EVOLVING CORE CAPABILITIES OF THE INTERNET

J. SCOTT MARCUS*

ABSTRACT

Historically, the Internet has served as an enormous hotbed of innovation. Nonetheless, deployment of a number of potentially beneficial and important Internet capabilities appears to be slowed or stalled for lack of sufficient commercial incentives. The primary concern is with *public goods*[1] where market forces alone might not be sufficient to drive widespread adoption. Timely and relevant examples are drawn primarily from the areas of network security and cybersecurity. How might government identify and prioritize those capabilities where intervention is warranted (if ever)? What actions on the part of industry and government are necessary and appropriate in order to ensure that societally significant problems, including network security and robustness, are addressed in the Internet?

* Author's current address: Federal Communications Commission (FCC), Office of Strategic Planning and Policy Analysis, 445 12th Street SW, Washington, DC 20554 and can be contacted at smarcus@fcc.gov. The author is affiliated with both the FCC and the European Commission, but the opinions expressed are solely those of the author, and do not necessarily reflect the views of either agency. The author is deeply indebted to his colleagues Richard Hovey and Jeffery Goldthorp, of the FCC; to Scott Bradner, of Harvard University; to Dale Hatfield, Gary Chapman and Andrew Johnson, of the University of Colorado; and to Scott Rose of the National Institute of Standards and Technology for a wealth of helpful and insightful comments.

1. The Economist, *Economics A-Z*, ECONOMIST.COM, *available at* http://www.economist.com/research/Economics (last visited May 10, 2004) (adapted from MATTHEW BISHOP, ESSENTIAL ECONOMICS (2004).

TABLE OF CONTENTS

INTRODUCTION

Many have argued that the Internet is far more hospitable to innovation than the traditional public switched telephone network (PSTN).[2] Not so long ago, it seemed that all things were possible in the free-wheeling entrepreneurial and unregulated culture of the Internet. Nonetheless, it now appears that many seemingly promising innovations have languished in recent years. Is it possible that the Internet is hospitable to some innovations, but not to others? Is it possible that pure free market mechanisms will fall short in cases that are of vital importance to society at large? Might there be a role for government to play in promoting societally valuable goals that the market alone would not achieve? If so, what measures are available to government or industry to attempt to promote adoption of important and beneficial innovations?

One federal report, the draft version of *The National Strategy to Secure Cyberspace,* posed the key question succinctly: "How can government, industry, and academia address issues important and beneficial to owners and operators of cyberspace but for which no one group has adequate incentive to act?"[3] The final version of that same report offers an answer: "The government should play a role when private efforts break down due to a need for coordination or a lack of proper incentives."[4]

---

2. *Cf.* David Isenberg, *The Rise of the Stupid Network*, COMPUTER TELEPHONY, Aug. 1997, at 16-26, *available at* http://www.hyperorg.com/misc/stupidnet.html.

3. THE PRESIDENT'S CRITICAL INFRASTRUCTURE PROTECTION BOARD, THE NATIONAL STRATEGY TO SECURE CYBERSPACE, DRAFT FOR COMMENT 47 (2002), *available at http://www.iwar.org.uk/cip/resources/c-strategy-draft* [hereinafter DRAFT NATIONAL STRATEGY TO SECURE CYBERSPACE].

4. THE PRESIDENT'S CRITICAL INFRASTRUCTURE PROTECTION BOARD, THE NATIONAL STRATEGY TO SECURE CYBERSPACE 31 (2003), *available at* http://www.whitehouse.gov/pcipb/cyberspace_strategy.pdf [hereinafter NATIONAL STRATEGY TO SECURE CYBERSPACE].

A particular concern here is with *public goods*. *The Economist* defines public goods as:

> Things that can be consumed by everybody in a society, or nobody at all. They have three characteristics. They are:
>
> - non-rival – one person consuming them does not stop another person consuming them;
> - non-excludable – if one person can consume them, it is impossible to stop another person consuming them;
> - non-rejectable – people cannot choose not to consume them even if they want to.
>
> Examples include clean air, a national defense system and the judiciary. The combination of non-rivalry and non-excludability means that it can be hard to get people to pay to consume them, so they might not be provided at all if left to market forces . . . . [5]

Most of the examples in this paper are drawn from the fields of network security and cybersecurity. In the aftermath of the events of September 11, 2001, there is a widespread recognition of the need to enhance the robustness and security of the Internet. Many security exposures exist. Techniques are available to prevent or at least mitigate the impact of the exploitation of certain of the known exposures; however, in certain instances, it is not clear that the organizations that would need to make investments to deploy the technologies are motivated to do so. This is especially likely where deployment costs would exceed the quantifiable economic benefits to the organizations that would have to bear those costs.

