarantee that players are not too motivated to change their actions despite the new information. The current paper expands on these ideas, and places this resilience in the broader context of fault tolerance.

Fault tolerance is one of the major topics of study in dis-

tributed computing, and much work in this field has focused on designing multi-agent protocols that are resilient against various types of faults. Game theory, however, is relatively unconcerned with these issues. In recent survey articles, Halpern [13, 14] has argued that perhaps these issues should be studied more extensively in a game theoretic framework.

Notions of resilience against Byzantine faults have appeared in a few implementation frameworks. Eliaz [11] shows how to implement a mechanism in such a way that if some players fault, the other players' strategies are still best-response (this is the notion we call tolerance). Aiyer et al. [2] implement a protocol in an asynchronous setting that is resilient against Byzantine, altruistic, and rational behavior. Additionally, the burgeoning field of rational cryptography (see [23] for a survey) is concerned with the implementation of protocols that are secure against both rational and Byzantine players. Finally, some recent work of Abraham et al. [3, 4] on the implementation of mediators has considered solution concepts that take into account the possibility of irrational play by some of the players.

Questions on how to deal with asynchrony have also come up in game theory. Specifically, Monderer and Tennenholtz [20, 21] present a framework that incorporates game theoretic analysis in a distributed setting. A very different approach to handling asynchrony in two-player games is taken by Ilcinkas and Pelc [15].

Solution concepts in game theory that deal with deviations by coalitions have a longer history. The first work is perhaps that of Aumann [1] – in his solution concept, there can be no coalitions of players in which all players benefit from a deviation. Bernheim et al. [8] present the notion of a *coalition-proof Nash equilibrium*, in which there can be no deviations in which all players do better, but also no deviations (by a subset of the coalition) from the deviation. Moreno and Wooders [22] extend this work to allow correlated strategies. Finally, in their work on implementing mediators, Abraham et al. [3, 4] also consider solution concepts in which some players may coordinate a deviation. Their notion requires that no such deviation will benefit any player in the coalition. One of the problems with these notions is that in general, a game need not have such a coalition-proof equilibrium (as opposed to a Nash equilibrium, which always exists).

Finally, a couple of recent papers quantify the difference between a regular equilibrium and one that is resilient against some adversarial players. Specifically, Moscibroda et al. [19] study the so-called *price of malice* in a virus inoculation game, and Babaioff et al. [7] study a different notion with the same name in congestion games. Additionally, Blum et al. [6] show that in some games, the ratio between a social optimum and the result obtained when players minimize their regret is unaffected by the presence of Byzantine players.

2. DEFINITIONS

We use the standard definition of a normal-form game (c.f. [24]).

DEFINITION 2.1 (GAME). *A game G is described by a triple $G = (N, A, u)$ as follows:*

- *$N = \{1, \ldots, n\}$ is the set of players.*
- *A is the set of player actions.*
- *$u = (u_1, \ldots, u_n)$ is the vector of payoff functions. For every i, the payoff of player i is $u_i : A^n \mapsto [0, 1]$. We will also slightly abuse notation, and denote by $u_i(\mathbf{A}) = \mathrm{E}_{a \sim \mathbf{A}}[u_i(a)]$, where $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_n)$ is a random variable over A^n.*

We also use the following standard notation. For a subset $I \subseteq N$ and some vector $X = X_1 \times \ldots \times X_n$, let $X_I = \bigotimes_{i \in I} X_i$, and $X_{-I} = \bigotimes_{i \notin I} X_i$. Furthermore, $\mathbf{A}_{-i} : \mathbf{A}'_i = \mathbf{A}_1 \times \ldots \times \mathbf{A}_{i-1} \times \mathbf{A}'_i \times \mathbf{A}_{i+1} \times \ldots \times \mathbf{A}_n$.

The most common solution concept for such games is a Nash equilibrium.

DEFINITION 2.2 (NASH EQUILIBRIUM). *A product distribution $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ over A^n is an* α-Nash equilibrium *if for all i and all $\mathbf{A}'_i$,*

$$u_i(\mathbf{A}_{-i} : \mathbf{A}'_i) \leq u_i(\mathbf{A}) + \alpha.$$

It is a Nash equilibrium *if $\alpha = 0$.*

We now define some properties of functions that are needed for this work. In the following, let $f : A^n \mapsto [0, 1]$.

DEFINITION 2.3 (λ-CONTINUOUS). *f is* λ-continuous *if for all $x, y \in A^n$,*

$$|f(x) - f(y)| \leq \lambda \cdot \Delta(x, y),$$

where

$$\Delta(x, y) = \frac{|\{i : x_i \neq y_i\}|}{n}.$$

For a player i and an action $a \in A$, define $u_i^a : A^{n-1} \mapsto [0, 1]$ as

$$u_i^a(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = u_i(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n).$$

Then a game G is λ-continuous if for every i and every a, u_i^a is λ-continuous.

DEFINITION 2.4 (ANONYMOUS). *f is* anonymous *if $f(x) = f(y)$ whenever*

$$|\{i : x_i = a\}| = |\{i : y_i = a\}|$$

for all $a \in A$. This occurs if and only if

$$f(x_1, \ldots, x_n) = f(\sigma(x_1), \ldots, \sigma(x_n))$$

for all permutations $\sigma : N \mapsto N$ and all $(x_1, \ldots, x_n) \in A^n$.

A game G is anonymous if for every i and every $a \in A$, u_i^a is anonymous.

We now define two properties of distributions that are relevant to this work.

DEFINITION 2.5 (ε-UNIFORM). *A distribution $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_n)$ is* ε-uniform *if for all i and all $a \in A$, $\Pr[\mathbf{A}_i = a] \geq \varepsilon$.*

A weaker notion is that of weak ε-uniformity, in which only elements in the support occur with probability at least ε, and not every element.

DEFINITION 2.6 (WEAKLY ε-UNIFORM). *A distribution $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_n)$ is* weakly ε-uniform *if for all i and all $a \in \mathrm{supp}(\mathbf{A}_i)$, $\Pr[\mathbf{A}_i = a] \geq \varepsilon$.*

Suppose $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ are the strategies of a Nash equilibrium. If $\mathbf{A}$ is ε-uniform, then no player i has a pure strategy (unless $|A| = 1$). Weak ε-uniformity, on the other hand, only states that **if** a player mixes, then his strategy is very mixed (every element occurs with probability at least ε).

