Journal of Cyber Policy

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rcyb20

The influence of internet architecture on centralised versus distributed internet services

Jari Arkko

To cite this article: Jari Arkko (2020) The influence of internet architecture on centralised versus distributed internet services, Journal of Cyber Policy, 5:1, 30-45, DOI: 10.1080/23738871.2020.1740753

To link to this article: https://doi.org/10.1080/23738871.2020.1740753

Published online: 22 Apr 2020.

JOURNAL OF CYBER POLICY
2020, VOL. 5, NO. 1, 30–45
https://doi.org/10.1080/23738871.2020.1740753

OPEN ACCESS

The influence of internet architecture on centralised versus distributed internet services

Jari Arkko

Oy LM Ericsson Ab, Jorvas, Finland

ABSTRACT

The internet evolves rapidly as innovations in technology, applications and business emerge. In some cases, the changes have also given rise to the creation of centralised service deployment models and industry consolidation. This paper focuses on the question of how internet technology and open interfaces may affect the ability to deploy services in a collaborative fashion vs in a centralised fashion. The paper presents a categorisation of factors influencing these choices and discusses these factors in the context of several case studies. Some aspects of centralisation and consolidation are direct consequences of physics. A large organisation can provide a short round-trip time to users around the globe, due to service instances in many locations. Other aspects are due to economics. For instance, network effects cause the value of a service to grow per Metcalfe's law. But technology and deployment choices also have an effect. Federated, collaborative networks with open standardised interfaces (such as email) allow multiple service providers to interact with each other. Closed systems may not allow this. Many popular social networks fall in this category. While technology is not the main driving force behind what are often business and economic decisions, awareness of technology choices makes it easier to understand the likely impacts of a chosen model.

ARTICLE HISTORY
Received 28 August 2019
Revised 23 December 2019
Accepted 30 January 2020

KEYWORDS
Internet architecture; centralised and distributed systems; internet consolidation; open interfaces; DNS

1. Introduction

Many of us have held a vision of the internet as the ultimate distributed platform that allows communication, the provision of services and competition from any corner of the world (See e.g. Litan and Rivlin 2001). But as the internet has matured, it seems to have also given rise to the creation of centralised service deployment models and industry consolidation (Dolata and Schrape 2018).

The term *centralized service deployment* refers to serving users with the help of a network-based function which may be aware of the users' communications and data. Such services are typically replicated but under the control of a single administrative domain.

The term *internet consolidation* (Arkko et al. 2019) refers to the process of increasing control over internet infrastructure and services by a small set of organisations.

CONTACT Jari Arkko ✉ jari.arkko@ericsson.com; jari.arkko@piuha.net

Studying these trends is important, as they have a potentially big effect not only on the competitive landscape in offering internet services but also on internet users and even the evolution of internet technology.

Depending on the competitive landscape, it may be either easy or hard to develop competitive social media applications, deploy cloud platforms or device ecosystems. Internet users are of course affected by the kind of choice the market offers. They may also affect relative bargaining positions, e.g. with plenty of choice, users would typically have more bargaining power to choose the kind of service that fits their needs.

The technology impact comes from developments that are naturally tailored to the most common deployments. In a market with many centralised services, technology development also tends to focus on improving those rather than, e.g. collaborative or distributed alternatives. Further, in a market with vertical consolidation – such as entities that control both browsers and the services they connect to – it becomes possible for the largest players to introduce technological changes much more rapidly than in other situations.

Centralisation and consolidation trends can be studied from various perspectives. A key question is to what extent they are the natural consequence of economics in globally competitive services. The purpose of this paper is not to analyse the degree of consolidation in different markets or to make any recommendations on whether some specific situations are problematic from a business or legal perspective. Rather, this paper focuses on the much narrower question of how technology interacts with changes along the collaborative vs consolidated and distributed vs centralised dimensions. The paper presents a categorisation of factors influencing the choices and discusses these factors in the context of an example from a currently evolving part of internet technology.

The choice of internet architecture and technology has an effect: for instance, email systems are federated, collaborative networks with open, standardised interfaces, allowing multiple service providers to connect with each other. However, not all systems have such interfaces, or federation may not be permitted per policy or business reason: instant messaging systems, for instance, are less well connected than email. Interestingly, the dependency on spam filtering mechanisms has started to erode interconnectivity within email systems as well, due to the fact that email acceptance from a particular provider is often a policy decision involving checking IP address ranges and black and white lists.

The rest of this paper is organised as follows. Section 2 looks at general and technological factors driving centralisation and consolidation. Section 3 discusses the impact of security technology for these trends, and Section 4 highlights some of the issues with a case study in the Domain Name System (DNS) space. Finally, Section 5 draws some conclusions.

2. Factors driving consolidation

The main drivers behind consolidation are economic factors relating to scale (the ability to easily reach a large market of users over the internet) and network effects that increase the value of a service per Metcalfe's law (Gilder 1993). This kind of setting tends to enable winners to take large market shares (Noe and Parker 2005).

