Blockchain: the Emperor's New PKI?

Hilarie Orman
Purple Streak

I would like to jump on the blockchain bandwagon. I would like to be able to say that blockchain is the solution to the longstanding problem of secure identity on the Internet. I would like to say that everyone in the world will soon have a digital identity. Put yourself on the blockchain and never again ask yourself, "Who am I?" – you are your blockchain address.

> *"All the world's a blockchain,*
> *And all the men and women merely entries in it."*

Certainly, it is important to solve the online identity problem. Identity can be destiny. Today, the mere fact of a person's physical existence is not enough to guarantee some rights. Citizenship, refugee status, right to travel, government benefits, etc., are tied to a provable identity of the right type. Some have gone so far as to say that identity is a human right.[1] Is it feasible to solve this problem through digital technology? And if so, is blockchain part of the solution? And do we get privacy with that?

The digital divide leaves first-world citizens with the problem of having too many online identities to manage, while those on the other side of the chasm, some two billion people, perhaps, not only have no digital identity, but further have no documented identity at all. Everyone in the world needs at least one persistent and secure identity that is easily manageable by the individual.

Public key technology has always held the elusive promise of a universal digital identity system that makes paper obsolete, but attempts to do this have never measured up. Fraught with complex management, lookup, and trust issues, public key has been largely relegated to a way to protect Internet communication channels. Although some countries use digital identification via public key and electronic ID cards, there is no widespread movement towards this kind of solution.

Blockchain technology seems like an odd savior, especially because its main use to date has been for digital cash with anonymity features. A verified identity is nearly the opposite of anonymity, so what does blockchain have to do with it? And how can we have verified identities that preserve privacy? The technology for doing this is in its infancy, and the management problems are largely unexplored, but when has that stopped technology ebullience? Let's look at what identity means and how life might be if lived in the blockchain bubble.

Published by the IEEE Computer Society
1089-7801/18/$33.00 USD

WHITHER COMEST (OR GOEST) IDENTITY?

Identity is a psychological, philosophical, and legal concept that is carried over into the digital world in a variety of ways. For an interesting discussion of desirable principles in the digital realm, see the work of Kim Cameron.[2]

The simple story is that the various authorities that control the resources that we need to participate in life today need to know who we are, which means they want enough information to disambiguate all people from one another. Somehow, there must be a short index from a person into a table of data about many people, potentially all people on earth. Some authorities want to attach attributes to identities, other authorities need to examine those attributes. At the center of it all is a unique human being. In order to avoid some kinds of dystopian nightmares about authoritarians dictatorships, it seems important to let the person, the identity subject, control how and when the attributes are disclosed. For example, people need to communication their immunization status to a health organization, residence to local housing authorities, financial status to lending organizations, etc. A recent trend is to use cryptographic zero-knowledge proofs for this purpose. But this is not new. What is new is using blockchains for bootstrapping the trust process and for the persistence and availability of the information.

We all know that our identity is more than our name and appearance, and that name recognition and personal relationships are not enough to maintain one's status in an impersonal digital world. People need to document their place of origin and residence, their health status, and their economic ties if they are going to participate in the global economy and make use of the rights to which they are entitled.

Today, identity is established in a variety of ways for people and for governments and businesses and almost all organizations. On social media, identity means a username, a communication history, and interactions with some kind of online community. For email, it is simply an email address. For a website it is a URL and perhaps a certificate. For most website interactions, it is a username and a password, perhaps augmented by a cell phone number. For newer mobile devices, various kinds of biometrics identify the device owner. The essential government issued identity documents are still based on paper: driver's licenses, passports, birth certificates, etc. (I told my doctor's office to address me by my date of birth, as it seems to be essential for them to verify it multiple times in any conversation.)

We may be on the verge of a major change, one in which public keys are of central importance, users have total control over disclosure of their identity attributes, and blockchains are the root points of trust for obtaining credentials. Or, we may be on the verge of discovering that blockchains recapitulate the well-known problems of public key infrastructure.

WHAT IS A BLOCKCHAIN?

Blockchains offer the promise of a trustworthy way to record shared data. It is a publicly verifiable ledger that maintains the integrity of the individual entries.

