4-30-2015

Cloud Computing, Contractibility, and Network Architecture

Christopher S. Yoo
University of Pennsylvania Law School

Cloud Computing, Contractibility, and Network Architecture

Christopher S. Yoo*

ABSTRACT

The emergence of the cloud is heightening the demands on the network in terms of bandwidth, ubiquity, reliability, latency, and route control. Unfortunately, the current architecture was not designed to offer full support for all of these services or to permit money to flow through it. Instead of modifying or adding specific services, the architecture could redesigned to make Internet services contractible by making the relevant information associated with these services both observable and verifiable. Indeed, several on-going research programs are exploring such strategies, including the NSF's NEBULA, eXpressive Internet Architecture (XIA), ChoiceNet, and the IEEE's Intercloud projects

* John H. Chestnut Professor of Law, Communication, and Computer & Information Science and Founding Director, Center for Technology, Innovation and Competition at the University of Pennsylvania. I would like to thank Bill Boebel, Daniel Burton, Robert Krauss, Bill Lehr, Antonio Nicolosi, Don Norbeck, Todd Proebsting, Jonathan Smith, Jeff Vagle, Michael Walfish, and Joe Weinman, as well as participants in the conference on "Cloud Computing: Economic and Regulatory Implications" held at the University of Pennsylvania Law School on February 24, 2011, and the Research Roundtable and Public Policy Conference on "The Law and Economics of Privacy and Data Security" held at George Mason University School of Law on December 12–13, 2012, and June 19, 2013, for input on earlier drafts. This work is partially supported by National Science Foundation Grant CNS-10–40672. It explores solutions to problems identified in Yoo (2010, pp. 83–87) and Yoo (2011b).

INTRODUCTION

One of the fundamental changes effected by the emergence of cloud computing is to take functions previously performed by resources contained within a user's personal computer or laptop and transfer them to servers located in a distant data center. For users running virtual desktops, every single keystroke must pass through the network and be remotely registered by a virtual machine.

Such a drastic rearrangement of where particular tasks are performed creates pressure for the architecture to evolve to meet these new demands. For example, the fact that data that previously did not need to leave the personal computer sitting on the user's desk must now pass through a transmission network may lead to increased demand for bandwidth and decreased tolerance for latency. Moreover, because the path over which the data may travel may not be secure and the legal requirements of various jurisdictions may be diverse, cloud computing may lead users to demand the ability to verify the source of a packet and to exercise a greater degree of control over the path over which their data will travel and the locations where their data will be hosted.

In short, the advent of cloud computing is placing new demands on the Internet and other communications networks that users are employing to access cloud-based services and resources. Unfortunately, the services that the Internet is designed to provide are rather limited in both number and scope. Moreover, the types of services being provided over the cloud are constantly changing. Any attempt to modify the Internet to incorporate a discrete set of new services risks being rendered obsolete by the next wave of innovation.

Rather than try to restructure the network to provide a particular set of services needed by the current iteration of the cloud, an alternative approach would seek to reengineer the

architecture to provide the necessary primitives sufficient to enable present and future cloud service providers to support a wide range of policies. If properly designed, these primitives should allow both users and providers to create and monitor the services that they need by contract. Basic principles of contract theory have revealed two prerequisites that must be satisfied if contracts are to be effective. First, the information that is subject to the contract must be *observable*, in that both parties can perceive the relevant states of the world with respect to that information. Second, the relevant information must be *verifiable*, in that the parties must be able to prove after the fact in a court of law or to some other third party that the relevant state of the world did or did not occur (see, e.g., Hölmstrom 1979; Hart and Moore 1988). A classic example of a matter that is observable but not verifiable is effort exercised by an employee in the context of an employment contract. Both the employer and the employee may be well aware of the employee's level of effort or lack thereof, but it may be unable to be prove in a court of law.

If the architecture is to allow private actors to contract for certain levels of quality of service or data security, the minimum information needed for parties to be able to enforce their bargains must be observable and verifiable by the parties. The logical solution is to locate the primitives to make Internet transactions contractible in the network layer, which is the spanning layer visible to all network participants. The problem is that the Internet's current architecture does not satisfy either criterion. Users can observe or verify neither the source of data nor the precise path that a given transmission will take through the network. Moreover, the network layer of the Internet, which consists of the Internet Protocol (IP), is notoriously difficult to change.

Ongoing research sponsored by the National Science Foundation (NSF) and the Institute of Electrical and Electronics Engineers (IEEE) is exploring ways to redesign the architecture to

provide the primitives necessary to make the network services demanded by cloud computing contractible. Such a revision would enable network providers and users to make their own arrangements regarding new network services required by the cloud, which would be more consistent with the Internet's traditional approach of relying on decentralized decision making and would preserve the flexibility to adapt to new developments and increases in scale.

