On the duality of resilience and privacy†

Jon Crowcroft

Computer Laboratory, University of Cambridge, Cambridge CB3 0FD, UK

Protecting information has long been an important problem. We would like to protect ourselves from the risk of loss: think of the library of Alexandria; and from unauthorized access: consider the very business of the ‘Scandal Sheets’, going back centuries. This has never been more true than today when vast quantities of data (dare one say lesser quantities of information) are stored on computer systems, and routinely moved around the Internet, at almost no cost. Computer and communication systems are both fragile and vulnerable, and so the risk of catastrophic loss or theft is potentially much higher. A single keystroke can delete a public database, or expose a private dataset to the world. In this paper, I consider the problems of providing resilience against loss, and against unacceptable access as a dual. Here, we see that two apparently different solutions to different technical problems may be transformed into one another, and hence give better insight into both problems.

1. Introduction

Society has become increasingly dependent on information systems, now embedded in the Internet and the Cloud. The Internet is a vast, decentralized communications system. The Cloud (in most of its forms, as Infrastructure as a Service or Platform as a Service, for example) is a small set of very large centralized storage and processing systems. Many of today’s information systems make use of small devices (laptops, smartphones, tablets), together with the network, to access these centralized resources. It was not always so.

†An invited Perspective to mark the election of the author to the fellowship of the Royal Society in 2013.
There seems to be a trend for utilities to centralize as civilization becomes urban, but then decentralize over time. The communications world centralized on postal, then telephone networks, but the Internet has moved us back initially towards decentralization. As we build micro-generation installations for power, we are seeing a move towards decentralization for the grid. Transportation systems are slowly becoming automated, and autonomic control may mean less central planning (e.g. of flight paths) and a more distributed approach to resource allocation.

One reason for centralization is the ability to aggregate computation and storage facilities to get economies of scale, coupled with resource pooling efficiency gains. For the customer, this takes away the burden of management, but at the cost of higher delays and possible loss of service when any part of the network access path fails. This lack of fault tolerance or resilience motivates me to look at decentralization again. However, it seems that these decentralized systems can also offer better privacy, specifically, in terms of guarantees of integrity and confidentiality, and, more generally, availability.

I claim that there is something fundamental about the link between resilience and privacy, and that it derives from properties of symmetric networks of all kinds. The key seems to be diversity, which is going to be maximized by mixing resources from many potential locations in the network at any time, rather than putting all one’s eggs in one basket. In communications systems, this distributed alternative is typically referred to as the peer-to-peer approach. Metcalfe’s Law (http://www.en.wikipedia.org/wiki/Metcalfe’s_law) restates the network effect that the value of a system grows faster than linearly with the number of connected contributors: since all producers are also potential consumers, each added node gives the new producer as many customers as are on the network already, hence there is a notional square-law increase in value. This symmetric network, with all participants offering as well as demanding resource, quickly overtakes any centralized system. It also maximizes diversity.

Central government desire for networks being built on a centralized world view has negative consequences for fault tolerance, privacy and consequent other aspects (energy and other resource use) which may be counter productive.

The classical argument against the peer-to-peer approach is encapsulated in this paper by Charles Blake & Rodrigo Rodrigues [1]: choosing three (perhaps orthogonal) properties so that one can show the infeasibility of a satisfactory solution to a problem by arguing not all three can be provided and is a common rhetorical trick in systems papers, as well as in other disciplines.

However, the argument in that paper is neither necessary nor sufficient for a proof that we cannot construct viable decentralized systems. For example, I argue that given the higher reliability and symmetric performance of fibre to the home, and the deployment of LTE (so-called ‘4G’ cellular networks) and IPv6, it seems reasonable to distribute all services entirely to the edge where statistically availability and scalable storage will be achievable in a dynamic peer network [2]. The dynamics come from the user carrying most of the data they wish to, with them in mobile devices, but the home acting as the master point of contact and persistence (but providing low latency for the home user). For the mobile user, storing data on the device instead of the cloud (except, perhaps for backup) will hugely improve access speed. However, replicating data (encrypted and diversely) at home and in the cloud as well would regain possibly lost resilience (against loss or theft of mobile devices).

I further argue that when we add other (perhaps orthogonal) properties of interest, such as privacy, energy and fault tolerance, even without the deployment of faster net access, we can continue to better provide all but one property (e.g. high availability for others to access my data) in the peer-to-peer mode than in centralized systems.

Finally, based on approaches to quantifying resilience, and using the observed dual, I propose a new interpretation of a common metric for the attack surface of a system.

---

1The Internet is surely a utility.

2According to the Oxford English Dictionary the phrase peer-to-peer protocols goes back to the Magna Carta, albeit with a slightly different connotation.
2. What kind of duality?

There are several different notions of duality in the world, including:

Philosophical. The Platonic Ideal versus the ‘real’ world.
Electrical. Voltage versus current based model of circuits.
Parallel and distributed. Multiprocessor versus networked systems.
Operating systems organization. Message and thread versus RPC and client-server.
Distribution and diversity. This last is what this paper is about, and my claim is that the use of distribution mechanisms and diversity mechanisms for storage and processing, normally associated with achieving resilience against faults in communications, also serves to improve the privacy of our data.

