Journal of Cyber Policy

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

What lies beneath: transparency in online service supply chains

Jennifer Cobbe , Chris Norval & Jatinder Singh

To cite this article: Jennifer Cobbe , Chris Norval & Jatinder Singh (2020) What lies beneath: transparency in online service supply chains, Journal of Cyber Policy, 5:1, 65-93, DOI: 10.1080/23738871.2020.1745860

To link to this article: https://doi.org/10.1080/23738871.2020.1745860

Published online: 07 Apr 2020.

JOURNAL OF CYBER POLICY
2020, VOL. 5, NO. 1, 65–93
https://doi.org/10.1080/23738871.2020.1745860

Routledge
Taylor & Francis Group

What lies beneath: transparency in online service supply chains

Jennifer Cobbe *, Chris Norval and Jatinder Singh

Compliant and Accountable Systems Group, Department of Computer Science & Technology (Computer Laboratory), University of Cambridge, Cambridge, UK

ABSTRACT

There is a noticeable trend towards the increased centralisation of Internet-based services. Though much focus is on the dominance of organisations such as Facebook, Google and Netflix, popular consumer-facing services, there has been considerably less discussion regarding the organisations providing the infrastructure that supports online services. This bears consideration, given that many online services rely on a range of platforms and services operated by third-parties.

As such, this paper explores issues of consolidation as regards the systems supply chains that underpin and drive online services. Specifically, we note that while there are trends towards the increased centralisation and dominance in the provision of supporting technical infrastructure, the nature of these technical supply chains are relatively hidden. We explore the broader societal implications of this with regards to power and resilience, emphasising the lack of means, legal or technical, for uncovering the nature of the supply chains on which online services rely. Given society's ever-growing reliance on data-driven technology, we argue that more can be done to increase levels of transparency over the supply arrangements of technical infrastructure. This is a necessary precursor to determining what interventions, if any, may be required to deal with issues of consolidation in online infrastructure.

ARTICLE HISTORY
Received 16 August 2019
Revised 27 December 2019
Accepted 6 February 2020

KEYWORDS
Internet consolidation; infrastructure services; platforms and service providers; systems-of-systems; system supply chains

1. Introduction

There is a noticeable trend towards the increased centralisation of online services (Internet Society 2019). From a user perspective, the move in the last decade has been away from the more decentralised Internet of the past to dominance by a few main companies. Much of the related policy discourse tends to focus on a number of well-known, user-facing organisations featuring a large user-base, such as Google (including YouTube), Facebook (including Instagram), Amazon, Microsoft, Twitter, Reddit, Netflix and Spotify to name a representative few.

However, less discussed are potential issues of centralisation and consolidation in terms of the infrastructural software and services that underpin and ultimately drive online services. Many applications, websites and other online/Internet-based services (we use

CONTACT Jatinder Singh jatinder.singh@cst.cam.ac.uk
*Equal author contributions

'*applications*' to encompass all of these) are built upon a range of technologies and services provided by others. These infrastructure services often include, for example, cloud services (storage, compute, [Anything]-as-a-Service), content distribution networks, software, processing and support platforms, analytics services, advertising brokers, and so forth.[1] This means that applications are ultimately reliant on a data-driven supply chain of technologies and providers.

Consolidation at this service infrastructure level warrants greater attention, not least because these services provide the 'building-blocks' that underpin and enable a wide range of applications. The potential for consolidation at this level towards a small number of providers raises various issues. For instance, dominant players would have the power to directly or indirectly influence the nature of the applications that use them. This influence could be realised not only by way of the functionality provided, but also through various conditions of use, e.g. contracts, terms of service, and so forth. Consolidation at this level would also raise questions of resilience, as a change, update, bug, failure or outage in a supporting technology or service has the potential to impact all applications that rely on them (see §3.1).

Indeed, consolidation in the technical infrastructure services that underpin applications would mean that the nature of that infrastructure (and the practices of the organisations that run them) can have systemic impacts, affecting a wide-range of applications. Such effects are amplified in an environment of few alternatives. Naturally, the greater the consolidation and centralisation of infrastructure services, the more pronounced these issues would be.

However, the extent to which there is consolidation, and, indeed, the degree of reliance on third-party services, is currently unclear. There is little in the way of detailed, publicly-available information on the true extent to which applications use other platforms and providers. Though it is recognised that Amazon, Google and Microsoft dominate the cloud services market (Logic Monitor 2017), this represents only a partial picture. In practice, technology supply chains may be far more complex, a potential *system-of-systems*. Applications may rely on third-parties not only for storage and compute, but also for database services, content distribution and management, data brokerage, analytics, logging and authentication services, interfaces to machine learning models (e.g. for language processing and image recognition), and so on (see §4.3.2). Indeed, the fact that a wide range of these services have been readily available for some time is evidence that they are used by applications. But in general terms, the detailed technology supply chains that support applications are largely opaque; it is often unclear on which providers an application relies.

We argue that greater attention should be paid to these service supply chains. This aligns with the directions of current policy discussions concerning lower-level technical infrastructure (a prominent example being the use of Huawei technology in 5G networks (Swinford 2019)). Without a better understanding of the higher-level technical services supply chains that are the focus of this paper, the extent to which there is consolidation at this level will remain unclear. And, without a better understanding of the extent to which there is consolidation in this context, it will remain unclear where the problems that we identify as arising from consolidation will manifest.

However, a key challenge, as our analysis shows, is that there are currently few means for uncovering the nature of these technical supply arrangements. As we demonstrate, current technical and legal mechanisms for uncovering these system supply chains are

insufficient to provide the kind of insight required to identify consolidation. This hinders the policy discourse by limiting the ability to determine whether any interventions are required, and if so, what interventions are appropriate.

In all, this paper considers consolidation in the online supply chains that support, underpin and drive online applications. We explore the challenges that consolidation raises and argue the resulting need for increased transparency over these supply chains in order to enable appropriate policy responses. Our analysis begins with a brief overview of run-time supply chains (§2), highlighting their complexity and opacity. In §3 we explore how consolidation at this service infrastructure level could lead to substantive societal considerations regarding power and resilience, emphasising that these concerns are particularly relevant as we become ever-increasingly reliant on online and data-driven services. We then (§4 and §5) undertake an analysis considering some technical and legal (data protection) means for exposing service supply chains. This analysis indicates a degree of consolidation within application supply chains, and also highlights the limitations of current approaches for exploring consolidation. We conclude with a discussion of potential ways forward in tackling the lack of visibility in infrastructure supply chains, to enable a better informed policy discourse.

2. Run-time supply chains

Applications are comprised of a range of technologies. They are designed and developed using various software, libraries, languages and development frameworks (often termed a '*technical stack*'), and execute in particular operating environments (operating systems/platforms), on particular hardware (CPU architectures, mobile devices, etc.), and leverage communication infrastructure, often the Internet (e.g. see Charland and Leroux 2011; Jazayeri 2007). In recent years, online services have increasingly formed part of the technology mix that underpin and drive applications. These services, which operate at run-time, support application functionality, delivery and deployment. That is, they provide some technical functionality that supports applications '*as a service*'. 'Cloud' is a collective term often used to encapsulate many such service offerings. Traditionally, cloud is discussed in terms of three service models:

- *Infrastructure-as-a-Service* (IaaS), where the service provider effectively provides the 'servers' for which tenants (cloud users, i.e. application operators) deploy and manage their own platforms and applications;
- *Platform-as-a-Service* (PaaS) where tenants use (aspects of) the provider's stack to manage their own applications; and
- *Software-as-a-Service* (SaaS) where the service provider offers the tenant the application itself.

This categorisation (in which the boundaries can blur) focuses on the degree to which a provider is involved in supporting a tenant's application and the role in which they play in doing so (Millard 2020). In this way, cloud services are understood from the perspective of a tenant, who has a direct relationship and directly interacts with the service provider.

The term *[Anything]-as-Service* (XaaS) is sometimes used to reflect the fact that there are a vast range of services available that support applications (Duan et al. 2015). Some

indicative examples, beyond that of more traditional application storage and compute, include services for: content distribution; identity, authentication and credential management; distributed compute; database management; analytics; logging; machine learning (both model building and using pre-built models, such as for image recognition); and even to assist the integration of emerging technologies such as Blockchain and quantum computing. Applications will use and rely on these services to provide various functionality at *run-time* (during operation).

As such, applications will have an underpinning *service supply chain*, of which data (exchange) is the driver. A common example indicating the composite nature of applications is Dropbox, which for a period relied on Amazon cloud for storage (Millard 2020; Miller 2017). In practice, however, the supply chain underpinning an application can be far more complex, as the number and range of available so-called 'XaaS' offerings indicate. Even a fairly basic application may make use of an external single sign-on service for managing user logins, a customer relationship management (CRM) system, a service for managing application deployment and scalability to give performance guarantees, logging services to help manage security concerns, and so forth.

Of course, many goods and services in the physical world also rely on supply chains, which may themselves be complex, opaque and consolidated. However, supply chains for online applications are somewhat different as they are driven by data flows. As a result, they are more intangible and thus potentially more invisible. Moreover, the nature of data flows in online services and of the online economy more generally means that they are potentially more dynamic and more rapidly formed than real world supply chains. Ultimately, however, comparisons with offline services are somewhat beside the point – regardless of whether real world supply chains are complex, opaque and consolidated, the complexity, opacity, and consolidation of online supply chains are in and of themselves potential problems worthy of attention, as we indicate.

2.1. Complexity

Many applications will have a complex supply chain, representing a *system-of-systems*. Note that this complexity is not only in terms of technical integration, but also regarding the range of organisations who provide and operate the supporting services on which applications rely. This raises further complications, given that each service provider may also rely on the services of others, representing chains of providers and sub-providers — which can entail complex relationships (Millard 2020).