I. THE CLOUD'S NEED FOR NEW NETWORK SERVICES

As noted previously, cloud computing takes functions that used to involve the interaction of physical resources connected directly to a personal computer or a laptop and distributes them to a remote data center. This reconfiguration will place new demands on the access network in terms of bandwidth, ubiquity, reliability, latency, and route control. The access networks' ability to meet these demands will go a long way toward determining cloud computing's attractiveness as an option.

A. Bandwidth

Cloud computing is likely to increase the demands that are placed on the local access network. As an initial matter, new cloud computing customers must have some means for uploading their data to the data centers when setting up new applications. At this point, however, the access network does not have sufficient bandwidth to support this level of utilization. Because data sets in the terabyte range would take weeks to upload, cloud computing providers currently recommend that customers download their data onto a physical storage medium and send it via an overnight mail service (Brodkin 2010). Eventually, the hope is that network capacity will increase to the point where large data sets can be provisioned through the network itself.

Even after data has been provisioned to a new cloud computing facility, the fact that processing that used to occur locally is now being performed in the data center typically means that a greater volume of traffic must pass to and from the client that the user is operating. Cloud computing may thus cause an increase in the total bandwidth required from the network on a day-to-day basis as well.

B. Ubiquity

Cloud computing requires a higher degree of ubiquity than traditional computing solutions. When the software and the data needed to run a particular application reside on the users' hard disk, the unavailability of a network connection may inconvenience them and reduce the application's functionality, but it does not necessarily stop them from being productive in any way. When the software and data reside in the cloud, however, the absence of a network connection has more serious consequences, effectively preventing them from running the application at all. As a result, cloud computing customers regard ubiquitous access to network connections as critical.

C. Reliability

A related concern is access network reliability. The availability of an access network connection is meaningless if it is not functioning properly. Even when the application and the data reside on a user's hard disk, the failure of a network connection can severely limit the user's ability to perform productive work. Network failure becomes an even more serious obstacle when these elements are hosted in the cloud. Indeed, Gmail, Salesforce.com, and Amazon's Simple Storage Service (S3) and Elastic Compute Cloud (EC2) have suffered from well-publicized service outages that imposed severe difficulties on their customers. These higher

stakes mean that some customers are likely to demand that access networks offer higher levels of guaranteed uptime.

D. Quality of Service

Users' willingness to offload services that used to be provided locally into the cloud depends in no small part on how quickly the cloud is able to perform those functions. Aside from bandwidth, the most frequently discussed aspect of quality of service is latency, which is the delay that an application takes to register a change. Someone typing on a virtual desktop is likely to insist on latencies that are no more than a few hundred milliseconds. Other relevant aspects of quality of service include reliability (measured in terms of the accuracy of records) and jitter (measured in terms of variations in the spacing between packets). Different applications have different tolerances for each aspect of quality of service. As a result, cloud computing customers are likely to insist on service level agreements (SLAs) that guarantee them certain minimum levels of quality of service on those dimensions that matter most to them. These demands will likely vary in different cases. For example, financial services companies typically require perfect transactions, with latency guarantees measured in microseconds. In addition, these companies require the cloud provider to audit the accuracy and delivery time of every transaction after the fact.

Cloud computing is likely to require sophisticated network management techniques to provide minimum levels of quality of service. One way that cloud computing systems can improve the quality of service of network services is by taking advantage of the presence of multiple connections between two points. The Internet currently relies on protocols such as the Border Gateway Protocol (BGP) to determine the route that any particular stream of packets may take between domains. BGP is limited in its ability to manage multiple paths, routing all traffic

along a single route instead of balancing traffic across multiple paths. BGP, moreover, is controlled by the core routers rather than by users. A new architecture for cloud computing could improve network performance by providing greater ability to allocate traffic across multiple paths and to allow faster recovery from congestion and network failure. It could also increase functionality by giving users control over the particular routes taken by their traffic.

E. Control Over Routing

Cloud computing necessarily requires large amounts of data that previously did not leave a company's computer or internal network to be transported via a series of networks to a data center. The fact that this data must pass outside the company's firewall and through the access network renders it vulnerable to attack vectors that are different from those that plague corporate campuses.

As a result, cloud-based solutions must be able to assure these institutions that their data are being handled in a way that preserves confidentiality by giving users greater ability to control which networks their traffic passes through. Because cloud computing requires that sensitive information must pass over a network connection, users may demand the ability to verify a packet's source, as well as the means to ensure that their data will pass only over networks they deem trustworthy.

Moreover, the ability to shift data from one data center to another potentially makes that data subject to another country's privacy laws. Current data protection requirements vary widely across jurisdictions. For example, US law holds all institutions that maintain health or educational records responsible for maintaining their privacy. The fact that such records are now housed in the cloud does not obviate those responsibilities. However, in the European Union, the law requires that data be retained only for limited purposes and for limited times. Because

customers are ultimately responsible for any such violations, they are likely to insist on a significant degree of control over where their data reside at any particular moment. In addition, cloud computing may require an architecture that permits the exact routes that particular traffic takes to be auditable and verifiable after the fact.

