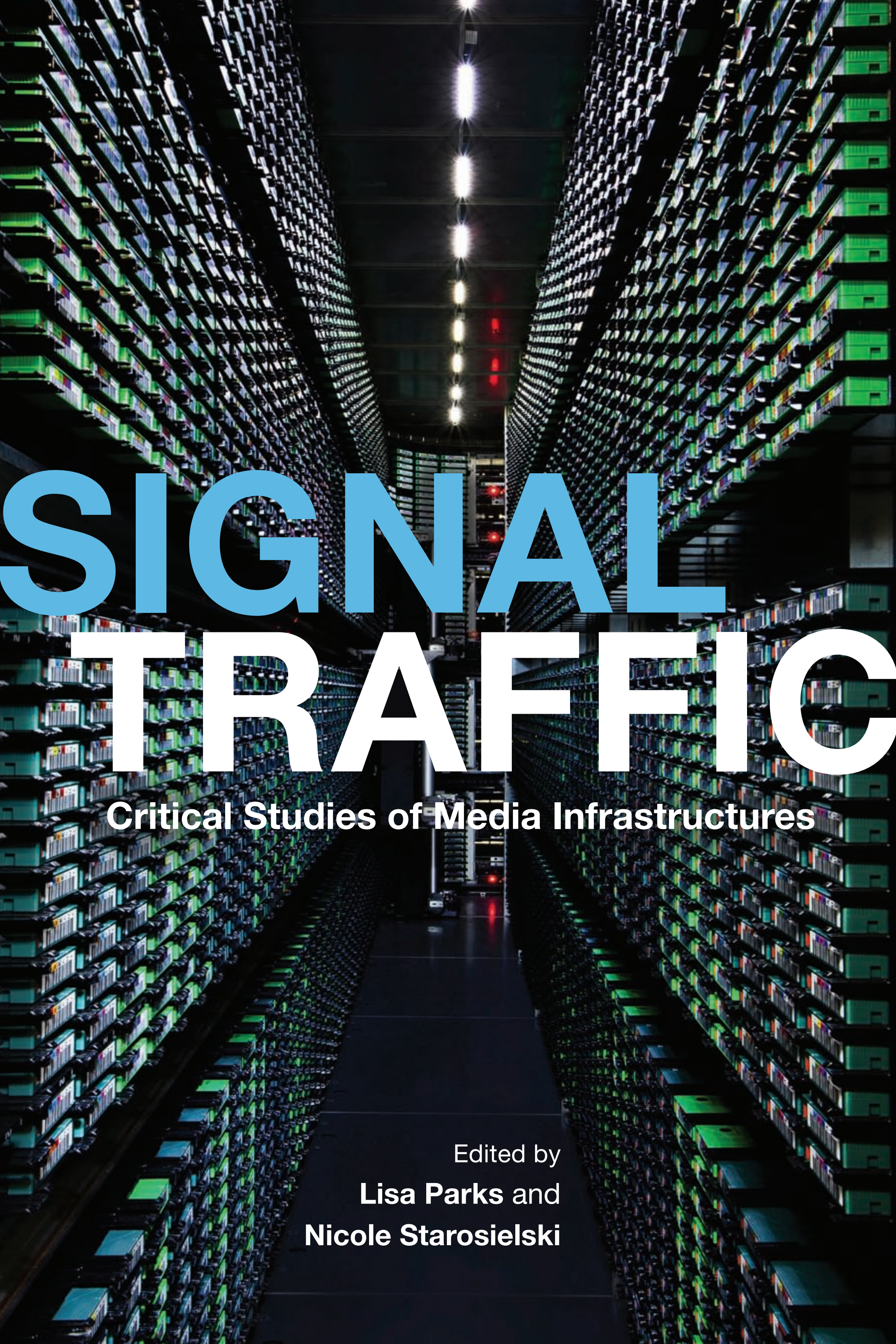

SIGNAL TRAFFIC

Critical Studies of Media Infrastructures

Edited by
Lisa Parks and
Nicole Starosielski

Signal Traffic

Critical Studies of Media Infrastructures

Edited by
**LISA PARKS AND
NICOLE STAROSIELSKI**

UNIVERSITY OF ILLINOIS PRESS
Urbana, Chicago, and Springfield

1 2 3 4 5 C P 5 4 3 2 1
This book is printed on acid-free paper.
Library of Congress Cataloging-in-Publication Data
Signal traffic: critical studies of media infrastructures / edited by
Lisa Parks and Nicole Starosielski.
pages cm. — (The geopolitics of information)
Includes bibliographical references and index.
ISBN 978-0-252-03936-2 (hardback : acid-free paper)
ISBN 978-0-252-08087-6 (paper : acid-free paper)
ISBN 978-0-252-08087-6 (e-book)
1. Telecommunication systems—Social aspects. 2. Digital
media—Social aspects. 3. Mass media—Social aspects.
4. Information superhighway. 5. Computer networks—
Social aspects. 6. Information networks—Social aspects.
7. Telecommunication—Traffic. 8. Signal processing.
I. Parks, Lisa. II. Starosielski, Nicole, 1984–
TK5102.5.S5434 2015
303.48'33—dc23 2014049422

CHAPTER 8

Protocols, Packets, and Proximity

The Materiality of Internet Routing

PAUL DOURISH

On the corner of Wilshire Boulevard and South Grand Avenue in downtown Los Angeles stands a stark, imposing, white office tower, thirty-five stories tall, named simply "One Wilshire." It looks much like any other downtown office building, although a careful listener might notice that the muted rumble of its air conditioning, audible even over noise of the downtown traffic, seems to go beyond what might normally be expected. The building directory at the security desk in the marbled lobby begins to hint, though, at what might be unusual about this building, as the companies it lists are uniformly telecommunications providers—Verizon, Covad, Level 3 Communications, and more. For a thirty-five-story building, as it turns out, One Wilshire houses very few people, but it is nonetheless quite full. The building is given over, almost entirely, to data centers and computer server rooms, colocation facilities and network equipment, with the building's high-speed network spine as critical to its operation as its architectural supports. The nerve center of the building, and its raison d'etre, is the "meet me room" on the fourth floor, an oppressive warren of telecommunications equipment in locked cages, connected overhead by a dense, tangled mesh of bright yellow single-mode fiber-optic cables. One Wilshire's meet-me room is a major Internet connection point. It is the physical site where the digital networks of corporations like Covad and Level 3 are connected together, the point at which digital messages flow from one operator's network to another. It is where the "inter" of the "internet" happens.[1]