This complexity makes mapping such supply chains challenging. As discussed, these supply chains are data-driven: it is the flow of data between services that drives functionality. But in practice, it can be difficult, at both technical and organisational levels, to discern what and how that data is actually used, processed and transferred once it moves to another administrative domain (Singh et al. 2015).

The complexity of the systems-of-systems that underpin online applications could lead to emergent properties and behaviours that cannot be easily predicted or readily understood (Singh et al. 2018). Essentially, emergence describes a situation whereby a complex system made up of individual component parts leads to properties or behaviours that are not realised by those individual components when operating independently. Emergence as a general phenomenon in complex systems has been observed and described (Mogul 2006).

As a result, systems can become unpredictable and unstable, leading to problems that are difficult to diagnose or correct. Even where the properties or behaviours that emerge are considered to be beneficial or desirous, or simply not worthy of concern, the mere fact of not knowing precisely why they have developed is a problem when it comes to important services or critical infrastructure. Emergence can make complex systems difficult to understand, and it is plainly undesirable to not know why and how these kinds of systems are functioning.

2.2. Lack of transparency

The lack of transparency in service supply chains is in large part a function of the complexity of these interconnected system-of-systems. As discussed, complexity makes technical explanations of data flows difficult, not least given the limited ability to 'see' what happens to data as it moves beyond a technical or administrative boundary. This opacity does not just extend to researchers studying the Internet from the outside; the developers and operators of those services themselves and of other applications that rely on them will also struggle to understand precisely what is going on, as will regulators and policymakers.

Yet transparency over service supply chains can benefit a range of parties: application and service providers have an interest in better understanding the systems and entities that they're dealing with; users have an interest in knowing what's happening in the services that affect them; researchers have an interest in informing debate; and regulators and policymakers have an interest in assessing the risks and societally relevant issues of consolidation.

In any case, there are generally few incentives for companies to provide information about how and by whom services are being used, nor are there strong incentives for application developers to publicise the infrastructure services on which they rely. Indeed, there may be strong *dis*incentives; for instance, commercial considerations (e.g. not wishing to discourage potential users or provide insights useful for competitors) and contractual provisions might preclude making such information available.

The challenges of emergence, discussed above, are amplified also by the lack of transparency over service supply chains. This could make it difficult to identify both the existence and the source of problematic or emergent properties or behaviours, and to assess the impact they may have. Without greater knowledge of the interconnections of infrastructure services, the capacity to respond to problems is limited. While it is likely impossible to eliminate emergent behaviour from such complex systems, transparency would potentially allow for more proactive oversight, seeking to ensure that the problems caused by emergence are addressed and can be investigated as far as possible (Singh et al. 2018).

3. Consolidation considerations and implications

As §2 describes, applications are often built on supply chains involving the services of others. These supply chains can tend towards consolidation, as application operators seek to rely on particular expertise, leverage economies of scale, and so forth. However, as we will discuss in §4 and §5, it can be difficult to get a clear picture of the level of consolidation and interconnection in service supply chains. The lack of transparency in supply

chains may be exacerbated by the fact that, with consolidation, a large number of services may be operating effectively behind the 'closed doors' of a particular company.

Perhaps as a result of this lack of transparency, the service infrastructure that supports applications has been little discussed (the online advertising industry aside (for example, Information Commissioner's Office 2019)). This is despite the fact that consolidation within these run-time supply chains has important implications, not least as they will underpin a range of applications on which society has increasingly come to rely. We argue that such issues therefore warrant consideration by developers, regulators and policymakers alike. In particular, we highlight two relevant implications of supply chain consolidation relating to application resilience and to the power of infrastructure providers.

3.1. Resilience

As discussed above, applications and infrastructure services are dependent on others in a complex, interconnected, and increasingly consolidated system-of-systems. In such a context, failures by one infrastructural component—whether as a result of emergent behaviours or otherwise—can propagate through components located downstream from the point of failure (Singh, Cobbe, and Norval 2019). This can result in problems affecting the range of other components that rely on that underlying infrastructure. Indeed, due to the consolidation of infrastructure services, problems with one infrastructure service can affect many applications. By contrast, with greater diversity in infrastructure, a failure in one infrastructure service would potentially have less severe knock-on effects, as fewer applications would be reliant on that particular service. As a result, applications may be built on top of consolidated infrastructural supply chains that are fragile in nature.

The complexity and lack of transparency of interconnected infrastructure services makes it difficult to know where some problems arise. This is because a failure or other issue may occur at a point far beyond the limits of the visibility available to the designer or operator of a particular component, service, or application relying on those services. As well as this, the potential for emergent behaviours and properties, previously discussed, exacerbates this problem. The lack of transparency and resilience in these interconnected systems-of-systems not only makes it difficult to mitigate against such failures, but, once a failure has occurred, may result in it being difficult if not impossible to diagnose the source of the problem, take corrective action, and restore an application to full functionality.

The failure of the O2 telecommunications network in the UK in December 2018 provides an example of these cascading consequences. A single expired security certificate in Ericsson infrastructure ultimately caused a global outage of 3G and 4G Internet services for O2 customers (BBC News 2018). The same expired certificate also led to a failure of Softbank mobile and fixed-line telephone services in Japan (The Japan Times 2018). Another example occurred in 2016 when a JavaScript library comprising just 11 lines of code was removed from a popular package manager, breaking many larger libraries and websites that directly or indirectly depended on it (Collins 2016).

We can also observe similar issues at higher-level service infrastructures (as is the focus of this paper). For instance, failures in cloud services do occur (Gunawi et al. 2016); a recent example involved an outage in Google Cloud that impacted a range of prominent applications, including Shopify (an e-commerce platform on which many online stores rely), in addition to Snapchat, Vimeo and Discord (Tung 2019). While the effects of these failures

were not catastrophic, the potential for severe consequences as a result of such a failure of critical Internet-based applications is clear.

That said, consolidation could also help improve the resilience of service supply chains. Where actions are taken to improve resilience, such as software updates, consolidation means that the resulting effects have impact at scale. Moreover, larger service providers may have better practices and a better security team than a smaller organisation with fewer resources. However, this is not necessarily the case; no system is perfect, and failures are possible in infrastructure suppliers of all sizes (see Gunawi et al. 2016). Moreover, the larger, better resourced companies do not inherently mean that supply chains are simpler overall. And even the larger companies may make use of a range of other providers to help support their infrastructure.

Greater transparency over service supply chains would improve the ability of both application developers and infrastructure service suppliers themselves to identify developing problems, locate and investigate the cause(s), and take corrective action to repair or restore services. Transparency would also assist in developing policy responses to these issues, which may seek to incentivize the development of mechanisms for improving the resilience of service supply chains.

3.3. Corporate power and accountability

It has been acknowledged for some time that architecture can constrain, restrict, permit, and otherwise have a significant influence on behaviour, giving it a regulatory role (Winner 1985). In the online world, as Lessig argues, code in many ways supplants the physical architecture of the real world as a dominant force influencing behaviour (indeed, Lessig describes 'the architecture of the Net, or its "code"' (Lessig 2006)). Control of the code-driven infrastructure underpinning online applications gives the corporations involved significant regulatory power—that is, the ability to 'durably affect or constrain the behaviour of others' (Delacroix 2019)—which can be expressed in various different ways.

Consolidation of infrastructure services means that a relatively small number of corporations are in a position to exercise that power with significant influence (Belli and Venturini 2019). This potentially allows those corporations to act as gatekeepers, raising the risk of such a corporation abusing its position to deny certain actors access to key services. It also raises the prospect of a corporation with a dominant position in a market artificially raising the price of a product or service beyond that which would be possible if there was effective competition. Consolidation among infrastructure services could work to influence and effectively define the nature of applications, by controlling the functionality offered and the means by which it is provided.

The provision of infrastructure services is therefore inherently value-laden. This is the case even when it is apparently done so neutrally, to all customers, without exception. As has been seen in the recent developments around the website 8chan (The Guardian 2019b), infrastructure providers are increasingly grappling with how to navigate this tricky terrain. 8chan prioritised user anonymity and the ability to communicate in largely unmoderated spaces. As a result, it became a forum for neo-Nazism, white supremacism, hate crimes and child sexual abuse (The Guardian 2019a). After several violent incidents were linked to 8chan, including multiple mass shootings, Cloudflare felt that continuing to provide infrastructure services to the site was unsustainable, and 8chan was

removed as a customer (Cloudflare 2019). Without Cloudflare's DDoS protection, 8chan was unable to reliably remain online.

The ability to offer or deny service to particular websites (even where it may seem to many people to be a straightforward choice) illustrates that service providers have significant power. This is particularly the case where services are heavily consolidated, with relatively few corporations having such a position. However, there are few legal restrictions and little motivation for infrastructure providers to account for their decisions or for how they exercise this power.

While the open nature of the Internet mitigates these possibilities to an extent, given that as others can develop and provide alternatives, the resources required to offer alternatives at scale, as well as the market power of dominant infrastructure providers, represent barriers. Greater transparency over service supply chains would give policy-makers and regulators a more accurate idea of the influence wielded by the corporations involved, potentially allowing for the development of legal and regulatory responses that seek to improve the accountability of those corporations.

4. Uncovering supply chains: a technical analysis

Often a range of services comprise an online application's supply chain. Common examples of such services include cloud computing, storage, content distribution, e-commerce, media, user analytics and advertising, to name but a few. While application developers/providers will themselves have some knowledge of the application's underlying infrastructure services (at least with respect to those with which they directly interact), many of those underlying services are not user-facing. This means that, at least from a technical perspective, information about the components and the organisations involved in supporting an application or service is often not readily available to users or other interested parties.

In light of the above discussion, we now examine technical approaches in an attempt to uncover the nature of system supply chains. We outline some methods from the literature, and conduct our own analysis focusing on the degree to which information about the consolidation of online infrastructure services can be obtained. Our findings indicate that while there does appear some consolidation in (web) application supply chains, shown by way of identifying a number of prominent and dominant firms, the technical methods for uncovering system supply chains provide only a limited picture.