The emergence of the cloud is thus causing users to place a different set of demands on the network. Moreover, not all cloud users will need the same combination of services. Word processing does not require significant bandwidth, but it is extremely sensitive to latency. Users who use the cloud to store video or music can tolerate the latency needed to buffer streaming media, but they demand significantly more bandwidth.

This heterogeneity has led many commentators to recognize that it is not appropriate to fold cloud computing into the conceptual framework traditionally applied to public utilities (Brynjolfsson, Hoffman, and Jordan 2010, Kushida, Murray, and Zysman. 2011, Bayrak, Conley, and Wilkie 2011; also see chapter 3, "Reliability and the Internet Cloud," in this volume). History has shown that public utility regulation is ill suited to technologies where the product attributes are complex and where the production technology varies and is undergoing rapid technological change (Yoo 2013b). It comes as no surprise, then, that early commentators who first conceived the computing utility typically acknowledged that it did not fit within the classic conception of public utilities (see, e.g., Irwin 1966, Parkhill 1966, Baran 1967, Barnett et al. 1967, President's Task Force on Communications Policy 1968, Smith 1969).

II. LIMITS OF THE CURRENT ARCHITECTURE

Cloud computing may demand quality of service guarantees, as well as the ability to control the routes that particular data will pass through the network. The existing architecture does provide some tools to facilitate the provision of these services. For example, the

engineering community has devised a wide range of protocols to help provide defined levels of quality of service. Indeed, the inclusion of a type-of-service field in the original Internet Protocol version 4 (IPv4) header reveals that quality of service through prioritization was part of the Internet's original design. Subsequently, the Internet community has developed a wide array of protocols to promote quality of service, including Integrated Services (IntServ), Differentiated Services (DiffServ), and MultiProtocol Label Switching (MPLS). More recent efforts include virtual circuit services such as Internet2's Interoperable On-demand Network (ION) and deprioritization regimes such as the Low Extra Delay Batch Transport (LEDBAT) from the Internet Engineering Task Force (IETF), both of which represent fairly substantial deviations from the principles around which the current Internet is organized (Yoo 2011a). The problem is that to date, none has attained sufficiently broad acceptance to support cloud services. The retention of the type-of-service field (renamed "Traffic Class") in Internet Protocol version 6 (IPv6) underscores the continuing importance of quality of service.

Regarding routing, the original IPv4 header includes an optional field to allow the source to determine the route that packets will take. This option is not mandatory and has largely been ignored. In addition, the BGP-based system responsible for routing traffic on the current Internet employs an algorithm that allows each router to make independent decisions about the path that particular packets will take through the network. In general, BGP by default simply sends traffic along the path that transverses the fewest autonomous systems. Although BGP is intended to permit networks to implement routing policies, these policies tend to be implemented by altering the routing table by hand by increasing the apparent length of a path in an effort to make certain routes unattractive. The fact that each firm makes such individualized adjustments on a

distributed basis can lead to interactions that make such policies difficult to implement and even harder to guarantee.

Cloud computing providers and users wishing to exercise greater control over the paths that are taken by traffic that passes between data centers may rely on MPLS or some other protocol to exercise control over the precise paths that are taken by particular traffic. In fact, IPv6 added a “Flow Label” field to incorporate this functionality into the network layer itself. Such control mechanisms are essential to ensuring that flows between data centers maintain the required levels of quality of service, protect network security, and maintain the privacy of users’ data. As of today, such services are provided as overlays rather than being designed into the network itself.

Most important in terms of contractibility, even if labels allow users to specify paths taken through the Internet, the Internet’s architecture does not provide any basis for verifying a packet’s source or the path that it traversed. Nor is there any basis for verifying whether any particular prioritization regime was followed. Stated in terms of contract theory, the information needed to make quality of service and route control contractible is neither observable nor verifiable.

In addition, with the Internet’s current architecture, the information needed to make Internet privacy contractible is not observable or verifiable either. First, consider information about the source of Internet transmissions. One of the foundational principles on which the Internet is based is that every administrative domain connected to the Internet exchange packets through a single, uniform spanning layer, represented by the IP, in which each machine is identified by a unique address that is visible to every other machine connected to the network (Cerf and Kahn 1974). One problem is that the source address included in the IP header is

insecure and can be misrepresented (spoofed), making it impossible to verify the source of any communication under the network's basic design. Although additional features, such as Internet Protocol Security (IPSec), have subsequently been developed that can support source authentication, they are not mandatory and are not widely deployed in IPv4. Although the specification for Internet Protocol version 6 (IPv6) makes IPSec mandatory, not all IPv6 implementations support IPSec.

