Technology Feasibility Research

Proof of Humanity (PoH) has been an ongoing field of research for over 20 years. This concept began to materialize with the introduction of one of its earliest forms: the CAPTCHA.

A CAPTCHA is a program that generates and grades tests which most humans can pass, but current computer programs cannot (von Ahn et al., 2003). It leverages the power of computational complexity classes to ensure that a sufficiently advanced AI cannot thwart PoH test. These types of systems have evolved in various forms but still rely on some form of a test that most humans can pass but is very challenging for a computer program to complete.

There are various forms of CAPTCHAs: text-based, image-based, audio-based, video-based, puzzle-based, and motion-based (Hasan, W.K., 2016). Each comes with its own advantages and disadvantages. Some of the most commonly found on the web in the past decade are the text and image versions, although newer forms of puzzle-based versions are arriving. These methods do work in practice but are bypassable with sufficiently advanced AI.

Some key takeaways from CAPTCHAs include the following: humans prefer to spend about 3 to 40 seconds solving a CAPTCHA; younger individuals typically solve CAPTCHAs faster than older individuals; users often favor click-to-pass, slide-to-pass, or rotate-to-pass CAPTCHAs; and CAPTCHAs are only about 80% accurate (Searles et al., 2023).

The next tier of Proof of Humanity (PoH) moves beyond automation via CAPTCHA to include human interaction, regulatory compliance, and on-chain solutions. Companies like Fractal ID and Proof of Humanity now offer advanced methods for PoH, which range from human certification and uniqueness to KYC/AML procedures, and verifiable digital credentials. These newer PoH systems typically rely on a combination of CAPTCHA, KYC (Know Your Customer)/AML (Anti-Money Laundering), and webs of trust. This combination should be very difficult to falsify and bypass but does require some form of centralization and potentially compromises privacy.

Most of the newer systems are designed either to upgrade existing systems like OAuth or to extend validation in blockchains such as Ethereum. As of now, there is no Proof of Humanity (PoH) system specifically designed for the Cardano ecosystem. The solution we seek must integrate seamlessly with Cardano and utilize advanced forms of CAPTCHAs. These two criteria will ensure that the project's primary goal is upheld, most notably: preserving user privacy. Forcing users to undergo KYC compromises privacy and introduces an inherent bias regarding who can obtain PoH proofs, as not everyone is able to pass KYC procedures. Since Cardano is designed to be decentralized and open to everyone, our PoH system should reflect those principles.

Cardano Integration

A Cardano integration can mean many things, depending on the use case. When building on Cardano, the question isn't typically whether the problem can be solved; rather, it becomes a matter of how decentralized—and thus trustless—the desired solution should be.

The position on the spectrum of decentralization will determine the nature of the integration required for a proof of humanity to exist on Cardano. Consequently, this could lead to the coexistence of various PoH systems, each with its own level of decentralization and mechanisms for proving human identity.

The Proof of Humanity (PoH) system could operate entirely off-chain, utilizing centralized services to issue certificates of humanity. In this model, the only on-chain element would be the ownership or reference to an NFT on the Cardano blockchain. This solution can serve as a proof of humanity but requires trust in and centralization of the NFT minting process.

Alternatively, the solution could be more integrated with the blockchain, requiring smart contract verification through cryptographic proof generation and validation methods. This Plutus-centric approach may result in higher costs due to the need for direct blockchain interaction, but it could offer enhanced security. The closer the solution is to a purely decentralized proof of humanity, the more it aligns with Cardano's ethos.

Limitations

An on-chain validation for Cardano is constrained by both space and time. The existing protocol parameters, coupled with the absence of bitwise primitives, render certain proof validation methods prohibitively costly. Consequently, this restricts the range of usable cryptographic proofs on Cardano, as any security parameters of real-world values would immediately exhaust the capacity of any transaction using the currently available methods in Plutus.