4.1. Visibility over data supply chains

When a user interacts with an online application (be it a mobile 'app', website, etc.), that application will typically direct or manage the *interactions* (data exchanges) with the infrastructural services on which the application directly relies. In some cases, this will involve the users—through their device, web browser, etc.—directly interacting with some of the supporting services that comprise part of the application's supply chain. For example, in some cases of content distribution, user authentication or audience analytics, these services obtain data directly from, or provide information directly to, users at the instigation of a particular application. Where there is a direct interaction, i.e. exchange of data, between the user and a service, it is generally possible to discover (to some degree)

that the service is involved in the application's supply chain. This, for instance, by monitoring or logging the communications taking place between the user and that service.

In this way, these services are notionally 'visible' from the user's perspective; one can see that data flows to and from those services by virtue of directly interacting with them (see Figure 1). This visibility has been leveraged to explore the use of third-party services in supporting a variety of different kinds of application (Binns et al. 2018a, 2018b; Englehardt and Narayanan 2016; Libert 2015; Starov and Nikiforakis 2017).

However, there are also cases where there is no direct interaction between the user and other services supporting that application. That is, once user data flows to a service, the subsequent data flows beyond that point are typically invisible from the user perspective (and *vice-versa* regarding the information a user receives). The lack of a direct interaction between the user and these 'deeper' supporting services represents a key challenge when attempting to technically uncover application supply chains.

4.2. Measuring user-oriented data flows

As discussed, it is possible to identify some of the third-party services supporting an application by recording and analysing the communications being made. In a web context, these communications are typically through the HyperText Transfer Protocol (HTTP) (Mozilla 2019)[2], involving *request* and *response* messages between a user's device (typically via a browser or app) and a server.

Since these HTTP messages travel to and from the user's system, monitoring these messages represents a means for identifying and measuring the use of certain third-party services that support online applications. Measuring these communications across many applications can reveal the prevalence of particular services and providers, and thereby indicate the level of consolidation within the applications' supply chains.

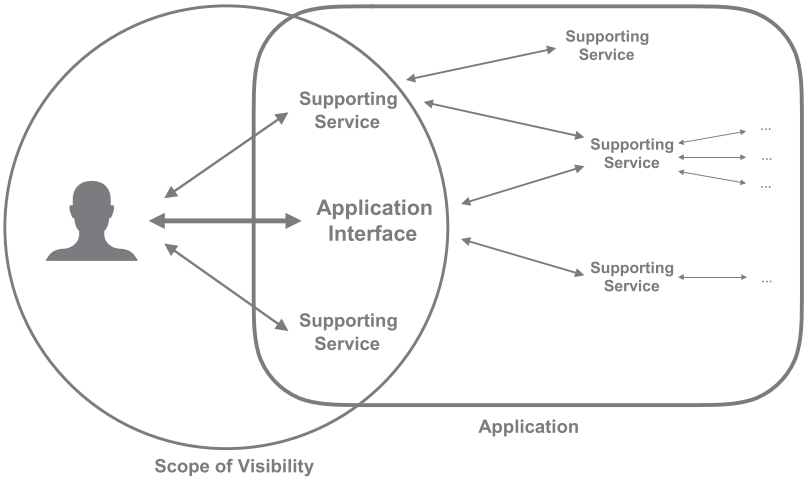

Figure 1. An example illustrating the visibility of user interactions with third-parties as they use an application. This represents the level of visibility for those methods that explore application interactions with third-party services.

We now explore some relevant literature that considers the HTTP communications of applications to identify interactions with third-parties. We then use similar data capture methods to undertake a more consolidation-oriented analysis regarding the underpinnings of popular websites.

4.2.1. Prior work on third-party domain call-outs

Much of the literature on identifying third-parties from HTTP traffic focused on behavioural tracking and privacy. The work of Libert (2015) and Englehardt and Narayanan (2016), for example, analysed the HTTP messages with third-parties from across approximately one million websites. In both studies, the researchers used an automated process to visit each of the websites, and analysed the third-party domain names[3] involved when visiting each site. By doing so, these papers set out to investigate and identify the key domains and organisations involved in online tracking and advertising.

Libert found that the majority of websites involved interactions with a third-party service – 88 per cent of the websites analysed had initiated HTTP requests between the user and one or more third-party domains (the average site making requests to around 9.5 third-party domains). HTTP requests to Google-owned domains (which include google.com, google.co.uk, youtube.com and others) were made by around 78 per cent of the websites – results which reflect Google's dominance. Other prevalent organisations included Facebook (requests to a Facebook-owned domain were made by approximately 32% of websites), Akamai (23%), and Twitter (18%). Englehardt and Narayanan (2016) also found evidence of Google's influence, finding that 12 of the top 20 third-party domains to which requests were being made were Google-operated domains, and that Facebook, Twitter, Amazon, AdNexus and Oracle were present on more than 10 per cent of sites analysed.

While both of these studies focused on websites, similar work has also explored and analysed the HTTP interactions with third-parties in browser extensions[4] (Starov and Nikiforakis 2017) and Android applications (Binns et al. 2018a, 2018b). As with websites, third-party domains operated by Google were found to be prevalent, as were domains operated by Amazon and Facebook.

4.3. A focus on consolidation

The results of those studies suggest a degree of consolidation regarding the third-party organisations associated with applications. However, that work did not focus on exposing consolidation *per se*. We therefore conducted a similar analysis to focus on the potential for such methods to identify consolidation; indicate some types of services comprising application supply chains and where consolidation can occur; and the limitations of such approaches for exploring consolidation in system supply chains.

4.3.1. Interactions with third-party domains

For our analysis, we identified a list of websites to observe by combining the Alexa top 500 rankings[5], the Moz top 500 list[6], and the top 1,000 from the Majestic Million list[7] as of July 2019. This left us with a list of 1,146 unique websites to investigate. We wrote a Python script to systematically visit each website, and collect information about the HTTP communications taking place.[8] Specifically, we captured the URLs for all outgoing HTTP

messages (e.g. the Google Analytics JavaScript: https://www.google-analytics.com/analytics.js), and the HTTP header data (additional information from the communication protocol) from the interactions with 1,127 websites.[9]

Generally, our findings are similar to that of the work discussed above. Of our 1,127 websites, the vast majority (89.3%) made requests to other domains (i.e. a different domain to that of the site being analysed). In total, we observed interactions with a total of 1,628 distinct third-party domains. Table 1 presents a list of the top 50 most prevalent domains that were accessed by the websites. From this we see that Google was prominent: google-analytics topped the list, and services operated by Google (Alphabet) made up eight of the top 10. Many of the prevalent domains appeared to be advertising and/or analytics oriented – we observed that many sites used one or more of such services, which indicates the prevalence of advertising within online ecosystems. However, we also see some infrastructure services including content distribution and cloud computing throughout the list.

Note, however, that the data of Table 1 represents an application resulting in a direct, user-oriented interaction with a particular domain. Importantly, this does not necessarily mean that these services are there for a particular purpose or operate in a particular way, and further, that the lack of a direct interaction with a particular domain does not imply that a particular service is *not* being used somewhere within the broader system supply chain. For example, 7.4 per cent of sites were observed making HTTP requests directly to amazonaws.com, belonging to Amazon Web Services (AWS). As a major cloud provider, we would expect many more applications to use AWS. However, the 7.4 per cent only reflects the number of applications that resulted in a direct user interaction with that particular Amazon domain. The numbers from the domain analysis would not capture, for instance, that a service provider could offer different services through different domains, and that applications (and third-party services) may be hosted or manage interactions with AWS services through their own domain. Indeed, on a deeper analysis, we saw evidence indicating that Amazon AWS was far more involved in application supply chains (as discussed below).

4.3.2. Consolidation within different types of third-party services

The supply chain underpinning an application can involve a range of tools and services, both involved directly (i.e. used by the application itself) or indirectly (used by a service on which the application depends). Some key categories of these types of services are presented in Table 2, alongside the typical role they play in supporting an application. We now consider some of these categories with respect to our data, and discuss evidence of consolidation which could be observed:

Advertising and Analytics: Advertising and analytics services both involve data collection about the user's browsing habits. From Table 1 we see that many different advertising and analytics services were prevalent, suggesting a number of dominant firms in the advertising and analytics ecosystem. Specifically, we see that Google dominated this list, taking eight of the top 10 places. Google Analytics was the most prominent third-party domain we observed by way of requests, while Google's DoubleClick platform was the most frequently occurring advertising domain. Facebook and other advertising entities also featured prominently. Our findings, consistent with that of other research

Table 1. The 50 most frequent interactions with (call-outs to) third-party domains from 1006 web applications. Note we broadly categorise these sites to provide context; a given domain may serve several purposes.