Another core principle of the Internet is that routers operating in the network's core operate on a store-and-forward basis, with each router making its own independent decision about the path that any particular packet should take. As a result, users typically cannot specify the path that a particular packet should take through the network, as the source routing option included in the Internet's original design is not mandatory and remains nearly universally unused. Even if users were able to do this, the localized nature of information on the Internet makes path information difficult to observe and verify. Any component network of the Internet can only observe the identity of the network residing immediately upstream and downstream of its location. The store-and-forward nature of the Internet's architecture provides no mechanism through which the user can observe the path actually taken through the network. Thus, even if an user enters into an SLA with its Internet service provider (ISP) specifying the paths that its traffic must traverse, it has no reliable way to determine whether the ISP actually honored the terms of the SLA with respect to a specific transaction. Even if all of the relevant parties were able to observe the relevant information, the absence of any authoritative record of the path traversed by any particular packet makes this information difficult, if not impossible, to verify in a court of law.

Finally, the architecture restricts the configuration of economic relationships on the Internet. Money enters the network from the edge, typically as a payment to a last-mile network provider, but the network architecture provides no flow-based mechanism to allocate it among the different entities involved in providing service. Simply put, the Internet's design does not place the information needed to allocate value in the network layer which, as noted previously, is the spanning layer visible to all network actors. Restated in terms of contractibility theory, the economic information needed to support complex contracts is not observable, let alone verifiable. Or, as MIT computer scientist and former DARPA chief protocol architect David Clark is reported to have quipped, "We never learned how to route money" (McTaggart 2006).

The absence of any mechanism within the architecture itself for accounting for and distributing value among different actors forces money to flow through mechanisms that exist outside the network. Constraining value to flow exclusively through contracts between ISPs makes ISPs the irreducible unit of economic analysis and reinforces the current hierarchy of transit providers. Turning ISPs into indivisible artifacts for economic purposes also severely limits the flexibility with which services can be provided and contracted. Early proposals to create smart markets that would have permitted the money flow to follow the traffic flow were never adopted (Mackie-Mason and Varian 1995).

III. Contractibility as a Potential Nonregulatory Solution

The current network architecture thus restricts the services that the network can provide, as well as the flexibility with which economic relationships among different network elements can be configured. The emergence of new phenomena such as the cloud has prompted a wide range of new research initiatives designed to make the Internet more contractible.

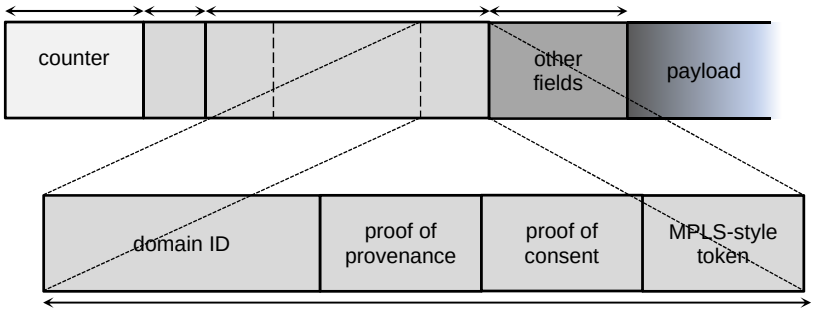
A. NEBULA

The need to explore ways that the Internet may need to evolve to respond to new use cases has led the NSF to launch its Future Internet Architecture (FIA) project. One new architecture known as NEBULA is specifically designed to support the new demands placed on the network by cloud computing (Anderson et al. 2013, 2014). The challenge is to strike a balance between flexibility and efficiency by creating a minimal spanning set of features that can support an arbitrarily broad set of transit policies. The aspects of the NEBULA design that make the services on which cloud computing depends contractible are a control plane, known as the NEBULA Virtual and Extensible Networking Techniques (NVENT), and the NEBULA Data Plane (NDP), which is being adapted from a technology known as ICING (Naous et al. 2011).

A communication begins when the sending network places a request for a path to NVENT. NVENT envisions that each administrative domain comprising the Internet will maintain a policy engine that reciprocally exchanges information with other policy engines about the available services, resources, and paths. The policy engine identifies paths to reach requested destinations in ways that comply with any preferred transit policies governing the entities permitted to constitute the path, bandwidth, latency, reliability, and other considerations. The path or paths and the associated policies are then embodied in a token in a manner similar to MPLS. Any noncompliant paths are considered rejected by default. Once a path has been discovered, NVENT consults a consent engine maintained by each administrative domain and collects cryptographic proofs of consent from each of the administrative domains comprising the path. The fact that the policy and consent engines are distributed across the administrative domains comprising the network facilitates this system's flexibility and extensibility. It also provides flexibility in the specific policies being implemented.