The basic idea of a blockchain is quite simple: it is a shared, replicated log file (sometimes called a ledger). The entries are sequential and time-stamped. A one-way function produces a short bitstring (for example, 512 bits) and depends on every item and its placement in the log. The function has mathematical properties that assure that it would be astronomically difficult to produce a different log with the same output. The output of the function is an abbreviation for the log itself. To add new entries, the function uses its current value and the contents of every new entry to compute a new output. The log maintainer publishes the log and the output value so that independent parties can verify the correspondence.

In itself, a blockchain offers no particular security. To get to an interesting trust model, two additional mechanisms are required. The first is to have a verifiable definition of correctness for each log file entry. This means that there is a correctness condition that is orthogonal to the correctness of the mere sequence of the entries; usually the log file entries are signed by public keys and have a standardized format. Secondly, the Important Idea that makes blockchains useful is to dis-

tribute the maintenance of the blockchain to many (possibly distrusting) parties, thereby achieving "distributed consensus" in real time. As long as the majority of the parties have a mutual interest in publishing a consistent view of the blockchain, it is a reliable record of transactions.

BitCoin has the interesting property that any party doing a sufficient amount of work can extend the blockchain. Anyone can become part of the peer-to-peer network and get the current state of the blockchain and list of transactions that are queued for inclusion. After checking that the transactions satisfy the correctness conditions, and then computing the validation function, a party can advertise the transactions and validation output to other nodes. All the other nodes will check the computation and begin publishing the blockchain with the new block of transactions. The salient property of the BitCoin validation function is that it is very difficult to compute but very easy to check. To incentivize validators (miners), the system automatically rewards them when their computation is accepted.

WHAT IS A BLOCKCHAIN IDENTITY?

There are proposals to create blockchain systems and communities that use blockchains in which the entries are somehow related to identification for an individual person or unique entity, such as a DNS name. In this model, the public key for a person or entity is represented in a blockchain entry, where it could be accessed by a relying party. In the simplest case, an entry would be signed by an identity provider of some kind, using its own public key, to assert: "The public key represented in this data item belongs to a person known as John Q. Public." This is similar to existing public key systems, such as X.509v3 and PGP.

Blockchains offer a couple of improvements, though. One is that the blockchain replication by a diverse set of entities ensure accessibility and reliability. Another is the immutability of the blockchain and the timestamps; there can be little argument that an entry was made, and that can be shown for as long as the blockchain exists.

In order to know if blockchains solve other fundamental identity problems, we need to look more deeply into why identity has become so important. Even though we still use passwords and special purpose apps, digital identities based on public keys are the only hope of getting out of the morass of relying on companies and governments to keep their customer or citizen data private. Their inability to do so is demonstrated on a regular basis.

There are two kinds of identity being discussed today. One is for most of the people in countries with good governmental record keeping, reliable power for computing devices, and nearly universal Internet service; the other is for people who have been cut off from government recognition, isolated by virtue of abuse or persecution, who are poor, or who do not have physical safety for possession. This can happen either because the their government is too weak to provide identity services, because they are ignored by their government, because they are victims of human trafficking, because they are refugees, etc.

The digitally dispossessed cannot rely on cellphones, cards with chips, or anything physical to provide their identity. What then identifies them? One way would be to record biometric data such as iris scans, facial features, fingerprints, or DNA. A cryptographic hash of the data could be an anchor point for attaching attributes such as country of residence, immunization status, etc. The blockchain would hold the cryptographic hash, and various authorities could certify possession of attributes. The feasibility of this is uncertain because biometric attributes are "fuzzy," while a cryptographic hash is anything but. Still, one could imagine a solution based on varying degrees of accuracy and multiple measurements to create attribute collections that could map to a unique match.

In an ideal world, none of the identity attributes could be released without permission from the person being identified. But if that person is digitally dispossessed, then she has no way to present a public key to unlock the record; instead, the biometric measurements alone are enough to unlock the information. This may well be the future identification situation not only for the digitally dispossessed, but perhaps for most of the world's population, if their governments insist on it.

There are many in the tech world today who are banking on the assumption that blockchain technology will be the way register identity and to redesign the entire concept of digital identity. The identity owner could get the unfettered freedom to manage access to his personal information, to his credentials, and to his organizational records. A whole economy of identity might arise, a libertarian utopia where the person and his keys are self-determining entities.

WHO IS DOING IT?

