Pretty Good Strategies for Benaloh Challenge

Wojciech Jamroga(✉)

Interdisc. Centre on Security, Reliability and Trust, SnT, University of Luxembourg
Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland
wojciech.jamroga@uni.lu

Abstract. Benaloh challenge allows the voter to audit the encryption of her vote, and in particular to check whether the vote has been represented correctly. An interesting analysis of the mechanism has been presented by Culnane and Teague. The authors propose a natural game-theoretic model of the interaction between the voter and a corrupt, malicious encryption device. Then, they claim that there is no "natural" rational strategy for the voter to play the game. In consequence, the authorities cannot provide the voter with a sensible auditing strategy, which undermines the whole idea.

Here, we claim the contrary, i.e., that there exist simple rational strategies that justify the usefulness of Benaloh challenge.

1 Introduction

Benaloh challenge [3,4] aims to give the voter the possibility to audit the encryption of her vote, and in particular to check whether the vote has been represented correctly. More precisely, the device that encrypts and sends the ballot must first commit to a representation of the vote. After that, the voter decides whether to cast it or "spoil" it, i.e., open the encryption and check its correctness. Intuitively, this should reduce the risk of altering the value of the vote by a malfunctioning or corrupt machine when it casts the ballot on the voter's behalf.

An interesting analysis of the mechanism has been presented in [6]. The authors propose a natural game-theoretic model of the interaction between the voter and a corrupt, malicious encryption device. Then, they claim that there is no "natural" rational strategy for the voter to play the game. More precisely, they claim that: (1) only randomized voting strategies can form a Nash equilibrium, (2) for audit sequences with bounded length, the voter gets cheated in all Nash equilibria, and (3) the Nash equilibria in the infinite game do not form an easy pattern (e.g., Bernoulli trials). In consequence, the voter cannot be provided with a sensible auditing strategy, which undermines the whole method.

In this paper, we claim that – on the contrary – there exist simple auditing strategies that justify the usefulness of Benaloh challenge. This follows from three important observations. First, we show that there *are* Nash equilibria in bounded strategies where the voter casts her intended vote with high probability. Based on this observation, we focus on a small subset of randomized strategies, namely the ones where the voter spoils the ballot with probability p in the first round,

M. Volkamer et al. (Eds.): E-Vote-ID 2023, LNCS 14230, pp. 106–122, 2023.
https://doi.org/10.1007/978-3-031-43756-4_7

and in the second round always casts. Secondly, we point out that the rationality of strategies in Benaloh challenge is better captured by Stackelberg equilibrium, rather than Nash equilibrium. Thirdly, a sensible Stackelberg strategy does not have to be optimal; it suffices that it is "good enough" for whatever purpose it serves. Fourthly, we prove that the Stackelberg equilibrium in the set of such strategies does not exist, but the voter can get arbitrarily close to the upper limit of the Stackelberg payoff. To show this, we formally define the concept of *Stackelberg value*, and show that it is always higher than the value of Nash equilibrium in the set of randomized strategies for the voter.

Related Work. Game-theoretic analysis of voting procedures that takes into account the economic or social incentives of the participants has been scarce. In [5], two voting systems were compared using zero-sum two-player games based on attack trees, with the payoffs representing the success of coercion. In [12], a simple game-theoretic model of preventing coercion was proposed and analyzed using Nash equilibrium, maxmin, and Stackelberg equilibrium. The authors of [22] applied Stackelberg games to prevent manipulation of elections, focussing on the computational complexity of preventing Denial of Service attacks. The research on *security games* [19], using Stackelberg equilibrium to design anti-terrorist and anti-poaching policies, is of some relevance, too.

2 Benaloh Challenge and Benaloh Games

We start by a brief introduction of Benaloh challenge. Then, we summarize the game-theoretic analysis of the challenge, proposed in [6].

2.1 Benaloh Challenge

Benaloh challenge [3,4] is a "cut-and-choose" technique for voter-initiated encryption audits, which proceeds as follows:

1. An empty ballot is generated and provided to the voter.
2. The voter fills in the ballot and transmits it to the encryption device;
3. The device encrypts the ballot with the election public key, and makes the encrypted vote available to the voter;
4. The voter decides to cast the encrypted vote, or to open and audit the encryption. If the encryption is opened, the ballot is discarded, and the voter proceeds back to step 1.

Benaloh challenge is meant to counter the threat of a malicious encryption device that falsely encrypts the ballot, e.g., in favor of another election candidate. Importantly, this should be done without compromising receipt-freeness of the voting protocol. In a broader perspective, the challenge can be applied in any communication scenario where the encryption mechanism is not trustworthy and plausible deniability is required on the side of the sender.

The idea behind the technique is that, if the voters audit the encryption from time to time, corrupt devices will be exposed and investigated. Thus, it does not pay off to tamper with the encryption in the long run, and the perpetrator would have little incentive to do that. At its core, this is a game-theoretic argument.

Condition	Voter payoff $u_V(n_{cast}, n_{cheat})$	Device payoff $u_D(n_{cast}, n_{cheat})$	Comment
$n_{cast} < n_{cheat}$	$Succ_V - (n_{cast} - 1)c_{audit}$	0	Voter votes as intended
$n_{cast} = n_{cheat}$	$-Fail_V - (n_{cast} - 1)c_{audit}$	$Succ_D$	Device successfully cheats
$n_{cast} > n_{cheat}$	$-n_{cheat} \cdot c_{audit}$	$-Fail_D$	Voter catches cheating device

Fig. 1. Inspection game for Benaloh challenge [6, Fig. 2]

2.2 Benaloh Challenge as Inspection Game

Intuitively, the interaction in Benaloh challenge can be seen as a game between the voter V and the encryption device D – or, more accurately, between the voter and the malicious party that might have tampered with the device. We will use the term *Benaloh game* to refer to this aspect of Benaloh challenge. In each round, the voter can choose between casting her intended vote (action *cast*) and auditing the encryption (action *audit*). At the same time, the device chooses to either encrypt the vote truthfully (action *true*) or cheat and encrypt another value of the vote (action *false*). Both players know exactly what happened in the previous rounds, but they decide what to do without knowing what the other player has selected in the current round.