The proofs of consent are then inserted into the NDP, which in effect replaces IP with a new spanning layer that supports source and path authentication. NDP reserves 42 bytes in its header (depicted in Figure 4.1) for four elements with respect to each administrative domain contained in the path constructed by NVENT: (1) the domain identifier; (2) a cryptographic proof that the domain has authorized being included in the path, called the *proof of consent* (*PoC*); (3) a cryptographic proof that the packet has actually followed that path, called the *proof of provenance* (*PoP*); and (4) the MPLS-style token associated with this path.

Figure 4.1: NDP packet format



These four elements are sufficient to allow networked entities to both express and enforce a broad range of policies about packet carriage. Decisions about policies are relegated to the control plane (NVENT), which remains distributed and flexible. Once those policies are embodied in the MPLS-style token, the four features embedded in the data plane (NDP) described previously are sufficient to allow these policies to be enforced. When a packet arrives at an administrative domain, that domain can check the PoC to verify that the packet was authorized by the policy. As the packet traverses the domain, the domain attaches its PoP to verify to all other network participants that it actually traversed that domain. The fact that both the PoC and PoP are cryptographically protected using public-key encryption allows every entity to authenticate the accuracy of the information contained therein. An added feature of NDP is

that the key for decrypting the PoC and PoP is the domain identifier, which obviates the need to manage keys or maintain certificate authorities.

The four basic primitives contained in NDP are sufficient to encompass the functions enabled by a wide range of other efforts to make networks more secure. Specifically, NDP can determine that all communications traverse an assured path, prevent the entry of unauthorized communications, support the establishment of multiple paths to ensure availability and reliability, permit each administrative domain to exercise autonomous control, and enhance privacy by ensuring that communications only traverse trusted providers. In so doing, they provide a parsimonious set of information that provides the same functions as a wide range of more complex designs already appearing in the literature. Preliminary experiments and prototypes indicate that the architecture is feasible (Naous et al. 2010, 2011).

The cryptographically protected PoCs permit each administrative domain to authenticate that the particular transaction has been authorized ex ante. At the same time, the cryptographically protected PoPs permit each administrative domain to authenticate ex post that the approved path was followed.

These innovations have a dramatic impact on contractibility. By making the specific administrative domains traversed and the policies being applied visible, NEBULA renders the information needed to enforce privacy observable. In addition, the application of cryptographic signatures as each packet crosses an administrative domain makes any agreements as to paths to be used and policies to be honored verifiable.

Like any architecture, NEBULA is not without its potential concerns. As an initial matter, NDP requires the incurrence of significant overhead. For example, the inclusion in the network layer of the additional information associated with the four elements described here in the

network layer header makes packets roughly 20 percent larger. Because NDP's functions are parallelizable, it appears that NDP can operate at backbone speeds. NDP-enabled routers would cost roughly 90 percent more than a bare-bones IP router but slightly less than a commercial IP router (Naous et al. 2010, 2011). These initial efforts are simply to provide proof of the concept. As was the case with the Internet itself, additional improvements on operating efficiency are likely to be realized once the network is refined and deployed. NDP is by no means the only approach to enhancing contractibility. In fact, NEBULA explored approaches based around alternative technologies(Anderson et al. 2013).

Although NDP enhances observability and verifiability, a trusted entity may nonetheless evade its limitations by subcontracting responsibility for transmission to another, unapproved entity or using tunneling to allow the communication to use unapproved routes. Moreover, although the information contained in the NDP header is designed to verify the path, it is less effective at revealing whether other policies were honored. NDP thus may require additional enforcement mechanisms outside its current design. Because contracting entities may not be in privity with these other parties, full enforcement may require recognizing a property interest in private information that is good against enforceable against all parties regardless of whether they have a contractual relationship.

The preliminary nature of the NEBULA project suggests that it is too soon to expect definitive answers to each of these issues. All of these considerations are the subject of future research. Any evaluation of its potential must bear in mind that it is the first implementation of such an architecture and is put forward as a proof of concept, not as an operational business model. One can reasonably expect performance improvements should the architecture be fully deployed.

B. IEEE Intercloud

Another research project that would provide the primitives needed to make cloud connectivity contractible is the IEEE's Intercloud working group (IEEE 2012, 2013). The primary goal of the Intercloud initiative is to standardize cloud services. Not only would this make them more portable; it would allow providers to combine cloud services offered by different providers and dynamically integrate them into a single offering.

The basic architecture of the Intercloud is modeled on the Internet. The architecture centers on community-governed Intercloud root providers that serve as the naming authority for the Intercloud in much the same manner as the domain name system does for the Internet. The Intercloud root providers also serve as the trust authority for the public key infrastructure (PKI) used to authenticate each entity. Presumably, the Intercloud root would not be a single entity. Instead, there would be multiple Intercloud roots that would host the cloud computing resource catalog in a federated manner.