The most visible aspects of consolidation involve well-recognized internet services such as advertising, social networks and search (ISOC 2019). But the diversity of many underlying services is as important as that of, e.g. consumer-visible social networks. For instance, the diversity of cloud services, operating systems, the Domain Name System, identity services or software components such as browser engines is important (Arkko et al. 2019; ISOC 2019).

The rest of this section first briefly discusses the impact of economics, then other, more technology-focused factors behind consolidation in greater detail.

2.1. Economics

Internet-based businesses are typically characterised by very low costs for serving additional customers, once a service for the initial customers exists (Mahedevan 2000). As a result, many internet-based businesses can potentially serve a very large customer base, leading to obvious scaling benefits.

Typically, network effects have an even more pronounced impact. Each additional user adds to the value of the network for all the users in a network. Depending on the type of application, the added value benefits users in different ways (as noted by Arkko et al. 2019): in the open web, the value grows for everyone, as the web is a globally connected, interoperable service that anyone with a browser can use.

This is not true of all applications, however. For instance, anyone with an account on any email server can use it globally, and, in theory, anyone can set up servers for this service. But even for such a highly distributed service as email, in recent years we have seen quite a bit of consolidation of email services into a few large email providers (Statista 2018). Email usage grows globally, and the biggest providers have rolled out innovative, high-quality services. Unfortunately, some of the concentration in large providers occurs because the running of email services by small entities is becoming difficult. The difficulties relate, for instance, to spam prevention practices that tend to recognise the largest entities better than smaller entities (Arkko et al. 2019). Smaller entities run a higher risk of not being recognised as well-known and well-working services. In contrast, no email administrator would block a larger entity, such as Google's Gmail, because such blocking would have immediate negative effects on the large proportion of users that the email administrator serves.

Network effects behave differently again in other, more closed applications, such as social media. Arkko et al. (2019) note that the value of being a customer of one social media service greatly depends on how many other customers that particular service has. Hence, the larger the service, the more valuable it is, without federation arrangements. The customers may not have many practical alternatives when the overall set of customers is very concentrated in a small set of services. Using a service outside this small set would imply not being able to connect with a majority of other customers.

When the value of a service is high and there are few alternatives to the service, this may also increase the relative power the service provider has compared to its customers. This may result in power asymmetry, which in turn may be reflected in the price or other aspects of the service.

2.2. Innovation in closed and open systems

The general 'permissionless innovation' principle is the idea that experimentation with technologies and business models should be allowed by default (Thierer 2016). Arkko et al. (2019) cast this general principle in a more network-oriented fashion as follows: a network can be simple but still powerful enough that essentially any application could be built on top of it without needing any special support from anyone else. An argument can be made that permissionless innovation has brought us many of the innovative applications that we enjoy today, and that the underlying internet, a highly interoperable network, has been the key component in this development (Arkko et al. 2019).

But the social media and email examples in the previous section show what role interoperability has. Paradoxically, if the underlying network is sufficiently powerful, the applications on top can become arbitrarily complex and capable, potentially leading to the closed services discussed above, prompting Arkko et al. (2019) to call this the Permissionless Completeness Problem.

The issue arises when there is no pressure for interoperability within the application. With a sufficiently powerful underlying network, there is no technical reason for interoperability, as anything that the application wishes to do can be done. There may, of course, be other reasons that create pressure. In many situations, customers demand standardised interfaces and modularity while suppliers wish to offer services that are entirely from them. This type of demand works well when the customers are businesses with the capability of making demands, but it may not work so well in consumer markets.

More generally, systems that have open interfaces and APIs tend to enable more opportunities for competition. Closed systems, on the other hand, may only provide a designated interface to the service itself, but no opportunity for further tailoring, improvement or interoperability with other service providers.

A recent networking technology trend causes new functionality to develop at the fastest pace high up in the protocol stack. For instance, new functionality is introduced at application layer faster than at transport layer, and transport layer changes are more common than IP layer changes. Recent advances in transport protocol design that puts much of the code in application programmes rather than in the kernel of operating systems is accelerating this trend (see e.g. QUIC (Iyengar and Thomson 2019)), as less agreement is needed on a new version across different developers; all that is needed is a new version of the application software and a peer that is willing to communicate with this new software. Some of the developments that are happening in the application software space are being standardised and have open interfaces – as the QUIC example shows. But in general, this is an area where there are not always standards, and this affects the ability for innovative competition to add features or interact with existing large systems.

However, the mere existence of these open interfaces does not guarantee actual interoperability in a business setting, as systems may choose not to connect for business reasons.

Interfaces need to be looked at more broadly than just as protocol interfaces. For instance, the ability to use training data is important in machine-learning, but may not be easily available except in large, centralised network functions.

2.3. The role of data

Many internet businesses depend on the availability of data concerning how their service is used, by whom, and in what environment (DeNardis and Hackl 2015). This data enables common monetisation approaches, such as targeted advertising or selling data about consumers and their behaviour patterns. However, this data is often also necessary to actually run the service, from targeting social media posts to users most likely to be interested in seeing them to having the users help to fine-tune a language translation service.