A very interesting analysis has been presented by Chris Culnane and Vanessa Teague in [6]. The authors model the interaction as an *inspection game* [2]. The idea is very simple: V chooses the round n_{cast} in which she wants to cast the vote, and D chooses the round n_{cheat} when it will fake the encryption for the first time. Consequently, the voter's plan is to audit the encryption in all rounds $n < n_{cast}$, and similarly the device encrypts truthfully for all $n < n_{cheat}$. The players choose their strategies before the game, without knowing the opponent's choice. Their payoffs (a.k.a. utilities) are presented in Fig. 1, with the parameters interpreted as follows:

- $Succ_i$: the reward of player i for succeeding with their task (i.e., casting the vote as intended for V, and manipulating the vote for D);
- $Fail_i$: player i's penalty for failing (i.e., getting cheated for V, and getting caught with cheating for D);
- c_{audit}: the cost of a single audit; essentially, a measure of effort and time that V needs to invest into encrypting and spoiling a spurious ballot;

It is assumed that $Succ_i, Fail_i, c_{audit} > 0$. Also, $c_{audit} < Fail_V$, i.e., the voter cares about what happens with her vote enough to audit at least once.

There are two variants of the game: finite, where the number of rounds is bounded by a predefined number $n_{max} \in \mathbb{N}_{\geq 1}$, and infinite, where the game can proceed forever. In the finite variant, the voter chooses $n_{cast} \in \{1, \ldots, n_{max}\}$, and the device selects $n_{cheat} \in \{1, \ldots, n_{max}, \infty\}$, with $n_{cheat} = \infty$ meaning that it always encrypts truthfully and never cheats. In the infinite variant, the voter and the device choose respectively $n_{cast} \in \mathbb{N}_{\geq 1}$ and $n_{cheat} \in \mathbb{N}_{\geq 1} \cup \{\infty\}$. The structure of the game is common knowledge among the players.

Discussion. One might consider a slightly richer game by allowing the voter to refuse participation ($n_{cast} = 0$) or to keep auditing forever ($n_{cast} = \infty$). Also, we could include a reward $Catch_V$ that the voter gets when detecting an attack and reporting it to the authorities. In this paper, we stick to the game model of [6], and leave a proper analysis of the richer game for the future.

2.3 Are There Simple Rational Strategies to Cast and Audit?

Culnane and Teague make the following claims about their model (and, by implication, about the game-theoretic properties of Benaloh challenge):

1. There is no Nash equilibrium in deterministic strategies [6, Lemma 1]. Thus, a rational voter must use *randomized strategies* in Benaloh challenge.[1]
2. A Nash equilibrium in the *finite Benaloh game* can only consist of the voter casting right away and the device cheating right away; the argument proceeds by backward induction [6, Lemma 2 and its proof]. Thus, by [6, Lemma 1], there are no Nash equilibria in the finite Benaloh game, and a rational voter should use *infinite audit strategies*.
3. In the *infinite Benaloh game*, there is no Nash equilibrium in which the voter executes a Bernoulli process, i.e., randomizes in each round with the same probability r whether to audit or cast [6, Theorem 2]. Quoting the authors, "this prevents authorities from providing voters with a sensible auditing strategy." In other words, there are no "easy to use" rational strategies for the voter in Benaloh challenge.

The above claims have two controversial aspects: a technical one and a conceptual one. First, while claims (1) and (3) are correct, claim (2) is not. By Nash's theorem [15], every finite game has a Nash equilibrium in randomized strategies, and this one cannot be an exception. We look closer at the issue in Sect. 4, show why backward induction does *not* work here, and demonstrate that a clever election authority can design the procedure so that the voters do have a simple Nash equilibrium strategy to cast and audit.

Secondly, the authors of [6] implicitly assume that "sensible strategies" equals "simple Nash equilibrium strategies." As we discuss in Sect. 5, Nash equilibrium is not the only concept of rationality that can be applied here. In fact, Stackelberg equilibrium [20] is arguably a better fit for the analysis of Benaloh challenge. Following the observation, we prove that generalized Stackelberg equilibrium [13] for the voter in the set of randomized strategies does not exist, but V can get arbitrarily close to the upper limit of the Stackelberg payoff function. Moreover, there is always a Bernoulli strategy for the voter whose Stackelberg value is higher than the payoff in Nash equilibrium. In sum, Stackelberg games better capture rational interaction in Benaloh challenge, provide the voter with simple strategies, and obtain higher payoffs for V than Nash equilibria.

[1] A concise explanation of game-theoretic terms is presented in Sects. 3 and 5.1.

Alice \ *Bob*	*bar*	*theater*
bar	3, 2	1, 0
theater	4, 0	2, 3

Fig. 2. A variation on the Battle of the Sexes game. The only Nash equilibrium is indicated by the black frame. Stackelberg equilibrium for Alice is set on yellow background. The players' best responses to the opponent's strategies are underlined (Color figure online)

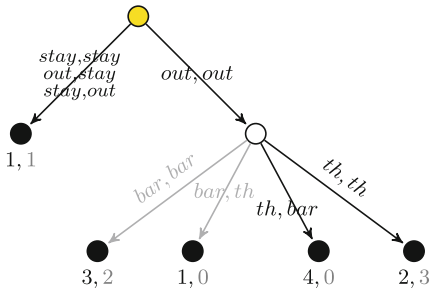

Fig. 3. Multi-step Battle of the Sexes. The initial state is filled with yellow, and terminal states with black. Transitions corresponding to dominated choices are shown in grey (Color figure online)

3 Intermezzo: Game Theory Primer, Part One

Here, we present a compressed summary of the relevant game-theoretic notions. For a detailed introduction, see e.g. [16,18].