In addition to performing these functions, the Intercloud root providers offer other services needed to support contracts for cloud services. Most important, they maintain cloud computing resource catalogs that make visible the available cloud resources. These advertisements are made using uniform semantics that describe not only physical resources, such as servers, disks, and connectivity, but also other less tangible attributes, including SLAs, pricing policies, and security and compliance policies. It thus provides visibility and access to the information needed to form contracts through the Intercloud (Bernstein and Vij 2010).

The architecture also envisions the Intercloud exchanges that provide locations where multiple cloud providers can interoperate in a manner similar to the role currently served by

Internet exchanges for the Internet. These exchanges draw on the catalog information hosted by the root providers to broker cloud transactions with users.

The typical transaction falls into the following pattern: Cloud 1 uses its Intercloud gateway to query the cloud computing resources catalog maintained by the Intercloud root server to determine if the resources needed to support a particular cloud service are available. If Cloud 1 finds out from the catalog that the resources needed to meet its requirements and constraints are available from Cloud 2, the Intercloud exchange brokers an agreement between the two clouds. If an agreement is successfully reached, the clouds bilaterally establish the web sockets and other protocols needed for Cloud 2 to provide services to Cloud 1. Cloud 1 then uses Cloud 2's resources as part of its federated architecture, while Cloud 2 meters resource usage and compliance with the terms of the SLA.

Once this happens, most of the primitives needed for contractibility will be met. The cloud computing resource catalog makes the necessary resources observable to others prior to performance. Importantly, this includes not just the physical elements of the cloud, but also the SLAs, pricing, and other terms needed for contacts to operate. Each cloud conducts its own metering.

The only thing missing is ex post auditability. The Intercloud's architects recognize that an audit trail constitutes an essential part of the architecture. These responsibilities are supposed to be borne by the Intercloud root servers, although the necessary implementations have not been designed yet. Deployment of this auditing capability would make these transactions observable and verifiable during and after performance.

C. eXpressive Internet Architecture

Other NSF-sponsored FIA projects are pursuing architectures that enhance contractibility in different ways. For example, the eXpressive Internet Architecture (XIA) project has developed an architecture known as Scalability, Control, and Isolation On Next-generation networks (SCION) that relies more on trust than explicit enforcement (Zhang et al. 2011; Naylor et al. 2014).

SCION envisions that each autonomous system (AS) joins with other ASes to form a trusted domain (TD) that separates trusted from distrusted entities. Each TD designates an AS to represent the TD in the TD core, which initiates path construction, disseminates policy-complaint paths, allows destinations to choose their preferred paths, and uses cryptographic methods to ensure the authenticity of end-to-end paths.

Once the route has been joined, the source domain embeds “opaque fields” into each packet that encode the path information. Because SCION does not require each domain to cryptographically sign each packet as it traverses the domain, this solution does not fully support verifiability. Instead, SCION relies on trust or some other mechanism outside the information contained in each packet to authenticate the source and to determine whether the domains actually honored the selected path.

D. ChoiceNet

The research program that most explicitly attempts to make Internet-based transactions contractible is ChoiceNet, which was also funded by the NSF’s FIA program (Wolf et al. 2014). The current architecture imposes a number of constraints on the network. Money enters the network only through its edges and flows outside the network architecture. Moreover, traffic is

constrained to move through the network in accordance with the economic relationships between network providers implicitly reflected in BGP routing policies. The result is that both users and content providers have little control over how their traffic is handled within networks.

ChoiceNet is designed to create an “economy plane” that enables actors to create contracts that dynamically integrate the offerings of multiple providers of network components into a single service offering. The goal is to allow the money flow to follow the traffic flow instead of vice versa. The resulting flexibility in configuring services from different providers should enable new entities to combine the underlying network elements in new ways, which can promote innovation and competition.

ChoiceNet provides a uniform address architecture and a minimal set of service semantics sufficient to permit network providers to advertise the nature of the services they are offering. The semantics must cover not only the physical services being provided, but also other contractual terms such as quality of service and price. ChoiceNet entities then act as intermediaries to combine and bundle these services into end-to-end offerings. ChoiceNet then provides a marketplace where customers can shop for different end-to-end service offerings that the intermediaries have created. The semantics of the marketplace advertisements must be general enough to be searchable, yet specific enough to specify the necessary dimensions of performance. Finally, ChoiceNet envisions a verification process through which customers can determine whether actual performance lived up to the contract specifications. While ChoiceNet’s designers are rather vague about the details of how to implement this function, suggesting that it might be fulfilled by third-party measurement providers or by the exchange of measurement data within the research community, they have acknowledged that verification plays an indispensable role in their scheme.