3. FAULTY BEHAVIOR

In this section we describe the various types of faulty behavior in which we are interested.

3.1 Byzantine Faults

When a Byzantine fault occurs, the player who faulted plays an arbitrary action. His action is not rational, and does not obey any known distribution. The only knowledge players have about such faults is that the number of players that fault is at most some t. There are two possible properties we may want: t-immunity and t-tolerance. Immunity means that even if some players fault, the utilities of the honest players are not affected by much. Tolerance means that even if some players fault, the original strategies of non-faulty players are still optimal, although their payoffs may be different from the case in which no players fault.

DEFINITION 3.1 (t-IMMUNITY). *In a game G, a Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is* t-immune *if for every set $S \subset N$, $|S| \leq t$, every $b_S \in A^{|S|}$, and every player $i \notin S$,*

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-S} : b_S).$$

The equilibrium is (ε, t)-immune *if*

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-S} : b_S) - \varepsilon.$$

DEFINITION 3.2 (t-TOLERANCE). *In a game G, a Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is* t-tolerant *if for every set $S \subset N$, $|S| \leq t$, every $b_S \in A^{|S|}$, every player $i \notin S$, and every strategy $\mathbf{A}'_i$,*

$$u_i(\mathbf{A}_{-S} : b_S) \geq u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, \mathbf{A}'_i).$$

The equilibrium is (ε, t)-tolerant *if*

$$u_i(\mathbf{A}_{-S} : b_S) \geq u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, \mathbf{A}'_i) - \varepsilon.$$

The notions of t-immunity and t-tolerance serve two purposes and are very distinct properties. However, the following definition provides a sufficient condition for both.

DEFINITION 3.3 (t-STABLE). *In a game G, a Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is* t-stable *if for every set $S \subset N$, $|S| \leq t$, every $b_S \in A^{|S|}$, every player $i \notin S$, and every action $a \in A$,*

$$\left|u_i(\mathbf{A}_{-i} : a) - u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, a)\right| = 0.$$

The equilibrium is (ε, t)-stable *if*

$$\left|u_i(\mathbf{A}_{-i} : a) - u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, a)\right| \leq \varepsilon.$$

The following lemma shows that stability is sufficient for both immunity and tolerance:

LEMMA 3.4. *Suppose that in a game G, a Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is (ε, t)-stable. Then $\mathbf{A}$ is (ε, t)-immune and $(2\varepsilon, t)$-tolerant.*

PROOF. Immunity follows almost directly. Now consider some alternate strategy $\mathbf{A}'_i$ for player i. Then for any $S \subseteq N$, $|S| \leq T$, and $b_S \in A^{|S|}$,

$$\begin{aligned}
&u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, \mathbf{A}'_i) \\
&\quad= \sum_{a\in A} u_i(\mathbf{A}_{-(S\cup\{i\})} : a, b_S) \Pr\left[\mathbf{A}'_i = a\right] \\
&\quad\leq \sum_{a\in A} \left(u_i(\mathbf{A}_{-i} : a) + \varepsilon\right) \cdot \Pr\left[\mathbf{A}'_i = a\right] \qquad (1) \\
&\quad= \left(\sum_{a\in A} u_i(\mathbf{A}_{-i} : a) \cdot \Pr\left[\mathbf{A}'_i = a\right]\right) + \varepsilon \\
&\quad= u_i(\mathbf{A}_{-i} : \mathbf{A}'_i) + \varepsilon \\
&\quad\leq u_i(\mathbf{A}) + \varepsilon \qquad (2) \\
&\quad= \left(\sum_{a\in A} u_i(\mathbf{A}_{-i} : a) \cdot \Pr\left[\mathbf{A}_i = a\right]\right) + \varepsilon \\
&\quad\leq \left(\sum_{a\in A} \left(u_i(\mathbf{A}_{-(S\cup\{i\})} : a, b_S) + \varepsilon\right) \cdot \Pr\left[\mathbf{A}_i = a\right]\right) + \varepsilon \qquad (3) \\
&\quad= u_i(\mathbf{A}_{-S} : b_S) + 2\varepsilon,
\end{aligned}$$

where (1) and (3) follow from (ε, t)-stability and (2) holds since $\mathbf{A}$ is a Nash equilibrium (and so $\mathbf{A}_i$ is a better strategy than $\mathbf{A}'_i$). Thus, $\mathbf{A}$ is $(2\varepsilon, t)$-tolerant. ■

3.2 Coalitions

Another type of fault occurs when sets of players collude in an attempt to increase their payoffs.

DEFINITION 3.5 (t-COALITIONAL NASH EQUILIBRIUM). *In a game G, we say that a Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is a* t-coalitional Nash equilibrium *if for every set $S \subset N$, $|S| \leq t$, every joint strategy $\mathbf{A}'_S$, and every player $i \in S$,*

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-S} : \mathbf{A}'_S).$$

The strategy is an (ε, t)-coalitional Nash equilibrium *if*

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-S} : \mathbf{A}'_S) - \varepsilon.$$

The notions of coalitional Nash equilibrium, immunity, and tolerance are different: For example, in a coalitional Nash equilibrium, the player i whose utility does not change is a member of the coalition S. In the case of immunity, on the other hand, i is **not** a member of S. However, the three notions are closely related, and in fact immunity and tolerance imply resilience against coalitions.

LEMMA 3.6. *Suppose that in a game G, a Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is (δ, t)-immune and (δ', t)-tolerant. Then $\mathbf{A}$ is a $(\delta + \delta', t + 1)$-coalitional Nash equilibrium.*

PROOF. Fix a player i, and consider some subset $S \subset N \setminus \{i\}$, $|S| \leq t$. By immunity, we know that for every $b_S \in A^{|S|}$,

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-S} : b_S) - \delta.$$

By tolerance we know that for every strategy $\mathbf{A}'_i$ of player i,

$$u_i(\mathbf{A}_{-S} : b_S) \geq u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, \mathbf{A}'_i) - \delta'.$$

Putting the two together implies that

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, \mathbf{A}'_i) - \delta - \delta'.$$

Since this holds for every $b_S \in A^{|S|}$, it also holds that

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-(S\cup\{i\})} : \mathbf{A}'_{(S\cup\{i\})}) - \delta - \delta'$$

for every joint strategy $\mathbf{A}'_{(S\cup\{i\})}$of the players in $S\cup\{i\}$. ■

3.3 Asynchrony

If players do not play simultaneously, then perhaps they are exposed to the chosen actions of other players. They may then be able to use this information to their advantage. Note that the difference between our notion of stability under asynchrony and that of tolerance is that in the former the actions of the faulting set S are sampled from the distribution $\mathbf{A}_S$ (and so the probability is over this sample as well), whereas in the latter the actions are chosen in a worst-case manner.

