BAR Nash Equilibrium and Application to Blockchain Design

Extended Abstract

Maxime Reynouard
Nomadic Labs
Paris, France
LAMSADE, Dauphine - PSL
Paris, France
maximereynouard@gmail.com

Olga Gorelkina
Mohammed VI Polytechnic University
Rabat, Morocco
University of Liverpool
Liverpool, United Kingdom
ogorelkina@gmail.com

Rida Laraki
Mohammed VI Polytechnic University
Rabat, Morocco
CNRS (Dauphine - PSL)
Paris, France
rida.laraki@dauphine.fr

ABSTRACT

This paper presents a novel solution concept, called BAR Nash Equilibrium (BARNE) and apply it to analyse the Verifier's dilemma, a fundamental problem in blockchain. Our solution concept adapts the Nash equilibrium (NE) to accommodate interactions among Byzantine, altruistic and rational agents, which became known as the BAR setting in the literature. We prove the existence of BARNE in a large class of games and introduce two natural refinements, global and local stability. Using this equilibrium and its refinement, we analyse the free-rider problem in the context of byzantine consensus. We demonstrate that by incorporating fines and forced errors into a standard quorum-based blockchain protocol, we can effectively reestablish honest behavior as a globally stable BARNE.

KEYWORDS

BAR setting; BAR Nash Equilibrium (BARNE); global stability; local stability; Verifier's dilemma; quorum based consensus protocols; blockchain

ACM Reference Format:
Maxime Reynouard, Olga Gorelkina, and Rida Laraki. 2024. BAR Nash Equilibrium and Application to Blockchain Design: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

Introduction. Security research in the field of Distributed Algorithms (DA) usually focuses on fault tolerance. However, the recent proliferation of blockchains shown that faults are not the only challenge: DA must also resist to selfish nodes that are neither faulty nor adversarial. In the case of Ethereum, [26] documented instances violation of the prescribed protocol in order to maximize mining rewards.[1]

BAR model. In 2005, [3] introduced the *Byzantine-Altruistic-Rational* (BAR) model where agents are prescribed a protocol $\tau \in T$ (T being an agent's strategy space) and the set $N = \{1, 2, ..., n\}$ is partitioned in the subsets:

- Set $F \subseteq N$ of *Byzantine* or ***F**aulty* agents deviate arbitrarily from τ, including individual and group deviations. They can be anything from faulty to collusive and adversarial.
- Set $G \subseteq N$ of *Rational*, *Selfish* or ***G**ain seeking* agents maximize their payoff in the game. In an incentivised distributed algorithms such as blockchain consensus protocols, gain seeking agents deviate from τ if it increases their payoff.
- Set $H \subseteq N$ of *Altruistic* or ***H**onest* agents always follow the protocol τ. Honest agents are unwilling or unable to change strategy. In the rest of the paper, all strategy profiles $s \in T^n$ are assumed to satisfy $s_H = \tau^{|H|}$ in line with this definition.

Since then rational deviations have been studied for different applied cases without a unified formal framework [2, 9, 11, 13, 15, 16, 19, 20, 24, 25]. In 2011, [1] informally introduced a BAR compatible game theoretical equilibrium: BAR-strong equilibrium, later formalized [22]. But from their own admission, it has the same downside as game theory's Strong Nash equilibrium: it rarely exists and is not always predictive [6].

BARNE. Our novel solution concept, the BAR-Nash Equilibrium (BARNE) is guaranteed to exist in a large class of games including mixed extensions of any finite games [23].

BARNE is set apart from the BAR-strong notion by its existence property inherited from Nash Equilibrium (NE)[17]. Existence is of considerable importance as it always allows to give a prediction as to rational agents' behaviour. Furthermore, it aims to introduce a unifying solution concept to the large literature that currently lacks a common methodology for analyzing incentive in protocols. For example, [10, 13, 25] essentially check whether the prescribed protocol is a (0,1)-BARNE, while [4, 11] restrict the setting in which the analysis of incentives is performed.

We use the following notations $i \in N$ denotes a single agent and $I \subset N$ a subset. For a (joint) strategy profile $s \in T^n$ of all agents, we note $s_i \in T$ the strategy of agent i, and $s_I \in T^{|I|}$ the sub-profile of agents in I. When $s \in T^n$ is played, i's payoff is $u_i(s_1, ..., s_n)$. With a slight abuse of notation, we write $(s_I, s_J, s_i, s_j) = s_{I \cup J \cup \{i,j\}}$, for disjoint subsets $I, J \subset N$ and distinct $i, j \in N \setminus (I \cup J)$. Similarly, in the case of utility functions, we write: $u_i(s) = u_i(s_1, ...s_n) = u_i(s_I, s_{N \setminus I})$.

