AN APPROACH TO DEPERSONALIZED IDENTITY, DECENTRALIZED PERSONAL DATA SHARING, AND PASSWORD MANAGEMENT WITH INTEGRATED CRYPTOCURRENCY WALLETS UTILIZING HIERARCHICAL DETERMINISTIC TREES AND ELLIPTIC CURVE CRYPTOSYSTEMS

Bryce Weiner
bryce@tao.network

Abstract

This identity ecosystem focuses on maintaining confidentiality using a cryptographic tree- based solution for key and password management. The advent of cryptocurrencies and the increased proliferation of PKE cryptosystems have increased the burden on consumers to manage secure data. In this paper we propose a novel solution for password and key management that incorporates cryptocurrency key pairs, providing the lowest possible number of potential attack vectors while maintaining consumer privacy. Presented is an infrastructure for secure, encrypted, sharing of user data which remains in the control of the issuer and, when combined with a unique implementation of depersonalization technology through hierarchical deterministic trees, provides the smallest possible profile in regard to cryptocurrency transaction data mining. This enables not only a level of increased privacy and security for the consumer and does so in a machine-readable context for the Internet of Things and smart contract applications. Traditional identity systems revolve around the certification and administration of credentials through a trusted third party, be that a government or an educational or financial institution. A deterministic identity system no longer requires the constant verification and re-certification of the authenticity of an individual's identity and achieves this through the application of modern cryptographic techniques.

What is identity?

A most common definition of identity is "the fact of being who or what a person or thing is". The establishment of such fact requires the authentication of any person or thing is indeed that which is being presented. The establishment of those facts is most often reserved to the province of government, however in an increasingly connected world governments may come and go and identities come and go with them. A consistent, stable identity solution capable of lasting a lifetime in a global context must be portable, secure, resistant to forgery, require only occasional internet connectivity, and an overall minimal amount of technology.

How is identity determined?

The relevance of an individual's identity is dependent upon the context in which that identity is applied. In the most basic context of interpersonal communication, the identity of an individual is determined based on visual cues, such as a photograph or a memory, and those personal affectations relating directly to the personality of the individual in question. Within the context of governance, a serialized identifier such as a Social Security Number or drivers' license

number is used to identify an individual and pertinent data is stored in a central repository. Such governmental applications of identity revolve around the execution of the functions of governance, such as the tracking of births and deaths for the purposes of governmental representation, licensure, or taxation. These systems have become a ubiquitous part of existence for those individuals who reside within a specific jurisdiction, but solutions must be harmonized to accommodate the interaction of individuals across jurisdictions. In many parts of the world, such information as births and deaths are still largely unrecorded, which leads to an imperfect system that creates significant barriers to participation in the development and consumption of goods and services. Future-thinking identity solutions should be easily implemented on the personal level without the necessity for government intervention.

Authentic Identity Through Cryptographic Determinism

Cryptographic protocols provide a means by which an individual in any location, with any level of technological sophistication, may generate and maintain a verifiable, authentic identity by simply following a set of mathematical rules. The term "cryptographic determinism" relates to just such a portability of implementation in the sense that an individual is *mathematically provable* to be as unique as the individual and may be reproduced and verified through the application of a specific protocol of mathematical rules. Such an identity may then address the issues such as governance and authenticity, as well as provide implementations simple enough to address the needs and requirements of low-tech and no-tech environments.

A Lifetime of Flexibility

Identity solutions based upon cryptographic proofs have the ability to not only be simply implemented, but may be used to provide the privacy, security, and regulatory compliance regardless of jurisdiction. A virtually unlimited number of identities could be generated from a single string of "identity DNA" which is unique to the individual. At no point need the individual share data between institutions or entities unless the individual so desires. This provides not only a level of security that is capable of being managed by the individual, but also allows for an individual to re-present themselves to a variety of institutions for the consumption of goods and services without the danger of a breech of confidential personal information. This may be achieved through the segregation of the provision of personal data through different channels, and those channels bear no association unless the individual so desires. Traditional identity solutions require that the individual be permanently associated with the full and complete history of the use of their identity, limited only by legislation. This is trust-based model which is very easily breeched and without the ability for appeal by the individual. An identity based on cryptographic determinism allows for an individual to make mistakes and "reinvent" themselves, while not adversely impacting the ability for commercial entities and institutions to apply risk assessments in regard to interacting with that individual.