DEFINITION 3.7 ((δ,p,t)-EX POST). *In a game G, a Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is* (δ,p,t)-ex post *if for every set $S \subset N$, $|S| \leq t$, and every player $i \notin S$,*

$$\Pr_{a_S \sim \mathbf{A}_S}[\exists a \in A \text{ s.t. } u_i(\mathbf{A}_{-(S\cup\{i\})} : a_S, a) > u_i(\mathbf{A}_{-S} : a_S) + \delta] < p.$$

If the equilibrium of a game is (δ,p,t)-ex post, then the original strategy of player i is nearly optimal (up to the additive δ) with probability $1-p$ even after he sees the chosen actions of players in any set S of size at most t.

4. λ-CONTINUOUS GAMES

It turns out that λ-continuous games are strongly fault-tolerant, as is shown in the following theorem. Note that we make no assumption on $|A|$ (except that it is discrete).

THEOREM 4.1. *Let G be an n-player λ-continuous game, and fix some $t \in [n]$. Then every Nash equilibrium of G is $(\lambda t/n, t)$-immune, $(2\lambda t/n, t)$-tolerant, and $(3\lambda t/n, t+1)$-coalitional.*

PROOF. Fix some player i and subset $S \subset N\backslash\{i\}$, $|S| \leq t$. Also, fix some Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ and some $b_S \in A^{|S|}$. For every $a \in A$,

$$\begin{aligned} u_i^a(\mathbf{A}_{-i}) &= \sum_{\overline{a}=(a_1,\ldots,a_{i-1},a,a_{i+1},\ldots,a_n)} u_i(\overline{a}) \Pr[\mathbf{A}_{-i} = \overline{a}_{-i}] \\ &\geq \left(\sum_{\overline{a}} u_i(\overline{a}_{-S} : b_S) \Pr[\mathbf{A}_{-i} = \overline{a}_{-i}]\right) - \frac{\lambda t}{n} \\ &= u_i^a(\mathbf{A}_{-(S\cup\{i\})} : b_S) - \frac{\lambda t}{n}, \end{aligned}$$

where the inequality follows from λ-continuity. An identical argument shows that

$$u_i^a(\mathbf{A}_{-i}) \leq u_i^a(\mathbf{A}_{-(S\cup\{i\})} : b_S) + \frac{\lambda t}{n}.$$

Thus, we have that for every $a \in A$,

$$\left|u_i(\mathbf{A}_{-i}, a) - u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, a)\right| \leq \frac{\lambda t}{n},$$

and so $\mathbf{A}$ is $(\lambda t/n, t)$-stable. Immunity and tolerance follow using Lemma 3.4, and Lemma 3.6 implies the resilience against coalitions. ■

Since tolerance is a strictly stronger property than resilience against asynchrony, we also get the following immediate corollary:

COROLLARY 4.2. *Let G be an n-player λ-continuous game, and fix some $t \in [n]$. Then every Nash equilibrium of G is $(2\lambda t/n, 0, t)$-ex post.*

5. ANONYMOUS λ-CONTINUOUS GAMES

A game G is anonymous and λ-continuous if for every player i and every action $a \in A$, u_i^a is anonymous and λ-continuous. Such games have the same immunity, tolerance, and coalitional resilience properties as general λ-continuous games. Their advantage, however, lies in their nearly optimal resilience to asynchrony – no player has much incentive to change his strategy even if he sees the chosen actions of **all** other players. This is shown in the following theorem of Kalai [16].

THEOREM 5.1 ([16]). *Let G be an n-player game that is both λ-continuous and anonymous, and fix some $s \in [n]$. Then every Nash equilibrium of G is*

$$\left(\frac{\lambda s|A|}{n}, 2|A| \cdot e^{-\frac{2s^2}{n}}, n-1\right)\text{-ex post.}$$

6. ANONYMOUS GAMES

Recall that a game G is anonymous if for every player i and every action $a \in A$, u_i^a is an anonymous function. We first state our main result on anonymous games, and then provide some intuition and discussion. The proof follows in Section 6.2. We note that our theorems in this section and the one on general games are for games in which $A = \{0,1\}$, but they can be generalized to any constant $|A|$.

THEOREM 6.1. *Let G be an anonymous n-player game with $A = \{0,1\}$ and fix some $\varepsilon = \varepsilon(n) > 0$. Let $c(n)$ be any complexity function (positive and nondecreasing), and fix*

$$t = \frac{\sqrt{\varepsilon n}}{c(n)}.$$

Then there exists a universal constant D such that every ε-uniform Nash equilibrium of G is (δ,t)-immune, $(2\delta,t)$-tolerant, and $(3\delta,t+1)$-coalitional, where

$$\delta = D \cdot \sqrt{\frac{\log c(n)}{c(n)}}.$$

Consider the following interesting ranges of possible parameters. If we are happy with δ as a small constant, then $c(n)$ can be a large constant. In this case, the number of faulting players can be as large as $\Omega(\sqrt{\varepsilon n})$. On the other hand, we can have δ almost as small as $O(1/n^{1/4})$ at the cost of having t only be some (possibly large) constant.

Theorem 6.1 relies on the following lemma, whose proof is omitted due to space constraints. The idea is that when considering independent Bernoulli random variables, the probability that their sum is x is close to the probability that their sum is $x+t$ (when x is close to the expectation and t is smaller than the standard deviation).

LEMMA 6.2. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent Bernoulli random variables with (possibly different) parameters $p_i \in [\varepsilon, 1-\varepsilon]$, where $\varepsilon > 0$, let $\mathbf{X}$ be their sum, and let $\sigma^2 = \mathrm{Var}[\mathbf{X}]$. Then for any $\delta > 0$ and positive integer t:*

$$|\Pr[\mathbf{X} = x] - \Pr[\mathbf{X} = x+t]| \leq \delta$$

with probability at least

$$1 - O\left(\frac{t}{\delta\sigma^2} + \frac{t}{\sigma}\right)$$

over the choice of x from $\mathbf{X}$.