The Internet is unquestionably one of the greatest technological successes of modern times. Among the many factors that contributed to its success is the *end-to-end model*, which enables innovation at the edge of the network without changes to the core; and the absence of central control or regulation, which has enabled the Internet to evolve largely through private initiative, without the restrictions of cumbersome governmental oversight. To a large degree, the Internet represents a triumph of unbridled capitalist initiative.

Today, most networking professionals would agree that the Internet would benefit from a number of evolutionary changes – changes which, however, appear not to be forthcoming. In many cases, the technology

---

5. The Economist, *supra* note 1. They go on to observe that, "public goods are regarded as an example of market failure, and in most countries they are provided at least in part by government and paid for through compulsory taxation." *Id.*

seems to be sufficiently straightforward, but deployment is stymied by a constellation of factors, including:

- the lack of sufficient economic drivers;
- the difficulty of achieving consensus among a plethora of stakeholders with interests that are either imperfectly aligned or else not aligned at all; and;
- the inability of government to foster change in an entity that is global in scope, and largely unregulated in most industrialized nations.

In other words, the very factors that fostered the rapid evolution of the Internet in the past may represent impediments to its further evolution. Historically, those Internet features that could be implemented through private initiative at the edge of the network emerged rapidly; those features that now require coordinated changes, and especially changes to the *core* of the network, are either slow to emerge or are not emerging at all.[6] One might now wonder whether the Internet has reached an evolutionary cul-de-sac.

This paper draws on examples associated with network security and cyber security; however, the issue of promoting public goods where market forces would otherwise be insufficient is a much larger topic. The author humbly asks the reader's indulgence as he frenetically jumps back and forth from the general to the more specific.

Readers who are well versed in the technology of the Internet may have an easier time following the issues, but this paper is not primarily about technology; rather, it focuses on the business, economic and regulatory factors that serve either to facilitate or to impede evolution. In any case, with the possible exception of Section II (which the reader could skip without loss of continuity), no prior knowledge beyond that of an intelligent layman is assumed as regards any of these disciplines.

This introduction provided a cursory overview of the issues. Section I provides background on factors that may militate against the deployment of certain kinds of enhancements to Internet functionality: the end-to-end principle, transaction costs, and the economics of network externalities (following the seminal work of Jeffrey Rohlfs).[7] Section II provides a brief technical overview of two emerging security

6. *Cf.* Christian Sandvig, Communication Infrastructure and Innovation: The Internet as End-to-End Network that Isn't (Nov. 2002) (unpublished manuscript, available at http://www.cspo.org/nextgen/Sandvig.PDF).

7. JEFFREY H. ROHLFS, BANDWAGON EFFECTS IN HIGH-TECHNOLOGY INDUSTRIES 3 (2001).

enhancements to the Domain Name Service (DNS), which collectively serve as an example of seemingly desirable security capabilities and the associated deployment challenges. Section III gingerly explores a topic that many in the Internet community will find uncomfortable: whether it is appropriate for government to play a more active role in fostering the further technical evolution of the Internet. Government intervention could be positive; it could be ineffective; or it could be counterproductive. What role, if any, should the U.S. Government play in the future technical evolution of the Internet? Section IV provides brief concluding observations.

I. BARRIERS TO ADOPTION

As part of the process of preparing the National Strategy to Secure Cyberspace, the President's Critical Infrastructure Protection Board (CIPB) convened a group of Internet experts. At a meeting of this group in May 2002, I commended them for their excellent and thoughtful recommendations.[8] I noted the importance of their work, and encouraged them to let their colleagues in government know if, as their work proceeded, they encountered difficulties in getting their firms to deploy the recommended facilities.

A moment of embarrassed silence followed. One of the attendees then timorously put up his hand and said:

> Scott, you don't have to wait a year or two to find out whether we are having problems getting this stuff deployed. We already know the answer. There is nothing new in these reports. All of this has been known for years. If we were able to craft business cases for our management, all of this would have been done long ago.