I am using a more informal definition of a duality than is used in the mathematical or engineering domains. The purpose remains to show how an approach to a systems problem in one domain provides solutions in a second.

3. Resilience and privacy

I will discuss how resilience and privacy can be achieved next.

(a) A naive attempt at resilience

Starting from a centralized system, one might attempt to provide tolerance to faults by simple replication. As well as being expensive, this fails to address common mode failures. At the least, a voter that chooses majority answers from copies could become faulty. However, all the copies could have the same flaw. In this sense, increasing the number of copies just increases the probability of common mode failure. Just as a small gene-pool places a population at risk from a single change in the environment (e.g. a new disease), a lack of variety in storage, computation or communication places information at higher risk.

By the same token, it increases the attack surface. The number of places where a single click could release all the data for a service is now many rather than just one. So availability has been increased at a considerable increase in risk.

Just as accidents can and will happen, so also will the enemy concentrate the attack at a single point—subverting a voter, he could deny service or steal copies of content. How can we remove this flaw?

(b) A better attempt at resilience

Diversity, whether in paths, computation, storage and even in implementation, was one of our goals for increasing resilience. In communications and storage systems, there have been techniques for this for many years. Essentially, data are encoded redundantly, but rather than making many copies, we weave a tapestry using the bits that represent data, so that threads making up particular pieces of information are repeated but meshed together with threads making up different pieces of information.

Then to disentangle a particular piece of information, we need to unpick several threads. Such coding techniques can be made as resilient as we wish, so that equipped with a thorough a priori knowledge of faults or the risk level we wish to sustain for attacks, we can provide a given robustness.

3 Common mode failures arise from multiple systems encountering faults due to statistically non-independent causes.
4 Needed for the naive approach, too.
In the radio world, this is extremely well known. The approach has also been proposed in the fixed Internet and in storage systems, but nevertheless still deployed typically, with centralized servers.

Diversity can apply in more basic ways than just encoding in transmission and storage. Taking inspiration from biodiversity, and the idea that the size of the gene-pool matters for survival of a species, diversity has been used as a general design principle for fault tolerance already (particularly in safety critical systems, for example, in aerospace). The idea has been extended to computation and even to approaches to writing software (with multiple programmers interweaving different versions of code). Perhaps it should just become routine.

(c) Revisiting duality

The two key techniques for increased resilience are also techniques for improved privacy, in the sense of reduced risk of loss of confidentiality. Both distribution and diversity are well-known techniques but have been lost for economic reasons—centralized storage has dominated the Cloud because it allows the site owner to run analytics. By contrast, decentralization (i.e. partitioning of data over distributed sites) means that the attacker has a much more complex target (has to know where something is, as opposed to brute forcing 1 place to get everything)—so a single point of failure is a single point of attack. In addition, coding for resilience in communications (as per multipath/net coding) has frequently been proposed and tried for storage (and even processing) but not taken up. In the philosophical sense discussed earlier, the more diversity used, the closer our system approaches an ideal in terms of availability and long-term persistence, but also in terms of privacy.

There is also the added benefit that we reduce the overall cost over a central system, since there is no ‘cache’ needed to store data nearer users for lower latency. The data are stored near users in any case, and only distributed (diversely) for resilience (i.e. restoration after local copy fails). The coordination costs of a decentralized service can be made tractable compared with past, excessively complex approaches, by taking advantage of the observation that the vast majority of data in the cloud today is immutable, and systems need largely to deal with append-only operations. Integrity checking in such a system is relatively simple. Newer algorithms such as RAFT [3] can be used for consensus in the rare case update is required.

An added bonus is that coercion (or natural disasters) on ‘backups’ does not work (to reveal or lose the data) since no single site can render all the data—this gains backup sites plausible deniability too.

In the philosophical, the more redundancy, the closer the system gets to the platonic ideal in terms of resilience and privacy protection. In OS (Lauer) [4] sense, there is a dual between the communications model and the storage model (parallel routes, multiple discs; coded transmission, coded storage). At a bit of a reach, one could even see an analogy with the circuit model of duality, looking at static storage overheads versus dynamic transmission overheads.

The dual nature of resilience and privacy arises from the reduction of risk of common mode failure, by introducing diversity and the resulting reduction of attack surface of the system. Distribution means that the risk of wholesale loss of confidentiality is reduced, while diversity makes attacks on retail privacy more complex. While multimodal failures (or attacks) may still occur, and succeed, the bar for those is higher.

5Going back, interestingly, to the invention by the Hollywood actress Hedy Lamarr of spread spectrum communications, during the Second World War. Indeed, the security properties of the pseudo-random chipping sequence used in some spread spectrum radio modulation schemes is exactly what we are exploiting here.