Since we are studying anonymous games, all utility functions are anonymous, and so their values are determined by the sum of the inputs. Now, if there are t Byzantine faults in such a game, then the sum is offset by at most t. Intuitively, however, since the probability of the sum being x and $x + t$ is roughly the same, these faults should not alter the expectation of the function by too much.

6.1 ε-Uniformity

In this section we discuss the notion of ε-uniformity. We begin with some examples of anonymous games with Nash equilibria that are **not** ε-uniform, and show that no fault tolerance is possible:

EXAMPLE 6.3. In the agreement game, every player has utility 1 if all players output 1, and 0 otherwise. One Nash equilibrium here is when every player outputs 1 with probability 1.

Note that this equilibrium is not 1-tolerant: If one player faults and outputs 0 instead of 1, then all players get 0 instead of 1.

EXAMPLE 6.4. In the minority game, players output 0 or 1, and win if and only if their action matches the minority of other players' actions. If n is odd, then a Nash equilibrium of this game is when one player flips a fair coin, half of the others always output 1, and the remaining players always output 0.

This equilibrium is not 2-immune: If two players that always output 0 fault and output 1, then it is beneficial for a player who is supposed to always output 1 to output 0. In this case his Nash strategy is very far from optimal.

Our conclusion from these examples is that we can not expect fault tolerance in general anonymous games without some restriction on the Nash equilibria. We now describe some situations in which it is reasonable to impose the assumption of ε-uniformity.

First, consider a situation in which a game is played over a noisy channel. Suppose players transmit their actions, but know that there is some small probability ε that their action will chosen at random instead of correctly communicated. Here all strategies of players are ε-uniform (for binary action spaces). It is also not difficult to show that there always exists a Nash equilibrium in this situation: conditioned on all players playing an ε-uniform strategy, there exists an ε-uniform best-response for all players. Consider a new game in which the set of actions of all players is the set of all ε-uniform strategies of the original game. Since this new action space is compact, there exists a Nash equilibrium. Note that an equilibrium in the new game consists of ε-uniform strategies over the original actions.

Second, adding some random noise can be a very easy way to make a mechanism fault tolerant. Suppose we implement a mechanism in Nash strategies. Then we can add random noise into the mechanism by flipping each player's actions with probability ε. In this case, the equilibrium will be ε-uniform and thus fault tolerant. Note that for some ranges of parameters ε can be as small as $O(1/n)$. In this case we are making the mechanism fault tolerant at the cost of only a constant number of random errors (on expectation).

Third, consider a situation in which faults of players who do not randomize their actions can be "fixed". That is, if a player has a pure strategy in a Nash equilibrium, then all other players know this. But then, if this player faults, then the other players know he faulted, and so they can "correct" his fault by treating his input as the original pure action. In such a situation, the equilibrium only needs to be weakly ε-uniform for the results to hold. Weak ε-uniformity arises naturally in two distinct contexts. First, suppose players have some bounded randomness (see [18]). If they have access to k random bits, then the Nash distribution is weakly 2^{-k}-uniform. Additionally, in the recent work of Daskalakis and Papadimitriou [10] on computing equilibria in anonymous games, they actually compute an α-Nash equilibrium that is weakly α^2-uniform.

Fourth, the notion of ε-uniformity is related to Selten's trembling-hand perfect equilibrium solution concept [27], in which a player's Nash strategy must remain nearly optimal even after small perturbations in the other players' strategies. Such an equilibrium always exists, but the size of the perturbation can be vanishingly small.

Finally, we note that for anonymous games, the requirement of ε-uniformity can be relaxed. All we require is that for every action $a \in A$, there are enough players who output a with probability between ε and $1 - \varepsilon$.

6.2 Proofs

In the following lemma, we denote by $f(j) = f(x)$, where $|\{i : x_i = 1\}| = j$. This is well defined because f is an anonymous function, and so its value depends only on the number of ones and zeros in the input.

LEMMA 6.5. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent Bernoulli random variables with (possibly different) parameters $p_i \in [\varepsilon, 1-\varepsilon]$, where $\varepsilon > 0$. Let $c(n)$ be any complexity function, and let $t = \sigma/c(n)$, where*

$$\mathbf{X} = \sum_{i=1}^{n-t} \mathbf{X}_i \qquad \text{and} \qquad \sigma^2 = \mathrm{Var}\left[\sum_{i=1}^{n} \mathbf{X}_i\right].$$

Finally, let $f : \{0,1\}^n \mapsto [0,1]$ be any anonymous function. Then for any $b, c \in \{0, \ldots, t\}$,

$$\left|\, \mathrm{E}\left[f(\mathbf{X} + b)\right] - \mathrm{E}\left[f(\mathbf{X} + c)\right] \,\right| \leq O\left(\sqrt{\frac{\log c(n)}{c(n)}}\right).$$

PROOF. Fix

$$\delta = \frac{1}{\sigma\sqrt{c(n)}},$$

and note that

$$\frac{t}{\delta} = \frac{\sigma^2}{\sqrt{c(n)}}.$$

Lemma 6.2 guarantees that

$$\left|\Pr\left[\mathbf{X} = x + b\right] - \Pr\left[\mathbf{X} = x + c\right]\right| \leq O\left(\frac{1}{\sigma\sqrt{c(n)}}\right)$$

with probability at least

$$1 - O\left(\frac{t}{\delta\sigma^2} + \frac{t}{\sigma}\right) \geq 1 - O\left(\frac{1}{\sqrt{c(n)}}\right)$$

over the choice of x from $\mathbf{X}$. Let $\mu = \mathrm{E}[\mathbf{X}]$ and $s = \sqrt{\log c(n)} \cdot \sigma$. Then there exists some constant $d > 0$ so that by a Chernoff bound,

$$\begin{aligned}\Pr\left[|\mathbf{X} - \mu| > d\sqrt{\log c(n)} \cdot \sigma\right] &\leq O\left(e^{-\log c(n)}\right) \\ &\leq O\left(\frac{1}{\sqrt{c(n)}}\right).\end{aligned}$$

For some k, let $\mu_k^- = \mu - ds - k$ and $\mu_k^+ = \mu + ds - k$.