Integration at Any Level

A properly designed identity protocol based on cryptographic determinism provides for the ability to become the underlying layer of identity verification and authentication for *all existing identity systems*. This provides the ability for the resulting complexity to be the ***choice of the***

individual, and not rely upon the permission of potentially corrupt entities or institutions. This freedom of complexity provides for a level of privacy, security, and personal sovereignty over identity information that is truly revolutionary. The lack of any central issuing authority allows for the maximum amount of freedom and flexibility in global implementation. Adherence to the protocol becomes the only rule in regard to the generation, provision, and utilization of identity.

Examples of Deterministic Seeds

For each application, a unique "key specification" ("keyspec") is provided. The keyspec is used to traverse the merkle tree to mathematically generate the PKE pair from the deterministic seed. The seed itself may be created from one of several different methods, described below. In each method, the keyspec and cryptographic signature resulting from the selected PKE pair is provided to guarantee authenticity of identity of the individual.

IMEI-based Identity

The simplest integration possible is utilizing the IMEI number of a wireless device as the deterministic seed for an identity. A user supplied password, or nonce, is required to prevent wireless carriers from interfering with usability. The IMEI and nonce are then combined to create a SPHINCS-256 pre-image (hash value) which forms the "seed" of the deterministic identity.

Biometric-based Identity

An individual's fingerprint or iris pattern may be scanned and rendered into a unique cryptographic pre-image, precluding the necessity for a password nonce. The resulting value may then be utilized as a deterministic seed.

Provisioned Serial Number

A serial number provisioned to an individual by a government, NGO, or other entity may be combined with a password nonce to generate a secured and appropriate value for deterministic identity in the same manner as with an IMEI.

Virtual Currency Wallets

Virtual currency wallet addresses may also be utilized to generate deterministic identities, as the primary operation of a VC wallet is as a PKE pair.

Examples of Use

Mobile Banking

As a user's identity seed may be generated from the IMEI of any given mobile device, each service utilized may have its own keyspec. To interface with a banking application, the application would read the IMEI and prompt the individual for their password. Once provided the deterministic identity is generated from a keyspec stored on the device. Best practices

would be to never store the password nonce on the device and only generate the deterministic PKE pair in memory, an action for which many mobile devices are well suited.

Emergency Assistance

A biometric-based identity could be used to assist in the provision of services to individuals displaced by war or natural calamity, as only the keyspec for the services rendered need be stored and could be easily memorized by the individual.

Accessing Government Services

A deterministic identity provides a means by which a government serial number, such as a social security number, could be utilized as a secure means of accessing government services. An individual would be provided a serial number for which they would then generate a password, keyspec, and cryptographic signature, of which only the keyspec and signature is submitted to the government for later authentication of identity.

A Working System

Each month seems to bring reports of a data breach of consumer information from a commercial source. While certain cryptocurrency and identity systems offer potential technological solutions, no implementation has been offered which employs these technologies with the experience of “Big Data” levels of behavioral and biostatistical analysis. Combining lessons learned in the healthcare IT field, where well-defined standards are set for biostatistical depersonalization, and the latest developments in elliptic curve cryptography and DSA technology, a novel solution can be derived which not only provides the security and privacy consumers demand but also leaves the consumer in sole control of which entities have access to specific pieces of their personal information. From this solution a truly cryptographically unique, depersonalized identity can be derived from which proof of ownership may be provided on demand.

While not yet in the public consciousness, the increased proliferation of blockchain technology makes such considerations a real and present privacy concern. The amount of data required to utilize cryptocurrency-to-fiat conversion services for the purposes of regulatory compliance is non-trivial, further complicated by the direct linkage to what is assumed to be a “pseudo-anonymous” settlement network. This represents a significant departure from the intended behavior of permissionless cryptocurrency networks and represents a potential threat to the privacy of future counterparties that does not exist in the majority of current fiat settlements. This methodology represents several fundamental changes in the way one considers how commercial entities access their personal information, and does so by providing a means to defeat very effective and popular biostatisitical analysis techniques that have been adapted for behavioral analysis in the Internet age, such as non-parametric inferential statistical analysis and propensity score matching. [1] Such analytical techniques may then be applied as an arbitrary means to discriminate financial services to individuals, limited only by regulatory compliance (if such regulation exists in a given jurisdiction).[2]

Furthermore, a successful solution should seek not to place further burden on, or attempt to fundamentally alter, established consumer and commercial behavior. A number of popular

consumer-level security products have established a baseline for consumer expectations and the proposed methodology has been offered with such considerations in mind, such as cross-platform development and browser extension and plugin capability, and does so without creating additional vulnerabilities.