Strategic Games. A *strategic game* consists of a finite set of *players* (or *agents*), each endowed with a finite set of *actions*. A tuple of actions, one per player, is called an *action profile*. The *utility function* $u_i(\alpha_1, \ldots, \alpha_n)$ specifies the *utility* (often informally called the *payoff*) that agent i receives after action profile $(\alpha_1, \ldots, \alpha_n)$ has been played. In the simplest case, we assume that each player plays by choosing a single action. This kind of choice represents a *deterministic strategy* (also called *pure strategy*) on the part of the agent.

The payoff table of an example strategic game is shown in Fig. 2. Two players, Alice and Bob, decide in parallel whether to go to the local bar or to the theater. The strategies and utilities of Bob are set in grey for better readability.

Rationality Assumptions. The way rational players choose their behaviors is captured by *solution concepts*, formally represented by a subset of strategies or strategy profiles. In particular, *Nash equilibrium (NE)* selects those strategy profiles σ which are stable under unilateral deviations, i.e., no player i can improve its utility by changing its part of σ while the other players stick to their choices. Equivalently, σ is a Nash equilibrium if each σ_i is a best response to the choices of the other players in σ. In our example, *(theater,theater)* is the only Nash equilibrium. Another solution concept (Stackelberg equilibrium) will be introduced in Sect. 5.1.

Multi-step Games. To model multi-step interaction, we use *concurrent extensive form games*, i.e., game trees where the players proceed in rounds, and choose their actions simultaneously in each round. The agents' payoffs are defined for

each *play*, i.e., maximal path from the root to a leaf of the tree. A multi-step variant of the Battle of the Sexes, where Alice and Bob first veto-vote on whether to go out and then decide on where to go, is shown in Fig. 3. In such games, a deterministic strategy of player i is a conditional plan that maps the nodes in the tree to i's actions. Each strategy profile determines a unique play.

Nash equilibrium is defined analogously to strategic games. Additionally, σ is a *subgame-perfect Nash equilibrium (SPNE)* if it is a Nash equilibrium in each subtree obtained by fixing another starting point for the game. *Backward induction* eliminates choices that are *weakly dominated*, i.e., ones for which there is another choice obtaining a better vector of payoffs. Backward induction preserves subgame-perfect Nash equilibria, and can be used to reduce the game tree if the agents are assumed to play SPNE. For example, Alice's strategy *bar* obtains payoff vector $\boxed{3}\,\boxed{1}$, while *theater* obtains $\boxed{4}\,\boxed{2}$. Thus, the former strategy is dominated by the latter, and can be removed from the game three.

Randomized Play. Randomization makes it harder for the opponents to predict the player's next action, and to exploit the prediction. Moreover, Nash equilibrium is guaranteed to exist for randomized strategy profiles (Nash's theorem). In multi-step games, players can randomize in two ways. A *mixed strategy* for player i is a probability distribution over the pure strategies of i, with the idea that the player randomizes according to that distribution, and then duly executes the selected multi-step strategy. A *behavioral strategy* assigns each game node with a probability distribution over the *actions* of i, with the idea that i randomizes freshly before each subsequent move. By Kuhn's theorem, every mixed strategy has an outcome-equivalent behavioral strategy and vice versa in games with perfect recall. Note that deterministic strategies can be seen as a special kind of randomized strategies that use only Dirac distributions, i.e., $s_i(\alpha) = 1$. In that case we will write $s_i = \alpha$ as a shorthand.

4 Benaloh According to Nash

In this section, we look closer at the claims of [6].

4.1 Deterministic Audit Strategies in Benaloh Games

The first claim is that Benaloh games have no Nash equilibrium where the voter plays deterministically [6, Lemma 1]. This is indeed true. To see that, consider any strategy profile (n_{cast}, s_D) where V deterministically chooses a round n_{cast} to cast her vote, and D chooses n_{cheat} according to probability distribution s_D. If $s_D \neq n_{cast}$, then the device increases its payoff by responding with $s_D = n_{cast}$, i.e., cheating with probability 1 at round n_{cast}; hence, (n_{cast}, s_D) is not a Nash equilibrium. Conversely, if $s_D = n_{cast}$, then the voter increases her payoff by changing her mind and casting at round $n_{cast} - 1$ earlier (if $n_{cast} > 1$) or at round $n_{cast} + 1$ (otherwise); hence (n_{cast}, n_{cast}) is not a Nash equilibrium either.

Ultimately, V must use randomized strategies, so that D cannot precisely predict in which round the vote will be cast.

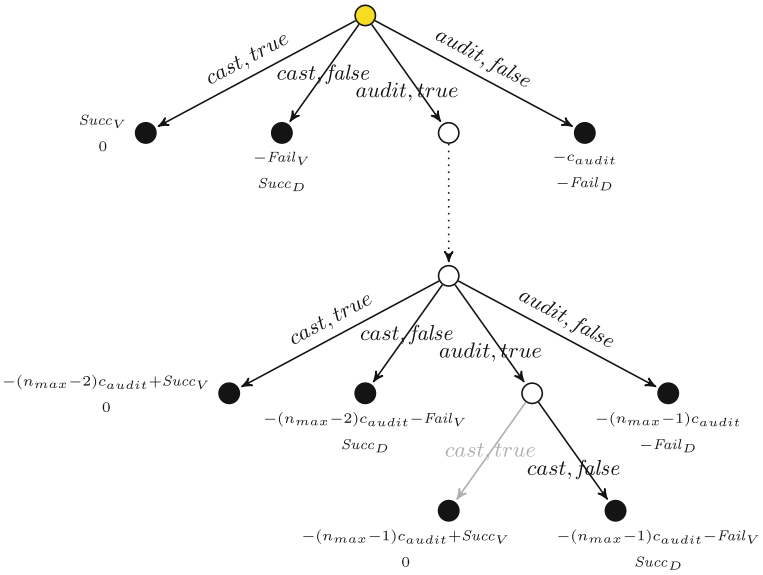

Fig. 4. Game tree for Benaloh challenge. V's payoffs are in black, D's payoffs in red (Color figure online)