$$\begin{aligned}&\left|\mathrm{E}\left[f(\mathbf{X}+b)\right] - \mathrm{E}\left[f(\mathbf{X}+c)\right]\right| \\ &= \left|\sum_{i=0}^{n-t} f(i+b)\Pr\left[\mathbf{X}=i\right] - \sum_{i=0}^{n-t} f(i+c)\Pr\left[\mathbf{X}=i\right]\right| \\ &\leq \left|\sum_{i\in[\mu_b^-,\mu_b^+]} f(i+b)\Pr\left[\mathbf{X}=i\right] - \sum_{i\in[\mu_c^-,\mu_c^+]} f(i+c)\Pr\left[\mathbf{X}=i\right]\right| \\ &\quad + \left|\Pr\left[\mathbf{X}\notin[\mu_b^-,\mu_b^+]\right] + \Pr\left[\mathbf{X}\notin[\mu_c^-,\mu_c^+]\right]\right| \cdot \max_x f(x) \\ &\leq \left|\sum_{i\in[\mu_0^-,\mu_0^+]} f(i)\Pr\left[\mathbf{X}=i+b\right] - \sum_{i\in[\mu_0^-,\mu_0^+]} f(i)\Pr\left[\mathbf{X}=i+c\right]\right| \\ &\quad + O\left(\frac{1}{\sqrt{c(n)}}\right) \\ &\leq O\left(\frac{2ds}{\sigma\sqrt{c(n)}} + \frac{1}{\sqrt{c(n)}}\right) \\ &\leq O\left(\sqrt{\frac{\log c(n)}{c(n)}}\right).\end{aligned}$$

■

We now prove Theorem 6.1.

Proof. Fix some ε-uniform Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ of G. Also fix some player i and subset $S \subset N\setminus\{i\}$, $|S| = t \leq \sigma/c(n)$, where

$$\sigma^2 = \mathrm{Var}\left[\sum_{j\neq i} \mathbf{A}_i\right].$$

Note that $\sigma^2 = \Omega(\varepsilon n)$ due to ε-uniformity and the fact that the $\mathbf{A}_i$ are independent. Also, fix some $b_S \in A^{|S|}$, and let $b = \sum_{i\in S} b_i$. That is, b is the number of 1's in the vector b_S. For any $a \in A$,

$$\begin{aligned}&\left|u_i^a(\mathbf{A}_{-i}) - u_i^a(\mathbf{A}_{-(S\cup\{i\})} : b_S)\right| \\ &= \left|\sum_{j=0}^{n-1} u_i^a(j)\Pr\left[\sum_{k\neq i}\mathbf{A}_k = j\right]\right. \\ &\quad \left. - \sum_{j=0}^{n-t-1} u_i^a(j+b)\Pr\left[\sum_{k\notin\{S\cup\{i\}\}}\mathbf{A}_k = j\right]\right| \\ &= \left|\sum_{j=0}^{n-t-1}\sum_{c=0}^{t} u_i^a(j+c)\Pr\left[\sum_{k\notin\{S\cup\{i\}\}}\mathbf{A}_k = j\right]\Pr\left[\sum_{k\in S}\mathbf{A}_k = c\right]\right. \\ &\quad \left. - \sum_{j=0}^{n-t-1} u_i^a(j+b)\Pr\left[\sum_{k\notin\{S\cup\{i\}\}}\mathbf{A}_k = j\right]\right| \\ &= \sum_{r=0}^{t}\Pr\left[\sum_{k\in S}\mathbf{A}_k = c\right]\cdot\left|\sum_{j=0}^{n-t-1} u_i^a(j+c)\Pr\left[\sum_{k\notin\{S\cup\{i\}\}}\mathbf{A}_k = j\right]\right. \\ &\quad \left. - \sum_{j=0}^{n-t-1} u_i^a(j+b)\Pr\left[\sum_{k\notin\{S\cup\{i\}\}}\mathbf{A}_k = j\right]\right| \\ &\leq \delta\end{aligned}$$

by Lemma 6.5.

Because this holds for every $a \in A$, $\mathbf{A}$ is (δ, t)-stable. Immunity and tolerance follow from Lemma 3.4, and resilience against coalitions from Lemma 3.6. ■

6.3 Resilience to Asynchrony

In this section we show that anonymous games are highly resilient to asynchrony by proving Theorem 6.6. For one possible setting of the parameters, the theorem states that with probability approaching 1, no player can improve his payoff by more than an arbitrarily small constant even after being exposed to the actions of nearly a constant fraction of all players.

Theorem 6.6. *Let G be an anonymous n-player game with an ε-uniform Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$, for $\varepsilon > 0$. Let $c(n)$ and $d(n)$ be any complexity functions, and fix*

$$t = \frac{n}{d(n)}.$$

Then there exist universal constants C and D such that G is

$$\left(\delta, e^{-\frac{\varepsilon C\cdot d(n)}{c(n)^2}}, t\right)\text{-ex post},$$

where

$$\delta = D\cdot\sqrt{\frac{\log c(n)}{c(n)}}.$$

Proof. Fix some player i and subset $S \subset N\setminus\{i\}$, $|S| \leq t$, and let

$$\mathbf{B} = \sum_{i\in S}\mathbf{A}_i.$$

Fix $k = \sigma/c(n)$, where $\sigma^2 = \mathrm{Var}[\sum_{j\notin S}\mathbf{A}_j]$. Note that $\sigma^2 \geq \Omega(\varepsilon(n-t))$ by ε-uniformity and the independence of the $\mathbf{A}_i$. By a Chernoff bound,

$$\Pr\left[|\mathbf{B} - \mathrm{E}[\mathbf{B}]| > k\right] \leq e^{-\frac{ad(n)\sigma^2}{c(n)^2 n}} \leq e^{-\frac{2\varepsilon C\cdot d(n)}{c(n)^2}}$$

for some constant C. Assume this does not happen (if it does then it is captured by the second parameter of the ex post property). In this case, we can continue the proof as in that of Theorem 6.1: The maximum offset given by the faulting players is $k = \sigma/c(n)$, which is the same size as the number of faulting players t in Theorem 6.1. Further details are omitted due to space constraints. ■

7. GENERAL GAMES

In general, it is impossible to show the same kind of robustness to Byzantine faults for general games, even if we assume the equilibrium is ε-uniform. The same holds for resilience to asynchrony (see the examples in Section 7.1). However, if we make some assumptions about how faulty or exposed players are chosen, then there is some innate fault tolerance even in general large games.

