A knightly quest for privacy

***Frank Stajano** suggests ways distributed ledger technology could be used to store personal data securely and to protect the privacy of consumers*

I have been invited to comment on whether distributed ledger technology (DLT) generates any particular issues concerning personal data. I confess that I am still leaning towards the sceptic's viewpoint on the general topic of DLT, and these are personal views.

You might have some objections to the bitcoin model; I do too, primarily on an economic basis, but don't worry, we will not be talking about that or looking at cryptocurrencies, just at the underlying technology when applied to other purposes. You might also object to the efficiency of the blockchain that bitcoin introduced, particularly its energy consumption. So do I. But don't worry, we won't be talking about that either. We will be talking about other flavours of blockchain that do not rely on "proof of work".

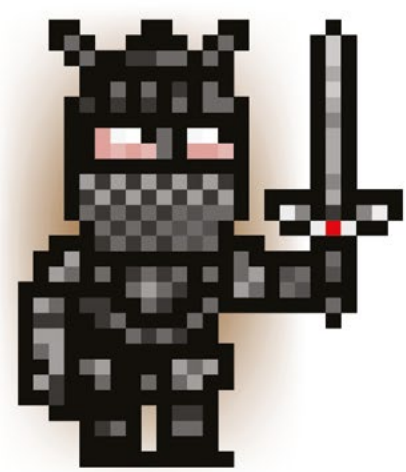
The technology we will be looking at, generically labelled as "distributed ledger technology", without saying anything concrete on its implementation or application domain, is some kind of distributed database with two defining properties. One is high availability, meaning it is difficult for an adversary to deny you access because there are many synchronised copies around. The other is tamper-evidence, meaning it is difficult for an adversary to make unauthorised changes without being detected.

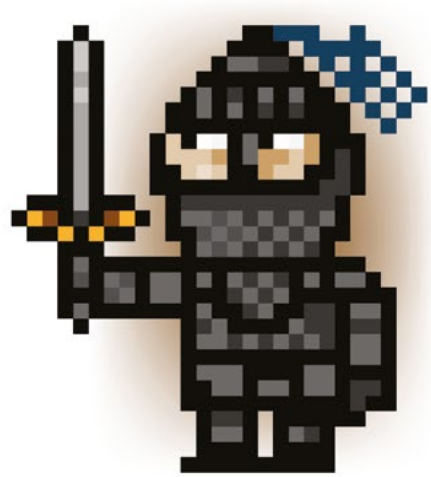
There is no need to assume that this technology provides any of the features that have made bitcoin and its blockchain famous: anonymity, untraceability, lack of central control and so forth. A cynic might be excused for thinking that, having stripped out all the new, intriguing and controversial features of bitcoin, a data structure with the remaining features could be implemented without any reference to blockchains – although it would probably fare poorly with investors for not using the fashionable buzzwords.

Supporters claim that one of the main benefits of this sort of DLT is that if a consortium of institutions uses it, none of them individually is in a position to make fraudulent changes to the ledger, as would be the case if one company ran the master server that hosted it. This opens the door to any inter-institution applications that require an unforgeable audit trail.

> *DLT could be used to store an audit trail of data accesses in addition to the information itself*

How would such a DLT interact with personal data? The question is too general to answer until someone says what kind of personal data would end up on the DLT, and for what application. Clearly a tamper-evident open ledger provides availability and integrity but not confidentiality, so we would not want to store any personal data in it, at least in plaintext. A data controller organisation might encrypt





the personal data of a customer before storing the details on the ledger, but then why store them on the ledger if the intention is that no other organisation should read them, rather than storing the details locally within just that organisation? For availability, maybe? But this has to be balanced with the cost of replicating the information, and why would the other participants in the DLT want to pay for the replication of data that by design they would never be able to read? One view is that data storage is so cheap that we don't care; another is that data always expand, like a gas, to fill up however much space you grant the information and that we should exercise restraint.

It might be possible, however, for the organisation to store the encrypted customer data privately but store just a hash of it on the public ledger, so that there would be a record of what the organisation claimed it stored about that customer at that time. In theory, the individual attributes of the data record held about that customer could be independently hashed, and their hashes stored in the DLT. That way, third parties observing the public log would not be able to see much, but the customer could check that the information kept was as expected and that any changes to it were applied as promised, because the customer, knowing the data, would be able to verify the hash.

An even more interesting usage pattern that has been suggested is to use the DLT to store an audit trail of data accesses rather than, or in addition to, the information itself. The data controller/processor organisation would store, in an irrevocable and tamper-evident fashion, a time-stamped record of every access to every item of the customer data (identified by its hash, so we could later verify exactly what had been accessed), together with the evidence presented by the accessor that proved that it was authorised for access.

This sounds like a neat way of protecting the privacy of personal data. You could even imagine forcing the accessor (say, an insurance company) to have to present an authorisation token obtained from the customer (the individual requesting insurance) before being authorised by the data controller (say, the NHS) to access a certain part of the record of that customer. Corporations, both those requesting access and those granting access, would be less likely to misuse the personal data of their users if evidence of the access patterns were recorded for ever in this global audit trail and if they, therefore, had to provide evidence that they had been given that consent for every access.

But, in practice, I am not particularly optimistic that this is how it will play out. The various megacorps have their own interests at heart regarding using and misusing the personal data of their and others' customers.

A small-scale demonstration of how the tussle is likely to play out can be seen by observing how the request for user consent works for cookies on websites. In theory, regulation says that websites may only install cookies with the consent of the user. In practice, websites extort this consent by making their pages essentially unusable until the user grants it. They make the workflow to say "No" much more cumbersome than the streamlined one to say "Yes", therefore complying with the letter but not the spirit of the law.

> *Websites extort consent for installing cookies by making their pages almost unusable until the user grants it*

The General Data Protection Regulation has brought in an unprecedented level of consumer protection, but has been fought at every stage by lobbyists from major companies. A recent news report stated: "Facebook has targeted politicians around the world...[pressuring] them into lobbying on Facebook's behalf against data privacy legislation, an explosive new leak of internal Facebook documents has revealed."[1]

The fact that DLT might be used to audit data accesses and deter misuse does not necessarily mean it will be used that way, unless some strong-hearted and persistent white knight in shining armour takes on the powerful megacorps to protect the privacy of personal data for consumers. ■

Frank Stajano *is a full professor of security and privacy at the University of Cambridge, where he is the head of the Academic Centre of Excellence in Cyber Security Research. He is also the managing director and co-founder of Cambridge Cyber, a provider of penetration testing and security consulting services*

1. ('Social Network Targeted Legislators Around the World, Promising or Threatening to Withhold Investment'. The Observer, 2 March, 2019. Available at: www.theguardian.com/technology/2019/mar/02/facebook-global-lobbying-campaign-against-data-privacy-laws-investment.)