That the vauntedly virtual world of Internet communications is in fact grounded in physical and material realities is not, of course, a novel observation. Lisa Parks has written about spatiality and territorialization in terms of the footprints of the satellites by which global communications are made real (and distinctly local);[2] Nicole Starosielski has examined the network of undersea cables snaking across the Pacific Ocean floor and the flows of capital, expertise, and people that follow their paths;[3] Steven Graham and Simon Marvin have discussed what they call the "splintering urbanism" that comes about when different networks and infrastructures create different experiences of urban space;[4] and Kazys Varnelis has examined Los Angeles in terms of its manifestations of networked infrastructures (including, specifically, One Wilshire).[5] I, too, take the materiality of digital networks as my topic here, but my concern is somewhat different. The digital materiality that concerns me is not the materiality of the infrastructures and wires but the materiality of the digital signals that cross them. I argue that data and their protocols are also material, both in their consequences for the organization of infrastructures and in their specific manifestations as flows of electrons and signals that spread out over the wires and channels that make places like One Wilshire work. While writers like Alexander Galloway have examined the politics of network protocols, and others such as Milton Mueller have written about the institutional arrangements that bring them into being, I want here to consider protocols as material that needs to be matched with and understood in relation to the brute infrastructural materialities that we encounter in places like One Wilshire.[6]

To make these questions more concrete, I will focus here in particular on the topic of Internet routing—the protocols and mechanisms that, first, allow digital data to traverse a complex, heterogeneous, and dynamic Internet and that, second, distribute the information that allows this traversal to be accomplished. In doing so, I want to suggest a new line of inquiry for examinations of digital materiality, one that moves from a study of physical infrastructural arrangements to consider the materialities at work in the protocols, representations, models, and interactions that take place within and through those infrastructures. It is for this reason that my focus here is on the materiality of Internet *routing*, not the materiality of Internet routers or the materiality of Internet routes. That is, my concern is not with the physical infrastructure as such—the cables, the servers, the switches, the buildings, and so on—but with the processes at work.

What does it mean to think of these as material? It requires first that we adopt a methodological skepticism toward the separation of domains of practice and expertise that disciplinary and institutional boundaries typically break apart—communication infrastructures, computational platforms, protocols, and applications. It involves, instead, seeing network protocols as things that are designed to serve applications, to run on computational platforms, and to control infrastructures, bound up with and contributing to the material realization of them all. It requires too that we take a historically and geographically situated view that examines the Internet not as a Platonic ideal but as a practical and political object, one that has been shaped by many different considerations and is just one of a range of possible Internets.

In his influential book *Mechanisms: New Media and the Forensic Imagination*, Matthew Kirschenbaum draws the distinction between formal and forensic materiality as aspects of media analysis.[7] He encourages media scholars to go beyond "the event on the screen" as the object of analytic attention and to examine the specific forms of technical practice that produce those events. This is not simply a call to examine the technological and material foundations of digital experiences, although that is an important consideration, particularly from the perspective of archival studies. Rather, I take it to be a call for an examination of the *relationship* between infrastructure and experience, with attention to the processes by which digital experiences are produced; and, further, to warrant an investigation of the practices of technological design that generate these arrangements. Galloway examines the notion of protocol as a manifestation of relational power, taking his cue from the pattern of technological arrangements but in general without examining their specifics. In this study, I want to maintain the relationality that both Galloway and Kirschenbaum draw our attention to, but in a manner that, first, draws on the dual nature of protocols as mechanism and inscription à la Galloway, and, second, addresses the production of the event on the screen à la Kirschenbaum.

With these perspectives in mind, I will begin by illustrating the broad approach and then proceed by degrees through the details of internetwork routing and its materialities, before returning at the end to the broader programmatic question of how this informs a more general inquiry into the materialities of information.

The Material Analysis of Protocols

In the late 1980s and early 1990s significant research attention in computer networking was devoted to ATM (Asynchronous Transfer Mode) networking as a technology for high-speed digital communications. Unlike the TCP/IP protocols familiar to Internet users, which had been developed primarily in the academic community, ATM networking was a product of government and commercial interests, in particular the large telecommunication companies, which, in many countries, operated as government-regulated monopolies. ATM networking was standardized through the International Telecommunications Union (ITU), a UN body whose members were not individual technical experts but member countries. In this forum, technological considerations and national interests were quite explicitly bound together.

Unlike TCP/IP, ATM is not a packet-switching technology, but like TCP/IP, which divides messages into smaller, individually routed units called "packets," ATM also divides messages or message streams into small, fixed-size units known as "cells," each of which carries its own addressing and control information. A key design decision in this approach concerns how small or large these fixed-size units should be. In general, larger cell sizes make more efficient use of limited transmission capacity because they increase the ratio of payload (the digital content in each cell) to header (the fixed size control information with which each cell begins). This means that more of the bits being transmitted along a cable are bits that carry digital content. On the other hand, however, larger cells mean fewer different messages can be carried along a channel in a fixed amount of time, since larger cells take longer to transmit; smaller cells can more easily be interleaved. Similarly, in the face of this difficulty, larger cells require more buffer capacity at switching units, where they may have to be stored awaiting transmission. So, larger cells make better use of limited *transmission* capacity but smaller cells make better use of limited *switching* capacity.

Arguments about transmission and switching capacity and their relative merits are important and have economic consequences, but some of the other topics that framed the debate about cell size have a more fundamental connection to questions of existing fixed infrastructure. One of these concerns the length of the transmission cables along which ATM cells would travel. When signals travel down a wire, they have a tendency to reflect off the end of the cable and the terminating equipment, sending an "echo" back down the cable along which they have traveled. In order to reduce interference between a signal and its own echo, then, it is better if the transmission is relatively short, so that by the time the echo reaches a given point, the transmission has already ended. "Relatively" short, in this case, means "short with respect to the length of the cable"—on a longer cable, a larger message can be sent without an echo interfering with the message itself. (Echo cancellation hardware can be used to reduce this problem, but it adds significant fixed costs—an important consideration when telecommunication companies are looking for new technologies that can be implemented on their existing line infrastructure.)