A recent technology trend is the introduction of AI and machine-learning as a part of optimising or even building services. This trend further emphasises the importance of data. The more users a service has, the more data is available for training machine learning models, and the better the service becomes, attracting yet more users (Arkko 2019). This reinforcement loop provides an advantage for large entities to succeed in their business.

A pessimistic view of data collection within the internet is that far too many parties focus on collecting data, and, in some cases, protecting the user in the best way seems a mere secondary goal. To be clear, many changes and new systems in the internet are about helping its users, including improving their security and privacy. But in many cases protecting the user involves passing more data to the party that is doing the protecting. Examples of this situation include, for instance, better authentication systems discussed in Section 3 or the encrypted name resolution discussed in Section 4.

2.4. Fundamentals of communication

The choice between local and centralised solutions can also be impacted by fundamental issues, such as physics and the speed of light. For instance, some classes of popular low-latency applications relate to local communications, such as industry automation in a factory setting (ITU 2015; NGMN 2015) or some automotive applications (Arena and Pau 2018). These types of applications are perhaps best built with local solutions, and present no specific concerns relating to consolidation.

However, other applications combine global services with low-latency requirements, such as low-latency Content Delivery Networks (CDNs), Virtual Reality (VR) or Augmented Reality (AR) solutions (Elbamby et al. 2018). In these systems it is typically necessary to be able to provide a local presence for content delivery functionality or edge computing while at the same being able to offer the service to a worldwide user base. Large organisations have built global, distributed networks of data centres to respond to these needs (Arkko et al. 2019). Collaborative or federated data centres or cloud computing services are other approaches to serving these needs but are not widely deployed today.

2.5. Attack resistance

The security and privacy of our technology also has an impact on consolidation and centralisation. There are some fundamental issues as well as technological choices that affect competitive situations between different parties.

A fundamental issue is the ability to defend against attacks. Distributed denial-of-service attacks have become more large-scale. For instance, the attacks on the network infrastructure provider Dyn in 2016 caused an outage that affected several commonly used internet services (Hilton 2016). This particular attack was launched by over 100,000 hijacked bots from badly secured Internet-of-Things devices. The overall volume of the attack partially overcame some functions at Dyn, even though they are a relatively large provider. But the effect of such attacks on smaller entities can be devastating. Indeed, the biggest cloud- and content-delivery providers are best positioned to deal with these attacks due to their scale, leading many customers to employ the services of the largest providers (Arkko et al. 2019). For a business that depends on its web presence it is not acceptable for that presence to be down for any appreciable length of time. Running their own server is not sufficient in this case, but rather one has to delegate the service to a global, large provider that can provide for both the latency and attack resistance needs that the business has.

This situation is driven by technical reasons but has obvious business implications. It also has potential impacts in other areas. For instance, a large content-delivery provider may have different content policy practices to the organisation that provides the content (Prince 2017). Similarly, the legal regime – such as privacy regulation – may differ between different locations. This can lead to a legal impact if technical reasons dictate the use of a service provider from another jurisdiction, although there are cases where regulations have extraterritorial effects (Bennett 2018) which may dampen the impact.

3. Trusting endpoints

The denial-of-service attacks discussed above are not the only security issue that has an effect. The selection of the parties involved in a communication system or application has a significant impact on the security of the system. There are often more parties – such as centralised nodes – than the communicating users' devices. The creation of such centralised nodes is an obvious example of centralisation and can also contribute to further consolidation of services.

For the purposes of this paper, endpoint trust is relevant both due to the role of new and potentially centralised nodes as well as the impact of pervasive monitoring on any centralised or consolidated internet infrastructure.

3.1. Third parties

Many systems involve the delegation of some authority to a third party. The delegation of user authentication on many applications and websites via protocols such as OAuth (Hardt 2012) is one example of this. Popular social media systems and other large entities have large databases of registered users and it is easy to employ these databases for handling user registration and authentication in other applications too. It is also convenient for users. Certification authorities, application shops and DNS resolution services are other examples of similar third parties.

These parties may be necessary but can also become control points or data sources for gathering more data about users. And from a user perspective, it may not be easy to switch one's identity in an application if it is already bound to a specific social media identity, for instance (Pasquale 2017). The selection of the used third parties in an application can also be problematic, as sometimes the choice of third party is made by someone else, at least by default. This can lead to a particularly dominating actor gaining yet more dominance.

The third parties are also problematic in the sense that they may not be entirely trusted, or at least their trustworthiness may be perceived differently by different parties. The compromise of an entire third party is also not unheard of. For instance, DigiNotar was a Dutch certification authority that was compromised (Van der Meulen 2013). In general, third parties employ a set of defences and those defences are continuously improved. Various certificate transparency mechanisms can be useful guards against attacks related to certification authorities (Farrell 2019; Laurie, Langley, and Kasper 2013). Nevertheless, security problems or even the full compromise of some of these systems cannot be entirely ruled out.