4.2 The Rise and Fall of Backward Induction

Now, we turn to randomized voting strategies in Benaloh games with finite horizon n_{max}. It was claimed in [6, proof of Lemma 2] that all V's strategies where the voter does not cast immediately cannot be part of a Nash equilibrium. The argument goes by backward induction: D knows that V must cast in round $n = n_{max}$, so it can safely cheat in that round. Thus, the voter should cast in rounds $1, \ldots, n_{max} - 1$ to avoid being cheated, in which case the device can actually safely cheat in round $n_{max} - 1$, and so on. Unfortunately (or fortunately from the voters' point of view), the argument is incorrect.

To begin with, backward induction *cannot* be applied to games in strategic form nor to inspection games; it requires a proper representation of the sequential nature of the game. We propose the concurrent EF game in Fig. 4 as a model of Benaloh challenge with horizon n_{max}. Each level in the game tree corresponds to a subsequent round of the game. The players choose their actions simultaneously; if V casts, or V audits and D submits false encryption, then the game ends and the payoffs are distributed. If V audits and D encrypts truthfully, the game proceeds to the next round. At $n = n_{max}$, the voter can only cast.

Let us start with the final round of the procedure (i.e., the lowest level in the tree). D has two available choices: *true* and *false*, promising the payoff vectors of $\boxed{0}$ and $\boxed{Succ_D}$, respectively. Indeed, the choice to encrypt truthfully is dominated and can be removed from the tree, leaving only the right-hand branch. We can also propagate the payoffs from the remaining leaf to its parent (i.e., $-(n_{max} - 1)c_{audit} - Fail_V$ for V, and $Succ_D$ for D).

Consider now the second-to-last level of the tree. Again, the device has two choices: *true* promising $\boxed{0}$ $\boxed{Succ_D}$, and *false* promising $\boxed{Succ_D}$ $\boxed{-Fail_D}$. It is easy to see that none of them dominates the other: *false* works strictly better if the opponent decides to cast, whereas *true* obtains better payoff if the opponent does *audit*. Also the voter has now two available choices: *cast* with the payoff vector $\boxed{-(n_{max}-2)c_{audit}+Succ_V}$ $\boxed{-(n_{max}-2)c_{audit}-Fail_V}$ and *audit* with $\boxed{-(n_{max}-1)c_{audit}-Fail_V}$ $\boxed{-(n_{max}-1)c_{audit}}$. Clearly, the former vector obtains better payoff in the first dimension, but strictly worse in the second one. Thus, no choice of the voter is dominated. Since we cannot eliminate any choices, the backward induction stops already at that level.

Why is the intuitive argument in [6] wrong? After all, if the voter assigns a positive probability p to auditing in the round $n_{max}-1$, she knows she will be cheated (in the final round) with exactly that probability. The problem is, if she sets $p=0$, she is sure to get cheated right away! Thus, the voter should use p to keep the opponent uncertain about her current action, which is the usual purpose of randomizing in strategies.

4.3 Mixed Nash Equilibria in Finite Benaloh Games

We know from Sect. 4.2 that backward induction does *not* eliminate randomized audit strategies in finite Benaloh games. The next question is: what Nash equilibria do we obtain? We start with *mixed strategies*, i.e., ones represented by probability distributions $s_V=[p_1^V,\cdots,p_{n_{max}}^V]$ and $s_D=[p_1^D,\cdots,p_\infty^D]$, where p_n^V is the probability that the voter casts her vote in round n, and p_n^D is the probability that the device cheats for the first time in round n.

Support sets of Nash Strategies. First, observe that there are no subgames outside of the main path in the game tree. Thus, all Nash equilibria are subgame perfect. Moreover, backward induction eliminates the possibility that the device encrypts truthfully in the last round, hence $p_\infty^D=0$ in any Nash equilibrium. Consequently, we can represent s_D by $[p_1^D,\cdots,p_{n_{max}}^D]$.

Secondly, all the other probabilities must be nonzero, see the following lemma.[2]

Lemma 1. *If* $s_V=[p_1^V,\cdots,p_{n_{max}}^V]$ *and* $s_D=[p_1^D,\cdots,p_{n_{max}}^D]$ *form a Nash equilibrium, then for all* $i=V,D$ *and* $n=1,\ldots,n_{max}$ *we have* $p_n^i>0$.

Calculating the Audit Probabilities. We compute $p_1^V,\ldots,p_{n_{max}}^V$ using the standard necessary condition for Nash equilibrium in mixed strategies [16, Lemma 33.2]. If (s_V,s_D) is a Nash equilibrium with $p_n^V>0$ and $p_n^D>0$ for all $n=1,\ldots,n_{max}$, then the following conditions must hold:

1. Every deterministic strategy of V obtains the same payoff against s_D, in other words: $\forall n_{cast},n'_{cast}\in\{1,\ldots,n_{max}\}\ .\ u_V(n_{cast},s_D)=u_V(n'_{cast},s_D)$

[2] The proofs of the formal results can be found in the extended version of the paper [11].

2. Every deterministic strategy of D obtains the same payoff against s_V, in other words: $\forall n_{cheat}, n'_{cheat} \in \{1, \ldots, n_{max}\} \;.\; u_D(s_V, n_{cheat}) = u_D(s_V, n'_{cheat})$

Consider condition (2). Using the payoffs in Fig. 1, we get:

Lemma 2. *If $s_V = [p_1^V, \cdots, p_{n_{max}}^V]$ is a part of Nash equilibrium then $p_{n+1}^V = \frac{Succ_D}{Succ_D + Fail_D}\, p_n^V$ for every $n \in \{1, \ldots, n_{max} - 1\}$.*

Theorem 1. *The mixed voting strategy $s_V = [p_1^V, \cdots, p_{n_{max}}^V]$ is a part of Nash equilibrium iff, for every $n \in \{1, \ldots, n_{max}\}$:*

$$p_n^V \;=\; \frac{(1-R)R^{n-1}}{1-R^{n_{max}}}, \qquad where\ R = \frac{Succ_D}{Succ_D + Fail_D}.$$

Indeed, the mixed equilibrium strategy s_V provides no *simple* recipe for the voter. This is evident when we consider concrete payoff values.