A working example, albeit minimal, of how a blockchain can enable a user to control his personal data can be seen in the MIT Media Lab's credentials experiment.[3] This is a blockchain-based system for issuing certification of academic achievement (it is confusing to security-minded folks that they use the term "certificate" to mean "credential"). There is a central authority (The Media Lab) that manages a blockchain. This system encodes information about a student's completion of a course and the student's public key. The signed hash of the data structure is put on the blockchain, and the student receives a copy. The student can prove that he has the certificate by sending his copy of the certificate and the blockchain address to, for example, a potential employer. The employer can see that the hash of the certificate matches the blockchain entry and that it is signed by MIT. This system gives control over the information to the learner, and MIT does not need to communicate with the potential employer. Indeed, taken to the extreme, MIT would not even need to retain any of the information; it belongs to the learner, and with great autonomy comes great responsibility. The learner must not lose his public key pair, and he must not lose track of his signed course completion document.

An additional advantage is that the Media Lab does not need to grant any special access provisions to the student. Not only does the granting institution not have to retain the record, they do not have to protect any sensitive login information for the student.

Those familiar with the trials and travails of managed public key infrastructure (PKI) may note that there are similarities between this approach and traditional ones, and there are some striking shortcuts. In this approach, the learner does not present a public key to the granting institution. Although the experimenters intended to add that in the future, the reality of the situation is that only a very sophisticated user can generate a public key pair, so it is up to the granting institution to generate a key pair for the learner and to communicate it to him securely. Of course, in the future, that will all be taken care of "in the app." Neither does the learner, or anyone else, have a certificate identifying the granting institution, nor any way to determine the legitimacy of the site with the blockchain. When the learner presents the document and its associated blockchain address to a potential employer, the employer can verify that the hash of the document is in the blockchain and signed by the granting institution's key.

While this seems straightforward, one cannot but wonder what advantage it offers over using PKI certificates for the granting institution and the learner. The granting institution can give the signed copy of the course completion document to the student, and then any third party can receive the document from the student and validate the signature.

The Media Lab system does include a provision for that bugaboo of public key systems, the revocation of keys. In this case, the revocation of the course completion document can be achieved by adding an entry to that effect onto the blockchain.

Finally, the Media Lab felt that Merkle trees offered advantages over blockchains.[4] They envision a hybrid system in which users store their identity credentials on a blockchain, but the granting institutions use Merkle trees for storing their endorsed documents for users. Though minimal, this kind of system might be a solution for storing biometric data for the digitally dispossessed. There is still a great challenge to be met in providing user consent, but there is an imminent need to help the world's vulnerable people, and circumstances may overtake technology in this sector.

There are much more ambitious projects that promise users detailed control over their identifying information and credentials. IBM and SecureKey are working with Canadian partners for an identity system based on the Linux Hyperledger software core (one of many open source blockchain projects). Microsoft got on their bandwagon recently with an announcement from their

identity division.[5] Another open source blockchain project, Sovrin, is the foundation for "self-sovreign" identity technology being developed by Evernym (https://www.evernym.com).

These systems appear to promote the concept of releasing all or part of the attributes in a personal credential. The canonical example is a driver's license. People use their licenses for a variety of purposes that have nothing to do with driving: age (to get a drink), residence (to vote, to use a public library), date of birth (for health care, sometimes), appearance (to prevent identity fraud). Many people would prefer to release only the minimal necessary information to the requestor. In the last decade or so, cryptographic techniques for achieving this kind of controlled disclosure have been developed in whole or in part: secret handshakes, hidden credentials, oblivious attributes, and the general technique of homomorphic encryption. The developers of these new identity systems would like to build in the privacy protections afforded by controlled release of information.

While this seems laudable, there are many questions to be answered, not the least of which is an overall analysis of security and usability. Some of the advanced cryptographic techniques are impossibly slow, for example. Another risk is that there may be a hidden reliance on a trusted third-party underlying the cryptography. If that third-party is shown to be untrustworthy or corrupted, then the privacy of the system might be undermined in ways that are difficult to detect. The problem of generating secure public keys remains a crucial issue, one that seems to trip up hardware vendors again and again.[6]

More practical issues center on whether or not the user should be able to hide essential information like the expiration date of a driver's license or special notations about wearing glasses. And, how can the user give consent to organ donation in the event of a fatality? Working out each contingency, even for ordinary credentials, may take us some years.

