Centralization, Decentralization, and Digital Public Infrastructures

Jon Crowcroft and Hazem Danny Nakib

with apologies to Mark Nottingham, https://www.rfc-editor.org/rfc/rfc9518.html

Through the Control/Power Lens.

Governments have typically centralized control of the things that governments do - raising taxes to fund provision of certain services (that that government has deemed for itself it part of its function or purpose) - education, health, transportation, defence, and even in the not too distance past, telecommunications. Decentralised government (e.g. syndicalism) has been rare. On the other hand, most governments in recent history have left certain domains outside of government to markets or communities, although not without some (perhaps limited) regulation or control of governance, such as markets.

In the past, communities have built cooperative ventures (shared barns, greenhouses and farms, shared savings and loans, coops, and even community banks) and more recently community networks and power grids to share electricity. These are examples of communities seeking to source the provision of services for themselves and for each other, as opposed to relying on government.

Originally, and across a number of different apparatuses, the function of government is a combination of determining the provision of services on the basis of solving complex coordination problems. The evolution of digital infrastructure points heavily to more efficient ways of solving some of these complex problems through more decentralised structures, that may be government supported or sponsored, to deliver better to the user base (i.e., citizens).

Through the Economic Lens

Markets often espouse competition where multiple providers offer equivalent substitutable products and services where parties need to compete on value, whether to do with price or utility, and that can push them to innovate. Various models exist for central versus decentralised economies.

There's interaction between government and economy, through regulation, especially with the threat of monopoly, or even just oligopolies, through coercion, government v. companies w.r.t making sure market operates transparently, efficiently and fairly and avoiding collusion and cartels forming. (see later feds v. apple, and in the UK, IPA v. GDPR.). Centralisation comes through, for example, regulation, to help keep markets fairer, otherwise consumers get locked into a commoditised single provider oligopoly of all services that extracts rents.

Through law, government may also provide citizens with Agency, Representation, Redress. Of course, there are good and bad government, and typically this can show up in terms of poor practice, or deliberate removal of rights (to agency or redress, e.g. concerning unfair treatment, including exclusion etc

The problem with centralisation when it comes to the economic apparatus is that governments may be good now, but bad later, or vice versa. It is not an accident that Germany has the strictest privacy laws in the world, it was as a result of their past experience of East Germany under the Stasi. They are not so naive as to believe that that couldn't happen again one day (sadly).

Through the Technology Lens

The Internet is probably the best example of something that had been largely a decentralised system for decades.

Horizontal - services - interoperation/federation

E-mail, web, name spaces

Vertical - stacks - silos

cloud, social media, online shopping, entertainment

Horizontal systems are somewhat decentralised (or at least distributed) whilst in some informational sense, vertical systems are somewhat centralised.

Through the Information Lens

Where is data, is orthogonal to who can read it, who can alter it. ownership and control depends on access rights, and access control, and legibility. So whether I can get at my, or your data depends on my role and my privilege level but also whether I can then actually decode your data depends on my having the right software.

At some level, we can expect most data today to be encrypted , at rest, during transfer and even while processing. Protection through access control is not sufficient, since there are mistakes, insider attacks and coercion. Software has vulnerabilities. Hence we employ keys, and encryption/decryption depends on having both data and relevant keys.

One step further, who has accessed the data, and who has been able to decode the data is part of audibility (who can see who can see - Quis custodiet ipsos custodes? etc)

If the user controls the keys, they may not care too much where the data is being held or being stored (except for potential denial of access) since others copying the data will not be able to decode it anyways. On the other hand, if the government keeps copies of keys but tells you that the data cannot be decrypted, they can still access any relevant data, whether it is central or decentralised as they are acting as a “trusted” third party. Of course, if the government accesses my data on my computer, I may be aware of that (through audit trails). but that might not do my much good in the face of a "bad" government.