We begin with a result on resilience to asynchrony, which is essentially a theorem of Gradwohl and Reingold [12] on partial exposure. We then proceed to state a new theorem on tolerance and immunity in general large games – Theorem 7.3.

DEFINITION 7.1 (RANDOM m-EXPOSURE (p,α)-EX POST). *For each $i \in N$, let $\mathbf{R}_i$ be a uniformly random subset $R_i \subset N \setminus \{i\}$ of size $|R_i| = m$. Then a Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is* (p,α)-ex post Nash with random m-exposure *if for every i:*

$$\Pr_{\substack{r\sim\mathbf{R}_i \\ a_r\sim\mathbf{A}_r}}\Big[\exists a \in A \text{ s.t. } u_i(\mathbf{A}^{-i} : a|\mathbf{A}_r = a_r) > u_i(\mathbf{A}|\mathbf{A}_r = a_r) + \alpha\Big] < p.$$

In words, the definition states that even if a player sees the actions of a random subset of m players, the probability that he will be able to improve his payoff by more than α is at most p. The following theorem shows that m can be rather large:

THEOREM 7.2 ([12]). *In any n-player game G, for any $p, \alpha > 0$, and positive integer m, every Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is*

$$\left(p + \frac{4m \cdot |A|}{np\alpha^2}, \alpha\right)\text{-ex post Nash with random } m\text{-exposure.}$$

Note that for $\mathbf{A}$ to be (p,α)-ex post Nash with random exposure we can take

$$m = \Omega\left(\frac{p^2\alpha^2 \cdot n}{|A|}\right).$$

If $|A|$, p and α are constants, then m can be as large as a constant fraction of all the players.

Note that this property of ex post Nash is a weaker form of resilience against asynchrony than what we have for λ-continuous or anonymous games. In particular, Theorem 7.2 implies that "most" players (on expectation, $(1-p)n$) will not be able to improve their expected payoffs by more than α even after being exposed to the actions of a random subset of m players. This is necessarily so, since in general games it is impossible to obtain fault tolerance for all players simultaneously – see Example 7.6 below.

If $\mathbf{A}$ is ε-uniform, then there are similarly weakened forms of t-immunity and t-tolerance as well.

THEOREM 7.3. *Let G be any n-player game with actions $A = \{0,1\}$ and an ε-uniform Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$, for $\varepsilon > 0$. Let $c(n)$ be any complexity function, and fix*

$$t = \frac{\sqrt{\varepsilon n}}{c(n)} \quad \text{and} \quad \delta = \frac{1}{\sqrt[3]{c(n)}}.$$

Finally, let $k \geq 1$. Then the following statements hold for any player i, where the probability is over a uniform random choice of $S \subset N \setminus \{i\}$ of size $|S| = t$:

$$\Pr\Big[\exists a \in A \text{ s.t. } \max_{b_S \in A^t}\left\{u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, a)\right\} > u_i(\mathbf{A}_{-S} : b_S) + 2k\delta\Big] < \frac{2}{k}.$$

Also,

$$\Pr\left[\min_{b_S \in A^t}\left\{u_i(\mathbf{A}_{-S} : b_S)\right\} < u_i(\mathbf{A}) - k\delta\right] < \frac{2}{k}.$$

Theorem 7.3 states that if a set of t faulting players is chosen uniformly at random, then for any player, his strategy is nearly best-response (up to the additive $k\delta$) with probability at least $1-2/k$ even conditioned on the faults. Additionally, the same holds for immunity – t randomly chosen faulting players do not change a player's utility by more than $2k\delta$ with probability more than $2/k$.

7.1 Impossibility of Fault Tolerance with Arbitrary Faulting Players

In this section we give some examples of general games and show that it may be impossible to have fault tolerance if the faulting players are chosen arbitrarily, as opposed to randomly. The first example shows that in general games it is impossible to have 1-tolerance or resilience to asynchrony if a faulting player is chosen in an adversarial manner.

EXAMPLE 7.4. In the expert game, a player wins (has utility 1) if he matches the action of player 1 (and player 1 always wins). One of the Nash equilibria of this game is for every player to flip a fair coin.

If player 1 has a fault and plays action 0 (with probability 1), then no other player's strategy remains optimal. The same holds if some player is exposed to the chosen action of player 1.

The following somewhat degenerate game shows that if the faulting player is chosen in an adversarial manner, then 1-immunity also fails.

EXAMPLE 7.5. In the dictator game, player 1 always wins, and every other player wins if player 1 outputs a 1 (and loses otherwise). One of the Nash equilibria of this game is for every player to flip a fair coin.

Clearly, if player 1 faults and outputs a 0, all other players go from an expected utility of 1/2 to an expected utility of 0.

The next example shows that in general games, it is impossible to obtain 1-tolerance or resilience to asynchrony for all players simultaneously (even if the faulting players are chosen randomly):

EXAMPLE 7.6. In the many-experts game, player i wins if and only if his action matches the action of player $i+1 \pmod n$. There are 3 Nash equilibria in this game: all players play 1, all players play 0, and every player flips a fair coin.

Note that none of the Nash equilibria is fault tolerant. Consider the third equilibrium. If **any** player j has a fault and plays action 1 (with probability 1), then player $j-1 \pmod n$ no longer wishes to flip a coin. The same argument shows

that if player $j-1 \pmod n$ is exposed to player j, then his Nash strategy will no longer be optimal.

From these examples, we conclude that the choice of faulting players needs to be restricted somehow. Additionally, it is necessary to relax the requirement that the equilibrium remain best-response for all players simultaneously after some fault has occurred.

7.2 Proof of Theorem 7.3

The proof utilizes the following lemma, which bounds the number of sets of players who have large influence on the expectation of a function. The proof of the lemma expands on some ideas of Al-Najjar and Smorodinsky [5], and is a reduction to the case of anonymous functions (about which we already have explicit statements – see Section 6).

LEMMA 7.7. *Let $f : \{0,1\}^n \mapsto [0,1]$ be any function, and let $\mathbf{X} = \mathbf{X}_1 \times \ldots \times \mathbf{X}_n$ be an ε-uniform product distribution over $\{0,1\}^n$. Let t and δ be as in Theorem 7.3. Then*

$$\mathop{\mathrm{E}}_{S}\left[\max_{b_S \in \{0,1\}^t} \left| \mathrm{E}\left[f(\mathbf{X})\right] - \mathrm{E}\left[f(\mathbf{X}_{-S} : b_S)\right] \right| \right] \leq \delta,$$

where the expectation is over $S \subseteq N$ of size $|S| = t$ chosen uniformly at random.