The different perceptions of trustworthiness are particularly problematic in cases where a dominant entity – such as a large operating system or browser vendor – makes a selection of a third party that does not align with the wishes of the users. In some cases individual users may have an opportunity to change the settings, but in general, there is no way to force a particular player to employ a different set of default settings and trusted third parties even if there was a widespread concern that the settings are problematic for the users.

Lack of trust may also take the form of doubting the ability of a commercial entity to withstand legal or technical attacks by national surveillance agencies (Arkko 2019). The Snowden revelations provided plenty of examples of pressures and tactics that resulted in widespread surveillance being conducted across many different services.

As a result of the Snowden revelations, the IETF adopted a policy to 'strive to produce specifications that mitigate pervasive monitoring attacks' (Farrell and Tschofenig 2014). New protocol work at the IETF is also required to consider the pervasive monitoring attacks and take this into consideration during architecture and protocol design.

3.2. Protecting communications vs endpoints

The proportion of internet traffic that is cryptographically protected has grown tremendously in the last few years. Several factors have contributed to this change, from the Snowden revelations to business reasons and to better available technology such as HTTP/2 (Belshe, Peon, and Thomson 2015), TLS 1.3 (Rescorla 2018), QUIC (Iyengar and Thomson 2019). In many networks, the majority of traffic has flipped from being cleartext to being encrypted. Reaching the level of (almost) all traffic being encrypted is no longer something unthinkable but rather a likely outcome within a few years (Arkko 2019).

This does not imply that all problems in communications security have been resolved. They have not, and work on any remaining issues needs to proceed. The IETF and other standards organisations continue to work on communications security. However, the broad application of encryption in internet communications has changed the situation considerably. Today it is much harder to attack communications or glean information from them than it was just a few years ago.

There are, however, significant issues beyond communications security within the internet (Arkko 2019; Farrell 2019). To begin with, it is not necessarily clear that one can trust all the endpoints.

There never was full trust on the endpoints, of course. But the pressure against a compromise through endpoints seems to be increasing. One reason for this is communications security improvement. In addition, manufacturer-controlled operating system installations and tightly controlled applications leave little room for the user to be in control of their own devices, particularly when popular applications come with excessive rights to access the user's media content, location and the peripherals of the user's devices (Arkko 2019).

Server-side devices also have issues. Arkko (2019) identifies a typical pattern of communications as the key issue:

> The pattern of communications in today's internet is almost always via a third party that has at least as much information as the other parties have. For instance, these third parties are typically endpoints for any transport layer security connections, and able to see any communications or other messaging in cleartext. There are some exceptions, of course, e.g. messaging applications with end-to-end protection.

On top of these structural and business-driven issues there are also governmental pressures, e.g. for service providers to grant access to users' data (Stilgherrian 2019).

3.3. Designing for endpoint compromise

Given all the above, precautions seem desirable to protect users against endpoints that are compromised or malicious. It is also quite common that the business interests of one party (such as a website) may not be entirely aligned with the interests of its users, such as when a website collects information about users which users would not wish to have collected (Farrell 2019).

From the perspective of protecting the user in the best possible way, a critical issue is that more attention needs to be given to how the user's data is handled. The right system design can prevent or minimise damages associated with data leaks, whether they are caused by communications security, attacks on data at rest or endpoint problems. The right design can also reduce the temptation for commercial misuse of information.

Similarly, the right design can reduce the risk of external coercion to reveal information to authorities or surveillance organisations. Of course, whether one believes authorities deserve to have access to information is a matter of opinion and not the subject of this paper. However, from a technical perspective the chosen design affects how easy or hard such access may be. Other things being equal, data collection from a number of distributed entities is harder than from a single entity. Of course, other things are often not equal. For instance, there may be legal differences in jurisdictions; authorities in one jurisdiction may have a harder time acquiring information from another, and governments may attempt to create regulation that has extraterritorial reach.

While users themselves cannot design their systems, application developers that are interested in the best protection of their users can take these considerations into account when designing the application system. They can take the following aspects into account, for instance:

(1) Avoiding application design patterns that result in cleartext information relating to the user passing through a third party or the application owner (Arkko 2019). And where a third party is absolutely necessary, attempt to provide a design where the function of the third party is provided by multiple different entities and implemented in a distributed fashion.

(2) Involving only network entities that need to be involved in order to provide the service.

(3) Avoiding architectures that result in unnecessary information collected at a central location (or information collected in a distributed manner but under the control of a single entity).
(4) Using control points that can be selected or effectively monitored by the users or end-user device owners.

In short, there needs to be strict control on what data is collected to begin with, and with whom this data is shared. Of course, these measures, or the work of application designers, are only a part of the overall defences to protect users. Many other areas of systems also deserve attention. For instance, there can be issues in internet and computing infrastructures, in the security of users' devices, or the ability of the user to control those devices.