What has emerged is a future-proof and extremely straight forward solution which addresses the security and privacy concerns of both consumers and commercial entities. When one considers the scale of the Target data breach [3] with the publicly available data of the Bitcoin blockchain, and that Bitcoin conversion giant Coinbase at the time of this writing report a collective user base of over 2,900,000 individuals, [4] the problem is quite significant.

Related Work

Depersonalized identity (DI) draws its primary utility from the hierarchical deterministic traversable tree for the generation of public/private key pairs along an elliptic curve as described in Bitcoin Improvement Proposal #32. [5] Though this solution provides a convenient means for traversing public/private key pairs, it is not itself a sufficient identity solution. The application of the same process towards creating a structure for a provable identity suitable for commercial applications requires a protocol for negotiating key specifications between entities in various scenarios and a means for cryptographically secure data sharing. It has, however, demonstrated overall soundness of such a scheme in public code review and numerous successful commercial applications.

Provided a sufficient protocol for key specification exchange is established, the latter concerns are addressed in a means similar to BitTorrent or OpenBazaar: employing a distributed hash table (DHT) in a decentralized, peer-to-peer network for trustless, cloud-based storage of encrypted data. Through the unique implementation of the Kamedlia DHT by the Entangled Python project, information may be deleted from the cloud by network consensus upon request of the issuing authority. Further, a consumable distributed tuple space (DTS) is employed as a peer-to-peer encrypted messaging queue for DHT data operations.

Design

The assumptions and requirements are stated firstly, and in doing so is proposed a NOVEL DEPERSONALIZED IDENTITY SCHEME involving establishing a provably unique, analysis resistant, cryptographic identity and the provisioning of that identity into four distinct applications: personal data sharing, security authentication, website-specific identity management, and cryptocurrency wallets.

Definitions

Identity - A 128-bit string of system-generated entropy. [6] This entropy is used as the mathematical seed for the derivation of a hierarchical deterministic tree (HDT). To protect the HDT ecosystem, a passphrase is required for AES encryption of the entropy bits. 50,000 rounds of SHA-256 hashing are applied to the unencrypted entropy and the resulting first 6 bytes are used as a fingerprint.

Key Specification (“keyspec”) - The map to a particular public/private key. Within this system each keyspec consists of exactly six (6) branches, each branch selected as a random integer value between 0x000000 and 0xFFFFFF providing a provably unique deterministic public/private key pair for each required operation, a suitable universe for one-time use applications, and potential recovery of data if only an entropy is present.

Data Providing Entity (DPE) – The entity provisioning data for consumption, identified by public key.
Intended Consuming Entity (ICE) – Each data provision has an intended entity as the recipient, identified by public key.
Transaction Identity – A keyspec negotiated between the DPE and ICE. The DPE suggests a keyspec, the ICE accepts the keyspec if it is not already in use. If it is in use, the ICE suggests an alternative. The DPE may then accept the suggested keyspec, or reject it with their own suggestion. This process is repeated until a suitable keyspec is found.
Transaction – The provision or revocation of a single point of data by a DPE.
Transaction Set – A collection of transactions between a DPE and ICE. Each transaction set represents the interaction between one individual and one service provider.

Identity specialization

As each identity is itself a self-contained ecosystem of personal information, each identity may be employed within a specific use case to provide increased resistance to data analysis.
As such, login and transaction information associated with fiat conversions of cryptocurrencies can be separated from economic transactions to purchase goods and services with a single intermediate step (discussed later).

Per-Transaction Set Addressing and Encryption

Each transaction set involving a DPE and ICE requires per-set addressing for identification purposes. This is achieved through an implementation of the Bitcoin Base58 address system. The address prefix is altered from 0x00 to 0x3C to depart from the protocol and prevent confusion.
As a unique keyspec is negotiated, the PKE pair associated with that keyspec is used to encrypt all data between parties.