Consequently, debates about the best size for a cell inevitably involve differences between groups who have largely "long" wires and those who have largely "short" ones. Telecommunications operators in large countries, such as the United States, where long-haul networks of the sort on which ATM would be deployed often cover large distances, were inclined to favor larger cell sizes, whereas those in smaller countries, and particularly those who didn't have echo cancellation installed, favored smaller ones. In the standardization process, then, the United States advocated a relatively large payload size of sixty-four bytes. France and other European countries, on the other hand, argued for much smaller size cells with a thirty-two byte payload. Each country's position incorporated the particular perspectives that one might have on questions of protocol design and efficiency in the context of their own fixed infrastructures and geographical realities. Sixty-four bytes simply made for cell sizes too "long" for small countries. A compromise position was eventually struck, and ATM cell size was fixed at forty-eight bytes of payload along with five bytes of header for a total cell size of fifty-three bytes—a size deemed equally inconvenient for all parties.

The example of the debate around ATM cell sizes troubles questions of sociotechnical analysis as they appear both in technological and in sociocultural academic circles. For the technologists, it undermines a conventional idea that while networks and technological objects are used in ways that are governed by the social, they remain themselves simply technical objects. This is what Kling et al. have called the "layer-cake model" of sociotechnical analysis—the idea that the social is something that happens "after" and "over" the technical, a consequence of material arrangements that are themselves solely technical, with the social presented as being "at the top of the network stack" by analogy with the OSI network stack, an oft-used pedagogical device.[8] Sociocultural analysts, on the other hand, are skeptical of the technological determinism at the heart of those analyses and see technological arrangements as always already social. However, in these analyses, "the technical" is rarely opened up to critical scrutiny; while technological systems are understood to be amenable to (indeed, to require) sociological analysis, the specific technological arrangements and their alternatives are rarely examined in detail. What is more, where they are, the focus is often on hardware and infrastructure. While we might laugh at the poorly informed political discussion that suggests "the Internet is a series of tubes," an understanding of internetworking as more than simply just that remains rare in sociocultural analysis;[9] protocols, representations, and their dynamics as effective media are largely unexamined. Consider a second example in which this question of dynamism and effectiveness is central. Protocols need to be designed not only to fit the sizes and properties of fixed digital infrastructures. They must also be computationally feasible—that is, fit for the computational infrastructures that will process them. When thinking about the network core, this takes on particular resonance.

Craig Partridge's book *Gigabit Networking* provides a comprehensive overview of the technical issues involved in running networks at gigabit speeds (that is, at data transmission rates of more than one billion bits per second).[10] However, at the outset of his discussion of the use of the Internet TCP/IP protocols the author finds himself presented with an odd challenge: while it is clear that network technologies can transmit data at gigabit speeds—indeed, much faster—it was not at the time universally accepted that IP-based networks could run at that speed. Skeptics argued that the protocol itself—the set of conventions and rules about how data should be formatted and how it should be processed when being sent or received—made gigabit speeds impossible. In IP networks, data is processed as "packets"—discrete chunks of data that, together with some information about where the data should be sent and how it should be interpreted, are processed independently by the network. The challenge for running IP (or any protocol) at gigabit speeds is whether a network router can process a packet before the next one arrives. As network transmission rates increase, the time available to process any given packet decreases. Advocates for alternatives to IP would argue that IP couldn't "scale"—that is, that the overhead of processing the protocol data was too great, and that IP packets could not be processed fast enough to make effective use of a gigabit network. Partridge begins his discussion, then, by laying out specific software code that can interpret IP packets and can be shown to operate fast enough to enable gigabit speeds.

The very fact of Partridge's demonstration—and, more to the point, its necessity as a starting point for his discussion—highlights some significant issues in how we think about networking technologies. These are issues that might seem self-evident to computer scientists and engineers, although their very obviousness might blind us to their importance; to others who write, talk, and think about networking technologies, though, they may seem unusual. First, it highlights the idea that, although we often talk about them as though they are the same thing, what the network can do is not the same as what the transmission lines can do—that is, a transmission line might be able to transmit data more quickly than the "network" can. Second, it highlights the fact that different network protocols can have different properties, not just in terms of their formal properties but also in terms of their practical capacities—a protocol does not simply imply rules and conventions (see Galloway) but is also subject to analyses that focus on weight and speed. Third, it draws our attention to the relationship between the "internals" of a network—the practical manifestations of how it operates—and the "externals"—that is, what it can do for us and how.

The broader, programmatic point to be made here is that a materialist concern with the Internet needs to be engaged not just with what "networks" are but rather with what this particular network—as a specific case, and as one among a range of alternatives—might be. It is not enough to argue for the critical role of decentralization, to examine the formalized disengagement afforded by protocols, or to note the geographical siting of infrastructure. Rather, I argue that a materialist account must examine just how that decentralization becomes possible, what specifically the protocols do and how, and how specific forms of spatial and institutional arrangements become possible in *just this* Internet—one that is materially, geographically, and historically specific. It is for this reason that I want to take as my focus here the question of Internet routing, as arguably one of the key conditions on what the Internet—our Internet, our current Internet—actually is, highlighting its material specificities.

Fundamentals of Internet Routing

Although we talk casually about "the Internet" or "the network," the very terms "Internet," "Internet protocol," and "internetworking" point to a more complicated reality, which is that there are multiple networks. In fact, this is the point. The crucial feature of Internet technology is that it provides a way for data to move across multiple networks, potentially of quite different sorts. The Internet links a series of networks together—local area networks on a college campus, long-distance networks operated by commercial providers, cellular networks maintained by mobile service operators, Wi-Fi networks in homes, and so on—in such a way that data can move easily from one to another.

Routing refers to the function whereby Internet messages or packets get from their source to their destination or, more accurately, from the network to which their source computer is connected to the network to which their destination computer is connected. Packets might have to traverse multiple other networks in order to get from one to the other. Those networks might be of dissimilar types, they might be owned and managed by different authorities, and there might be multiple alternative routes or paths. In this model, a "network" is an individual span of some kind of digital medium, and so it might be quite small. For instance, it is not uncommon for an average American home network to incorporate three different networks—a Wi-Fi network, a wired network (into which your Wi-Fi router is connected), and another "network" which constitutes the connection between your broadband modem and the service provider. Each transmission from your laptop to the outside world must start off by traversing these three separate but connected networks. Similarly, there are many networks inside the Internet infrastructure. For example, from where I sit writing this text in a hotel lobby in Paris, the UNIX *traceroute* utility reveals that a connection to my university's web server traverses more than twenty networks, including the local Wi-Fi network, several networks operated by international Internet provider Proxad (including networks in London, New York, and Palo Alto), several networks operated by the California nonprofit academic network operator Cenic, and finally the multiple networks of my own university. Routing is the process by which packets are correctly directed across these different network connections to reach their destinations.