PROOF. We first assume that all $\mathbf{X}_i$ are distributed identically. We will then show how to reduce the problem to this case. Let $\sigma : N \mapsto N$ be a permutation of the players. For a set $S \subset N$, let

$$V_S(f) = \max_{b_S \in \{0,1\}^t} \left| \mathrm{E}\left[f(\mathbf{X})\right] - \mathrm{E}\left[f(\mathbf{X}_{-S} : b_S)\right] \right|.$$

Also, for some positive integer t let

$$V_t(f) = \sum_{\substack{S \subset N \\ |S|=t}} V_S(f).$$

Now, let $f^\sigma : \{0,1\}^n \mapsto [0,1]$ be such that $f^\sigma(x_1, \ldots, x_n) = f(\sigma(x_1), \ldots, \sigma(x_n))$. Observe that $V_t(f) = V_t(f^\sigma)$: this holds because in both cases we are summing over all subsets S of size t, and the only difference is that the sum is in a different order. Here we also use the fact that the $\mathbf{X}_i$ are distributed identically.

Now consider the function $g : \{0,1\}^n \mapsto [0,1]$, where

$$g(x_1, \ldots, x_n) = \frac{1}{n!} \sum_{\sigma} f^\sigma(x_1, \ldots, x_n).$$

First of all, note that g is an anonymous function, since $g(x_1, \ldots, x_n) = g(\sigma(x_1), \ldots, \sigma(x_n))$ for all permutations σ of the players. Additionally,

$$V_t(g) = \sum_{\substack{S \subset N \\ |S|=t}} \frac{1}{n!} \sum_{\sigma} V_S(f^\sigma) = \sum_{\substack{S \subset N \\ |S|=t}} \frac{1}{n!} \sum_{\sigma} V_S(f) = V_t(f).$$

Since g is an anonymous function, every S of size t has the same $V_S(g)$. Thus, on expectation (over the random choice of S), $\mathrm{E}[V_S(f)] = V_S(g)$. Now, the proof of Theorem 6.1 shows that for any anonymous function g, $V_S(g) \leq \delta$, and this completes the proof of the claim (actually, Theorem 6.1 is stated only for binary random variables. A similar claim for trinary random variables is omitted due to space constraints).

Recall that in the above, we assumed that all $\mathbf{X}_i$ are distributed identically, and this is not necessarily the case. We now show how to get around this problem by reducing the distribution $\mathbf{X}$ and the function f to a different function and a distribution $\mathbf{Y}$ that is identically distributed, without affecting the influences of sets of players.

For each i, let $p_i = \Pr\left[\mathbf{X}_i = 1\right]$, and recall that $p_i \in [\varepsilon, 1-\varepsilon]$. Now, consider the function $f' : [0,1]^n \mapsto [0,1]$ as follows:

$$f'(r_1, \ldots, r_n) = f(\mathbf{X}_1(r_1), \ldots, \mathbf{X}_n(r_n)),$$

where $\mathbf{X}_i(r_i) = 1$ if $r_i \leq p_i$ and 0 otherwise. Let $\mathbf{U} = \mathbf{U}_1 \times \ldots \times \mathbf{U}_n$ be the uniform distribution over $[0,1]^n$. Then when the inputs to f' are drawn independently from $\mathbf{U}$ we have that $\mathrm{E}[f] = \mathrm{E}[f']$. Also, if we define $V_S(f')$ similarly to $V_S(f)$ (with sup instead of max), then we have $V_S(f') = V_S(f)$.

Now let $\mathbf{Y} = \mathbf{Y}_1 \times \ldots \times \mathbf{Y}_n$ be independent and identically distributed random variables over $\{0, 1, \perp\}$ such that for every i, $\Pr\left[\mathbf{Y}_i = 0\right] = \Pr\left[\mathbf{Y}_i = 1\right] = \varepsilon$, and $\Pr\left[\mathbf{Y}_i = \perp\right] = 1 - 2\varepsilon$. Consider the function $h : \{0, 1, \perp\} \mapsto [0,1]$ such that

$$h(y_1, \ldots, y_n) = \mathrm{E}\left[f'\big(\mathbf{U} | (\mathbf{U}_1 \in y_1), \ldots, (\mathbf{U}_n \in y_n)\big)\right],$$

where

$$(\mathbf{U}_i \in y_i) \Rightarrow \begin{cases} \mathbf{U}_i \in [0, \varepsilon] & \text{if } y_i = 1, \\ \mathbf{U}_i \in [1-\varepsilon, 1] & \text{if } y_i = 0, \\ \mathbf{U}_i \in (\varepsilon, 1-\varepsilon) & \text{otherwise.} \end{cases}$$

We now claim that for every S, $V_S(h) = V_S(f')$. First note that

$$\begin{aligned} \mathrm{E}[h(\mathbf{Y})] &= \sum_{y \in \mathrm{supp}(\mathbf{Y})} \Pr\left[\mathbf{Y} = y\right] h(y) \\ &= \sum_{y \in \mathrm{supp}(\mathbf{Y})} \Pr\left[\mathbf{Y} = y\right] \mathrm{E}\left[f'\big(\mathbf{U} | (\mathbf{U}_1 \in y_1), \ldots, (\mathbf{U}_n \in y_n)\big)\right] \\ &= \mathrm{E}[f'(\mathbf{U})]. \end{aligned}$$

Also, let $b_S \in \{0, 1, \perp\}^t$ be such that

$$\left|\mathrm{E}\left[h(\mathbf{Y})\right] - \mathrm{E}\left[h(\mathbf{Y}_{-S} : b_S)\right]\right|$$

is maximal. Then $b_S \in \{0,1\}^t$, and

$$h(\mathbf{Y}_{-S} : b_S) = f'(\mathbf{U} | (\mathbf{U}_i \in b_i \ \forall i \in S)).$$

Hence,

$$\begin{aligned} V_S(h) &= \max_{b_S \in \{0,1,\perp\}^t} \left|\mathrm{E}\left[h(\mathbf{Y})\right] - \mathrm{E}\left[h(\mathbf{Y}_{-S} : b_S)\right]\right| \\ &= \max_{b_S \in \{0,1\}^t} \left|\mathrm{E}\left[h(\mathbf{Y})\right] - \mathrm{E}\left[h(\mathbf{Y}_{-S} : b_S)\right]\right| \\ &= \max_{b_S \in \{0,1\}^t} \left|\mathrm{E}\left[f'(\mathbf{U})\right] - \mathrm{E}\left[f'\big(\mathbf{U} | (\mathbf{U}_i \in b_i \ \forall i \in S)\big)\right]\right| \\ &= V_S(f'). \end{aligned}$$

So now we have an i.i.d. distribution $\mathbf{Y}$ and a function h such that

$$\mathop{\mathrm{E}}_{\mathbf{Y}}[V_S(h)] = \mathop{\mathrm{E}}_{\mathbf{U}}[V_S(f')] = \mathop{\mathrm{E}}_{\mathbf{X}}[V_S(f)]$$

as desired. ■

We can now prove Theorem 7.3.