	Domain	Operated by	Purpose	Prevalence # sites/1006
1	google-analytics.com	Google (Alphabet)	Analytics	679 (67.5%)
2	doubleclick.net	Google (Alphabet)	Advertising	669 (66.5%)
3	google.com	Google (Alphabet)	Search & Other	632 (62.8%)
4	google.co.uk	Google (Alphabet)	Search & Other	548 (54.5%)
5	googletagmanager.com	Google (Alphabet)	Analytics	498 (49.5%)
6	googleapis.com	Google (Alphabet)	Developer tools	429 (42.6%)
7	gstatic.com	Google (Alphabet)	Offload static content	418 (41.6%)
8	facebook.com	Facebook	Social Media & Other	382 (38.0%)
9	facebook.net	Facebook	Social Media & Other	354 (35.2%)
10	googletagservices.com	Google (Alphabet)	Analytics	250 (24.9%)
11	googlesyndication.com	Google (Alphabet)	Advertising	224 (22.3%)
12	googleadservices.com	Google (Alphabet)	Advertising	205 (20.4%)
13	adnxs.com	AppNexus (Xandr)	Advertising	199 (19.8%)
14	scorecardresearch.com	Full Circle Studies (Comscore)	Analytics	190 (18.9%)
15	cloudfront.net	Amazon	Content Distribution	181 (18.0%)
16	twitter.com	Twitter	Social Media	178 (17.7%)
17	adsrvr.org	The Trade Desk	Advertising	160 (15.9%)
18	everesttech.net	Adobe	Advertising	150 (14.9%)
19	rubiconproject.com	Rubicon Project	Advertising	147 (14.6%)
20	demdex.net	Adobe	Analytics	143 (14.2%)
21	casalemedia.com	Casale Media	Advertising	125 (12.4%)
22	amazon-adsystem.com	Amazon	Advertising	123 (12.2%)
23	openx.net	OpenX	Advertising	122 (12.1%)
24	ampproject.org	Google (Alphabet)	Web Development	119 (11.8%)
25	quantserve.com	Quantcast	Analytics	117 (11.6%)
26	youtube.com	Google (Alphabet)	Multimedia	114 (11.3%)
27	yahoo.com	Yahoo! (Verizon)	Search & Other	112 (11.1%)
28	pubmatic.com	PubMatic	Advertising	107 (10.6%)
29	linkedin.com	LinkedIn (Microsoft)	Social Media	105 (10.4%)
30	newrelic.com	New Relic	Analytics	104 (10.3%)
31	nr-data.net	New Relic	Analytics	104 (10.3%)
32	rlcdn.com	RapLeaf (TowerData)	Analytics	104 (10.3%)
33	bing.com	Microsoft	Search & Other	103 (10.2%)
34	mathtag.com	MediaMath	Advertising	102 (10.1%)
35	bidswitch.net	BidSwitch (IPONWEB)	Advertising	101 (10.0%)
36	ytimg.com	Google (Alphabet)	Content Distribution	101 (10.0%)
37	ads-twitter.com	Twitter	Advertising	98 (9.7%)
38	cloudflare.com	Cloudflare	Content Distribution	98 (9.7%)
39	criteo.com	Criteo	Advertising	98 (9.7%)
40	t.co	Twitter	Social Media	97 (9.6%)
41	consensu.org	IAB Europe	Cookie Notice	94 (9.3%)
42	adobedtm.com	Adobe	Analytics	88 (8.7%)
43	criteo.net	Criteo	Advertising	87 (8.6%)
44	hotjar.com	Hotjar	Analytics	81 (8.1%)
45	turn.com	Amobee	Advertising	81 (8.1%)
46	omtrdc.net	Adobe	Analytics	80 (8.0%)
47	adform.net	Adform	Advertising	78 (7.8%)
48	moatads.com	Moat (Oracle)	Analytics	77 (7.7%)
49	advertising.com	Advertising.com (Verizon)	Advertising	75 (7.5%)
50	amazonaws.com	Amazon	Cloud Computing	74 (7.4%)

(Englehardt and Narayanan 2016; Libert 2015), indicate some consolidation in the area of advertising and analytics.

JavaScript Libraries: We observed that requests for JavaScript files were called by over 93 per cent of the websites observed, with 85 per cent of websites requesting JavaScript files from third-party domains. While many of these were related to the advertising and analytics

Table 2. A selection of different types of services that can support online applications.

Category	Description
Advertising	Provides adverts to the users of the web application.
Analytics	Records information about the users' habits when using the web application.
Authentication	Allows users to log in to the site using their account details from another service.
Content Delivery Network (CDN)	A collection of servers which redistribute web content to users according to demand and geographic location.
Content Management System (CMS)	An application to create and manage web content.
eCommerce	A platform for managing online commerce, sales, etc.
Framework	A software framework to assist developers in the creation and deployment of web applications.
HTTP Server	Software which manages the distribution of files on web server hardware.
JavaScript Library	A script executed within the user's browser which can provide more sophisticated functionality to the web application.
Operating System	Software installed on the server which manages its hardware and software.
Payment	A platform for managing monetary transactions online.
SSL Certificate Provider	A service which provides a digital certificate, authenticating a website and enabling encrypted communications.
Web Hosting Provider	A service which provides infrastructure for hosting web applications.

services above, there were also requests for other types of libraries; e.g. jQuery—a popular library for simplifying a number of common JavaScript tasks—was requested by 46 per cent of websites. While many of these sites (60%) requested the jQuery library from a third-party domain, we also observed that the library came from a range of different sources – with over 100 unique domains serving the jQuery script. In other words, despite the prevalence of this library, its source was highly distributed. Components 'distributed' as such differ from other supporting services that are more clearly under the direct control or management by a particular entity, and may raise additional considerations.

Hosting providers: A website may have the technical infrastructure provided by one or more cloud-based services (virtual machines/servers, file storage, databases, etc.). From an external viewpoint, it may not always be obvious which organisations are involved in hosting a website or online resource. This is because the user typically only interacts with the domain name of the website as opposed to the supporting host(s); e.g. a user visits netflix.com, despite Netflix making extensive use of Amazon AWS services (Amazon 2017). As such, we also investigated the *headers*[10] of HTTP interactions in order to have more information beyond just that of the URL being accessed. We observed that in some instances the hosting services were reported within the optional parameters of HTTP headers.

Looking throughout the visible supply chains of each site, we could explore the use of third-party hosting providers on which a given website relies (i.e. used either by the site themselves or by a third-party resource on which each site relies). Unsurprisingly, Google was prevalent, supplying content to 79 per cent of our 1,127 sites. Similarly, Amazon S3 (storage) could be identified in the supply chains of 53 per cent of our sites, Microsoft Azure by 29 per cent, and Fastly by 13 per cent.

Again, in terms of consolidation, it appears that the services of a small group of companies support a large number of sites. Even in instances where a particular application does not itself use a particular web hosting provider, that does not mean the provider is not involved somewhere in the application's supply chain. This is because that provider could be used by other third-party services on which the application relies.

Content Delivery Networks (CDNs): Similar to web hosting providers, looking at the domains being accessed does not itself provide a clear indication of CDNs being used (again, given that the user will interact with the domain of a website, such as netflix.com, rather than its underlying infrastructure services). However, we can once again explore the HTTP headers in order to gain some insight into the use of CDNs. In some cases, CDNs were named within the server header (e.g. CloudFlare and Akamai), and we also observed optional (e.g. 'x-cache: Hit from CloudFront') or custom headers (e.g. 'x-akamai-transformed') being used. By analysing the supply chain of each site, we observed that 54 per cent of our sites relied on content distributed through the CloudFront CDN. Similarly, CloudFlare and Akamai were present within the supply chains of 30 per cent and 33 per cent of sites respectively.

HTTP servers and Content Management Systems (CMSs): The server field in a HTTP response often reported information about the HTTP server software used. We observed some 60 per cent of web applications reporting the use of one of two servers: NGINX or Apache, used by 35 per cent and 25 per cent of sites respectively. Regarding CMSs, we identified the use of WordPress and Drupal by 171 sites (88 sites (8%) and 83 sites (7%) respectively) by analysing URLs and HTTP headers.

While there appears a degree of consolidation with regards to HTTP servers and CMSs, note that the examples mentioned represent open source software. Again, this may raise different policy considerations, given that the technology may not be managed and operated by an organisation in the same way as they might be, for example, through the use of a particular cloud service, and because open source can be customised and extended.

4.3.3. Service types in practice: single domain analysis

Looking to how these components and services can come together to support a website, we now outline an illustrative 'deep-dive' into one application: www.thesun.co.uk, a UK-based news website. We undertook a manual exploration to indicate its composition.[11]

When visiting www.thesun.co.uk we recorded many HTTP requests to third-party domains. The third-party HTTP requests included calls to an extensive list of **advertising** (including Pubmatic, DoubleClick, and AppNexus) and **analytics** (such as Google Tag Manager, Krux Analytics, and RapLeaf) domains. The site also requests a number of **JavaScript** files, including the Facebook SDK for **social integration**, and for https://www.thesun.co.uk/wp-content/themes/thesun/js/promise.min.js, indicating the use of WordPress as the **CMS**. Inspecting the headers of the responses indicated more information about the infrastructure, including that the **web hosting provider** was WordPress VIP, the **HTTP server** was NGINX, and we observed the use of Akamai as a **CDN**.

We also observed different third parties being included while navigating to different sections of the site. Visiting news stories initialised call-outs to Spot.IM for comments, and login.thesun.co.uk requested scripts from Auth0 for **authentication**. The page help.thesun.co.uk indicated within response headers that it used SalesForce as a Customer Relationship Manager (**CRM**), and requested **JavaScript** files such as jQuery and Bootstrap. We also noted that mypreferences.thesun.co.uk used Amazon CloudFront – a different **CDN** to that of the homepage.

Generally, the site was dynamic, whereby the use of third-party services depended on the webpage or subdomain that the user was visiting. This illustrates the potential complexity and multi-party nature of application supply chains, which can involve numerous

services which may or may not always be observable without deeper investigation. In terms of consolidation, it highlights some of the challenges faced when attempting to understand the entities involved when visiting a website or using an application.

4.4. The technical challenges in uncovering consolidation

Our analysis aligns with that of related work, and confirms a general intuition, by finding that (a) applications entail a complex supply chain, and (b) a small number of corporations play a significant role in the supporting of popular online applications. While providing some evidence of consolidation in online infrastructure service supply chains, our results are only indicative. There are a number of challenges in technically uncovering the nature of supply chains, as we now explore. We argue that more is needed to enable a more accurate measurement of consolidation in online services.

4.4.1. Limitations of analysis

We observed several limitations of measuring data flows between users and third-parties as a means of assessing the consolidation of service supply chains.