Internet routing depends on three key ideas—gateways, routing tables, and network addresses. Gateways (also known as routers) are central elements in routing. A gateway is essentially a computer that is connected simultaneously to two or more networks and thus has the capacity to receive packets via one network and then retransmit them on (or "route them" to) another. For instance, a domestic broadband modem is connected both to your home network and to your service provider's network, and so it can act as a gateway that moves packets from one to the other; when it receives a message on your local network that is destined for the wider Internet, it will retransmit the message on the connection that links it to your service provider, where another router will see the message and retransmit it appropriately. Most laptops and desktop PCs have the capability to act as gateways if they are connected to more than one network simultaneously, although most gateways are actually dedicated devices (like your broadband modem or, on a larger scale, routers made by companies such as Cisco). A gateway, then, can route packets from one network onto another. To do so, though, the gateway requires information about how the network is organized.

The second key element is the information a gateway needs in order to successfully route packets. In general, this information is captured by a gateway's "routing tables," a list that associates destinations with networks. These can be thought of as rules that say, for example, "If you see a packet that is destined for X, send it to network Y." When a gateway receives a packet, it looks up these rules to determine which of its connected networks should receive it. Note that a rule of this sort does not imply that destination X is directly connected to network Y; it might simply be that another gateway connected to network Y is one step "closer" to destination X. Routing, in other words, is decentralized in TCP/IP. There is no central authority where the topology of the network is entirely known, nor any single point from which all networks can be reached. Rather, a packet makes its way across, say, the twenty "hops" from my Paris hotel lobby to UC Irvine's servers through a series of individual decisions made at each gateway it passes. A packet is passed from gateway to gateway according to routing tables that move it closer to its destination until it is finally delivered to a network to which its destination host is directly connected.

This brings us to our third concern with network addresses. Routing tables would become impossibly large if they needed to store routes for each host connected to the Internet. Instead, they record routes to networks, not to individual hosts. This requires, in turn, that some means be found to identify and name networks. A typical IP address—the familiar four-byte number like 128.195.188.233—identifies a particular host, but some part of the address is seen as numbering the network to which that host is connected. The precise mechanism by which networks are numbered will be discussed in more detail later, but for now it is sufficient to note that routing tables do not record routes for every single host address, but rather for networks (for example, 128.195.188.0, a network to which the host 128.195.188.233 might be connected.)

Before looking more directly at the protocol issues involved in distributing routing information, it is worth pausing to note some materialist concerns at work even at this foundational level. First, we should be attentive to the questions of topologies and temporalities. In small and stable internetworks, routing is a relatively straightforward operation. However, as the internetwork grows larger, the information needed to produce effective routes also grows, as does the computational power needed to process it. Similarly, in a network that is often changing, the potential paths are also highly variable. Technically, the topology of the Internet changes every time someone unplugs a cable or powers down a Wi-Fi hotspot. In a network of the scale and geographical distribution as the Internet, those sorts of changes are happening constantly. Particular solutions to the problem of routing embody assumptions about the significance, the pace, and the consequences of change and disruption.

Second, we need to be concerned as well with issues of bounds and scale. The question of routing—and in particular, its decentralized decision-making process, which we will revisit—draws attention to how particular kinds of boundaries and particular scales of operation and significance arise in the network-as-practiced. That is, the question of the temporality of changing topologies also creates zones of social, organizational, and institutional autonomy and dependence, and forces the emergence of scales and structures of control. This issue will become more important in the discussion to follow.

Third, it suggests that we might need to distinguish between protocol, implementation, and embodiment. The distinction between protocol and implementation is well recognized: we understand, analytically, the distinction between those formal descriptions of how systems interoperate on the one hand and the specific pieces of software that implement those protocols on the other—the fact that protocols are rarely complete, for example, or at least that implementers have discretion in the choices they make about the extent to which deviations from the protocol will be accepted. The distinction between protocol and embodiment speaks to a different issue. It highlights the fact that a protocol might be clear, well-defined, and effective in design and yet ineffective or inoperable in practice—when routing tables are too large, for example, or when network connections are too slow, or when routing hardware lacks the computational resources needed to process the protocol, or where the protocol is poorly matched to local conditions. A failure of protocol-connected systems is not in itself a failure of protocol, or even necessarily of implementation; specific embodiments, not just of infrastructure but crucially also of the protocol itself—data on the wire—also matter.

The fourth consideration involves decentralization, deferment, and delegation. The decentralization of Internet operations—the fact that packets are routed without appeal to a central authority, and that Internet policy is driven by what is called the "end-to-end" model, which argues for placing control at the edges of the network—is one of the most widely acknowledged features of the Internet as a specific technology.[11] However, one of the things that an examination of Internet routing reveals is that the flexibility of decentralized routing depends on many other components that may not have that same degree of decentralized control. Galloway has noted what kinds of commitments to collective agreement are implied by decentralization within a regime of protocol-driven interaction.[12] We might also point to questions of network addressing and topology as places where decentralization operates within a framework of deferment of authority and delegation to others.

Fifth and finally, an understanding of the operation and specific manifestations of routing and routing protocols needs to be seen within the context of conventions of use and practice. This point will come to be of central importance below, but it should be clear even in the discussion so far that the effectiveness of Internet routing depends not simply on the operation of the protocols but on the relationship between protocol and conventions of use—conventions that govern patterns of network addressing, for example, or topologies and practices of connectivity among service providers, or our conventional patterns of distinction between those services that are provided "close to the core" or "at the periphery" of the network. One can never rely purely on what the protocol defines or how the mechanisms operate for an account of the specifics of how our networks work.

Bearing these considerations in mind, let's proceed to another level of the analysis of routing: the protocols that govern the distribution of routing information.