No one who has dealt with these issues in industry should be surprised by this answer. Certain Internet innovations have achieved widespread use with no market intervention, perhaps the most noteworthy being the World Wide Web. A great many other Internet innovations have languished, even though the underlying technology appeared to be sound.

8. For a public summary of their major findings, see AVI FREEDMAN, AKAMAI TECHS., ISP WORKING GROUP INTERNET VULNERABILITY SUMMARY & DISCUSSION (2002), *available at* http://www.nanog.org/mtg-0206/avi.html.

In addition to the DNS security facilities described in this report, similar deployment concerns might be raised about:

- Internet Protocol (IP) version 6[9]
- Differentiated services (DiffServ)[10]
- IP multicast
- Operational tools and protocol enhancements to enhance the security of BGP-4 routing protocols

Engineers tend to conceptualize these deployment delays in terms of engineering concerns, such as incomplete protocol specifications, immature protocol software implementations, and insufficient interoperability testing. It may well be that these engineering problems are symptomatic of deeper business and economic impediments that militate against deployment and use of certain *kinds* of innovations in the Internet today.

This section of the paper discusses a constellation of economic factors that impede deployment of certain kinds of Internet facilities. The detailed interplay among these factors, and perhaps among other factors not considered here, may vary from one service to the next, but much of the observed behavior can apparently be explained by a small number of underlying economic factors.

A. Transaction Costs

Transaction costs are the economic costs associated with effecting a transaction.[11] Some transactions involve far higher transaction costs than others. If a customer buys a candy bar in a luncheonette, she typically hands the cashier some money, receives her change, and walks out the door with the desired item. Transaction costs are low. If that customer

---

9. The National Telecommunications and Information Administration (NTIA), which is a part of the U.S. Department of Commerce, is currently conducting a Notice of Inquiry regarding IP version 6. Public comments are available at http://www.ntia.doc.gov/ntiahome/ntiageneral/ipv6/commentsindex.html. The parallels to DNS security are quite striking.

10. Within the network of a single service provider, differentiated services are readily achievable. In the general, multiple-provider case, there is no significant deployment.

11. Various definitions exist in the literature. *See, e.g.,* ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT, TRANSACTION COSTS AND MULTIFUNCTIONALITY, *available at* http://www1.oecd.org/agr/mf/doc/Transactioncosts32.pdf (last visited May 26, 2004) (citations omitted). It defines transaction costs in this way: "Transaction costs are 'the costs of arranging a contract *ex ante* and monitoring and enforcing it *ex post*' . . . 'the costs of running the economic system' . . . and 'the economic equivalent of friction in physical systems . . . .' " *Id.* at 2 (citations omitted).

purchases by credit card, the merchant pays a fee for the use of that credit card – transaction costs are higher. If a person buys or sells a house, transaction costs (broker's fees, loan initiation, and various fees) might consume a hefty 5-10% of the value of the transaction.

Transaction costs thus represent sand in the gears, a form of economic friction. Where a large number of parties must independently come to terms with one another on a single transaction, and particularly where those terms require substantial discussion or negotiation, transaction costs will tend to be very high.

High transaction costs cut into the *surplus* (the degree to which the value to a purchaser exceeds the cost) associated with a transaction. High transaction costs can literally be prohibitive – they can make the transaction as a whole uneconomic. Those who claim that the Internet is a hotbed of innovation are implicitly arguing that transaction costs to deploy new innovations on the Internet are low. In the pages that follow, this paper suggests that this is true only for certain kinds of innovations.

B. Network Externalities

The value of a network is largely a function of who can be reached over that network. Robert Metcalfe, the co-inventor of the Ethernet Local Area Network, attempted to roughly quantify this in *Metcalfe's Law*, which claims that the value of a network is roughly proportionate to the square of the number of users.[12]

Most industries experience economies of scale – bigger is better. Networks, however, are subject to additional effects of scale that go far beyond traditional economies of scale. Every time that someone in North Dakota obtains telephone service for the first time, it enhances the value of *everyone's* telephone service – there is one more person who can be reached by phone. Economists refer to these effects as *network externalities*, or informally as *bandwagon effects*.