PROOF. Fix some $a \in A$ and a player i. Lemma 7.7 implies that

$$\mathop{\mathrm{E}}_{S}\left[\max_{b_S \in \{0,1\}^t} \left|u_i^a(\mathbf{A}_{-i}) - u_i^a(\mathbf{A}_{-S} : b_S)\right|\right] \leq \delta,$$

where S, $|S| = t$, is chosen uniformly at random from N. Now, a Markov bound implies that

$$\Pr\left[\max_{b_S \in \{0,1\}^t} |u_i^a(\mathbf{A}_{-i}) - u_i^a(\mathbf{A}_{-S} : b_S)| > k\delta\right] \leq \frac{1}{k}.$$

The probability that this is true for either u_i^0 or u_i^1 is at most $2/k$. Thus, we have that with probability at least $1 - 2/k$,

$$\left|u_i(\mathbf{A}_{-i}, a) - u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, a)\right| \leq \delta$$

for all $a \in \{0,1\}$ and all $b_S \in A^t$, and so $\mathbf{A}$ is (δ, t)-stable. Immunity and tolerance follow using Lemma 3.4. ■

8. ACKNOWLEDGMENTS

We would like to thank Ariel Yadin for helpful discussions throughout the course of this research.

9. REFERENCES

[1] R. J. Aumann. Acceptable points in general cooperative n-person games. *Contributions to the Theory of Games, Annals of Mathematical Studies*, IV, 1959. Pages 287–324.

[2] A. Aiyer, L. Alvisi, A. Clement, M. Dahlin, J.-P. Martin, and C. Porth. BAR fault tolerance for cooperative services. *Proceedings of 20th ACM Symposium on Operating Systems Principles*, 2005. Pages 45–58.

[3] I. Abraham, D. Dolev, R. Gonen, and J. Halpern. Distributed computing meets game theory: Robust mechanisms for rational secret sharing and multiparty computation. *Proceedings of the 25th Annual ACM Symposium on Principles of Distributed Computing*, 2006. Pages 53–62.

[4] I. Abraham, D. Dolev, and J. Halpern. Lower bounds on implementing robust and resilient mediators. To appear in *5th IACR Theory of Cryptography Conference*, 2008.

[5] N. I. Al-Najjar and R. Smorodinsky. Pivotal players and the characterization of influence. Journal of Economic Theory 92, 2000. Pages 318–342.

[6] A. Blum, M. Hajiaghayi, K. Ligett, and A. Roth. Regret minimization and the price of total anarchy. To appear in *STOC*, 2008.

[7] M. Babaioff, R. Kleinberg, and C. Papadimitriou. Congestion games with malicious players. *Electronic Commerce*, 2007.

[8] B. D. Bernheim, B. Peleg, and M. Whinston. Coalition proof Nash equilibrium: Concepts. *Journal of Economic Theory*, 42(1), 1989. Pages 1–12.

[9] S. Chien and A. Sinclair. Convergence to approximate Nash equilibria in congestion games. SODA, 2007.

[10] C. Daskalakis and C. Papadimitriou. Computing equilibria in anonymous games. *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, 2007. Pages 83–93.

[11] K. Eliaz. Fault-tolerant implementation. *Review of Economic Studies*, 69(3), 2002. Pages 589–610.

[12] R. Gradwohl and O. Reingold. Partial exposure in large games. Submitted.

[13] J. Halpern. A computer scientist looks at game theory. *Games and Economic Behavior* 45:1, 2003. Pages 114–131.

[14] J. Halpern. Computer science and game theory: A brief survey. To appear in *The New Palgrave Dictionary of Economics*.

[15] D. Incinkas and A. Pelc. Impact of asynchrony on the behavior of rational selfish agents. Fundamenta Informaticae 82(1-2), 2008. Pages 113–125.

[16] E. Kalai. Large robust games. *Econometrica*, Vol. 72, No. 6, November 2004. Pages 1631–1665.

[17] E. Kalai. Partially-specified large games. *Lecture Notes in Computer Science* 3828, 2005. Pages 3–13.

[18] S. Kalyanaraman and C. Umans. Algorithms for playing games with limited randomness. *Proceedings of the 15th Annual European Symposium on Algorithms*, 2007. Pages 323–334.

[19] T. Moscibroda, S. Schmid, and R. Wattenhofer. When selfish meets evil: Byzantine players in a virus inoculation game. *25th Annual Symposium on Principles of Distributed Computing*, 2006.

[20] D. Monderer and M. Tennenholtz. Distributed games. *Games and Economic Behavior* 28, 1999. Pages 55–72.

[21] D. Monderer and M. Tennenholtz. Distributed games: from mechanisms to protocols. *AAAI*, 1999. Pages 32–37.

[22] D. Moreno and J. Wooders. Coalition-proof equilibrium. *Games and Economic Behavior*, 17(1), 1996. Pages 80–112.

[23] J. B. Nielsen (editor): Summary Report on Rational Cryptographic Protocols. ECRYPT Report D.PROVI.7, March 2007.

[24] M.J. Osborne and A. Rubinstein. A Course in Game Theory, MIT Press, Cambridge, MA, 1994.

[25] R. Peled, A. Yadin, and A. Yehudayoff, The maximal probability that k-wise independent bits are all 1. arXiv:math.PR/0801.0059.

[26] S. M. Samuels. On the number of successes in independent trials. *The Annals of Mathematical Statistics*, Vol. 36, No. 4, August 1965. Pages 1272–1278.

[27] R. Selten. A reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* 4, 1975. Pages 25–55.