Definition 0.1. Given F and G, two disjoint subsets of N, the joint strategy profile $s^*_G \in T^g$ is

(1) BARNE at (F, G) if $\forall i \in G$:
$s^*_i \in argmax_{s_i \in T} \, min_{s_F \in T^f} \, u_i(s_F, s_i, s^*_{G \setminus \{i\}}, s_H)$.
(2) BARNE at (f, g) if $\forall F, G \subset N$, disjoints, and with $|F| = f$ and $|G| = g$; s^*_G is a BARNE at (F, G).

[1] Theoretically, selfish mining attacks were studied in [9, 19, 20, 24]; while other types of selfish behaviour were studied in [2, 13, 15, 16, 25]. A block creator maximising extractable value, e.g. front-running, is another type of selfish behavior.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

Table 1: Properties of different equilibria notions

	BAR-strong	BARNE	locally stable BARNE	globally stable BARNE
existence in a large class of games		✓		
anti coalition & dominant best reply	✓			
anti individual deviations of rationals	✓	✓	✓	✓
dominant strategy best-reply wrt Byzantines	✓			
max-min best-reply wrt Byzantines	✓	✓	✓	✓
locally stable	✓		✓	✓
globally stable	✓			✓

BARNE stability. We introduce two refinements of BARNE: local and global stability. They are of particular relevance for when designing distributed algorithms that aim at being able to tolerate a certain number of deviations (either byzantine fault tolerance or selfishness tolerance). Stability refers to the robustness of a BARNE to changes in the number of agents of each type and aims at being able to predict rational behavior without perfect information on agents' types. *Locally* stable BARNE are the strategy profiles that remain BARNE in spite of local perturbations in the numbers of Byzantine and selfish agents, this is meant to be an intermediary step toward traditional tolerance. *Globally* stable BARNE is concerned with the stability of the equilibrium profile for all parameters below certain thresholds in the numbers of Byzantine and selfish agents, it is in line with the traditional fault tolerance in DA. The notion of BAR-strong equilibrium that was defined in [1] is an even stricter refinement requiring a dominant strategy and being anti coalition. We believe each equilibrium notion to have their use and table 1 illustrates the properties of the various solution concepts.

Verifier's Dilemma. To ilustrate the workings of BARNE and its refinements, we study a pressing blockchain problem: the Verifier's Dilemma, in the context of Quorum Based Consensus Protocols (QBCPs, [5, 7, 27]).[2] The Verifier's dilemma arises because multiple agents must verify and validate transactions to maintain blockchain integrity; Since verification is individually costly, it can be rational to forego verifying altogether and rely on the others' verification effort.[3]

BARNE analysis. Since the QBCP happens at each block proposal, it results in a repeated game, however, to better illustrate our solution concepts but also be closer to what we expect to observe in practice, we focus on the stationary BARNE equilibria (e.g. where the rational agents repeatedly play, iid, the same strategy profile of the stage game). Our analysis, shows that following the prescribed strategy of the standard QBCP is almost never a BARNE and is never a stable BARNE.[4] A previous work [4] formally studied a similar problem, they proposed an amendment where the designer sends personalised, yet correlated recommendations to the agents. Notwithstanding its ingenious design, their amendment is vulnerable because instead of prescribing a protocol that is a (globally) stable BARNE, they propose different protocols for different values of f and g. So their protocol is not stable as their construction relies on the exact knowledge of the number of the Byzantine and Selfish agents, and it is subject to Single Points Of Failure (SPOF). Moreover their analysis only encompassed **B**yzantine and **R**ational agents, and did not consider the full BAR spectrum.

To restore honest behavior as a stable BARNE, we consider two simple amendments to the classical QBCP. In particular, we show that (1) applying monetary penalties for observed deviations from the protocol restores honest verification as a locally stable BARNE, and (2) injecting trap errors *à la* [15, 25] in addition to the penalty, results in the prescribed protocol becoming a globally stable BARNE. In the sense of BARNE, the amendments solve the Verifier's dilemma, including both one-shot and repeated setting.