From the point of view of a credential issuing or attribute using organization, there are advantages to being able to reduce the attack surface for privacy, but there are uses for collections of personal data that are at odds with privacy. Organizations need to do analyses of their user base in order to tailor their services. How many users are under the age of 25? How many live in cold climates? Will companies really be motivated to take privacy more seriously in a blockchain world?

Another concern is the complications for the user of managing a large number of secure blockchain "blobs." Many new systems tout the ability of the user to create different "identities" (keys) for different uses, but that gets complicated. Moreover, it remains to be seen if credential issuers will go along with such a model. For example, if someone uses a driver's license to get a special TSA boarding privilege, will TSA issue the credential to an arbitrary key provided by the user, or will they issue it to the same key that is bound to the driver's license?

Realistically, in the fullness of time, all of the PKI management problems will devolve to the blockchain. Companies acquiring other companies will want to re-sign all the credentials issued by subsidiaries, teenagers will want to have credentials that their parents controlled reissued to their "grown up" key, etc. It seems that these issues, and similar ones, might lead to the same kind of inconvenient X.509 certificate chains that bog down PKI; will we have blockchain certificate chains that replicate the same problems?

Public keys are essential for safe identities, but everything hinges on generating "good" keys, ones that do not have hidden pitfalls. That, in turn, depends on the quality of the software and the systems that generate them.

A further risk is that there will be no reason for institutions to stay in the game without substantive advantages. If members drop out of a blockchain, then the guarantees of availability and non-malleability become weakened. The remaining members may decide to abandon the previous blockchain, either by not maintaining it at all and moving to a new blockchain, or by forking the previous blockchain. At some point, the Wayback Machine may be littered with the charred remains of blockchains that fell out of favor.

Nonetheless, almost anyone in the identity game today probably is considering a move to blockchain for the convenience that it could bring in widespread availability and risk reduction.

THE FUTURE

An underlying current in all the blockchain identity fervor is the desire to start over with digital identity. A fresh beginning, unhampered by stodgy standards and the shackles of complex software, a new identity ecosystem is beginning to be born. I think that vision, vague as it is, might be the future of identity for the world. We could have the forced Orwellian government-issued privacy-depriving biometric credentials coexisting with subversive underground identities. free flowing reputation systems, and multiple personality disorder do-it-yourself identities for the people on the fringes.

It is clear that eventually, every human will have some representation as digital attributes. This is good and bad, like civilization itself. As humans, we seem to enjoy building bigger and bigger societies, even though that has a huge effect on how we view ourselves as individuals. It is destiny.

Based on the nonproliferation of PKI, my prediction is that when we do finally reach the point of having digital identities instead of paper documents and login passwords, then they will be built on some kind of distributed ledger, something resembling a blockchain. I only hope that it improves security, rather than making it ever more obscure and tenuous.

REFERENCES

1. *ID2020, identity for the world*, ID 2020 Initiatve, 2017; https://id2020.org.
2. K. Cameron, "The Laws of Identity," Microsoft Corporation, 2005; https://msdn.microsoft.com/en-us/library/ms996456.aspx.
3. "What we learned from designing an academic certificates system on the blockchain," MIT Media Lab, 2016; https://medium.com/mit-media-lab/what-we-learned-from-designing-an-academic-certificates-system-on-the-blockchain-34ba5874f196.
4. R.C. Merkle, "A Digital Signature Based on a Conventional Encryption Function," *Advances in Cryptology — CRYPTO '87*, Springer, 1987.
5. A. Simon, "The Future as We See It," blog, 2018; https://cloudblogs.microsoft.com/enterprisemobility/2018/02/12/decentralized-digital-identities-and-blockchain-the-future-as-we-see-it.
6. N. Heninger et al., "Mining Your Ps and Qs: Detection of Widespread Weak Keys in Network Devices," *Proc. 21st USENIX Security Symposium*, 2012; https://factorable.net/weakkeys12.extended.pdf.

ABOUT THE AUTHOR

Hilarie Orman is a security consultant and president of Purple Streak. Her research interests include applied cryptography, secure operating systems, malware identification, security through semantic computing, and personal data mining. Orman has a BS in mathematics from the Massachusetts Institute of Technology. She's a former chair of the IEEE Computer Society's Technical Committee on Security and Privacy. Contact her at hilarie@purplestreak.com.