HTTP messages are only indicative of technologies used: Much of our analysis involved identifying the use of technologies through their HTTP messages, either through analysing the third-party domains (URLs) being requested, or by exploring the header information within the HTTP responses being returned. While indicative, in both cases these are insufficient for providing accurate measurements of consolidation. The former approach means that many infrastructure providers are masked, hidden behind the domains with which there is a direct interaction. For the latter, we often relied on optional HTTP headers, which are dependent on the way that site or associated services are built and configured. As the nature of the headers can vary, their utility as reliable indicators are limited.[12] Further, undertaking the analysis involved a degree of intuition and inference on behalf of the researchers. As such, these approaches do not provide a reliable way of exploring consolidation within application supply chains.

Third-parties contacted may vary: Our data was captured by visiting the homepage of each site once. However, previous work has found that the third-party services called by a particular website can often fluctuate from visit to visit (Sørensen and Kosta 2019). The third-parties involved in a supply chain could vary greatly depending on the part of the website being accessed (as we saw in §4.3.3), the user's physical location, whether they are logged in or profiled, and what they are attempting to do on the website, etc. In other words, the third-party services invoked by a website can be highly contextual, and automating the analysis of these application interactions will often be limited in coverage.

Identifying organizations and their role in the supply chain is difficult: While we identified a large number of third-party domains being contacted, it wasn't always possible to identify (i) the organisation behind that domain, nor (ii) the nature of their service in the context of the particular application.

One challenge is that a given organisation may operate a number of domain names. For example, Google (Alphabet) operates well-known domains like youtube.com, but also more technical domains which may be less recognisable, such as 1e100.net (Google 2019b). Particularly for smaller, less well-known organisations, it can be difficult to map

a domain to the organisation which operates it, which results in opacity. Some in the literature created annotated lists of organisations and their associated domains (e.g. Libert 2015, 2018), and others used WHOIS records in an attempt to identify organisations (Binns et al. 2018b).[13] We explored both of these options, and while helpful on occasions, neither were consistent and reliable in identifying the organisations operating a domain in order to investigate consolidation.

Moreover, even identifying that a given domain is involved in a website's supply chain does not necessarily indicate the extent to which it is involved, nor the services being provided. For example, we observed a large number of call-outs to facebook.net, but given that Facebook operates a number of services (social media integration, authentication, developer tools and advertising, etc.), knowing what services are relied upon may not always be clear.

4.4.2. The technical challenges in uncovering supply-chain transparency

As outlined in §4.1, an inherent limitation of the methods that we have used is that they can only typically detect third parties with whom the user has a direct interaction, i.e. by way of a network request. However, there are also infrastructural services that do not interact directly with the user, but instead only with the application (or with other infrastructure services). Common examples might include storage, databases, and so on. Analysis predicated on measuring communications between users and third parties will not reveal these services, instead only detecting the direct ('first-hop') interaction—i.e. between users and third-parties—in the supply chain. This means that the other infrastructure providers remain hidden from view. Aside from the wider issues of consolidation and transparency, this also has the potential to cause problems to service providers themselves, as we discussed in §3.1, such as for risk management. Applications may come to rely on third-party services, libraries, etc. beyond their (and their operator's) scope of visibility, which represent points of failure where issues can propagate throughout and across systems.

In short, while there are some technical methods that can assist in shedding light on supply chains, these only appear capable of providing a partial view over the technical supply chain underpinning an application. Nevertheless, the results of such analyses are indicative of some consolidation taking place. From a technical perspective, more transparency would assist in gaining a better understanding of the implications of infrastructure consolidation. Provenance-based approaches to tracking data flows throughout systems may indicate one potential way forward (Singh, Cobbe, and Norval 2019). However, any such solutions need more than just technology; important are the motivations and incentives for application and service providers to make their supply chains more transparent in the first place. Legal mechanisms may offer some assistance in this regard, which we explore next.

5. Legal mechanisms for investigating infrastructure supply chains

Beyond technical means, some legal mechanisms exist that may provide a degree of information about infrastructure supply chains through transparency obligations. In investigating these, we found that these are generally limited in application and utility.

We note that the legal frameworks applying directly to the providers of some types of infrastructure services—such as the EU's Networks and Information Systems Directive (European Union 2016a)—do not typically provide for the kind of transparency that

would assist in investigating technical supply chains or their consolidation. But other frameworks that provide for transparency in technical systems may be of benefit.

In particular, data protection law appears to be one route for gaining more information about these supply chains. Data protection law, such as the EU's General Data Protection Regulation (GDPR) (European Union 2016b), governs the processing[14] of personal data (that is, any information about an identified or identifiable natural person[15]) and typically provides for individuals to be adequately informed about what is happening with their data and with whom it is being shared. Since supply chains often involve a flow of personal data between applications and services deep into the chain of system-of-systems that support them, these transparency requirements in data protection law should, in theory, provide a mechanism for investigating these data flows – not just at a higher level of the supply chain, but at multiple levels.

However, while personal data is defined broadly in GDPR[16], it is possible that much of what will flow through supply chains between applications and infrastructure services will not fall within that definition. As GDPR would not apply to that processing, neither would its transparency mechanisms.

Where infrastructure services *do* process personal data and therefore come within the remit of data protection law, GDPR would typically consider infrastructure providers to be data *processors*[17] who act under the instruction of the providers of applications[18] (who are considered to be data *controllers*; the entities who determine the means and purposes of the processing).[19] Multiple controllers, processors and subprocessors may be involved in a given supply chain, and therefore, GDPR's various transparency mechanisms could potentially provide insight into the data flows between them.

Some of these transparency mechanisms relate to data subjects (the individuals to whom personal data relates – *users* in this context), and oblige data controllers to provide them with certain information where their personal data is obtained or upon request. Other mechanisms relate to the relationship between data controllers and data processors, and to obligations to provide information to data protection regulators for oversight purposes. Some of these transparency mechanisms will be of less utility to considerations of consolidation than others. We now consider these in turn.

5.1. Transparency to data subjects

GDPR provides that data controllers should give certain information to data subjects at certain points. This includes, among other things, an obligation to inform data subjects of the recipients or categories of recipients of personal data, if any, and of the purposes for its processing. Potentially, this could allow data subjects to be informed not only of which infrastructure providers are receiving their data throughout all levels of the supply chain, giving them some idea of the data flows involved, but also to have some insight into which kinds of services those providers are offering. GDPR establishes obligations to provide data subjects with this information upon first obtaining the personal data[20] (in practice, typically through privacy policies), and to provide this information to data subjects where it is subsequently requested[21] (typically through a 'right of access' to personal data, exercised through a subject access request).

To investigate consolidation in online services supply chains by obtaining information about data flows, we analysed the privacy policies of several websites and exercised the

right of access in relation to a selection of prominent applications. We found that, in practice, these legal mechanisms were generally insufficient to provide the kind of transparency necessary to understand the data flows or supply chains underpinning an application, or consolidation in any detail, as we now discuss.

5.1.1. Privacy policies and transparency requirements

Privacy policies contain information about how the company will treat and disseminate customer information, which may help consumers decide whether to disclose information (Metzger 2007). Policies such as these are often required under data protection and privacy laws, including GDPR. While one might expect that privacy policies would outline the third parties involved in the delivery of an online application or service, Libert (2018) studied the policies of approximately 1 million websites looking for the names of third parties identified through the monitoring of HTTP requests. They found that only 14.8 per cent of data transfers to identified third parties were disclosed, with data transfers to Google being disclosed 38.3 per cent of the time, whereas those 'without consumer-facing services' were said to have an average rate of disclosure below 1 per cent. While these HTTP requests likely involved some form of personal data (such as a user's IP address), it is worth noting that some third party service providers within an application's supply chain may not deal with personal data. In such cases, there is no obligation under the GDPR for those services to be disclosed in the application's privacy policy.

To gain a wider understanding of the extent in which third parties are disclosed in such documents, we manually explored the terms of service and privacy policies of 12 popular websites.[22] The websites were selected to represent a broad range of services, selected for (i) making a large number of third-party requests, as determined from our technical work in §4; (ii) having a high (global) Alexa web ranking[23]; (iii) for being UK-based (thus subject to the GDPR and where policies are written in English) and having a high (UK) Alexa web ranking (and therefore, given their prominence, likely familiar with the GDPR); or (iv) having a high Alexa web ranking within a health-related category[24] and thus an example where information can be particularly sensitive.[25] These 12 websites are listed in Table 3.

In most cases, we found that the names of third parties being used were not disclosed in either the terms of service or privacy policy documents – though often a list of subsidiary organisations (from the main organisation) was provided. Rather, the websites tended to rely on providing (fairly) vague categories of third parties, such as for advertising, analytics, social media, and those which 'provide, manage and improve our websites'.

Information about other third-party organisations providing technical infrastructure was similarly vague. A representative example includes '[data may be shared] with others (companies, contractors) to provide services relating to technology, data analysis, research, … '. Such statements provide little visibility over an application's technology stack, nor do they facilitate one in determining the third-parties involved in its supply chain. That said, on a few occasions, some external organisations were explicitly named where they related to tracking and advertising; we observed that Google Analytics appeared to be directly named as a third party on a few sites.

In all, we found that privacy policies as currently used are insufficient, providing inadequate information about an application's use of third-party services. As discussed,

Table 3. The 12 websites for which we investigated their terms of service and privacy policies.