Routing Protocols

Efficient and effective Internet routing depends on information about local network topology being available to routers. However, as we have seen, the decentralized nature of the Internet means that there can be no single source of information about the structure of the Internet. Routing is wayfinding without a map; it is based instead on local decisions using the best-available information. Just as routing decisions are made in a decentralized manner, the sharing of information on which those decisions are based is similarly decentralized. Routers periodically exchange information about available network connections and routes to update each other's tables. Routing protocols define how information about network topology and available routes—the information that routers store in their routing tables and use to make routing decisions—spread through the network.

There is no single, universal routing protocol. Different protocols exist for different needs, and different protocols have predominated at different historical moments. I will examine two protocols here—the Routing Information Protocol (RIP) and the Exterior Gateway Protocol (EGP).

RIP: The Routing Information Protocol

One of the earliest Internet routing protocols, and one of the most widely deployed, is RIP: the Routing Information Protocol. RIP's widespread adoption in the 1980s and 1990s derived not from its technical superiority but from the fact that it was implemented by the *routed* (pronounced "route-dee") software distributed with the 4BSD Unix software distribution, popular in academic environments, and later with Sun Microsystem's SunOS operating system. In fact, this software was for some time the only available specification of the protocol: there was no formal description or definition, merely the behavior of the implementation. Not only was RIP an early protocol for exchanging Internet routing information, but it was heir to a longer legacy; RIP is a variant of the routing information protocol that formed part of Xerox's Network Service protocol suite (XNS), which itself embodied ideas originally developed as part of Xerox's PUP and Grapevine systems.[13]

A router running RIP will periodically broadcast a description of its routing tables to other routers that it knows about. The information that an RIP router provides is essentially its own perspective on the network—that is, it describes everything relative to itself. RIP provides two pieces of information about each of the routes—its destination and the hop count. Hop count—the number of routers or network segments to be traversed in order to reach the destination—serves as an approximate metric of distance. Some networks might be fast, some slow; some might cover long distances, some shorter ones. These distinctions are not, however, captured in the hop count, which provides a more rough-and-ready basis for decision making about the most efficient route.

RIP uses just four bits to record the hop count, allowing it to indicate a range of values from zero to fifteen. A value of fifteen indicates an "infinite" hop count, used to signal that a network is unreachable. Accordingly, a network in which routes are communicated solely via RIP can be no more than fifteen networks "wide"—in other words, a packet cannot be routed across more than fifteen segments unless other measures are taken. In practice, this makes RIP useful primarily in smaller networks that are themselves connected using different routing protocols; the global reach of "the Internet," in other words, is premised on such specificities.

EGP: The Exterior Gateway Protocol

EGP, the Exterior Gateway Protocol, is a more complex protocol than RIP. Itself an evolution of an earlier protocol called GGP, it is designed for communication between routers that connect so-called "autonomous systems"—networks that span particular organizations, corporations, or institutions. Intuitively, if you imagine that a university such as UC Irvine or a corporation such as Intel each runs its own networks according to the institution's own conventions and procedures, then each is designated as an autonomous system; EGP is the protocol by which routing information about one of these networks is communicated to routers for the other.

As with RIP, the core of the EGP is a mechanism by which routes are shared from one router to another. Also like RIP, hop count is used as a rough-and-ready measure of distance, although unlike RIP, routes of more than fifteen hops can be expressed. EGP expresses the "distance" to particular destinations relative to specific, identified gateways (rather than implicitly from the sending router). The protocol is also more fully featured than that of RIP; for instance, there is an explicit component by which new neighboring gateways can be identified and polled. By contrast, this structure in RIP is left as a matter of configuration.

The intended purpose and conventional use of EGP differs from those of RIP, which is not committed to particular forms of use, although its constraints limit the conditions under which it can be deployed. EGP, on the other hand, is designed specifically to connect autonomous systems. Accordingly, EGP is designed to be used with particular conventions regarding which routes should be advertised and which should not; these conventions are not encoded directly in the protocol, but rather the protocol is designed under the assumption that administrators will be conscious of them.

Material Considerations

Network protocols are shaped by material constraints. ATM cells have not just an abstract size but also a length when transmitted along cables. IP packets do not simply have a format, they have a format that has consequences for the speed of processing in network routers and so can be limited by switch fabrics. Similarly, the centrality of routing to the Internet can be understood materially in terms of the arrangement of network nodes, the cost of routing, the structure of networks, the size of routing tables, and the dynamics of connectivity. Critically, this materiality cuts across apparently different domains of concern—from the practice of network operations to the rhetoric of democratic access. I will consider four aspects here.

Routing Tables, Classless Routing, and the Politics of Address Exhaustion

In 1993, changes were introduced to the way that Internet routing worked. The new model—called CIDR (pronounced “cider”), or Classless Inter-Domain Routing—extended and built upon conventions of “subnet routing” that had been in operation for some time but were adopted as the basis of a new model of routing that would apply across the Internet.[14] CIDR was a response to a growing problem for Internet routers, particularly core routers: the size of routing tables. The large number of networks meant that routing tables were growing, with three consequences: the storage demands on each router were growing significantly, beyond what had ever been imagined; the processing time necessary to sort through the routes was also growing; and the process of transmitting the routing tables (via protocols like EGP) was becoming unwieldy because the routing tables were so large.

CIDR was known as “classless” routing because it replaced an earlier scheme of network addressing that distinguished between three “classes” of networks, A, B, and C. Class A network addresses were distinguished by just their first eight bits, so that, for instance, all addresses that begin 13.x.x.x belong to Xerox Corporation. Holding a class A network allocation meant that one had control of a total of 16,777,216 separate Internet addresses (although a handful are reserved and cannot be used.) Class B addresses were distinguished by their first two bytes (sixteen bits). Addresses in the range 192.168.x.x, for example, are B addresses designated for private use. Each class B allocation includes 65,536 addresses (with the same caveat.) Class C network addresses were distinguished by their first three bytes and provide 256 addresses (again, minus a handful).

In the original scheme—which, by contrast with classless routing became known as "classful" routing—the three classes of addresses served two simultaneous functions. They were the units of routing because, as discussed above, routes direct packets toward networks rather than to individual hosts; and, at the same time, they were also the units of allocation. If a new entity needed some address space—perhaps a university, a government, a company, or an ISP—then it would be allocated one or more class A, class B, or class C addresses. This conflation of two functions then resulted in two interrelated problems that classless routing could, it was hoped, solve. The technical problem was the growth of routing tables; the political problem was the growing scarcity of address space.