For a product or service subject to substantial network externalities, nothing succeeds like success. One of the most common examples of a bandwagon effect is the competitive clash of two videocassette standards, VHS and Betamax. At a technical level, neither had a decisive advantage over the other, and for a time they coexisted in the marketplace. Over time, VHS acquired more customers. As a result, studios developed more programming in the VHS format. Consumers with Betamax

12. *Cf.* Andrew Odlyzko, *Content is Not King*, FIRST MONDAY, Jan 8, 2001, *at* http://www.firstmonday.dk/issues/issue6_2/odlyzko/ (arguing that ". . .Metcalfe's Law does not reflect properly several other important factors that go into determining the value of a network. However, the general thrust of the argument . . . [is] valid.").

equipment found less and less of interest in rental stores, and eventually nothing at all. "Eventually, all consumers – even those who preferred Beta[max]'s picture quality . . . – had no choice but to get on the VHS bandwagon."[13]

In some instances, network externalities manifest themselves by way of direct interactions with other users of the same network. In others, the bandwagon effects relate to complementary upstream or downstream industries, as was the case with VHS and Betamax (the player was valuable only if extensive content was available to play on it). These complementarities often lead to the classic "chicken and egg" problem, where two vertically related industries cannot succeed unless both are launched at once.

In a bandwagon marketplace, multiple stable equilibria are usually possible, and these equilibria can differ greatly. Rohlfs defines the *initial user set* as comprising "all individual entities . . . that can justify purchasing the service, even if no others purchase it."[14] If the demand for the service is enhanced by being taken up by the initial user set, then additional users will acquire the service until a higher equilibrium is reached, the *demand-based equilibrium user set*. The level of usage that is societally optimal, the *maximum equilibrium set*, may be much larger than the demand-based equilibrium user set.[15]

Unfortunately, "ordinary demand adjustments do not provide a path to the optimum."[16] Achieving the maximum equilibrium set often requires "supply-side activities or government intervention."[17]

New technology products and services have to get over an initial "hump" in order to reach critical mass. Different high-technology industries have achieved critical mass in different ways. Large numbers of videocassette recorders (VCRs) were sold to time-shift television programs on a stand-alone basis; subsequently, these VCRs established the necessary preconditions for the videocassette rental business that today represents the primary use of the VCR.[18] For CD players, necessary complementary products became available due to vertical integration – the same firms that were manufacturing CD players (Phillips and Sony) had significant ownership interests in producers of recorded music.[19] For black and white television, industry convergence on the National Television Standards Committee (NTSC) technical

13. ROHLFS, *supra note* 7. (The discussion of network externalities that follows draws heavily on Rohlfs's work.).
14. *Id.* at 23.
15. *Id.* at 24.
16. *Id.*
17. *Id.*
18. *Id.* at Ch. 10.
19. ROHLFS, *supra note* 7, at Ch. 9.

standard, coupled with its rapid adoption by the FCC, played a large role in overcoming the initial start-up problem.[20]

C. Misalignment of Incentives

In a largely unregulated, market-based system, firms make business decisions based on anticipated costs and benefits. Any decision to change a firm's existing operating environment will entail initial costs. If the firm is to incur those costs, it must believe that there will be corresponding benefits that exceed those costs.

A recent report by the Institute for Infrastructure Protection (I3P) describes the dilemma:

> In a market-based economic system, it is not surprising that the market for IT and cyber security products defines the state of cyber security. Two closely related questions appear to drive decisions on how security products and services are acquired and used: (1) what are the cyber security risks to the enterprise and how do they fit into the overall risk equation of a company, and (2) what is the value of cyber security – how much financial benefit it provides. There are no clear answers to these questions.[21]

Features that constitute public goods (such as enhancements to network security) do not in general reduce recurring operating costs, so the benefits must come from somewhere else. Many organizations find it difficult to justify these expenditures for one or more of a number of reasons. Notably, the benefits may be difficult or impossible to quantify,[22] or whatever benefits exist may accrue to a party or parties other than the firm that must make the investments. Collectively, these two factors mean that the organization is unlikely to be motivated to make the investment.

---

20. *Id.* at Ch. 12.

21. INSTITUTE FOR INFORMATION INFRASTRUCTURE PROTECTION, CYBER SECURITY RESEARCH AND DEVELOPMENT AGENDA 40 (2003), *available at* http://www.thei3p.org/documents/2003_Cyber_Security_RD_Agenda.pdf [hereinafter I3P REPORT].

