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 that 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 the test. These types of proof-of-humanity systems have evolved in various forms but still rely on some form of test that most humans can pass, which is very challenging for a computer program to do so.

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 they are passable 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 it does require some form of centralization and could potentially compromise privacy.

Most of the newer systems are designed either to upgrade existing systems like OAuth or to extend validation on 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 within 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. These integrations will range from controlled structures to pure hyperstructures. Structure in this context implies how much has to exist in the off-chain world for the application to exist on Cardano. A pure hyperstructure only requires Cardano to exist, while a controlled structure requires centralized infrastructure outside the blockchain to exist for the application to exist. Consequently, this could lead to the coexistence of various PoH systems, each with its own level of decentralization and mechanisms for proving humanity.

The proof-of-humanity 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 minting, ownership, or reference of the NFT on the Cardano blockchain. This solution can serve as a proof of humanity but requires centralization for the NFT minting process, as this would be the on-chain representation of the certificate of proof of humanity. Typically, controlled structure solutions are easier to implement and work well with already existing third-party PoH applications and protocols.

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 user costs and longer development time due to the need for direct blockchain interaction, but it could offer enhanced security and unparalleled customization. The closer the solution is to a purely decentralized proof of humanity, the more it aligns with Cardano's ethos. Though it may be difficult to create a hyperstructured PoH system, any proof of humanity needs to be updated to maintain the requirement that any sufficiently advanced AI cannot thwart the smart contract test. In turn, the solution will require some mediation, either in the form of a DAO or a federated party, to upgrade the tests used in the smart contract for the next generation of AI.

In practice, most solutions end up being some combination of off-chain and on-chain that meets the use case, resulting in a quasi-hyperstructure solution. The application could exist on its own as a program on the blockchain that anyone can use, but there does exist some group that handles any required maintenance to maintain the PoH requirements.

Limitations

Space and time constraints limit Cardano's on-chain validation. The existing protocol parameters, coupled with the absence of bitwise primitives, render certain proof validation methods prohibitively costly. This restricts the range of usable cryptographic proofs on Cardano, as any security parameters of real-world primes would immediately exhaust the capacity of any transaction using the currently available methods in Plutus. There are potentially more cryptographic primitives coming to Cardano in the future with Plutus V3, but the release date is too far into the future to be useful in this project.

The current methods available in Plutus include `ECDSA-SECP256k1`, `Ed25519`, `SCHNORR-SECP256k1`, arithmetic circuits, and customized transaction logic. This limited set of methods makes proving the validity of a signature on data trivial but makes computing large integers for zero-knowledge proofs unfeasible. Novel primitives like 3-colorings and bulletproofs can be implemented at the cost of development time and the risk of incorrect implementation. The default primitives of a PoH system may limit us, but we retain the ability to implement custom validation methods to meet the use case's demands.

Some significant constraints arise from on-chain randomness, on-chain data availability, and the tradability of native assets. A proof-of-humanity test must be generated from true randomness, forcing a trusted oracle to consistently update its random data. It must be non-tradable, necessitating its residence within a permanent lock smart contract. And it must be reference-able, requiring real-time on-chain data. While the perpetual locking of an asset is trivial from a development standpoint, it does demand that the permanent lock contract be a controlled or quasi-hyperstructure, as the ability to remove or void certificates of humanity may be required. Should this token exist within a contract, then any service requiring verification would need access to the latest on-chain information. Although new technologies are continually emerging to optimize the management of this specific data, making the process increasingly feasible, it still represents considerable overhead and centralization 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 PoH systems, a Cardano PoH will need to be an interacting solution in some form. A potential human will need to prompt the PoH system to start the verification process; they will need to receive the test, solve the test, and then submit the test for a result. The outcome of the test is the validity of the submitted transaction. The validity of the transaction will be based on existing primitives accessible in Plutus and new primitives that are time-based, one-time use, and integrate directly with the test itself. This allows the PoH-proof system to be comprehensive without sacrificing security. The general flow leads us towards an advanced CAPTCHA system that can be

mapped directly into a cryptographic 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 a verified proof of humanity—would need to initiate a certification UTxO for a random CAPTCHA state. The user will solve the CAPTCHA, generate the proof, and submit the transaction for verification within a specified time frame. The time frame is long enough that valid transactions will land on the chain, but it is short enough to prevent brute-forcing the solution to the CAPTCHA. If the transaction contains a valid proof within the allotted time for the given CAPTCHA, then 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 stakes, and 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 centralized 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. Smart contracts will store NFTs that represent these humanity proofs.

The envisioned ideal PoH system on Cardano would be Plutus-centric, utilizing novel and existing cryptographic primitives. The system will be 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. Technological improvements are making PoH on Cardano better by managing data more efficiently and making things easier for users, even though it's challenging to make the system secure and handle the data on the blockchain.

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.*