Website	Type	Selection Criteria
variety.com	Entertainment Magazine	Large number of third-party requests
thesun.co.uk	News	
timesonline.co.uk	News	
google.com	Search	High Alexa Ranking
facebook.com	Social Media	
baidu.com	Search	
bbc.co.uk	News & Entertainment	UK-based and High Alexa Ranking
ladbible.com	Entertainment	
gov.uk	Government	
nih.gov	Government	High Alexa Ranking (Health)
webmd.com	Health	
mayoclinic.org	Health	

examining privacy policies seem promising in theory since they require information to be provided about the recipients of personal data and the purposes for its processing. This seems to provide a mechanism for investigating data flows through the infrastructure service supply chain. But, in practice, the privacy policies we studied indicate that insufficient information is provided on personal data flows to determine which infrastructure services are used by the applications in question, even at the surface level, let alone deeper into their supply chains.

5.1.2. Right of access

As discussed, GDPR also provides data subjects with the right to obtain from data controllers, among other things, information relating to 'the recipients or categories of recipient to whom the personal data have been or will be disclosed'.[26] To explore whether third-party services would be disclosed (either by name or by category), we submitted right of access requests to a selection of prominent online applications. These requests specifically asked for information about 'the recipients or classes of recipients to whom that data may be disclosed'. We found that, in most cases, data controllers provided only the personal data held, and often did not provide extra information on recipients to whom that data might have been disclosed, aside from that provided in their privacy policies. As was also the case with privacy policies (above), some data controllers provided information only on the categories of recipients of personal data, and did not enumerate specific recipients.[27] This indicates that while access requests seem to be useful transparency mechanisms, investigation of supply chains is hindered where data controllers fail to respond in full to requests (which may put them in breach of their compliance obligations under GDPR).

That said, informed users, civil society groups and academic researchers may still find the right of access useful for investigating supply chains. Prior work has shown that information about a company's technological infrastructure (e.g. database schema, software being used, etc.) could be derived from subject access requests (Singh and Cobbe 2019). While the right of access might not provide complete visibility over the entire supply chain of third-party services, it may provide insights into some technological components that are in use. As such, this approach may assist in building an overall picture of the consolidation landscape.

5.1.3. Limitations

Although privacy policies and the right of access appear to be obvious routes to getting a better idea of the infrastructural supply chains of applications who are acting as data controllers, our research shows that they may in practice provide only limited benefit. An initial issue stems from the fact that, by virtue of these being data protection obligations, data controllers are only obliged to provide information about flows of *personal data*. Infrastructure services that do not process personal data would lie outside of the reach of these mechanisms.

Where personal data is involved, the utility of these transparency mechanisms is restricted by the fact that they rely on individual obligations – that is to say, information is to be provided only to *data subjects* where their personal data is being processed by a data controller. Data controllers are only obliged to tell data subjects about the recipients of any personal data relating to them specifically, rather than about their supply chains more generally. If the personal data of other individuals is processed in a different way (e.g. perhaps through different use of the application in question), then the supply chain for that processing will itself be different.

Moreover, these obligations only extend to informing data subjects about *categories* of recipient, rather than naming specific organisations, which in practice, can be vague and of limited utility. If data subjects are provided with information on the categories of recipient with which personal data is being shared (and only those that receive personal data), this would not be enough information to determine which infrastructure services in particular are involved. These mechanisms do not therefore establish a mechanism by which a more systematic investigation into the infrastructure supply chains of online applications may be undertaken.

5.2. Data controllers and processors

While privacy policies and the right of access are of only limited practical utility in many cases, GDPR does provide for some other requirements for the relationship between data controllers and processors that may be of more use in revealing the consolidation of service supply chains. As these controller/processor relationships (as well as processor/sub-processor relationships), as defined by GDPR, exist throughout service supply chains, these requirements may provide a way to investigate multiple levels of an application's underpinning system-of-systems.

Specifically, GDPR establishes that data controllers (in this context, primarily application developers and providers) may only lawfully use data processors (in this context, primarily infrastructure services) who provide sufficient guarantees that their processing will be undertaken in accordance with GDPR.[28] Additionally, GDPR requires that both data controllers and processors take technical and organisational measures to ensure the security of any personal data for which they are responsible.[29] And GDPR also establishes that data processors are not themselves permitted to contract with another data processor (as a sub-processor) to undertake processing on their behalf without written authorisation from the controller[30] – for instance, in this context, infrastructure providers would not be able to rely on the services of another infrastructure provider without written authorisation from the application developer in question.

In order to allow data controllers to ensure compliance with these provisions, GDPR contains several transparency mechanisms. Primarily, this involves requiring that data

controllers establish certain contractual obligations with data processors, which includes obliging them to comply with various obligations.[31] For example, processors have to provide to the controller any information necessary to demonstrate compliance and to facilitate auditing and inspections of their activities by the controller.[32]

The result of this is that if application developers establish those contractual obligations with processors, as required by GDPR, then this may provide a route towards gaining greater knowledge of the nature of the infrastructure services on which they are relying. *That is, a supply chain could be exposed by 'following' the contractual chain*. For example, where a data processor has no need to contract with further processors in order to provide particular functionality, this may indicate that the processor in question provides that functionality in-house, which, depending on the functionality in question, may be less likely with less consolidated supply chains.

However, the information provided by infrastructure providers to application developers by way of contracts may, in many cases, be only of limited use, even to application developers in understanding their supply chains. For instance, providing more general information on which categories of sub-processor are being used by the processor, would likely not in itself reveal the full complexity or consolidation of supply chains. Additionally, the obligations described here may not enable one to 'see' beyond one or two layers (i.e. only with those with which there is a direct contractual relationship), leaving application developers with little knowledge of what is occurring further downstream.

5.3. The role of data protection regulators

GDPR also contains powers for data protection regulators to require certain information from data controllers.[33] This includes a wide-ranging power to obtain any information from data controllers that is necessary for the performance of the regulator's tasks[34], and also includes the power to audit controllers.[35] While data protection regulators would therefore appear to be best placed to investigate the consolidation of service supply chains, given their auditing powers, this is not their function, and such investigations would typically be beyond their remit.

Further, the information available to data protection regulators would likely not be available to entities other than the supervisory authority itself, despite the fact that those other entities may themselves have an interest—not only corporate competitors, but perhaps as part of a public, journalistic or academic function—in having more information about infrastructure supply chains. Nor would it be available to other regulators or oversight bodies who would benefit from greater information about infrastructure supply chains.

Considering the above, it is clear that while GDPR does establish some requirements that may be of some use, its benefit in providing visibility over supply chains is limited. As discussed previously, the transparency mechanisms for data subjects are limited in practical utility. And the mechanisms available for application developers (as data controllers) and data protection regulators to assess the consolidation of infrastructure services and supply chains are not necessarily mechanisms by which a holistic view of service supply chains can be obtained. Nor are they mechanisms that may be of use to interested parties other than those entities, including regulators, who may have a particular interest in getting a better understanding of supply chain consolidation.

Moreover, as noted previously, the transparency mechanisms provided for by GDPR—whether in relation to data subjects or to data controllers and processors— only exist where the data being processed in the supply chain is personal data. Where data is not personal, these mechanisms will not be available. GDPR itself is therefore not sufficient to provide for full transparency over service supply chains, as would be necessary to investigate their consolidation.

6. Interventions for increased transparency

As the previous sections show, there are many aspects in which a lack of transparency and a lack of disclosure can effectively obfuscate the nature of any consolidation within online service supply chains. This lack of transparency around infrastructure services poses a problem for understanding and overseeing the service supply chains of online applications. Without greater visibility over the actors involved within these complex supply chains (and the nature of these data flows), the true nature of any consolidation is masked, as are the related implications.

We therefore argue that *those interested in the policy implications of online consolidation should consider measures and interventions for increasing the transparency of the underlying infrastructure* on which many applications depend. This is to provide more information, visibility and a better understanding of the complexity, resilience, and corporate power issues previously discussed, thereby allowing for more meaningful policy responses.

6.1. Legal and regulatory interventions

There is much scope for legal and regulatory intervention, given (i) the current lack of effective mechanisms for providing transparency of infrastructural supply chains or for the accountability of infrastructure service providers, as well as (ii) the commercial incentives that favour opacity.

Legally, frameworks that provide mechanisms for individuals, organisations, and oversight bodies to obtain more information about the use and functioning of infrastructure services in service supply chains would be beneficial. This would be useful for regulators, application developers, academics, and others who are interested in better understanding the consolidation of infrastructure services and how this may affect the development and resilience of the Internet in future. Moreover, such information can be useful for users in general, to give some indication as to where their data flows, how their data is being used, and to indicate any other influences on the applications that they rely upon.

Establishing proactive transparency requirements or reporting mechanisms around the use of infrastructure services warrants consideration. While, as discussed above, GDPR does provide for some transparency obligations, they are insufficient for allowing a systematic investigation into supply chain consolidation and complexity. A relatively straightforward legal change that would bring immediate and significant benefits is to remove the words ‘or categories of recipients’ from GDPR Articles 13 and 14 (relating to privacy policies) and Article 20 (relating to the right of access). This would force a more explicit naming of the organisations a data controller deals with. Though these are still primarily mechanisms aimed at assisting individuals (data subjects), expanding and sharpening reporting obligations would go some way towards providing for the kind of transparency

needed to undertake a more systematic investigation than privacy policies and the right of access could permit. Moreover, additional obligations on infrastructure suppliers to provide relevant information to application developers or regulators represent a direct approach for bringing about greater transparency.

Accountability mechanisms for service infrastructure suppliers would also be beneficial, although it is not clear at present what shape these would take. There are reviews into the responsibilities of online platforms underway in several jurisdictions (Department for Digital, Culture, Media & Sport 2019; European Union 2019), but these typically concern user-facing issues and the shape of the regulatory regimes that may emerge from these reviews is generally uncertain at the time of writing. While, as discussed, some existing legal frameworks apply specifically to infrastructure providers (such as the Networks and Information Services Directive (European Union 2016a)), these are generally outside of the scope of potential reforms. Revision of those frameworks to include transparency obligations that would provide a greater understanding of supply chains would be a welcome development. However, effective application of existing legal mechanisms might be useful to some extent; e.g. where an infrastructure provider is abusing a dominant position, this would be a matter for competition regulators.