In classful addressing, networks are either class A, class B, or class C; the division between the network portion of an Internet address and the host portion occurs at an eight-bit boundary, so that in a class A address the thirty-two bits of an IP address are divided into eight bits of network identifier and twenty-four bits of host, rather than sixteen and sixteen for class B, and twenty-four and eight for class C. Classless addressing introduced a mechanism whereby the boundary between network and host portions could be made more flexible.

This mechanism would address the technical problem, the growth of routing tables, by allowing smaller networks to be coalesced into a single network that was still smaller than the next class up. This was especially important for class C addresses. A class C address covers only around 250 possible hosts. For many organizations that didn't need a whole class B address, a class C network was too small. So many organizations would be allocated many class C network addresses—each of which would require individual entries in routing tables. By creating the opportunity to have, say, ten bits of host space (for around twelve hundred hosts) rather than eight bits—a new arrangement not possible in traditional classful routing—classless routing could shrink the size of the routing tables by dealing with networks at a new level of granularity.

This also addressed, to some extent, the political problem. The introduction of classless routing may have been sparked not least by the troublesome growth of routing tables, but it directly addresses another regular concern around Internet management and governance, the question of address allocation and exhaustion. The original strategy for Internet address allocation was based on address classes, with class A addresses particularly prized among large organizations. This resulted in various well-known inequities of allocation, such as the fact that MIT (with the class A network address now known as 18.0.0.0/8) has more allocated addresses than China.

Classful address allocation suffers a second problem, which is address wastage. Xerox, for instance, was allocated the class A network 13.0.0.0, although it seems unlikely that they would use all 16 million available addresses; however, at around sixty-five thousand addresses, the next step down (class B) was deemed likely to small. No effective mechanism was available for smaller allocations. It remains the case, even in a world of classless routing, that the IP address space is continually approaching exhaustion, even as we know that wastage goes on within the already allocated blocks.[15] Again, this is also happening within the context of deep historical inequities in address allocation, as noted above.

The fact that routing presents both technical problems and political problems is not surprising. What is important here is the material entwining of these problems—the fact that the politics of network-address space allocation and the dynamics of routing-table growth and exchange are dual aspects of the same material configurations. The political and technical issues are not so much twin problems as different facets of the same problem, which is that in the interests of efficiency and effectiveness, networks are the primary units of route determination. When we see these in a material light—that is, when we see route advertisements as things that have size, weight, lifetimes, and dynamics, then the problems become material too.

Granularity and Networks as User Objects

This discussion has been based on a set of technical conditions that govern what "a network" is, in Internet terms—not an abstract large-scale entity ("the Internet" rather than the ATM network) or an autonomous system ("UC Irvine's network") or even the entities of common experience ("my home network"), because none of these are the sorts of "networks" of which the Internet is a connective fabric. Rather, the "networks" that the Internet connects are particular media arrangements—lengths of coaxial or fiber-optic cable, wireless signal fields, tangles of twisted-pair wires, and so on. These networks are technical entities but not user entities.

Or, at least, not usually. The vagaries of routing can also result in the infrastructure of network arrangements becoming visible and even important as a question of user interaction. Voida et al. document an interesting and unexpected case—the case of music sharing within a corporate network.[16] A network in IP is not only the unit of routing, but also the unit of broadcast—that is, messages can be delivered to all the hosts on a particular network. This in turn means that networks can become visible as marking the boundaries of broadcast-based sharing. In the case that Voida and colleagues document, participants in a corporate network come to realize aspects of its internal structure—that is, the way that it is composed of multiple subnetworks connected via routers—through the patterns of visibility and invisibility that they can perceive through the use of music sharing in the iTunes application. iTunes can allow people on a network to share music with each other, and the corporate participants in the study took advantage of this facility. However, they began to encounter problems that resulted from the network topology—that they couldn't see a friend's music, for example, because the friend was on a different network, or that someone's music might "disappear" if that person relocated to a different office, even within the same building. The network topology itself became visible as a routine part of their interactions, and suddenly the material arrangements that underlay the notion of "the network" became an aspect of the user experience.

What is significant here is the collapse of the network stack—the tower of abstractions that separates the physical medium from internetworking protocols, and internetworking from applications. Suddenly, in this model, the specific material manifestation of IP networks—as runs of cable governed by signal degradation over distance and the number of available sockets on a switch—need to be managed not only for the network but also for the users. Network engineers and researchers have long recognized that the abstractions of the network stack may be a good way to talk about and think about networks but a less effective way of managing or building them.[17] However, the intrusion of the material fabric of the network and its routing protocols into the user experience is a different level of concern. If granularity is a core material property, one that shapes the decomposition of the network into effective units of control, then Voida and colleagues highlight the way that this materiality manifests itself in multiple different regimes.

The Emergence of Centralized Structure

The mythology of the Internet holds that it was designed to withstand nuclear assault during the Cold War era, both by avoiding centralized control and by using a packet-switching model that could "route around" damage, seeking new routes if any node were lost. Whatever the status of that claim,[18] it is certainly true that one of the defining features of the Internet is its variable and amorphous topology. Unlike networks that rely on a fixed arrangement of nodes (such as a ring-formation in, for example, the Cambridge Ring), the Internet allows connections to be formed between any two points, resulting in a loosely structured pattern of interconnections (what computer scientists call "a graph").[19] The absence of formal structure and the avoidance of critical "choke points" or centralized points of control is arguably one of the essential characters of the Internet.