22. *Id.* at 34-45.

> Decision makers lack a foundation of data about the current investment and risk levels: metrics that express the costs, benefits, and impacts of security controls from an economic perspective, technical perspective, and risk perspective; and ways to predict the consequences of risk management choices. . . . Risk assessment and dependency modeling for cyber security remain in an immature state with only little momentum in the marketplace.

Id.

D. The Time Frame of Risks and Rewards

Après moi, le déluge! (After me, the flood!)[23]

Firms fund business cases where the expected return exceeds the expected investment within some defined period of time.

Many cyber vulnerabilities relate to potential exploits that have very high cost, but very low probability of occurrence. These are "thirty year flood" events. Firms may resist funding solutions to thirty year flood problems for some combination of reasons, including:

- The business case takes too many years to prove in;
- The risks are too speculative, and thus too difficult to quantify;
- The risks are born primarily by their insurers, or possibly by the government;
- They may believe, rightly or wrongly, that even if the event takes place, they are unlikely to be viewed as negligent if their competitors were similarly unprepared;
- The current managers may consider it unlikely that the event will happen while they are still with the firm. They bequeath the problem, if indeed it proves to be a problem, to their successors.

E. The TCP/IP Reference Model

The underlying architecture of the Internet has significant implications for the transaction costs associated with the deployment of new capabilities. This part of the paper describes the architecture of the Internet in order to motivate the discussion of the economics associated with the *end-to-end principle* that appears in the subsequent section.

Perhaps the most significant advance of the past thirty years or so in data networking is the advent of *layered* network architectures. A layered network architecture breaks the functions of a data network up into functional layers, each of which communicates with its peer layers in other communicating systems, while deriving services from the layer

23. Attributed to Louis XV, king of France from 1715-1774. Some sources instead attribute this quotation to his mistress, Madame de Pompadour.

beneath. This layering helps insulate one layer from another, providing many benefits – a topic we return to later in this section of the paper.

The TCP/IP protocol family, or *protocol suite*, is the preeminent example today of such a layered network architecture.[24] The TCP/IP protocol suite is based on a conceptual model that characterizes the communications hardware and software implemented within a single communicating system – for instance, the personal computer (PC) on your desk – as being comprised of a protocol stack containing multiple layers (see Figure 1).[25]

Levels 1 and 2, the *Physical* and *Data Link Layers* respectively, represent the realization of the "wire" over which communication takes place and the management of that wire. For instance, the Data Link Layer might determine which of several computers is authorized to transmit data over a particular local area network (LAN) at a particular instant in time.

Level 3, the *Network Layer*, forwards data from one interconnected network to the next. For the Internet, the Network Layer is the *Internet Protocol* (IP), which independently routes and forwards small units of data (datagrams).

Level 4, the *Transport Layer*, processes those datagrams and provides them to whichever application needs them, in the form that the application requires. For the Internet, the *Transmission Control Protocol* (TCP) supports applications that need a clean and reliable stream of data with no omissions or duplicates. The *User Datagram Protocol* (UDP) represents an alternative Transport Layer protocol that supports applications that do not require the tidy delivery that TCP provides. E-mail uses TCP, while Voice over IP (VoIP) uses UDP.

---

24. The evolution of the TCP/IP protocol suite was influenced by earlier layered network architectures, and influenced in turn the subsequent evolution of a number of those network architectures. Among the layered network protocol families that emerged during the Seventies and Eighties were CCITT's X.25, IBM's System Network Architecture (SNA), Digital Equipment Corporation's DECnet, and Xerox Network Systems (XNS). Perhaps the most influential layered network architecture was the Reference Model for Open Systems Interconnection, usually referred to as the *OSI Reference Model*. The OSI Reference Model was developed jointly by the International Organization for Standardization (ISO) and the ITU/CCITT. The most readable descriptions of the OSI Reference Model appear in Hubert Zimmerman, *OSI Reference Model – The ISO Model of Architecture for Open Systems Interconnection*, 4 IEEE TRANSACTIONS ON COMM. 425 (1980), and in ANDREW TANENBAUM, COMPUTER NETWORKS (Prentice Hall 3d ed. 1996).