4. Case study: centralised DNS

The Domain Name System (DNS) has a long history but has seen a relatively slow pace of technical change. Most queries to the DNS are still made through the original UDP port 53 protocol (Mockapetris 1987) to a resolver that typically resides in the local Internet Service Provider (ISP). A fraction of DNS queries are made through more global services, such as those offered by the so-called 'Quad-N' services such as Google's 8.8.8.8 or several similar others. These services provide a high-quality, globally-available DNS resolution service that typically does not perform any local filtering except where mandated by the home location of these services themselves.

Geoff Huston's research on DNS resolver centrality (Huston 2019) indicates that at the time of his initial research, roughly 87% of users employed a primary resolver from the ISP. Google's public resolver was used by roughly 9% of users, while other public resolvers were used by much smaller percentages. Huston's conclusion was 'It is challenging to make a case that this level of use represents some form of centralization of the DNS'. That describes the current situation, however. Huston's research continues to track the use of different resolvers throughout the internet.

The internet market space is quite dynamic, however. Two of the most popular browsers have a roughly 80% market share (Statcounter 2019). Changes in the behaviour or defaults of commonly used software such as browsers may change the situation quickly.

Nevertheless, the current situation with DNS resolver consolidation is not alarming. The kinds of network effects (Section 2.1), closed ecosystem innovation (Section 2.2) or fundamental limits of communication (Section 2.4) do not seem to be current issues in this area. For instance, latency does not bring a large benefit to public resolvers, as the main alternative resolver is typically close by at the user's ISP. The largest public resolvers may benefit from data collected from the systems in various ways, but there are no visible signs that this has caused any ill effects.

The situation may of course change, either through changes in browsers or the DNS technology itself. Recently, the IETF has developed new technology for DNS queries, such as DNS-over-HTTPS (DoH) (Hoffman and McManus 2018) and DNS-over-TLS (DoT) (Hu et al. 2016). These technologies allow queries to be performed confidentially, using a modern and flexible web protocol framework. So far, deployment of DoT/DoH as a replacement for the plain UDP-based approach has been slow, as it requires updates to both operating systems as well as millions of DNS resolvers in the world's ISP networks.

However, some browsers have adopted DoH, given this is an HTTP-based service and as such a natural fit for browsers. Browsers are considering the use of DoH as the default resolution mechanism, with the consequence of directing all queries to a DoH-based resolver. Depending on the specific deployment models used, this can be a single resolver service within a geographical area (Mozilla 2019) or some small set of services. This solution is relatively easy to deploy, as changes are only required in a particular browser and the global services that are used. This kind of a deployment results in potentially quick adoption, helping to avoid local filtering and DNS tampering on the path.

While there are benefits to this deployment model, the downside is that centralising all users' DNS queries to a single or small set of places, no matter how trustworthy, has some issues.

A record of a user's DNS queries is sensitive information, because it provides a history of what websites the user has visited. The use of an encrypted DNS query protocol such as DoH helps keep this history private. DoH also makes it hard to use DNS query information at the ISPs.

It is clear that DoH perfectly performs the communications security aspect discussed in Section 3.2. But at the same time, it offers no protection against lack of trust in the endpoints. As Section 3.1 notes, third parties may not always be sufficiently trustworthy. Putting many users' DNS service in a central location creates new issues:

(1) A centralised service that handles information for a very large number of users becomes valuable from a commercial perspective. As discussed in Section 2.3, the value of data comes not only from monetising that data directly, but also from its usefulness for improving the service itself. A centralised service is also much easier to use for that purpose than the millions of different DNS resolution services used throughout the internet today in different ISP and enterprise networks. As a result, there is a risk that a centralised DNS resolution service will at some point be used for data mining, irrespective of the original good intentions of the people who set this system up.

(2) As with commercial data mining, a centralised service becomes very interesting for governments to tap as a source of information for intelligence operations or pervasive surveillance. Governments may not be able to easily compel millions of current DNS resolution services to provide information to them, but a centralised service under the control of one commercial entity in one jurisdiction is much more practical, at least for a government in that jurisdiction or having good relations with that jurisdiction. As a result, there is a risk that this tapping is either already happening or will be happening in the future. Again, this may well happen irrespective of good intentions or careful operational procedures (such as deletion of logs). In the long run, governments that have the political agreement to pass laws to enable particular types of surveillance will be able to do so, and often while requiring the targeted parties to be silent about the surveillance. On the other hand, centralisation effects may also lead to different implications for other governments, depending on the final location of services. As noted in Section 3.3, differences in jurisdictions may have an impact. For instance, some governments are concerned that centralised services not within

their jurisdiction do not allow them to perform the tasks those governments or their citizens see as necessary. As a result, centralisation can lead to a situation where some governments have greater access internet-based information than some others.

(3) The centralised service may be hosted under a jurisdiction that the user is not aware of. The jurisdiction may have a significant impact on what privacy laws apply, for instance. Citizens also have some ability to influence local laws regarding privacy and surveillance but have no such ability with regard to foreign services.