However, our examination of routing and routing protocols reveals a more complicated situation at work. The contrast between the operation and the operating context of RIP and EGP is educational in this respect. RIP is a simple protocol that predates the Internet; EGP, on the other hand, is a protocol that emerged over time and evolved in response to the conventional practices of the Internet as a set of practical institutional arrangements. As described, EGP is based around the idea of autonomous systems—the idea that different networks will belong to different institutional entities and will be managed autonomously by different authorities. It is also based on the idea that access to each autonomous system will be brokered by one or a small number of authoritative gateways. The conventions that govern the use of EGP—for instance, the rule that no gateway may advertise a route to a network other than those within the autonomous system it represents—foster this concentration of authority. In other words, what we see, within the framework of the open connection, open routing, and amorphous structure afforded by the Internet's fundamental technologies, is the emergence of authority, control, institutional structure, and local points of centralization. Centralization may not be inscribed in the basic protocols of TCP/IP but may emerge at other points as a consequence of practicality.[20]

This raises some interesting questions. One is: Precisely which Internet do we talk about when we celebrate openness, diversity, and decentralization as characteristics of the Internet when compared to mass media as forms of communication? Certainly, we can celebrate the potential for these properties, but perhaps not their practical embodiment within the Internet as we know it—our Internet rather than a possible Internet. It is not at all clear that the Internet, our Internet, is in fact the decentralized, open, and democratic tool of connection and communication that technolibertarian rhetoric applauds. Second, it is important to see the kinds of centralizing tendencies and emerging structure and conventions that EGP represents and encodes as material consequences of the Internet's form—the dynamics of its topological organization, the pragmatics of routing, the consequences of bandwidth provision, the economics of access, and so forth. That is, they are not consequences of Machiavellian dabbling, of corporate subversion, of capitalist corruption, or of state intrusion. In the spirit of Kelty's recursive publics, these are protocols and conventions, after all, that have been developed by the very people who hold dear the Internet's independence from these constraints.[21] Rather, as I have attempted to show here, they are material consequences of the relationship between infrastructure and protocol, between representation and practice, and between encoding and practical action.

Historical Patternings

As we have seen, the routing protocols implemented on the Internet reflect a historical pattern of evolution. EGP grew out of GGP; it was itself superseded by BGP (the Border Gateway Protocol), which implements CIDR and has been in use since the mid-1990s. RIP was derived from earlier protocols developed at Xerox—first PUP and then XNS; the XNS routing protocol similarly became the basis for routing in Novell's NetWare product suite. From one protocol generation to the next, certain ideas and expectations are inherited in the technical design. What else is inherited along with those technical features? Each protocol is designed to capture both what has worked well in a protocol that came before it and to correct or respond to problems that have arisen. Assessments of success and failure, and the identification of effective and ineffective properties, are made relative not to designs but to deployments.

One of the central considerations that arises when we see the protocols as emerging out of deployments rather than simply as technical designs is the issue of control, authority, and management. The question essentially becomes to what extent the network should operate as a self-organizing, adaptive entity (which is the principle embedded in the routing algorithms themselves) and to what extent it is an entity that is actively managed (the principle increasingly embedded in the protocols by which routing information is distributed). In the evolution of GGP to EGP to BGP, we see similarly an evolution in understandings of the degree of control and management needed within core networking routing. Within the history of the protocols of which RIP was a part, we also see some important considerations in the expectations of deployment. PUP and XNS were protocols deployed within formal organizations (PUP internal to Xerox, and later XNS as a product for Xerox's corporate customers); similarly, Novell's later NetWare product was also a product for corporate networking. Corporate network management generally implies the presence of network administrators and a policy function to manage how networks are designed, deployed, and operated. What we find embedded in the protocols are organizational expectations about structure and management—the constraints within which the flexibility of an adaptive, evolving, self-managing network can operate. Indeed, the question we might want to ask at each juncture where the open and self-managed nature of the network appears, such as in routing, is: What structures or constraints are needed to allow this flexibility?

Conclusion

The mathematical computer scientist Edsger Dijkstra is reputed to have remarked, "Computer science is no more about computers than astronomy is about telescopes."[22] We may read this as a comment that computation should be seen as independent of its material manifestations. By contrast, recent interest in "information infrastructure studies" has demonstrated the importance of turning attention to the infrastructures of contemporary information systems in order to examine the processes and conditions under which information and information systems are produced, maintained, and put to work.[23] Whether or not this is a consideration for computer science as a discipline, it is most certainly an important consideration for the way computer science and its products manifest themselves in our everyday world—and for the way that computer science as a discipline evolves and develops.

In framing this article, I distinguished between the materialities of Internet routes, of Internet routers, and of Internet routing as distinct topics of investigation. Of course, at places like One Wilshire, these come together. In July 2013 the One Wilshire building was sold for more than $430 million—the highest price per square foot ever commanded for real estate in downtown Los Angeles.[24] The materialities of the routes signaled by the spray-painted markings on the street outside the building (tracing conduits below) and the materialities of the routers and devices the building hosts, powers, and cools are important ways into understanding the realities of contemporary digital life, but the protocols that tie these things together—that make the conduits effective, that enliven the servers, that allow them to operate productively—must also be part of the picture.

This chapter arises as part of a larger investigation of what colleagues and I have been calling the "materialities of information."[25] The twin foundations of this project are, first, that recent interests in materiality arising in the social sciences and humanities provide an important basis for understanding contemporary technological phenomena, with an attentiveness not just toward infrastructure but toward information itself;[26] and, second, that to do so effectively requires a foundational engagement with the computational objects and processes that make up the technological landscape.

Our interest in materiality is not taxonomic—that is, our goal is not to redraw the boundaries that separate "the material" from "the immaterial." Our concern instead is to examine the material considerations within the body of technical and social practice that constitutes the contemporary regime of information. Internet routes and Internet routers have been productively examined from the perspective of materiality;[27] however, turning a materialist eye upon Internet routing reveals the entangling of protocol, politics, and pragmatics that come together not only at physical sites like One Wilshire but in the materialization of protocols like EGP as embedded within systems of practice and technological artifacts. Indeed, I would argue that an examination of the materialities of information must engage with information systems not simply as metaphors of virtuality but as historically and geographically specific configurations of technology and practice. This provides an opportunity to frame an investigation of the materialities of information as what Pickering has called a "real-time understanding."[28] Routing—as manifested in *our* Internet, in *the* Internet, in *this* Internet, rather than in *an* Internet—provides an example of doing so.

Notes

1. The explorations of materiality presented here have arisen primarily in conversation with Melissa Mazmanian, whose contributions are central. Nicole Starosielski provided thoughtful and useful comments on an earlier draft. This work is supported in part by the National Science Foundation under awards 0917401, 0968616, and 1025761, and by the Intel Science and Technology Center for Social Computing.

2. Lisa Parks, "Satellites, Oil, and Footprints: Eutelsat, Kazsat, and Post-Communist Territories in Central Asia," in *Down to Earth: Satellite Technologies, Industries, and Cultures*, ed. Lisa Parks and James Schwoch (New Brunswick: Rutgers University Press, 2012).