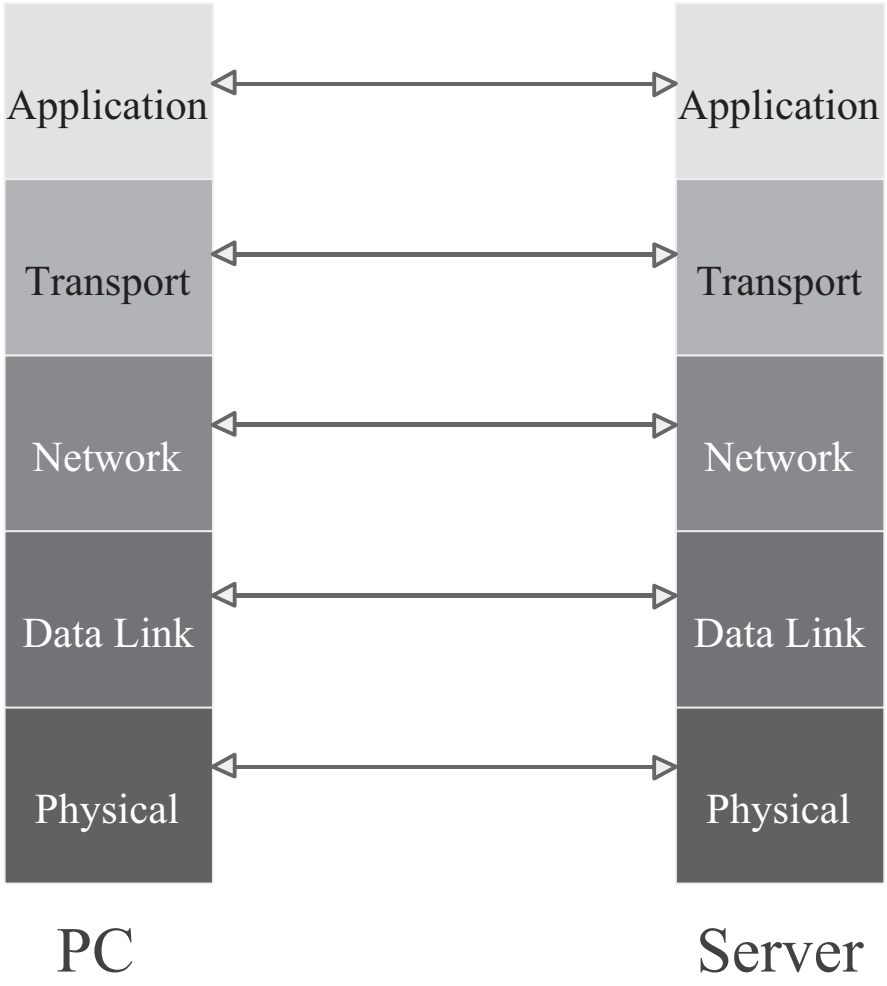
25. Rigid adherence to protocol layering tends to impose a high overhead on protocol software. In reality, TCP/IP implementations often combine layers or take short-cuts as a means of reducing this overhead. *See* DAVID D. CLARK, RFC 0817: MODULARITY AND EFFICIENCY IN PROTOCOL IMPLEMENTATION (Internet Engineering Task Force, July 1982), *at* http://www.ietf.org/rfc.html.

FIGURE 1
PROTOCOL LAYERS IN THE OSI / INTERNET REFERENCE MODEL

Level 5, the *Application Layer*, performs useful work visible to the end user, such as the browser or e-mail client (SMTP, HTTP) on your PC.

In this reference model, a layer logically interacts with its peer in a communicating system (see Figure 2). Thus, an Application Layer, such as the web browser in your PC, communicates with its peer process, a web server in a distant computer.

FIGURE 2
PEER LAYERS LOGICALLY INTERACT WITH ONE ANOTHER



Each layer within a communicating system implements this logical interaction by requesting services from the next lower layer. Thus, the Application Layer requests data from the Transport Layer. In doing so, it uses an interface that intentionally hides the details of how the lower layer implements its service. This information hiding is a key beneficial property of a layered network architecture – it enables the implementation of a layer to change without impacting the layers above or below.