There are two separate aspects of identity systems where visibility of data matters, in terms of threats to citizens from bad actors: firstly, foundational id provides linkability across surveillance of actions (voting, signing on to services, etc) so exposes the individual's digital footprint to long term analysis; secondly, functional id includes

particular attributes (age, gender, race, religion, licenses to operate vehicles, medical, academic qualifications etc etc), which offers opportunity to discriminate (treating groups preferentially or excluding or reducing rights of other groups etc etc). A bad actor doesn't need the whole government (or its service providers) to misbehave - just that systems are poorly designed so that insiders can exploit vulnerabilities). The perception of this possibility is enough to create distrust, and disengagement, which itself will mitigate against vulnerable groups in society more than privileged.

Through the Efficiency Lens

We can put all the data in the world in one place, or we can leave it where it was originally gathered. This is a choice that represents two points on a spectrum of centralized versus decentralized data. One can also copy the data to multiple places at the same time.

There are efficiency considerations in making this choice, which entail more energy, higher latency, lower resilience, worse attack surface, and potential for catastrophic mistakes, when taking the centralised path. The decentralised path reduces these risks, but still requires one to consider copies for personal data resilience but with better attack surfaces, lower latency, less energy and greater resilience.

These choices are orthogonal to the access choices, which merely concern who has rights and keys, and where they keep those, not who actually holds the data where. Instead, they are separate questions that need to be tallied up, overall when considering the centralised v decentralised conundrum across different dimensions.

Conclusions, regarding Alternative Solutions in the Digital Identity Space

A digital public infrastructure such as an identity system needs to be trusted by people using it and trusted in a way that make people comfortable using it (so people use it), and therefore considerations about whether the user base trust the government or not matter in the design, including the access choices.

If we don't trust the government, we might choose a decentralised system, or at least a system with decentralised keys (like the Apple iCloud eco-system).

A government may also be trustworthy, but the userbase may be generally distrusting because of some incident, whether of government or of the private sector, giving further reason for decentralised architecture, at least along some or many of the dimensions such as the economic apparatus or across access.

The question of whether there should be one provider, or six, or 10 billion is orthogonal to this trust, although it does impact resilience and latency, i.e. efficiency.

If the keys are owned by users, then this impacts governments' ability to use identity data (attributes, and identity usage) to plan, whether for good or for bad. That said, some privacy technology (e.g. FHE or MCS) combined with decentralised learning might allow non privacy invasive statistics to be gathered by a centralised agency (i.e. government) without actual access to individually identifying attributes. A good example of this was

the Google Apple Exposure Notification system designed for use for digital contact tracing during Covid, which could have been adapted to offer statistical information (e.g. differentially private) if necessary (though it wasn't used that way in practice).

All of this leads to the question about who provides key management, and a related question of certification (i.e. why should we trust the key management software too). One solution to this is to provide a small (e.g. national scale) set of identity services, but a decentralised key management system that can also be used to federate across all the identity services (cross-border/ or between state and private sector).One technology that we built to provide that independent key management for identity systems is trustchain[1], which is a prototype that services to replace a (somewhat) centrally owned platform such as Web PKI.

An interesting oligopolistic system that offers somewhat decentralised certificates is the Certificate Transparency network (of about 6 providers) that sign keys for the Internet -- this arose because the previously centralised CAs were hit by attacks which caused major security breaches in the Internet. We would argue that a similar scale system for key management and certification for digital identity is evidentially the bare minimum for acceptability for any trustworthy system.

Whether the system infrastructure itself is decentralised or not is a separate question which concerns efficiency, and, perhaps, some types of resilience (Estonian Digital Citizenship systems are distributed over several countries for backup/defensive/disaster recovery reasons).

[1] trustchain is a prototype that is based on ION and makes parsimonious use of the bitcoin proof-of-work network to provide decentralised trustworthy time, and then can create/issue keys in a way not dependent on any central provider or service, resilient to coercion, collusion and sybil attacks. We are currently investigating replacing the proof-of-work component with TimeFabric, which itself depends on a ledger, but can use a proof-of-stake or proof-of-authority and is therefore massively more sustainable.