Example 1. Take $n_{max} = 5$ and assume $Succ_D = 1, Fail_D = 4$, i.e., the opponent fears failure four times more than he values success. Then, $R = 0.2$, and hence $s_V = [0.8, 0.16, 0.032, 0.006, 0.001]$ is the unique equilibrium strategy for the voter. In other words, the voter should cast immediately with probability 0.8, audit once and cast in round 2 with probability 0.16, and so on.

4.4 Towards Natural Audit Strategies

So far, we have considered *mixed strategies* for the voter. That is, the voter draws n_{cast} before the game according to the probability distribution s_V, and then duly follows the outcome of the draw. An alternative is to use a *behavioral strategy* $b_V = (b_1^V, \ldots, b_{n_{max}}^V)$, where the voter does a *fresh* Bernoulli-style lottery with probability of success b_n^V in each subsequent round. If successful, she casts her vote; otherwise, she audits and proceeds to the next round.

Behavioral Nash Equilibria. First, we observe that the game in Fig. 4 is a game of *perfect recall*, i.e., the players remember all their past observations (in our case, the outcomes of all the previous rounds). Thus, by Kuhn's theorem, mixed and behavioral strategies are outcome-equivalent. In other words, the same outcomes can be obtained if the players randomize before the game or throughout the game. Below, we characterize the behavioral strategy that corresponds to the mixed strategy of Theorem 1.

Theorem 2. *The behavioral voting strategy $b_V = [b_1^V, \cdots, b_{n_{max}}^V]$ is a part of Nash equilibrium iff, for every $n \in \{1, \ldots, n_{max}\}$:*

$$b_n^V \;=\; \frac{1-R}{1-R^{n_{max}-n+1}}, \qquad where\ R = \frac{Succ_D}{Succ_D + Fail_D}.$$

Example 2. The behavioral strategy implementing s_V of Example 1 is $b_V = [0.8, 0.801, 0.81, 0.83, 1]$. That is, the voter casts immediately with probability 0.8, else audits, randomizes again, and casts with probability 0.801, and so on.

n_{cast} \ n_{cheat}	**1**	**2**
1	$-Fail_V,\ Succ_D$	$Succ_V,\ 0$
2	$-c_{audit},\ -Fail_D$	$-c_{audit} - Fail_V,\ Succ_D$

n_{cast} \ n_{cheat}	**1**	**2**
1	$-3,\ 1$	$2,\ 0$
2	$-1,\ -4$	$-4,\ 1$

Fig. 5. Benaloh game for $n_{max} = 2$: (a) parameterized payoff table; (b) concrete payoff table for the values of Example 4

Behavioral Audit Strategies are Reasonably Simple. At the first glance, the above behavioral strategy seems difficult to execute, too. We cannot expect the voter to randomize with probability *exactly* 0.8, then *exactly* 0.801, etc. On the other hand, b_V can be approximated reasonably well by the following recipe: "in each round before n_{max}, cast with probability close to 0.8, otherwise audit, randomize freshly, and repeat; in the last round, cast with probability 1." This can be generalized due to the following observation.

In Benaloh games, we can usually assume that $Fail_D \gg Succ_D$. First of all, it is important to realize that the opponent of the voter is not the encrypting device, but a human or organizational perpetrator represented by the device. To be more precise, the strategies in the game are defined by the capabilities of the device, but the incentives are those of the perpetrator. Thus, the utility values defined by u_D should not be read as "the payoffs of the device," but rather the utilities of the external party who rigged the device in order to achieve some political, social, or economic goals. Secondly, the scope of the opponent's activity is not limited to the interaction with a single voter and to corrupting a single encryption device. Presumably, they must have tampered with multiple devices in order to influence the outcome of the vote. Consequently, the opponent is in serious trouble if even few devices are caught cheating. This is likely to attract attention and trigger investigation, which may lead to an audit of all the encryption devices, revision or voiding of the votes collected from those that turned out corrupt, and even an arrest and prosecution of the perpetrator. All in all, the penalty for fraud detection ($Fail_D$) is usually much higher than the reward for a successful swap of a single vote ($Succ_D$).

Theorem 3. *If* $\frac{Succ_D}{Fail_D} \to 0$*, then the equilibrium strategy* b_V *of the voter converges to the following behavioral strategy:*

$$\widehat{b_n^V} = \begin{cases} \frac{Fail_D}{Succ_D + Fail_D} & \text{for } n < n_{max} \\ 1 & \text{for } n = n_{max} \end{cases}$$

The finite Bernoulli strategy to audit with probability $R = \frac{Fail_D}{Succ_D + Fail_D}$ in each round except last seems reasonably simple. By Theorem 3, it is also reasonably close to the unique Nash equilibrium.

Making Things even Simpler for the Voter. In order to make Benaloh challenge even easier to use, the voting authority can set n_{max} accordingly. In particular, it can fix $n_{max} = 2$, i.e., allow the voter to audit at most once. That does not seem very restrictive, as empirical evidence suggests that voters seldom

audit their votes [1,7,21], and even fewer are able to complete it correctly [1,9,21].[3] The Benaloh game in strategic form for $n_{max} = 2$ is shown in Fig. 5a.

Theorem 4. *For $n_{max} = 2$, the behavioral NE strategy of the voter is:*

$$b_1^V = \frac{Succ_D + Fail_D}{2Succ_D + Fail_D}, \qquad b_2^V = 1.$$

To make the analysis intuitive, consider the concrete values in Example 1.