A ChoiceNet transaction can be envisioned as a five-step process. In the *advertisement* step, the intermediary advertises in the marketplace the end-to-end offering that it has created. In the *planning* step, prospective customers use algorithms to search the marketplace and explore the bundles of services created by the intermediaries. In the *provisioning* step, the customers and intermediaries establish contracts, exchange consideration, commit the resources to provide the agreed-upon service, and generate tokens that serve as credentials to obtain access to the contracted-for resources and to prove policy compliance. In the *usage* step, each resource examines the token to verify that usage has been authorized in the economy plane. In the *verification* step, the customer employs a verification service to confirm that it received the promised level of performance.

ChoiceNet was not designed specifically for the cloud, although cloud computing would be able to employ its architecture to the same extent as other Internet-based applications. Admittedly, the ChoiceNet architecture is not yet well developed. But the basic approach is fully consistent with the concerns raised in this chapter. The ChoiceNet architecture is designed to provide the elements needed for contractibility. Creating semantics and loci for sharing information about service attributes makes the necessary information observable prior to performance. The envisioned verification mechanism would make the actual performance both observable and verifiable.

At this early stage, one must be careful not to become overly distracted by debates over the relative merits of any particular proposal. The principle of making Internet transmissions contractible by making key information both observable and verifiable is ultimately more important than the details of how that is done in any particular implementation.

CONCLUSION

Cloud computing is creating new demands for network services that the current architecture was not designed to support. The increasing diversity of new technologies and applications and the accelerating rate of innovation suggest that direct regulation will continue to lag behind the current environment and will struggle to scale as use cases continue to proliferate.

A more fruitful approach may be to focus instead on supporting more contract-oriented governance by making Internet transactions more contractible. Rather than envisioning any particular type of contractual relationship, this approach revises the network layer visible to all Internet-connected actors to provide sufficient primitives to permit individual actors to construct a wide range of interconnection relationships. The resulting ability to authenticate the source of packets as well as the path they used to traverse the network should greatly enhance Internet privacy.

Clean-slate redesigns of network architecture are notoriously difficult to implement. The presence of a large installed base creates significant inertia behind the existing architecture. Such inertia can be overcome if the new architecture provides sufficient value. Even so, such transitions are likely to pose significant multilayer coordination problems (Yoo 2013a). The details of the specific implementation are less important than the need to refocus the privacy debate away from command-and-control regulation in favor of a more flexible, decentralized, and evolvable regime that is more consistent with the Internet philosophy and relies primarily on private ordering.

References

Anderson, Tom, Ken Birman, Robert Broberg, Matthew Caesar, Douglas Comer, Chase Cotton, Michael J. Freedman, et al. 2013a. A Brief Overview of the NEBULA Future Internet Architecture. *Computer Communication Review* 44 (3): 81–86. doi:10.1145/2656877.2656889.

Anderson, Tom, Ken Birman, Robert Broberg, Matthew Caesar, Douglas Comer, Chase Cotton, Michael Freedman, et al. 2013b. The NEBULA Future Internet Architecture: A Mid-Course Report. In *The Future Internet—Future Internet Assembly 2013: Validated Results and New Directions*, ed. Alex Galls and Anastasius Gavras, 16–26. New York: Springer.

Baran, Paul. 1967. The Future Computer Utility. *Public Interest* 8 (June): 75–87.

Barnett, C. C., Jr., B. R. Anderson, W. N. Bancroft, R. T. Brady, D. L. Hansen, H. Simmons, D. C. Snyder, D. Wechsler, and J. L. Wilcox. 1967. *The Future of the Computer Utility*. New York: American Management Association.

Bayrak, Ergin, John P. Conley, and Simon Wilkie. 2011. The Economics of Cloud Computing. *Korean Economic Review* 27 (2): 203–230.

Bernstein, David, and Deepak Vij. 2010. "Using Semantic Web Ontology for Intercloud Directories and Exchanges." 2010 International Conference on Internet Computing (ICOMP), Las Vegas, NV. Available at http://www.intercloudtestbed.org/uploads/2/1/3/9/21396364/using_semantic_web_ontology_for_intercloud_directories_and_exchanges.pdf.

Brodkin, Jon. 2010. “Amazon Cloud Uses FedEx Instead of the Internet to Ship Data,” *Network World,* June 10. http://www.networkworld.com/news/2010/061010-amazon-cloud-fedex.html.

Brynjolfsson, Erik, Paul Hoffman, and John Jordan. 2010. Cloud Computing and Electricity: Beyond the Utility Model. *Communications of the ACM* 53 (5): 32–34. doi:10.1145/1735223.1735234.

Cerf, Vinton G., and Robert Kahn. 1974. A Protocol for Packet Network Interconnection. *IEEE Transactions on Communications* 22 (5): 637–648. doi:10.1109/TCOM.1974.1092259.