The current methods available in Plutus include`ECDSA-SECP256k1`, `Ed25519`, `SCHNORR-SECP256k1`, and boolean logic. Consequently, while proving the validity of a signature on certain data is straightforward, computing large integers for zero-knowledge proofs is unfeasible. Moreover, the boolean logic method is not only restricted by the number of logical operations due to the maximum transaction size, but also may be limited in scope or in the ability to generate comprehensive proof systems usable on Cardano.

A significant constraint arises with on-chain data availability and the tradability of native assets on the Cardano blockchain. A proof of humanity must be non-tradable, necessitating its residence within a smart contract. While the perpetual locking of an asset is trivial from a development standpoint, it does demand that the lock contract be a hyper-structure beyond anyone's control. Should this exist within a contract, any service requiring verification would need access to the latest on-chain information. Although new technologies are continually emerging to optimize the management of specific data, making the process increasingly feasible, it still represents a considerable overhead for those wishing to utilize this system on Cardano.

Implementing Proof Generation and Validation

Getting something to prove humanity on Cardano will take some novel solutions in its current state. Due to the nature of proof of humanity systems, a Cardano PoH will need to be interacting, signature base, or boolean logic based. Potentially, new primatives can be created that are time-based and one-time use thus allowing the proof system to use more advance techniques but at the sacrafice of security within some given time window. The general flow leads us towards an advanced CAPTCHA system that can be mapped directly into a proof that can be verified on-chain in an interacting way.

For instance, a user seeking to acquire a 'token of humanity'—an NFT serving as verified proof of humanity—would need to initiate the contract for a CAPTCHA state, then calculate the proof and submit it for verification within a specified time frame. A potential source of centralization lies in the state prompt provided to the user, which would be the CAPTCHA manifested as a UTxO with a unique datum. After solving the CAPTCHA, the user must send their verification through a smart contract transaction. If the proof contained within the redeemer of the verification request is validated within the allotted time for the CAPTCHA, the user is acknowledged as human; otherwise, the funds in the certification UTxO are forfeited.

This method integrates various elements of other systems, including time-based, one-time-use, CAPTCHAs, monetary stake, and the verification culminating in the minting of the token of humanity. Simultaneously, the system maintains privacy owing to the decentralized nature of the validation.

Summary

The concept of PoH has evolved significantly over time, from simple CAPTCHA systems to more complex, multi-faceted solutions that can include human interaction and on-chain technologies. For Cardano, a blockchain known for its decentralized ethos, the ideal PoH system would leverage advanced forms of CAPTCHAs and maintain user privacy without relying on KYC processes.

Cardano's architecture supports various cryptographic methods suitable for on-chain validation, but faces limitations in terms of transaction size and computational resources. Despite these constraints, it is possible to create non-tradable proofs of humanity within smart contracts on Cardano. These proofs would be represented by NFTs.

The envisioned PoH system on Cardano would be Plutus-centric, utilizing existing cryptographic primitives. It would probably have to involve a user interacting with a smart contract to solve a CAPTCHA within a certain timeframe, with successful completion resulting in the minting of a “token of humanity” NFT, while failure would mean the forfeiture of funds.

This method incorporates time-based and one-time-use elements, a monetary stake, and preserves privacy through decentralized validation. Although there are challenges, such as ensuring the system is tamper-proof and managing on-chain data requirements, the potential for PoH on Cardano is bolstered by ongoing technological advances that optimize data management and reduce the overhead for users.

References

1. Luis von Ahn, Manuel Blum, Nicholas J. Hopper, and John Langford (2003). *CAPTCHA: Using Hard AI Problems for Security.* E. Biham (Ed.): EUROCRYPT 2003, LNCS 2656, pp. 294–311, 2003

2. Hasan, W.K. (2016). *A Survey of Current Research on CAPTCHA*. International Journal of Computer Science & Engineering Survey, 7, 1-21.

3. Searles, Andrew & Nakatsuka, Yoshimichi & Ozturk, Ercan & Paverd, Andrew & Tsudik, Gene & Enkoji, Ai. (2023). *An Empirical Study & Evaluation of Modern CAPTCHAs.*