(4) The centralised service places become critical infrastructure and a potentially weak point in the internet. In 2016, for example, DoS attacks were launched against a large DNS provider, Dyn, leading to outages in common internet services throughout the world. It is difficult to imagine that centralised DNS resolvers wouldn't be a target in future attacks.

In conclusion, while a centralised, encrypted DNS service provides some benefits, it also comes with its own significant issues. For instance, unless one is in a location that does excessive filtering, little has been gained except for moving the potential for tampering and data leakage to a different entity.

A common mistake in security analysis is to look too much at individual components and miss system-level issues. In this particular case, an analysis of individual components is not enough, and one must consider whether a centralised service becomes a more valuable attack target. Specifically, expected DoH deployments do not follow the guidelines from Section 3.3, and consequently, while communications security improves, the overall system security may not improve. Or at least, it may have some other issues remaining.

Different perspectives are also at play. What is considered a safe service by one person may look like a data leak to a foreign country to another. Or what is a global, securely encrypted DNS service for one person may look like a browser calling home for every user action to another.

As discussed in Section 3, improvements in communication security have made attacks other than capturing traffic much more significant. In the authors' opinion, avoiding central control points and databases as much as possible is necessary for continuing to protect users' privacy. For instance, there is little that even large service providers can do to refuse authority-sanctioned pervasive monitoring. The recommended defence against this is to ensure that no such information or control point exists.

Nevertheless, encrypted DNS is still important, as are global DNS resolution services. The above discussion applies only to deployments that attempt to run them in a centralised manner. Encrypted DNS should be deployed broadly. And browsers and other applications should look at ways to employ such encrypted DNS service at the local ISP or build global services out of distributed, collaborative networks. It is also important that applications are transparent to their users about where data relating to the user is going, be it about DNS queries or something else.

There are also new ways of architecting DNS services so that they leak less information about the users who query them. This would reduce the main drawbacks of centralised systems for DNS resolution. Oblivious DNS (ODNS) (Schmitt, Edmundson, and Feamster 2018) is a mechanism that splits the knowledge of who is asking from what is being asked between two different entities. This is possible by sending a query to one entity while encrypting it to another. The first entity does not know what name is being queried. The first entity forwards the encrypted query to the other entity, which responds but will not know who asked the question.

However, it is interesting that while this arrangement provides clear privacy benefits, it also introduces an extra hop and latency that may not be desirable. So here we have an example of a situation where a centralised, data-collecting service would be able to outperform a more privacy-sensitive arrangement. But perhaps collaborative solutions can be found that do not give all information about a particular user's queries to a single entity, without a latency penalty.

The above considerations are largely technical, as were the suggested guidelines in Section 3.3. The key questions for making improvements are, however, not technical, but rather relate to incentives and who can cause changes to be made. The earlier sections have argued for a holistic analysis of the impact of technical choices in the design of systems. There is a role for both the research community and end-users in pushing for this. Attention and awareness of potential issues is needed, as is tracking real-world traffic patterns or the impacts of different choices. Of course, both the research work and the work on solutions depend on sufficient user interest in the privacy of their queries.

The work on solutions naturally depends on the availability and technical or economic feasibility of the potential solutions. In the DNS resolver case, there is a number of different potential solutions such as ODNS or Adaptive DNS (Kinnear et al. 2019), but the industry as a whole is quite fragmented on the desired direction. There is, however, quite a lot of attention being given to this topic, and a large number of entities that provide DNS services today (including many that are considering deploying or have already deployed DoH). Optimistically, the industry can converge on suitable improvements to provide better privacy for DNS resolution.

5. Conclusions

The main purpose of this paper is to highlight the relationship of architectural choices in internet services and their impact on how the offered services can serve a distributed and competitive internet environment. The general considerations outlined in Sections 2 and 3 were accompanied by a case study about DNS resolution in Section 4, highlighting how the general considerations apply in a concrete situation.

The DNS example was about a simple service. But even that example shows interesting technical and policy challenges in making sure that the internet stays diverse and distributed, and that security and privacy can improve. There are certainly challenges to all of these aspects.

One lesson from the DNS case is that analysis for any solutions should always be performed at a system level, considering not merely technical issues but also policy and legal aspects, as well as incentives for both intended participants and potential attackers.

On the other hand, there is reason for optimism, and work on innovative, collaborative DNS resolution solutions seems like a fruitful direction. It is certainly not a situation where all avenues have been explored or where we have hit a fundamental limit.

More generally, besides the system-level analysis, continuing to ensure that key aspects of the evolving internet stay open, e.g. through open, standardised interfaces and that open source continues to be an important building block. For instance, as discussed in Section 2.2, new interfaces at the application level may be necessary, even if creating them may be difficult in the consumer market. And with open source, users and other organisations have been able to, for instance, run their own versions of browsers or other systems or provided extensions that suit their needs better than the default system (Arkko et al. 2019). The freedom and ecosystems provided by standards and open source solutions continue to be crucial to ensure competition and evolution within the internet.