Example 3. Take $Succ_D = 1, Fail_D = 4$. By Theorem 2, the behavioral Nash equilibrium strategy of the voter is $b_V = [\frac{5}{6}, 1]$. That is, the voter casts immediately with probability $\frac{5}{6}$, otherwise audits and casts in the next round – which is a rather simple strategy.

Also, recall our argument that, typically, $Fail_D \gg Succ_D$. In that case, p_V^1 becomes close to 1. In other words, the voter should *almost always* cast immediately, which is a very simple recipe to follow. Thus, contrary to what Culnane and Teague claim in [6], Benaloh challenge can be designed in a way that admits simple Nash equilibrium strategies of the voter.

4.5 Behavioral Audit Strategies are Simple Enough, but are They Good Enough?

We have just seen that finite Benaloh games do allow for simple and easy to use Nash equilibrium strategies. This seems good news, but what kind of utility do they promise for the voter? That is, how much will the voter benefit from playing NE in Benaloh challenge? For easier reading, we calculate the answer on our running example.

Example 4. Following Example 3, we take $n_{max} = 2, Succ_D = 1, Fail_D = 4$. Moreover, we assume $Succ_V = 2, Fail_V = 3, c_{audit} = 1$, i.e., the voter loses slightly more by getting cheated than she gains by casting successfully, and the cost of an audit is half of the gain from a successful vote. The resulting payoff table is presented in Fig. 5b.

We can now compute the Nash equilibrium strategy of the device using Lemma 1 and Condition 1 of Sect. 4.3. Consequently, we get $-3p_1^D + 2(1-p_1^D) = -p_1^D - 4(1-p_1^D)$, and thus $s_D = [\frac{3}{4}, \frac{1}{4}]$. Recall that the NE strategy of the voter is $s_V = [\frac{5}{6}, \frac{1}{6}]$. This yields the following expected payoffs of the players:

$$u_V(s_V, s_D) = -3\frac{15}{24} + 2\frac{5}{24} - 1\frac{3}{24} - 4\frac{1}{24} = -\frac{7}{6}$$
$$u_D(s_V, s_D) = 1\frac{15}{24} + 0\frac{5}{24} - 4\frac{3}{24} + \frac{1}{24} = \frac{1}{6}.$$

[3] In fairness, there is also some evidence that suggests the contrary [8, Section 5.6.1].

So, the voter gets negative expected utility, and would be better off by not joining the game at all! If that is the case, then a considerate election authority should forbid electronic voting *not* because there are no simple NE strategies to audit and vote, but because there is one and it is bad for the voter. The big question is: does Nash equilibrium really provide the right solution concept for rational interaction in Benaloh challenge? We discuss this in Sect. 5.

5 Benaloh According to Stackelberg

Nash equilibrium encodes a particular view of rational decision making. In this section, we discuss its applicability to Benaloh games, suggest that Stackelberg equilibrium is a much better match, and analyze Benaloh challenge through the lens of Stackelberg games.

5.1 Game-Theoretic Intermezzo, Part Two

Every solution concept encodes its own assumptions about the nature of interaction between players and their deliberation processes. The assumptions behind Nash equilibrium in 2-player games can be characterized as follows [17]:

1. Alice and Bob have common belief that each of them plays best response to one another, and
2. Alice believes that Bob has an accurate view of her beliefs, and that Bob believes that Alice has an accurate view of his beliefs,
3. ...and analogously for Bob.

Alternatively, NE can be characterized as a local optimum of strategy search with mutual adaptations. Informally, it represents collective behaviors that can emerge when the agents play the game repeatedly, and adapt their choices to what they expect from the other agents. Thus, it captures the "organic" emergence of behavior through a sequence of strategy adjustments that leads to a point where nobody is tempted to change their strategy anymore.

Is Nash equilibrium the right concept of rationality for Benaloh games? Note that the characterizations of NE are inherently symmetric. In particular, they assume that both players are able to form accurate beliefs about each other's intentions. This is *not* the case in Benaloh challenge. In line with the arguments of [6], the perpetrator has significant technological and motivational advantage over an average voter. For example, he can use opinion polls and statistical methods to get a good view of the voter's preferences. Even more importantly, machine learning techniques can be used to profile the frequencies with which the voter chooses to audit or cast. On the other hand, the voter has neither data nor resources to form accurate predictions w.r.t. the strategy of the encryption device. This seems pretty close to the Stackelberg model of economic interaction.

Stackelberg Equilibrium. *Stackelberg games* [20] represent interaction where the strategy of one player (called the *leader*) is known in advance by the

other player (the *follower*). The follower is assumed to play best response to that strategy. The *generalized Stackelberg equilibrium (SE)* [13] prescribes the leader's strategy that maximizes the guaranteed payoff against the follower's best responses. We define and analyze SE for Benaloh games in Sect. 5.2.

5.2 Pretty Good Strategies Against Best Response

For simplicity, we assume that $n_{max} = 2$ throughout this section, i.e., the voter can audit the encryption at most once. Thus, the strategy of the voter can be represented by the probability p_V of casting the vote in the first round. Similarly, the strategy of the device can be represented by the probability p^D of cheating in the first round. We first establish D's best response to any fixed p^V and the voter's guaranteed expected utility against best response. These can be formally defined as follows.

Definition 1. *The* best response *of D, given V's strategy represented by p^V, returns those strategies p^D for which the expected value of $u_D(p^V, p^D)$ is maximal:*

$$BR_D(p^V) = \text{argmax}_{p^D \in [0,1]}(Eu_D(p^V, p^D)).$$

Note that a best response always exists, though it does not have to be unique.