6.2. Commercial incentives

Compounding the lack of legal mechanisms compelling supply-chain transparency, commercial considerations and imperatives may incentivize the withholding of information about supply chains and the use of infrastructure services. This might be due to intellectual property and trade secrets concerns, or to maintain an advantage in terms of functionality, price or reputation.

Conversely, however, commercial considerations and imperatives may also provide application developers with an incentive for knowing more about the infrastructural supply chains on which their application relies. Without knowledge of those supply chains, the risks with regard to an infrastructure service becomes harder to assess and mitigate. This could potentially lead to failures of service or downtime, resulting in loss of customers and revenue.

Moreover, businesses are increasingly aware that they need to take data protection and security concerns seriously. Not only is there greater publicity around data breaches and other security issues, but, as discussed above, GDPR places obligations on data controllers and data processors in relation to transparency and the security of personal data. In particular, GDPR places the ultimate responsibility for compliance with its requirements on data controllers (in this context, typically the application operators), and provides them with mechanisms by which they can obtain greater information about data processors (i.e. here, typically infrastructure services) that are acting on their behalf. This may result in corporations taking a greater interest in infrastructure services, potentially wanting to know more about who is in their supply chains and what their security practices are like.

Further, aside from operational concerns, there may also be reputational implications for organisations by virtue of the involvement of downstream providers; e.g. if certain data flows to an organisation in a particular jurisdiction, or if an organisation perceived by some as 'untrustworthy' is somehow involved. Knowledge of the nature of supply chains can assist organisations in mitigating their exposure to such risks.

6.3. Technology-oriented interventions

From a technical point of view, we have observed that it is generally difficult for an application operator to see and understand what happens beyond the infrastructure with which they directly manage and interact. For example, the data flows beyond these relatively narrow boundaries are often unknown.

There is some work by the technical research communities towards increasing levels of accountability within complex systems (Singh et al. 2018). One example, for instance, are provenance methods—which involve the recording and analysis of the lineage and flow of data surrounding context (i.e. metadata about data and system interactions)—which show promise in providing greater visibility within complex systems and over systems supply chains. However, despite the potential, much work is required and many research opportunities in the space remain (Singh, Cobbe, and Norval 2019).

In line with this, there is a clear opportunity for legal, policy and regulatory interventions to provide incentives, on top of any commercial considerations, for the development of technical means to improve the transparency of supply chains. Regulation that influences technology development represents one approach, and has precedent, an example being GDPR's requirements for 'data protection by design/default'.[36] Policymakers can also leverage their ability to provide socio-economic and regulatory guidance and shaping around the design, deployment, and use of technical systems. This includes setting or encouraging standards development, or recommendations around the development of technical systems to establish 'best practice'. Again, considering GDPR, we have seen guidance on technical design provided by the data protection regulators of various EU member states. Further, to encourage new approaches, funding should also be directed towards research (academic or otherwise) that seeks to develop a better understanding of service supply chains and to develop technical means for transparency and accountability in those supply chains.

7. Concluding remarks

Modern society relies heavily on online applications. These applications are often built on top of the infrastructure services provided by others, with applications relying on a supply chain of interconnected services operating as a system-of-systems. The consolidation and complexity of those service supply chains raises several significant policy issues, particularly regarding the resilience of online applications on which many people rely, and in relation to the power of dominant information technology firms.

The general lack of transparency over service supply chains makes investigating these issues difficult, whether for academic researchers, regulators or policymakers. Moreover, this opacity poses a serious challenge for application developers and infrastructure providers, who may struggle to (i) identify the source of problems with their systems; (ii) take measures to respond to those issues and repair or restore functionality; and more generally (iii) have a clearer understanding of the influences impacting their applications and business. Without a better understanding of the underpinnings of online systems, the challenges of consolidation cannot be adequately addressed.

However, existing mechanisms for transparency are limited in utility. As discussed, there are few technical means for obtaining detailed and reliable information about

service supply chains. Mechanisms under data protection law are also insufficient for the reasons discussed.

From a policy perspective, to address these limitations and facilitate greater transparency over service supply chains, we argue that various interventions are necessary. Legally, establishing mechanisms for users, application providers, regulators and policymakers, to obtain more information about supply chains would be of significant benefit, as would establishing accountability mechanisms to put a check on the power of infrastructure providers. Technically, there is much scope to encourage responsible and appropriate technology design, use, standards and further research.

In all, the consolidation and complexity of service supply chains warrants more attention. While the technical and legal mechanisms for addressing the resulting problems are limited in effect, there are options available for lawmakers, regulators, policymakers and engineers to take concrete steps that could go a significant way towards providing more effective mechanisms for transparency. Such steps could help to improve the understanding of consolidation in service supply chains and its consequences.

Notes

1. There has been some consideration regarding issues of consolidation in the communications and networking space (Labovitz et al. 2010; Internet Society 2019). The focus of this paper, however, is the technical layers of abstraction above this, focusing on the range of online support services (at the application layer), which operate at run-time (see §2).
2. Our analysis included both HTTP & HyperText Transfer Protocol Secure (HTTPS) (Google 2019a). For simplicity, we will refer to these communications as HTTP, as is consistent within the literature.
3. Domain names are identifiers for a website, such as google.com or facebook.net. A third-party domain is one which is external to that of the website being directly accessed. As an example, one might watch an embedded YouTube video on the BBC website. In this case, though the user is visiting the BBC website, interactions would also occur directly between youtube.com (the third-party) and the user's browser.
4. Small 'addons' that typically provide some functionality in addition to that of the web browser.
5. https://www.alexa.com/topsites.
6. https://moz.com/top500.
7. https://majestic.com/reports/majestic-million.
8. Our script uses Selenium and ChromeDriver to create a new instance of the Chrome browser, visit one of the websites, wait 30 s for the page to load, and then return information about the HTTP requests made during that time (an approach similar to that taken by Libert (2015)).
9. We found that 39 (3.3%) of the sites failed to load, which is consistent with other similar analyses, e.g. Libert (2015) observed a 4 per cent failure rate.
10. A *header* is part of the broader networking protocol, providing information to support communication.
11. Note that this process is not easily automated given the uniqueness of each site and need to explore different pages, subdomains, usage patterns, etc (see §4.4).
12. Not least because header information is recommended to be obfuscated or removed for security purposes (Hunt 2012).
13. WHOIS records are publicly available and provide information on the owner of a particular domain, including names and contact details. However, they are not always a reliable source of information (Liu et al. 2015; Watters et al. 2013), as WHOIS anonymization services can be used to mask the identity of a website owner for privacy reasons. Further, WHOIS records are thought to face further uncertainty as a result of data protection laws, given that they involve personal data (Kulesza 2018).

14. GDPR art.4(2).
15. GDPR art.4(1).
16. Personal data includes potentially any data from which an individual can be identified, whether directly or indirectly, and whether from that data alone or when combined with other data (GDPR art.4(1), recital 26).
17. GDPR art.4(8).
18. GDPR art.28.
19. GDPR art.4(7).
20. GDPR arts.13-14.
21. GDPR art.15.
22. Note the goal was to gain some 'on the ground' insights into the GDPR in a consolidation context. Though a small sample size, our aim was not to conduct an exhaustive or conclusive analysis into the nature of GDPR transparency mechanisms in practice, but rather to indicate the potential of such for consolidation concerns.
23. https://www.alexa.com/topsites.
24. https://www.alexa.com/topsites/category/Top/Health.
25. Termed 'special category' data – see GDPR art.9.
26. GDPR art.15(1).
27. Though it is worth noting that if data subjects are provided with the names of each recipient organization then this may lead to an information overload where they are confronted with long lists of recipients without further information as to what they do. Such a list would not likely be of much use unless the data subject recognizes the names of some recipients or is prepared to go through each company in order to attempt to determine what kind of service they might provide.
28. GDPR art.28.
29. GDPR art.32.
30. GDPR art.29(2).
31. See GDPR art.28(3).
32. GDPR art.28(3)(h).
33. GDPR art.58.
34. GDPR art.58(1)(a).
35. GDPR art.58(1)(b).
36. GDPR art.25.

Disclosure statement

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

Funding

We acknowledge the financial support of the UK Engineering and Physical Sciences Research Council (EP/P024394/1, EP/R033501/1), the University of Cambridge (through the Trust & Technology Initiative), and Microsoft (through the Microsoft Cloud Computing Research Centre).

Notes on contributors

Jennifer Cobbe is a postdoctoral researcher in the Compliant and Accountable Systems research group in the Department of Computer Science and Technology at the University of Cambridge. She is also on the executive committee of Cambridge's Trust & Technology Initiative. Jennifer holds a PhD in Law from Queen's University Belfast. Her research focuses on law and regulation of new and emerging technologies.

Chris Norval is a postdoctoral researcher in the Compliant and Accountable Systems research group in the Department of Computer Science and Technology at the University of Cambridge. He completed his PhD in Human Computer Interaction at the University of Dundee in 2014. He has since worked as a data analyst in the games industry, as well as investigating ethical issues of machine learning as a postdoctoral researcher at the University of St Andrews.

Jatinder Singh is based at the Department of Computer Science and Technology (Computer Laboratory), University of Cambridge, where he leads the Compliant and Accountable Systems research group. The group focuses on the intersections of CS and law, which considers better aligning technology with legal concerns, and vice-versa. He also co-chairs Cambridge's Trust & Technology Initiative, and is a Fellow of the Alan Turing Institute. He received his PhD in computer science from the University of Cambridge and has some background in law.

ORCID

Jennifer Cobbe http://orcid.org/0000-0001-8912-4760
Chris Norval http://orcid.org/0000-0002-4331-7863
Jatinder Singh http://orcid.org/0000-0002-5102-6564

References

Amazon, A. W. S. 2017. "Netflix on AWS." Accessed December 21 2019. https://aws.amazon.com/solutions/case-studies/netflix/.