Discussing BAR-strong Equilibrium. Since [22] established that a punishing strategy (which our first amendment is) and a trap strategy (second amendment) are necessary conditions for the existence of a BAR-strong equilibrium; it seems natural to wonder whether they were sufficient in this case. The answer is dependent on the design of the first amendment: if at least part of the fine is paid to the accuser, then the honest strategy is a BAR-strong equilibrium in the stage game. However, contrary to BARNE, BAR-strong equilibria do not automatically hold as static equilibria in the repeated game. Indeed, the coalition can take revenge on a betrayer by denying them to join back, in doing so they deny them the benefits of the coalition, making the betrayal non-profitable in the long term.

The repeated game. Supposing the selfish agents have a discount factor we could look at what happens in the base protocol if they try to regulate each others instead of using our amendments. By playing a public perfect equilibrium with a punishing strategy, such as Tit-for-Tat, where deviations lead to punishment by blocking the chain for several blocks making free-riders lose rewards. This would technically emulate the first amendment by fining invalid endorsements. However, the issue is that Byzantines could then trigger the punishments of the rationals by acting as if they were ill behaved rationals. This makes such collaborative path unsustainable. Therefore, symmetric public repeated game equilibria of the three protocols are the sequences of stationary BARNE equilibria such as alternating between two static BARNEs. Thus, our main conclusions remain unaltered: the unique globally stable BARNE in the repeated game of the standard protocol is free-riding, while it is the honest behavior once amended.

[2] QBCPs such as Tendermint, Tenderbake, and Hotstuff use adaptations of pBFT [8], a prominent solution to the Byzantine consensus problem [14]. QBCPs have the advantage of deterministic block finality, see [18].

[3] The Verifier's dilemma is therefore a case of the free-rider problem studied extensively in economics, where it is known to cause a collapse in public good provision, see [12], [21]. See [10] for an excellent informal account of the Verifier's dilemma.

[4] When we say that honest behaviour is a BARNE we refer to the behaviour of rational (selfish) agents only.

REFERENCES

[1] Ittai Abraham, Lorenzo Alvisi, and Joseph Y. Halpern. 2011. Distributed Computing Meets Game Theory: Combining Insights from Two Fields. *SIGACT News* 42, 2 (2011), 69–76.

[2] Ittai Abraham, Dahlia Malkhi, Kartik Nayak, Ling Ren, and Alexander Spiegelman. 2016. Solidus: An Incentive-compatible Cryptocurrency Based on Permissionless Byzantine Consensus. *CoRR* abs/1612.02916 (2016).

[3] Amitanand S. Aiyer, Lorenzo Alvisi, Allen Clement, Mike Dahlin, Jean-Philippe Martin, and Carl Porth. 2005. BAR Fault Tolerance for Cooperative Services. In *Proceedings of the Twentieth ACM Symposium on Operating Systems Principles* (Brighton, United Kingdom) *(SOSP '05)*. Association for Computing Machinery, New York, NY, USA, 45–58. https://doi.org/10.1145/1095810.1095816

[4] Yackolley Amoussou-Guenou, Bruno Biais, Maria Potop-Butucaru, and Sara Tucci-Piergiovanni. 2020. Rational vs Byzantine Players in Consensus-Based Blockchains. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems* (Auckland, New Zealand) *(AAMAS '20)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 43–51.

[5] Lacramioara Astefanoaei, Pierre Chambart, Antonella Del Pozzo, Edward Tate, Sara Tucci Piergiovanni, and Eugen Zalinescu. 2020. Tenderbake - Classical BFT Style Consensus for Public Blockchains. arXiv:2001.11965 https://arxiv.org/abs/2001.11965

[6] Michel Balinski and Rida Laraki. 2011. *Majority Judgment: Measuring, Ranking, and Electing*. The MIT Press, Cambridge, MA, USA. https://doi.org/10.7551/mitpress/9780262015134.001.0001

[7] Ethan Buchman, Jae Kwon, and Zarko Milosevic. 2018. The latest gossip on BFT consensus. *CoRR* abs/1807.04938, "" (2018), "".

[8] Miguel Castro and Barbara Liskov. 1999. Practical Byzantine Fault Tolerance. In *Proceedings of the Third Symposium on Operating Systems Design and Implementation* (New Orleans, Louisiana, USA) *(OSDI '99)*. USENIX Association, USA, 173–186.

[9] Ittay Eyal and Emin Gün Sirer. 2018. Majority is Not Enough: Bitcoin Mining is Vulnerable. *Commun. ACM* 61, 7 (2018), 95–102.

[10] Edward Felten. 2019. The Cheater Checking Problem: Why the Verifier's Dilemma is Harder Than You Think. https://medium.com/offchainlabs/the-cheater-checking-problem-why-the-verifiers-dilemma-is-harder-than-you-think-9c7156505ca1 (Accessed on 10/01/2022).