Definition 2. *The* generalized Stackelberg equilibrium *for V is defined as the strategy that maximizes V's expected payoff against best response. In case of multiple best responses to some p^V, we look at the worst case scenario.*

$$SE_V = \text{argmax}_{p^V \in [0,1]} \inf_{p^D \in BR_D(p^V)}(Eu_V(p^V, p^D)).$$

For randomized strategies of the leader, the Stackelberg equilibrium does not have to exist (cf. Example 5). To characterize the leader's abilities in such games, we propose the notion of *Stackelberg value*.

Definition 3. *The* Stackelberg value *for V is the expected guaranteed payoff that V can obtain against best response* in the limit*:*

$$SVal_V = \sup_{p^V \in [0,1]} \inf_{p^D \in BR_D(p^V)}(Eu_V(p^V, p^D)).$$

Clearly, $SVal_V$ is always well defined. Moreover, the game has a Stackelberg equilibrium if V obtains the Stackelberg value for some strategy. Finally, for each $\epsilon > 0$, the voter has a strategy that ϵ-approximates the Stackelberg value, i.e., obtains at least $SVal_V - \epsilon$ against best response.

Lemma 3. *The best response of the device to any fixed strategy of the voter is*

$$BR_D(p^V) = \begin{cases} 0 & \text{for } p^V < p^V_{NE} \\ 1 & \text{for } p^V > p^V_{NE} \\ \text{any } p^D \in [0,1] & \text{for } p^V = p^V_{NE} \end{cases}$$

where $p^V_{NE} = \frac{Succ_D + Fail_D}{2Succ_D + Fail_D}$ is the NE probability of casting in round 1.

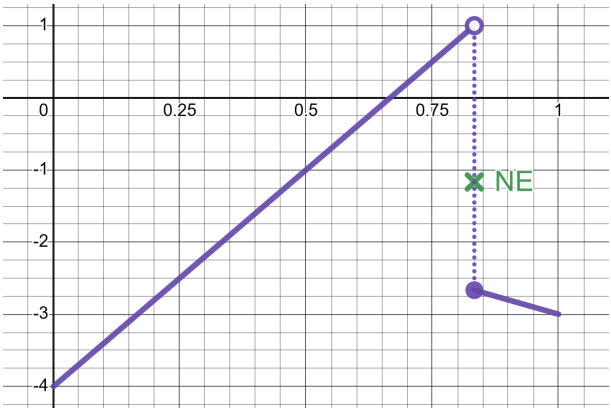

Fig. 6. V's payoffs against best response for the Benaloh game in Fig. 5b. The voter's payoff obtained by Nash equilibrium is shown for comparison

Lemma 4. *The voter's expected utility against best response is:*

$$Eu_V(p^V, BR_D(p^V)) = \begin{cases} p^V Succ_V - (1-p^V)(c_{audit} + Fail_V) & \text{for } p^V < p^V_{NE} \\ -p^V Fail_V - (1-p^V)c_{audit} & \text{for } p^V \geq p^V_{NE} \end{cases}$$

Example 5. The graph of $Eu_V(p^V, BR_D(p^V))$ for the parameters in Example 4 (i.e., $n_{max} = 2, Succ_D = 1, Fail_D = 4, Succ_V = 2, Fail_V = 3, c_{audit} = 1$) is depicted in Fig. 6. It is easy to see that the function does not reach its optimum, and hence the optimal p^V against best response does not exist. Still, the strategies based on p^V being *slightly smaller* than the Nash equilibrium strategy $p^V_{NE} = \frac{5}{6}$ are quite attractive to the voter, since they obtain payoff that is both positive and strictly higher than the Nash payoff.

The next and final theorem generalizes the example to arbitrary two-round Benaloh games. It shows that the voter has no optimal Stackelberg strategy in the game (point 1), but the value of $SVal_V = \frac{Succ_D(Succ_V - Fail_V - c_{audit}) + Fail_D Succ_V}{2Succ_D + Fail_D}$ can be approximated arbitrarily closely (point 2). That is, for each $\epsilon > 0$, the voter has a strategy that obtains at least $SVal_V - \epsilon$ against best response. Moreover, ϵ-approximating Stackelberg equilibrium is strictly better than playing Nash equilibrium (point 3). Lastly, approximate Stackelberg strategies obtain positive utility for the voter under reasonable assumptions (point 4).

Theorem 5. *The following properties hold for the Benaloh game with $n_{max} = 2$:*

1. *There is no Stackelberg equilibrium for V in randomized strategies.*
2. *The Stackelberg value of the game is $SVal_V = \frac{Succ_D(Succ_V - Fail_V - c_{audit}) + Fail_D Succ_V}{2Succ_D + Fail_D}$.*
3. *$SVal_V > Eu_V(p^V_{NE}, p^D_{NE})$, where (p^V_{NE}, p^D_{NE}) is the Nash equilibrium.*
4. *If $Fail_D \gg Succ_D$ and $Succ_V \geq aFail_V$ for a fixed $a > 0$, then $SVal_V > 0$.*

Thus, Stackelberg games capture the rational interaction in Benaloh games better than Nash equilibrium, and predict strictly higher payoffs for the voter.

6 Conclusions, or What Do We Learn from That?

In this paper, we analyze a simple game-theoretic model of incentives in Benaloh challenge, inspired by [6]. Contrary to [6], we conclude that the voters have at their disposal simple strategies to audit and cast their votes. This is especially the case if encryption audits are limited to at most one audit per voter. In that event, a pretty good strategy for the voter is to almost always (but not *exactly* always!) cast immediately in the first round. Interestingly, this is how voters usually behave in real-life elections, according to empirical evidence.

Moreover, we point out that rational interaction in Benaloh games is better captured by Stackelberg equilibrium, rather than Nash equilibrium. While the optimal Stackelberg strategy is not attainable for the voter, it can be approximated arbitrarily close by casting the vote immediately with probability *slightly lower* than for the Nash equilibrium. This is good news, because Stackelberg strategies (even approximate) promise strictly better payoffs for the voter than Nash strategies. And, under reasonable assumptions, they produce positive utility for V. Thus, using Benaloh challenge *is* beneficial to the voter, after all.