3. Nicole Starosielski, "Beaches, Fields, and other Network Environments," *Octopus Journal* 5 (2011): 1–7.

4. Steven Graham and Simon Marvin, *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition* (London: Routledge, 2001).

5. Kazys Varnelis, *The Infrastructural City: Networked Ecologies in Los Angeles* (Barcelona: Actar, 2008).

6. Alexander Galloway, *Protocol: How Control Exists after Decentralization* (Cambridge, Mass.: MIT Press, 2004); Milton Mueller, *Networks and States: The Global Politics of Internet Governance* (Cambridge, MA: MIT Press, 2010).

7. Matthew Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination* (Cambridge, Mass.: MIT Press, 2008).

8. Rob Kling, G. McKim, J. Fortuna, and A. King, "Scientific Collaborations as Socio-Technical Interaction Networks: A Theoretical Perspective," *Proceedings of the American Conference on Information Systems*, Long Beach, California, 2000; Andrew Tanenbaum and Davis Wetherall, *Computer Networks*, 5th ed. (Boston: Prentice-Hall, 2010).

9. Ken Belson, "Senator's Slip of the Tongue Keeps on Truckin' over the Web," *New York Times*, July 17, 2006.

10. Craig Partridge, *Gigabit Networking* (Boston: Addison-Wesley, 1994).

11. J. H. Saltzer, D. P. Reed, and D. D. Clark, "End-to-End Arguments in System Design," *ACM Transactions on Computer Systems* 2, no. 4 (1984): 277–88; Tarleton Gillespie, "Engineering a Principle: 'End-to-End' in the Design of the Internet," *Social Studies of Science* 36, no. 3 (2006): 427–57.

12. Galloway, *Protocol*.

13. D. Oppen and Y. Dalal, "The Clearinghouse: A Decentralized Agent for Locating Named Objects in a Distributed Environment," *Office Systems Division Tech Report OSD-T8103* (Palo Alto, Calif.: Xerox Corporation), 1981; A. Birrell, R. Levin, R. Needham, and M. Schroeder, "Grapevine: An Exercise in Distributed Computing," *Communications of the ACM*, 25, no. 4 (1982): 260–74; D. Boggs, J. Shoch, E. Taft, and R. Metcalfe, "Pup: An Internetwork Architecture," *IEEE Transactions on Communications* 28, no. 4 (1980): 612–24.

14. V. Fuller, T. Li, J. Yu, and K. Varadhan, "Classless Inter-Domain Routing (CIDR): An Address Assignment and Aggregation Strategy," *RFC 1519*, Internet Engineering Task Force, 1993.

15. Z. Wang and J. Crowcroft, "A Two-Tier Address Structure for the Internet: A Solution to the Problem of Address Space Exhaustion," *RFC 1335*, Internet Engineering Task Force, 1992; Mueller, *Networks and States* (Cambridge, Mass.: MIT Press, 2010).

16. A. Voida, R. Grinter, N. Duchenaut, K. Edwards, and M. Newman, "Listening In: Practices Surrounding iTunes Music Sharing." *Proceedings from the ACM Conf. Human Factors in Computing Systems CHI 2005*, Portland, Oregon, 2005, 191–200.

17. D. Clark and D. Tennenhouse, "Architectural Considerations for a New Generation of Protocols," *ACM SIGCOMM Communications Review* 20, no. 4 (1990) 200–208.

18. It is true that Paul Baran's original report on packet switching was inspired by the nuclear-assault scenario, and that the developers of the Internet recognized the originality and value of this approach to network design; however, the extent to which the Internet or its ARPANET predecessor was designed with this goal in mind is highly questionable. Paul Baran, "On Distributed Communications," *RAND Memorandum RM-4320-PR* (Santa Monica, Calif.: Rand Corp., 1964).

19. A. Hopper and R. Needham, "The Cambridge Fast-Ring Networking System," *IEEE Transactions on Computers* 37, no. 10 (1988) 1214–23.

20. EGP was replaced by BGP, which has different properties. Unlike EGP, which runs only at the edges of autonomous systems, BGP is also used internally, which provides some opportunities to defuse the pattern of centralization that arises with EGP. However, in BGP, routing is governed by explicit policies rather than by distance metrics, representing to some degree an assertion of authority and control over autonomy and adaptation. The same tendencies, then, remain at work.

21. Chris Kelty, *Two Bits: The Cultural Significance of Free Software* (Durham, N.C.: Duke University Press, 2008).

22. It is at best questionable whether Dijkstra himself ever said this; rarely at a loss for a pithy comment, Dijkstra, like Mark Twain or Winston Churchill, was one of those quotable characters to whom all manner of comments are often attributed.

23. Geoffrey Bowker, Karen Baker, Florence Millerand, and David Ribes, "Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment" in *The International Handbook of Internet Research*, ed. Jeremy Hunsinger, Lisbeth Klastrup, and Matthew M. Allen (New York: Springer, 2010), 97–117.

24. Roger Vincent, "Downtown L.A. Office Building Sells for Record $437.5 Million," *Los Angeles Times*, July 17, 2013.

25. Paul Dourish and Melissa Mazmanian, "Media as Material: Information Representations as Material Foundations for Organizational Practice," in *How Matter Matters: Objects, Artifacts, and Materiality in Organization Studies,* ed. Paul Carlile, David Nicolini, Ann Langley, and Haridimous Tsoukas (Oxford University Press, 2013), 92–118.

26. Daniel Miller, ed., *Materiality* (Durham, N.C.: Duke University Press, 2005); Matthew Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination* (Cambridge, Mass.: MIT Press, 2008); Wanda Orlikowski, and Susan Scott, "Sociomateriality: Challenging the Separation of Technology, Work and Organization," *Academy of Management Annals* 2, no. 1 (2008): 433–74.

27. Adrian Mackenzie, "Untangling the Unwired: Wi-Fi and the Cultural Inversion of Infrastructure," *Space and Culture* 8, no. 3 (2005): 269–85; Kazys Varnelis, *The Infrastructural City: Networked Ecologies in Los Angeles* (Barcelona: Actar, 2008); Starosielski, "Beaches."

28. Andrew Pickering, *The Mangle of Practice: Time, Agency, and Science* (Chicago: University of Chicago Press, 1995).