6See, for example, Ross Anderson’s Eternity Service, http://www.cl.cam.ac.uk/~rja14/eternity/eternity.html for an early example.
4. A novel interpretation of the attack surface metric

Bellovin [5] has argued that metrics for security are impractical due to the brittleness of software systems. On the other hand, Manadhata & Wing [6] present a single system attack surface metric. However, neither work addresses the distributed approach that I describe here. In contrast to the problems for security metrics, metrics for availability are in widespread use and can be used quantitatively to show how faults are mitigated by particular redundancy and diversity techniques. In the future, direct application of these metrics may show the reduction in the attack surface, if the same techniques are used for security purposes. Of course, an empirical evaluation depends on having a detailed model of the arrival process of faults in components, which itself is based on large-scale empirical studies. In the same way, future research should use detailed statistical models of the arrivals of attacks to quantify the improvements that our proposed decentralized resilient systems provide for security.

5. Discussion

It is a truth universally acknowledged that centralized cloud services offer the opportunity to monetize vast aggregates of personal data. This affords the subjects of that data ‘free’ use of the storage (and to some extent processing) in said central repositories. Thus, services like Gmail, Facebook, Dropbox and Flickr abound, and the only payment seen by the user is made with her attention (eyeball time).

At the same time, many are the attempts to build decentralized services that run in some edge nodes in a dynamic peer-to-peer mode. Hence systems such as Diaspora, Peerson, Persona, Footlights, vis-à-vis and others have been built, but seen little use.

Of course, this may be because they were poorly built, or one might conjecture what would happen if they were to offer a pay-per-use (or subscription) model of resource allocation. First of all, one could work out the cost of a central service, based conservatively on the current revenue streams of service providers who currently use two-sided business models (such as advertizing and analytics) as an alternative to actually billing the user. Estimates of the charge can be derived from the financial reports of the cloud service providers mentioned above—as little as a pound a month might cover most of what is needed, much less than people already typically pay merely to have Internet access at all, and far less than mobile tariffs. Having removed the need for advertisements and market research, the central system would be much less expensive to operate.

In particular, you could envisage merely supporting backup and archiving to these pay-per-use centralized archives. In the extreme, mutual benefit of peer-to-peer backup would remove the need for any central service, and, possibly, the need for any payment either. These types of models and potential charges were discussed in detail in a recent Dagstuhl seminar on decentralized privacy preserving systems, for which see the report archived there http://www.dagstuhl.de/en/program/calendar/semhp/?semnr=13062. One particular point of captured, there was the surprisingly low cost of a pay-per-use/subscription system, compared to the basic cost of Internet access at all (possibly only around 10% of that common price). Thus, the cost to users to replace time-consuming advertisements and loss-of-privacy from analytics, could be a fractional increase in what they pay today for just accessing the Cloud at all.

There has been a lot of speculation about why the decentralized systems have not seen massive take-up. Perhaps it is just premature: we have only had these centralized two-sided systems for about a decade, and there are some signs of development of decentralized alternatives in the technology sector. I would claim that the performance and symmetric market value of network economies make it timely to revisit the peer-to-peer approach, and that we will reap the benefits of massively improved resilience to failure and to attack on our privacy.

However, as my title implies, I claim that there are functional reasons why, in the very long run, I would favour (perhaps strongly) the decentralized approach. In contrast to centralized systems, dynamic peer networks offer better scalability, fault tolerance, privacy and lower energy costs, while statistically, aggregated over all users, affording higher availability. Additionally,
networks are rapidly transforming, so that performance bottlenecks may soon not be a barrier to this approach, even for the most demanding of applications [2].

In these new decentralized coded systems, how will we incentivize people to participate? This is where I think the current Internet services are misleading. The typical cloud of services today that many of us use appears free, but is in fact based in centralized businesses exploiting our personal data, both for targeting advertisements to each and all of us, and to carry out analytics (what used to be known as market research).

We can incentivize users to share resource simply by saving them time, and offering far stronger privacy and resilience to failures. It may be that early attempts to provide such systems have just been too complex to install. Bundling a new system with a network service might be another way to get early adoption to accelerate. If we insist on retaining the current Internet business models, then we might write a roadmap for technology to continue to support them in a decentralized, privacy-preserving world like this:

1. First, build a decentralized system, with a suitable resilient, privacy preserving coding scheme and see if people like it.
2. Second, employ a low-transaction-overhead system (such as M-PESA or BitCoin) for people to pay each other for service.
3. Third, make sure that the legal and regulatory system do not inhibit the right to pool resources in a peer-to-peer system.
4. Fourth, decentralize differential privacy techniques to allow aggregate data mining without loss of privacy (e.g. search and analytics).
5. Fifth, try to speed up homomorphic encryption to the point where it could be used realistically to support targeted advertising without knowledge of the subject.

6. Conclusion

There seems to be some duality between fault tolerance and privacy. This seems to apply not just in communications networks, but in other systems (transportation, energy grids, perhaps others).

Acknowledgements. I acknowledge constructive comments from Dr Hamed Haddadi, Dr Anil Madhavapeddy, Dr Richard Mortier and Dr Kuan Hon.

Funding statement. This work was funded by the RCUK Horizon Digital Economy Research Hub grant, EP/G065802/1.

References