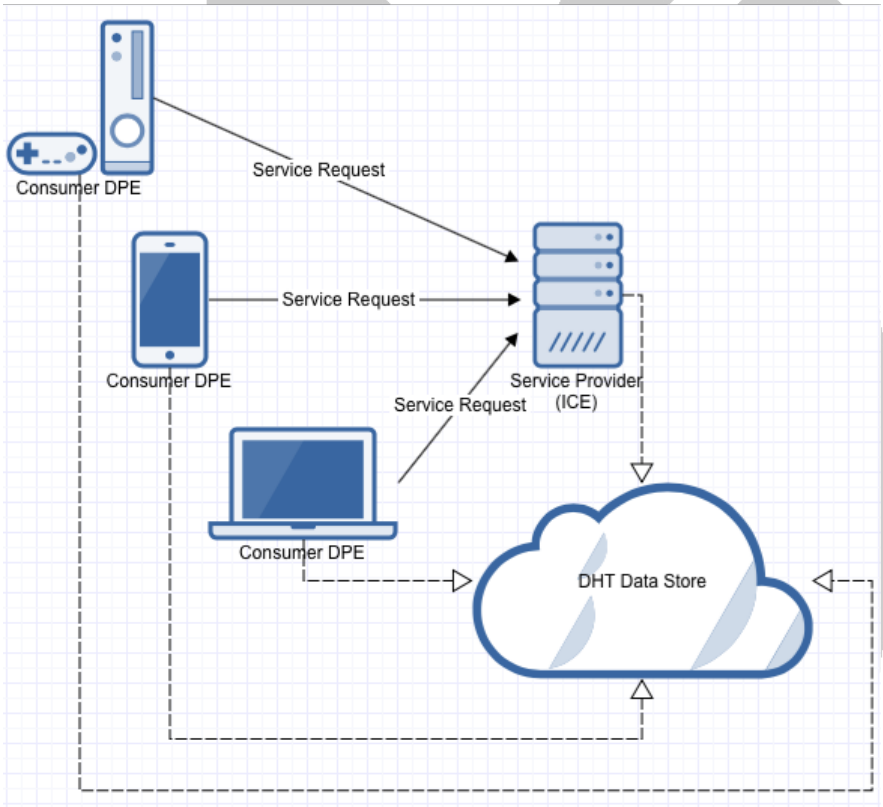
Provable data integrity

Data submitted to the DHT must be independently provable as unaltered by third parties. This is achieved through a transaction structure rendered as:
Network address derived from the public key of the ICE (Pb-ICE) for this transaction set.
Network address derived from the public key of the DPE (Pb-DPE) for this transaction set.
Binary blob of personal data encrypted with Pb-ICE
A creation timestamp
Recoverable Schnorr signature of DPE (Ss-DPE), which signs binary blob, ICE address, and timestamp.
A hash value which is a Merkle root derived from the SHA256 hash values of the previous five values
Prior to consumption, data is validated by matching the Merkle root derived from the collected data. A public key is recovered from the Schnorr signature from which a network address is generated. This address is compared to the Pb-DPE address as origin validation.
The first six bytes of the Merkle root may serve as a data fingerprint.

Provision unique contact information per website

a. A simple SMTP relay service will relay email messages sent to a Base58 network address derived from a transaction set keyspec to an address specified by the user.

5. ECDSA-based authentication a. A negotiated keyspec may then be used as ECDSA-based authentication.
A user requests a login to a website with which a keyspec was previously negotiated.
The website presents the user with a simple plain English passphrase.
The user signs the passphrase with the PKE pair indicated by the negotiated keyspec and presents the signature to the website.
The website recovers the public key, converts it to a Base58 network address, and compares it to the registration information. Message content is then validated and should both results return a positive result, the user is authenticated.
OAuth 1 and OAuth 2 authentication can then be supported, where the user is presented the opportunity to provision data on an as- needed basis and easily integrated into existing authentication systems.



6. HDT-based cryptocurrency wallets a. Each identity may employ a virtually unlimited number of cryptocurrency wallets.
A keyspec is selected as the “wallet root”.
Additional addresses are generated on demand from the wallet root from its own HDT.
Only the keyspec is stored on disk.
Cryptocurrency wallets can be exported in WIF format for import into other clients.

Implementation

The implementation of the described platform independent network topography is presented in Figure 1. The DHT data store is maintained by a peer-to-peer network consisting of ICE service providers.

Conclusion

As blockchain and IoT technology evolves, the need for tools which provide a consistent level of information security and privacy become paramount. The flexibility of the cryptosystems employed in blockchain technology can be effectively repurposed to provide for entirely new paradigms of personal data privacy.

References

[1] http://www.econ.ucla.edu/people/papers/Matzkin/Matzkin616.pdf [2] http://cointelegraph.com/news/113207/coinbase-is-tracking-how-users-spend-their-bitcoins
[3] http://www.forbes.com/sites/paularosenblum/2014/01/17/the-target-data-breach-is-becoming-a-nightmare/
[4] https://www.coinbase.com/about
[5] https://github.com/bitcoin/bips/blob/master/bip-0032.mediawiki [6] https://tools.ietf.org/html/rfc4086