[11] Hanna Halaburda, Zhiguo He, and Jiasun Li. 2021. *An Economic Model of Consensus on Distributed Ledgers*. Working Paper 29515. National Bureau of Economic Research. https://doi.org/10.3386/w29515

[12] Garrett Hardin. 1968. The tragedy of the commons. *Science* 162, 3859 (1968), 1243–1248.

[13] Harry Kalodner, Steven Goldfeder, Xiaoqi Chen, S. Matthew Weinberg, and Edward W. Felten. 2018. Arbitrum: Scalable, private smart contracts. In *27th USENIX Security Symposium (USENIX Security 18)*. USENIX Association, Baltimore, MD, 1353–1370. https://www.usenix.org/conference/usenixsecurity18/presentation/kalodner

[14] Leslie Lamport, Robert Shostak, and Marshall Pease. 1982. The Byzantine Generals Problem. *ACM Trans. Program. Lang. Syst.* 4, 3 (1982), 382–401.

[15] Loi Luu, Jason Teutsch, Raghav Kulkarni, and Prateek Saxena. 2015. Demystifying Incentives in the Consensus Computer. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (Denver, Colorado, USA) *(CCS '15)*. Association for Computing Machinery, New York, NY, USA, 706–719. https://doi.org/10.1145/2810103.2813659

[16] Mohammad Hossein Manshaei, Murtuza Jadliwala, Anindya Maiti, and Mahdi Fooladgar. 2018. A Game-Theoretic Analysis of Shard-Based Permissionless Blockchains. *IEEE Access* 6 (2018), 78100–78112.

[17] J.F. Nash. 1951. Non-cooperative Games. *Annals of Mathematics* 54, 2 (1951), 286–295.

[18] Pontem Network. 2022. A detailed guide to blockchain speed | TPS vs. time to finality | Solana, Aptos, Fantom & Avalanche compared — which chain has sub-second finality? https://pontem.medium.com/a-detailed-guide-to-blockchain-speed-tps-vs-80c1d52402d0. (Accessed on 05/03/2023).

[19] Michael Neuder, Daniel J. Moroz, Rithvik Rao, and David C. Parkes. 2019. Selfish Behavior in the Tezos Proof-of-Stake Protocol. *CoRR* abs/1912.02954 (2019).

[20] Michael Neuder, Daniel J. Moroz, Rithvik Rao, and David C. Parkes. 2020. Defending Against Malicious Reorgs in Tezos Proof-of-Stake. In *Proceedings of the 2nd ACM Conference on Advances in Financial Technologies* (New York, NY, USA) *(AFT '20)*. Association for Computing Machinery, New York, NY, USA, 46–58. https://doi.org/10.1145/3419614.3423265

[21] Elinor Ostrom. 2016. *Tragedy of the Commons*. Palgrave Macmillan UK, London, 1–5.

[22] Alejandro Ranchal-Pedrosa and Vincent Gramoli. 2022. TRAP: The Bait of Rational Players to Solve Byzantine Consensus. arXiv:2105.04357 [cs.DC]

[23] Maxime Reynouard, Rida Laraki, and Olga Gorelkina. 2024. BAR Nash Equilibrium and Application to Blockchain Design. (Jan. 2024). https://hal.science/hal-04424991 working paper or preprint.

[24] Ayelet Sapirshtein, Yonatan Sompolinsky, and Aviv Zohar. 2017. Optimal Selfish Mining Strategies in Bitcoin. In *Financial Cryptography and Data Security*, Jens Grossklags and Bart Preneel (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 515–532.

[25] Jason Teutsch and Christian Reitwießner. 2019. A scalable verification solution for blockchains. *CoRR* abs/1908.04756, "" (2019), "".

[26] Aviv Yaish, Gilad Stern, and Aviv Zohar. 2022. Uncle Maker: (Time)Stamping Out The Competition in Ethereum. Cryptology ePrint Archive. Retrieved January 8, 2023 from eprint.iacr.org/2022/1020

[27] Maofan Yin, Dahlia Malkhi, Michael K. Reiter, Guy Golan Gueta, and Ittai Abraham. 2019. HotStuff: BFT Consensus with Linearity and Responsiveness. In *Proceedings of the 2019 ACM Symposium on Principles of Distributed Computing* (Toronto ON, Canada) *(PODC '19)*. Association for Computing Machinery, New York, NY, USA, 347–356. https://doi.org/10.1145/3293611.3331591