As discussed in Section 3, it is also necessary to have more awareness of security issues that are invoked by centralised systems (as an attack vector). As the Section 4 example shows, it is important to understand the impact of consolidation with respect to trust, privacy-sensitive data, and user profiling and tracking.

From a security perspective, the following principles need to be employed in design:

(1) Provide end-to-end protection for all information passed via other parties. Many systems are designed around an architecture where a central entity connects users together or passes information from one user to another. But if the party who passes the information does not need to know the information to perform its passing, they should not get access to the information. The obvious solution is to protect such information with end-to-end encryption to its intended recipient. The situation is similar to general protocol design. Trammell and Kühlewind (2019) note that it is a useful design rule to avoid 'accidental invariance' (the deployment of on-path devices that over-time start to make assumptions about protocols). However, it is also useful to avoid 'accidental disclosure' in the same manner (Arkko 2019). Passed information that may have originally been thought to be benign and untapped may actually become a significant information leak at a later time.

(2) Minimise passing of control functions to others. Any passing of control functions to other parties should be minimised to guard against the potential misuse of those control functions (Arkko 2019).

(3) Be careful with the introduction of centralised resources or functions. Many systems require centralised functions, for instance, as rendezvous points or for data storage. But there are also risks associated with centralised functions, for instance due to their ability to control the other protocol participants or see information about them. When designing a system, it is important to consider whether a centralised function is actually appropriate. And even when it is appropriate, steps may be needed to mitigate the risks associated with it. One key issue with centralised functions is whether the users will be able to choose which particular resource will actually implement this function. Making centralised resources selectable can be beneficial (Arkko 2019).

(4) Employ explicit agreements as suggested by Arkko (2019). When users and their devices provide information to network entities, it would be beneficial to have an opportunity for the users to state their requirements regarding the use of the information provided in this way. Of course, the actual willingness of network entities to agree to such requirements is unknown. But today we lack even the technical means of doing this, so even among willing users and network entities this cannot be done.

(5) Treat any action as potentially dangerous, even if that action takes place over an encrypted communications channel. For instance, there is no guarantee that an

entity who you communicate with over an encrypted channel won't leak any information you give to it. This is a particular concern when communicating with other parties outside your control. But it is true even for nodes that you yourself have set up, as even such nodes can become compromised and no longer faithfully executing the role that you set them up to perform. Being conservative in distributing information is also important in your network.

In summary, systems need to be analysed for their potential impacts for centralisation, consolidation and security impacts, and many different aspects and potential failure scenarios need to be considered.

As the Section 4 example highlights, this awareness needs to go across not just the research community, but also the industry providing the services and the user community: there needs to be a user demand for competitive and secure services.

Acknowledgments

The author would like to thank the members of the Internet Architecture Board (IAB) and the participants of the Design Expectations vs Deployment Reality (DEDR) IAB workshop and of the EuroDIG workshop on 'Internet consolidation – opportunities and challenges' for interesting discussions in this problem space.

The author would also like to thank in particular Brian Trammell, Mark Nottingham, Christian Huitema, Martin Thomson, Jeff Tantsura and Niels ten Oever as the co-authors of Arkko et al. 2019 which greatly influenced this paper. The author would also like to thank his colleagues Mirja Kühlewind, Simone Ferlin, Gonzalo Camarillo, Göran Eriksson and Magnus Westerlund for inspiring thoughts and feedback, and Stephen Farrell for interesting discussions about security matters beyond communications security and his work (Farrell 2019) that has influenced the author's thinking. Finally, the author would also like to thank the anonymous reviewers for their feedback and suggestions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Jari Arkko is a Senior Expert with Ericsson Research. He has also served as the Chair of the Internet Engineering Task Force (IETF), the internet technology standards development organisation, has worked on routers, software development tools and cellular networks. He likes to personally build and use the technology that he works with. Today he works on internet evolution and 5G. He is a frequent contributor on matters relating to internet architecture, trends and administration.

References

Arena, Fabio, and Giovanni Pau. 2018. "An Overview of Vehicular Communications." *Future Internet* 2019 (11): 27.

Arkko, Jari. 2019. "Changes in the Internet Threat Model." Internet Draft draft-arkko-arch-internet-threat-model-01.txt, Work in Progress, IETF.

Arkko, Jari, Brian Trammel, Mark Nottingham, Christian Huitema, Martin Thomson, Jeff Tantsura, and Niels ten Oever. 2019. "Considerations on Internet Consolidation and the Internet Architecture." Internet Draft draft-arkko-iab-internet-consolidation-01.txt, Work in Progress, IETF.

Belshe, Mike, Roberto Peon, and Martin Thomson. 2015. “Hypertext Transfer Protocol Version 2 (HTTP/2).” RFC 7540, IETF.

Bennett, Colin. 2018. “The European General Data Protection Regulation: An Instrument for the Globalization of Privacy Standards?” *Information Polity* 23: 239–246.

DeNardis, L., and A. Hackl. 2015. “Internet Governance by Social Media Platforms.” *Telecommunications Policy* 39 (9): 761–770.