Hart, Oliver, and John Moore. 1988. Incomplete Contracts and Renegotiation. *Econometrica* 56 (4): 755–785. doi:10.2307/1912698.

Hölmstrom, Bengt. 1979. Moral Hazard and Observability. *Bell Journal of Economics* 10 (1): 74–91. doi:10.2307/3003320.

Institute of Electrical and Electronics Engineers (IEEE). 2012. “IEEE P2302/D0.2: Draft Standard for Intercloud Interoperability and Federation (SIIF).” Available at http://www.intercloudtestbed.org/uploads/2/1/3/9/21396364/intercloud_p2302_draft_0.2.pdf.

Institute of Electrical and Electronics Engineers (IEEE). 2013. “IEEE Intercloud TestBed: Technical Overview and Engineering Plan.” Available at http://www.intercloudtestbed.org/uploads/2/1/3/9/21396364/ieee_intercloud_testbed_technical_overview_engineering_plan_v6.pdf.

Irwin, Manley R. 1966. The Computer Utility. *Datamation* 12 (11): 22–27.

Kushida, Kenji E., Jonathan Murray, and John Zysman. 2011. Diffusing the Cloud: Cloud Computing and Implications for Public Policy. *Journal of Industry, Competition, and Trade* 11 (3): 209–237. doi:10.1007/s10842-011-0106-5.

Mackie-Mason, Jeffrey K., and Hal R. Varian. 1995. Pricing the Internet. In *Public Access to the Internet*, ed. Brian Kahin and James Keller, 269–314. Cambridge, MA: MIT Press.

McTaggart, Craig. 2006. "Was the Internet Ever Neutral?" Paper presented at the 34th TPRC, Arlington, VA. Available at http://ssrn.com/abstract=2117601.

Naous, Jad, Arun Seehra, Michael Walfish, David Mazières, Antonio Nicolosi, and Scott Shenker. 2010. "Defining and Enforcing Transit Policies in a Future Internet." University of Texas at Austin Department of Computer Sciences Technical Report TR-10–07. Available at www.cs.utexas.edu/~mwalfish/icing_tr_1007.pdf.

Naous, Jad, Michael Walfish, Antonio Nicolosi, David Mazières, Michael Miller, and Arun Seehra. 2011. "Verifying and enforcing network paths with ICING." *Proceedings of the 7th International Conference on emerging Networking Experiments and Technologies (CoNEXT)*, art. 30. doi:10.1145/2079296.2079326. Available at http://dl.acm.org/ft_gateway.cfm?id=2079326&ftid=1066769&dwn=1&CFID=54306638 4&CFTOKEN=90992542.

Naylor, David, Matthew K. Mukerjee, Patrick Agyapong, Robert Grandi, Ruogo Kang, and Michel Machado. 2014. XIA: Architecting A More Trustworthy and Evolvable Internet. *Computer Communication Review* 44 (3): 50–57. doi:10.1145/2656877.2656885.

Parkhill, Douglas F. 1966. *The Challenge of the Computer Utility*. Reading, MA: Addison-Wesley.

President's Task Force on Communications Policy. 1968. *Final Report*. Washington, D.C.: Government Printing Office.

Smith, Delbert D. 1969. The Interdependence of Computer and Communications Services and Facilities: A Question of Federal Regulation. *University of Pennsylvania Law Review* 117 (6): 829–859. doi:10.2307/3310940.

Wolf, Tilman, Jim Griffioen, Ken Calvert, Rudra Dutta, George Rouskas, Ilya Baldin, and Anna Nugurney. 2014. ChoiceNet: Toward an Economy Plane for the Internet. *Computer Communication Review* 44 (3): 58–65. doi:10.1145/2656877.2656886.

Yoo, Christopher S. 2010. The changing patterns of Internet usage. *Federal Communications Law Journal* 63 (1): 67–89.

Yoo, Christopher S. 2011a. Rough consensus and running code: Integrating engineering principles into Internet policy debates. *Federal Communications Law Journal* 63 (2): 341–356.

Yoo, Christopher S. 2011b. Cloud computing: Architectural and policy implications. *Review of Industrial Organization* 38 (4): 405–421. doi:10.1007/s11151-011-9295-7.

Yoo, Christopher S. 2013a. Protocol layering and Internet policy. *University of Pennsylvania Law Review* 161 (6): 1707–1771.

Yoo, Christopher S. 2013b. Is there a role for common carriage in an Internet-based world? *Houston Law Review* 50 (2): 545–608.

Zhang, Xin, Hsu-Chun Hsiao, Geoffrey Hasker, Haowen Chan, Adrian Perrig, and David G. Anderson. 2011. "SCION: Scalability, control and isolation on next-generation networks." Paper presented at the TRUST Autumn 2011 Conference, Washington, DC. Available at http://sparrow.ece.cmu.edu/group/pub/SCION.pdf.