The takeaway advice based on this study can be summarized as follows:

1. Using Benaloh challenge is practical and beneficial to the rational voter.
2. Putting a strict limit on the number of allowed audits makes things easier for the voter. The election authority might design the voting system so that each voter can audit the vote encryption at most once.
3. The voters should not try to adapt to the strategy of the attacker, the way Nash equilibrium prescribes. Instead, they should stick to auditing the votes with a fixed (and rather low) frequency, thus approximating the Stackelberg optimum and putting the opponent on the defensive.

Discussion and Future Work. An obvious limitation of the current study is the assumption of *complete information* about the structure of the game. In particular, it is dubious to assume that the voter knows how much the adversary values the outcomes of the game. In the future, we plan to extend the analysis to an incomplete information game model of Benaloh challenge, e.g., in the form of a Bayesian game [10].

Moreover, the analysis in this paper is performed as a 2-player game between a single voter and the voter's device. It would be interesting to see how this extends to scenarios where the adversary controls multiple devices and plays multiple rounds with different voters. Last but not least, the players' payoffs for either failing or succeeding need further discussion. In particular, we assume that the costs of failure for the opponent are much higher than the benefits of success; this should be better justified or refuted.

Acknowledgments. The author thanks Stanisław Ambroszkiewicz, Peter B. Roenne, Peter Y.A. Ryan, and the anonymous reviewers of E-VOTE-ID for their valuable comments, suggestions, and discussions. The work has been supported by NCBR Poland and FNR Luxembourg under the PolLux/FNR-CORE projects STV (POLLUX-VII/1/

2019 and C18/IS/12685695/IS/STV/Ryan), SpaceVote (POLLUX-XI/14/SpaceVote/2023 and C22/IS/17232062/SpaceVote) and PABLO (C21/IS/16326754/PABLO).

References

1. Acemyan, C.Z., Kortum, P., Byrne, M.D., Wallach, D.S.: Usability of voter verifiable, end-to-end voting systems: baseline data for Helios, Prêt à Voter, and Scantegrity II. In: Proceedings of EVT/VVOTE. USENIX Association (2014)
2. Avenhaus, R., Von Stengel, B., Zamir, S.: Inspection games. In: Handbook of Game Theory **3**, 1947–1987. North-Holland (2000)
3. Benaloh, J.: Simple verifiable elections. In: USENIX Electronic Voting Technology Workshop (2006)
4. Benaloh, J.: Ballot casting assurance via voter-initiated poll station auditing. In: USENIX/ACCURATE Electronic Voting Technology Workshop (2007)
5. Buldas, A., Mägi, T.: Practical security analysis of e-voting systems. In: Miyaji, A., Kikuchi, H., Rannenberg, K. (eds.) IWSEC 2007. LNCS, vol. 4752, pp. 320–335. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75651-4_22
6. Culnane, C., Teague, V.: Strategies for voter-initiated election audits. In: Zhu, Q., Alpcan, T., Panaousis, E., Tambe, M., Casey, W. (eds.) GameSec 2016. LNCS, vol. 9996, pp. 235–247. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47413-7_14
7. Ehin, P., Solvak, M., Willemson, J., Vinkel, P.: Internet voting in Estonia 2005–2019: evidence from eleven elections. Gov. Inf. Q. **39**(4), 101718 (2022)
8. Gjøsteen, K.: E-voting in Norway. In: Hao, F., Ryan, P.Y.A. (eds.) Real-World Electronic Voting. Design, Analysis and Deployment. CRC Press (2016)
9. Gjøsteen, K., Lund, A.S.: An experiment on the security of the Norwegian electronic voting protocol. Ann. Telecommun. **71**(7), 299–307 (2016). https://doi.org/10.1007/s12243-016-0509-8
10. Harsanyi, J.C., Selten, R.: A generalized Nash solution for two-person bargaining games with incomplete information. Manage. Sci. **18**(5/2), 80–106 (1972)
11. Jamroga, W.: Pretty good strategies for Benaloh challenge (2023). arXiv:2307.03258, https://arxiv.org/abs/2307.03258
12. Jamroga, W., Tabatabaei, M.: Preventing coercion in e-voting: be open and commit. In: Krimmer, R., et al. (eds.) E-Vote-ID 2016. LNCS, vol. 10141, pp. 1–17. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52240-1_1
13. Leitmann, G.: On generalized stackelberg strategies. J. Optim. Theory Appl. **26**(4), 637–643 (1978)
14. Marky, K., Kulyk, O., Renaud, K., Volkamer, M.: What did I really vote for? In: Proceedings of the Conference on Human Factors in Computing Systems CHI, p. 176. ACM (2018)
15. Nash, J.F.: Equilibrium points in n-person games. Proc. Nat. Acad. Sci. U.S.A., **36**, 48–49 (1950)
16. Osborne, M., Rubinstein, A.: A Course in Game Theory. MIT Press, Cambridge (1994)
17. Perea, A.: A one-person doxastic characterization of Nash strategies. Synthese **158**(2), 251–271 (2007)
18. Shoham, Y., Leyton-Brown, K.: Multiagent Systems - Algorithmic, Game-Theoretic, and Logical Foundations. Cambridge University Press, Cambridge (2009)

19. Tambe, M.: Security and Game Theory: Algorithms Deployed Systems Lessons Learned. Cambridge University Press, Cambridge (2011)
20. von Stackelberg, H.: The Theory of the Market Economy. Oxford University Press, Oxford (1952)
21. Weber, J.-L., Hengartner, U.: USAB. study of the open audit voting system Helios (2009). http://www.jannaweber.com/wpcontent/uploads/2009/09/858Helios.pdf
22. Yin, Y., Vorobeychik, Y., An, B., Hazon, N.: Optimally protecting elections. In: Proceedings of SECMAS. IFAAMAS (2016)