Dolata, Ulrich, and Jan-Felix Schrape. 2018. *Collectivity and Power on the Internet: A Sociological Perspective*. Heidelberg: Springer International Publishing.

Elbamby, Mohammed, Cristina Perfecto, Mehdi Bennis, and Klaus Doppler. 2018. “Towards Low-Latency and Ultra-Reliable Virtual Reality.” arXiv:1801.07587.

Farrell, Stephen. 2019. “We’re Gonna Need a Bigger Threat Model.” Internet Draft draft-farrell-etm-03.txt, Work in Progress, IETF.

Farrell, Stephen, and Hannes Tschofenig. 2014. “Pervasive Monitoring is an Attack.” RFC 7258, IETF.

Gilder, George. 1993. “Metcalf’s Law and Legacy.” *Forbes ASAP* 152 (6): 158–159.

Hardt, Dick. 2012. “The OAuth 2.0 Authorization Framework.” RFC 6749, IETF.

Hilton, Scott. 2016. “Dyn Analysis Summary of Friday October 21 Attack.” https://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack/.

Hoffman, Paul, and Patrick McManus. 2018. “DNS Queries over HTTPS (DoH).” RFC 8484, IETF.

Hu, Zi, Liang Zhu, John Heidemann, Allison Mankin, Duane Wessels, and Paul Hoffman. 2016. “Specification for DNS Over Transport Layer Security (TLS).” RFC 7858, IETF.

Huston, Geoff. 2019. “DNS Resolver Centrality.” https://blog.apnic.net/2019/09/23/dns-resolver-centrality/.

ISOC (Internet Society). 2019. “Consolidation in the Internet Economy.” Internet Society Global Internet Report.

ITU. 2015. “Framework and Overall Objectives of the Future Development of IMT for 2020 and beyond.” ITU Recommendation M.2083-0.

Iyengar, Jana, and Martin Thomson. 2019. “QUIC: A UDP-based Multiplexed and Secure Transport.” Internet Draft draft-ietf-quic-transport-22.txt, Work in Progress, IETF.

Kinnear, Eric, Tommy Pauly, and Chris Wood. 2019. “Adaptive DNS: Improving Privacy of Name Resolution.” Internet Draft draft-pauly-dprive-adaptive-dns-privacy-01.txt, Work in Progress, IETF.

Laurie, Ben, Adam Langley, and Emilia Kasper. 2013. “Certificate Transparency.” RFC 6962.

Litan, Robert, and Alice Rivlin. 2001. “Projecting the Economic Impact of the Internet.” *American Economic Review* 91 (2): 313–317.

Mahedevan, B. 2000. “Business Models for Internet-Based E-commerce: An Anatomy.” *California Management Review* 42 (4): 55–69.

Mockapetris, Paul. 1987. “Domain Names – Implementation and Specification.” RFC 1035.

Mozilla. 2019. “Firefox DNS-over-HTTPS.” https://support.mozilla.org/en-US/kb/firefox-dns-over-https.

NGMN Alliance. 2015. “5G White Paper.” https://www.ngmn.org/wp-content/uploads/NGMN_5G_White_Paper_V1_0.pdf.

Noe, Thomas, and Geoffrey Parker. 2005. “Winner Take All: Competition, Strategy, and the Structure of Returns in the Internet Economy.” *Journal of Economics and Management Strategy* 14 (1): 141–164.

Pasquale, Frank. 2017. “When Antitrust Becomes Pro-Trust: The Digital Deformation of U.S. Competition Policy.” University of Maryland Francis King Carey School of Law Legal Studies Research Paper No. 2017-24.

Prince, Matthew. 2017. “Why We Terminated Daily Stormer.” https://blog.cloudflare.com/why-we-terminated-daily-stormer/.

Rescorla, Eric. 2018. “The Transport Layer Security (TLS) Protocol Version 1.3.” RFC 8446, IETF.

Schmitt, Paul, Anne Edmundson, and Nick Feamster. 2018. “Oblivious DNS: Practical Privacy for DNS Queries.” *Computing Research Repository (CoRR)*: arXiv:1806.00276.

Statcounter. 2019. “Browser Market Share Worldwide.” https://gs.statcounter.com/browser-market-share.

Statista. 2018. “Market Share of the Most Used E-mail Clients in 2018.” https://www.statista.com/statistics/265816/most-used-e-mail-service-by-market-share/.

Stilgherrian. 2019. “The Encryption Debate in Australia.” International Encryption Brief, Carnegie Endowment for International Peace.

Thierer, Adam. 2016. *Permissionless Innovation: The Continuing Case for Comprehensive Technological Freedom*. Arlington, VA: Mercatus Center at the George Mason University.

Trammell, Brian, and Mirja Kühlewind. 2019. “The Wire Image of a Network Protocol.” RFC 8546, IETF.

Van der Meulen, Nicole. 2013. “DigiNotar: Dissecting the First Dutch Digital Disaster.” *Journal of Strategic Security* 6 (2): 46–58.