FIGURE 3
LOGICAL AND PHYSICAL INTERACTIONS BETWEEN NETWORK PROTOCOL LAYERS

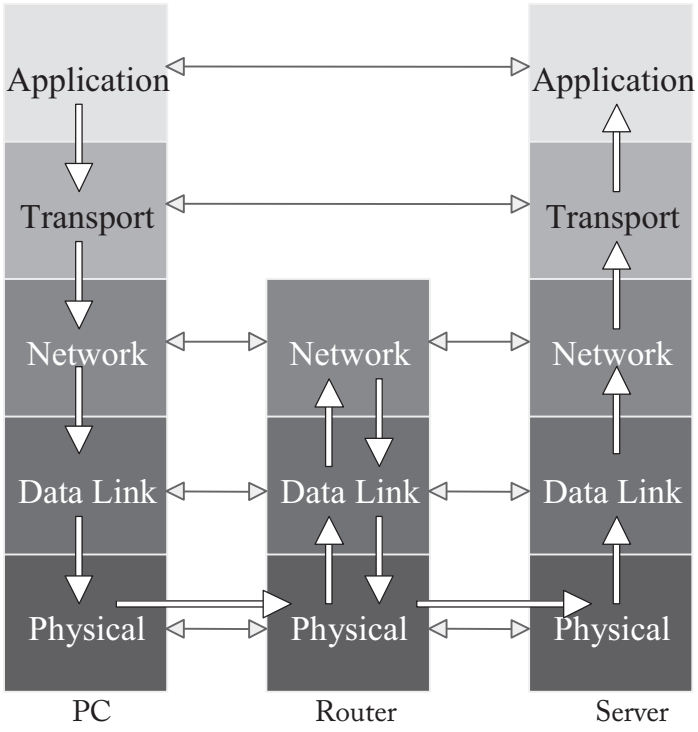


Figure 3 shows the relationship between logical and physical interactions in the Internet layered network architecture. It also adds another element to our understanding – a *router*, which is a device that exists solely to forward traffic in the Internet.

The information hiding property of a layered network architecture facilitates technical innovation over time. It also enables network applications to be written once to operate over any underlying transmission technology, or combination of technologies, thus simplifying the application creator's job. Conversely, the creator of a new transmission technology need only ensure that adequate interfaces exist to enable upper layers of the network to *use* the new communications layer – there is no need to make network applications specifically *aware* of a new underlying transmission technology. Phrased differently, a new network application will work with existing networks, and no changes are

needed to underlying network transmission technologies. A new network transmission technology will work with existing networks, and no changes will be needed to existing applications. These properties greatly simplify the evolution of the network over time, and thereby reduce the transaction costs associated with network evolution.

F. The End-to-End Principle

In the early Eighties, a number of distinguished computer scientists at MIT propounded the *end-to-end principle*.[26] They noted that certain communications capabilities were most appropriately associated, not with the underlying network, but rather with the application that used the network. End-to-end reliability of transmission, for instance, could truly be assured only at the end points themselves. They further argued that, if the function could only be correctly implemented in the end points of the network, that it was a bad idea to also implement these functions in intermediate systems—doing so introduced not only inefficiencies, but also an increased possibility of error. Internet engineers have generally accepted the end-to-end principle as a basic tenet of network design. Moreover, they have sometimes advanced the further argument that the end-to-end principle fosters the evolution of the Internet, in that it enables new applications to be developed at the edges of the network, without disrupting the underlying core.[27]

There is much to be said for this view. For example, the creation of the World Wide Web initially depended primarily on the creation of a browser that could read and interpret existing file formats, and secondarily on servers for HTTP. No prerequisite changes were needed to the underlying TCP/IP protocols, the IP addressing system, or the DNS—these already provided the necessary support. This absence of prerequisite changes in turn reduced the number of parties that had to change their infrastructure – no action was required, for instance, on the part of Internet Service Providers (ISPs). By reducing the number of parties who must act in order to implement a particular change to the Internet, the end-to-end principle reduces the transaction costs associated with the development of new applications, thus fostering the continuing evolution of the Internet.[28]

---

26. J.H. Saltzer et al., *End-to-End Arguments in System Design*, *in* ACM TRANSACTIONS ON COMPUTER SYSTEMS 2, 277 (1984), *available at* http://web.mit.edu/Saltzer/www/publications/endtoend/endtoend.pdf.