BBC News. 2018. "O2 4G Data Network Restored After Day-long Outage." Accessed August 15 2019. https://www.bbc.co.uk/news/business-46464730, 12.

Belli, Luca, and Jamila Venturini. 2019. "Private Ordering and the Rise of Terms of Service as Cyber-Regulation." *Internet Policy Review* 5 (4): 1–17.

Binns, Reuben, Ulrik Lyngs, Max Van Kleek, Jun Zhao, Timothy Libert, and Nigel Shadbolt. 2018a. "Third Party Tracking in the Mobile Ecosystem." *Proceedings of the 10th ACM Conference on Web Science*, WebSci '18, New York, NY, USA, 23–31. ACM. http://doi.acm.org/10.1145/3201064.3201089.

Binns, Reuben, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2018b. "Measuring Third-Party Tracker Power Across Web and Mobile." *ACM Trans. Internet Technol* 18 (4): 52:1–52:22. http://doi.acm.org/10.1145/3176246.

Charland, Andre, and Brian Leroux. 2011. "Mobile Application Development: Web vs. Native." *Communications of the ACM* 54 (5): 49–53.

Cloudflare. 2019. "Terminating Service for 8Chan." Accessed August 15 2019. https://blog.cloudflare.com/terminating-service-for-8chan/.

Collins, Keith. 2016. "How One Programmer Broke the Internet by Deleting a Tiny Piece of Code." Accessed November 28 2019. https://qz.com/646467/how-one-programmer-broke-the-internet-by-deleting-a-tiny-piece-of-code/, March.

Delacroix, Sylvie. 2019. "Beware of 'Algorithmic Regulation'".

Department for Digital, Culture, Media & Sport. 2019. "Online Harms White Paper, CP 57." Accessed December 10 2019. https://assets.publishing.service.gov.uk/government/uplo ads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf.

Duan, Y., G. Fu, N. Zhou, X. Sun, N. C. Narendra, and B. Hu. 2015. "Everything as a Service (XaaS) on the Cloud: Origins, Current and Future Trends." 2015 *IEEE 8th International Conference on Cloud Computing*. June, 621–628.

Englehardt, Steven, and Arvind Narayanan. 2016. "Online Tracking: A 1-million-site Measurement and Analysis." *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, New York, NY, USA, 1388–1401. ACM. http://doi.acm.org/10.1145/2976749.2978313.

European Union. 2016a. "Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 Concerning Measures for a High Common Level of Security of Network and Information Systems Across the Union." July.
European Union. 2016b. "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)." May.
European Union. 2019. "Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC OJ L 130.".
Google. 2019a. "HTTPS Encryption on the Web." Accessed December 21 2019. https://transparencyreport.google.com/https, 12.
Google. 2019b. "What is 1e100.net?" Accessed December 24 2019. https://support.google.com/faqs/answer/174717?hl=en.
The Guardian. 2019a. "8chan: The Far-right Website Linked to the Rise in Hate Crimes." Accessed December 20 2019. https://www.theguardian.com/technology/2019/aug/04/mass-shootings-el-paso-texas-dayton-ohio-8chan-far-right-website/.
The Guardian. 2019b. "Cloudflare Cuts off Far-right Message Board 8chan After El Paso Shooting." Accessed August 15 2019. https://www.theguardian.com/us-news/2019/aug/05/cloudflare-8chan-matthew-prince-terminate-service-cuts-off-far-right-message-board-el-paso-shooting/.
Gunawi, Haryadi S., Mingzhe Hao, Riza O. Suminto, Agung Laksono, Anang D. Satria, Jeffry Adityatama, and Kurnia J. Eliazar. 2016. "Why Does the Cloud Stop Computing?: Lessons from Hundreds of Service Outages." *Proceedings of the Seventh ACM Symposium on Cloud Computing*, SoCC '16, New York, NY, USA, 1–16. ACM. http://doi.acm.org/10.1145/2987550.2987583.
Hunt, Troy. 2012. "Shhh … Don't Let Your Response Headers Talk Too Loudly." Accessed December 20 2019. https://www.troyhunt.com/shhh-dont-let-your-response-headers/, 02.
Information Commissioner's Office. 2019. "Update Report into adtech and real time bidding." Accessed March 4 2020. https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf.
Internet Society. 2019. *Consolidation in the Internet Economy*. Technical Report. Internet Society.
The Japan Times. 2018. "O2 4G Data Network Restored After Day-long Outage." Accessed August 15 2019. https://www.japantimes.co.jp/news/2018/12/07/business/tech/ericsson-boss-apologizes-faulty-software-behind-huge-softbank-service-disruption/, 12.
Jazayeri, Mehdi. 2007. "Some Trends in Web Application Development." 2007 *Future of Software Engineering* 199–213. IEEE Computer Society.
Kulesza, Joanna. 2018. "Balancing Privacy and Security in a Multistakeholder Environment. ICANN, WHOIS and GDPR." *The Visio Journal* 3: 49–58.
Labovitz, Craig, Scott Iekel-Johnson, Danny McPherson, Jon Oberheide, and Farnam Jahanian. 2010. "Internet Inter-domain Traffic." *Proceedings of the ACM SIGCOMM 2010 Conference*, SIGCOMM '10, New York, NY, USA, 75–86. ACM. http://doi.acm.org/10.1145/1851182.1851194.
Lessig, Lawrence. 2006. *Code and Other Laws of Cyberspace, Version 2.0*. New York: Basic Books.
Libert, Timothy. 2015. "Exposing the Invisible Web: An Analysis of Third-Party HTTP Requests on 1 Million Websites." *International Journal of Communication* 9 (18): 3544–3561. https://ijoc.org/index.php/ijoc/article/view/3646.
Libert, Timothy. 2018. "An Automated Approach to Auditing Disclosure of Third-Party Data Collection in Website Privacy Policies." *Proceedings of the 2018 World Wide Web Conference*, WWW '18, Republic and Canton of Geneva, Switzerland, 207–216. International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3178876.3186087.
Liu, Suqi, Ian Foster, Stefan Savage, Geoffrey M. Voelker, and Lawrence K. Saul. 2015. "Who is.Com?: Learning to Parse WHOIS Records." *Proceedings of the 2015 Internet Measurement Conference*, IMC '15, New York, NY, USA, 369–380. ACM. http://doi.acm.org/10.1145/2815675.2815693.
Logic Monitor. 2017. "Cloud Vision 2020: The Future of the Cloud." Accessed August 14 2019. https://www.logicmonitor.com/wp-content/uploads/2017/12/LogicMonitor-Cloud-2020-The-Future-of-the-Cloud.pdf.

Metzger, Miriam J. 2007. "Communication Privacy Management in Electronic Commerce." *Journal of Computer-Mediated Communication* 12 (2): 335–361. https://onlinelibrarywiley.com/doi/abs/10.1111/j.1083-6101.2007.00328.x. .

Millard, Christopher. 2020. *Cloud Computing Law*. 2nd edition. Oxford: Oxford University Press.

Miller, Ron. 2017. "Why Dropbox Decided to Drop AWS and Build its Own Infrastructure and Network." Accessed August 15 2019. https://techcrunch.com/2017/09/15/why-dropbox-decided-to-drop-awsand-build-its-own-infrastructure-and-network, September.

Mogul, Jeffrey C. 2006. "Emergent (Mis)behavior vs. Complex Software Systems." *1st ACM SIGOPS/EuroSys European Conference on Computer Systems* 2006, 293–204.

Mozilla. 2019. "An Overview of HTTP." Accessed December 21 2019. https://developer.mozilla.org/en-US/docs/Web /HTTP/Overview, 12.

Singh, J., J. Cobbe, and C. Norval. 2019. "Decision Provenance: Harnessing Data Flow for Accountable Systems." *IEEE Access* 7: 6562–6574.

Singh, J., C. Millard, C. Reed, J. Cobbe, and J. Crowcroft. 2018. "Accountability in the IoT: Systems, Law, and Ways Forward." *Computer* 51 (7): 54–65.

Singh, J., J. Powles, T. Pasquier, and J. Bacon. 2015. "Data Flow Management and Compliance in Cloud Computing." *IEEE Cloud Computing* 2 (4): 24–32.

Singh, J., and J. Cobbe. 2019. "The Security Implications of Data Subject Rights." *Forthcoming, IEEE Security & Privacy* 17 (6): 21–30.

Sørensen, Jannick, and Sokol Kosta. 2019. "Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites." *The World Wide Web Conference*, WWW '19, New York, NY, USA, 1590–1600. ACM. http://doi.acm.org/10.1145/3308558.3313524.

Starov, Oleksii, and Nick Nikiforakis. 2017. "Extended Tracking Powers: Measuring the Privacy Diffusion Enabled by Browser Extensions." *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, Republic and Canton of Geneva, Switzerland, 1481–1490. International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3038912.3052596.

Swinford, Steven. 2019. "Theresa May Defies Security Warnings of Ministers and US to Allow Huawei to Help Build Britain's 5G Network." Accessed August 15 2019. https://www.telegraph.co.uk/politics/2019/04/23/theresa-may-defies-security-warnings-ministers-us-allow-huawei, April.

Tung, Liam. 2019. "Google Details 'Catastrophic' Cloud Outage Events: Promises to do Better Next Time." *ZDNet*. Accessed August 14 2019.

Watters, P. A., A. Herps, R. Layton, and S. McCombie. 2013. "ICANN or ICANT: Is WHOIS an Enabler of Cybercrime?" 2013 *Fourth Cybercrime and Trustworthy Computing Workshop*, Nov, 44–49.

Winner, Langdon. 1985. "Do Artifacts Have Politics?" *Daedalus* 109 (1): 26–38.