27. Isenberg, *supra* note 2.

28. For an interesting economic interpretation of the costs and benefits of this flexibility, see Mark Gaynor et al., *The Real Options Approach to Standards for Building Network-based Services* (2nd IEEE Conference on Standardization and Innovation in Information

More recently, a revisionist scholar, Christian Sandvig, has called this view into question.[29] He notes that this interpretation of the end-to-end principle presupposes that the underlying network already provides all of the functionality that will ever be necessary or desirable. In fact, it is difficult to know the impact of "missing" functionality – people develop applications to fit the functionality that is already available. Nobody takes the time to develop the applications that would have failed due to insufficient support in the underlying network; consequently, there is no obvious "graveyard" of failed applications.

Thus, while the end-to-end principle may tend to facilitate the development of new data networking *applications* (based in the Transport thru Application Layers of the familiar OSI Reference Model,[30] as described earlier in this paper),[31] it does nothing to foster the evolution of the underlying functionality associated with the Network Layer and below.

As it happens, this same OSI Reference Model has largely succeeded in decoupling and simplifying the evolution of its lowest layers. Below the Network Layer – which for TCP/IP is the Internet Protocol – datagrams can be transmitted over any Data Link Layer that is known to two systems that are topologically[32] adjacent. This is so because the lowest layers, the Physical and Data Link Layers, operate on a *point-to-point* basis.

Some years ago, the Dutch logician Edsgar Dijkstra conceived the notion of *structured programming*.[33] By a clean nesting of logical functionality, it was possible to contain the impact of changes to a program to a defined scope of statements within the program. This greatly enhanced the reliability of programs, and made it much easier to evolve programs (because a change in one part of the program was unlikely to cause unexpected and unpredictable adverse impact somewhere else).

A similar evolution took place for database management systems – by segregating functionality into various *schemas*, and hiding unnecessary details about how those schemas implemented their

---

Technology, Oct. 2001), *available at* http://people.bu.edu/mgaynor/papers/IEEE-standard-camera.pdf.

29. Sandvig, *supra* note 6.

30. Zimmerman, *supra* note 24 (the TCP/IP protocol suite that forms the foundation of the Internet broadly follows the OSI Reference Model, but with simplification in the upper layers).

31. *See supra* Section I.E.

32. Topology is the branch of mathematics that deals with the interconnectivity of the vertices and edges that comprise geometric figures, without considering their dimensions. It provides a useful way to visualize communications networks and to express their formal properties.

33. O.J. DAHL ET AL., STRUCTURED PROGRAMMING (1972).

respective functions, the database systems fostered greater reliability and ongoing functional evolution.

The OSI Reference Model attempted to apply similar principles to data networks. The functionality of the network was broken down into seven functional layers (five for the TCP/IP world). The upper layers were associated with the application, the lower layers with the transmission mechanism. Each layer communicated with its peer layer in another communicating system; however, each effectuated this communication by requesting services from the layer beneath it. A layer never needed to know *how* the underlying layer provided the functionality.

There is no need for the entire Internet to understand any particular Data Link protocol mechanism. A given system that participates in the Internet need only understand those Data Link protocols whereby it communicates with the systems with which it maintains direct point-to-point communications. These systems could be said to be *topologically adjacent*.

These properties provide a decoupling for the lower layers of the OSI Reference Model that is very similar in effect to that which the end-to-end principle provides for the upper layers. New applications can be implemented as communicating processes in any two cooperating systems. Likewise, new transmission facilities at the Data Link Layer and below can be implemented in any two adjacent cooperating systems. In both cases, the transaction costs associated with deployment are bounded.

All of this breaks down for the Network Layer, IP. IP provides global connectivity and interoperability for the Internet. There are, of course, ways to evolve the IP functionality of the Internet, but these tend to be complex. There is no assurance that a change made between a pair of systems will have no impact on other systems. There is no inherent mechanism for information hiding within the IP Layer. Any functional evolution must be orchestrated with exquisite caution, because there is no guarantee that the unintended consequences of a given change will be limited.

In sum, technology evolution tends to be complex and expensive for the IP Layer, and also for certain other elements of the Internet that are global in scope. Since the transaction costs associated with evolutionary change of these elements are high, the benefits of any proposed evolutionary change would have to be correspondingly high – otherwise, the deployment of the proposed change is likely to stall for lack of a sufficiently compelling business case.

II. THE TECHNOLOGY OF DNS SECURITY

There are a wide variety of Internet facilities that might logically fall within the scope of this discussion. In order to motivate the discussion, we focus on a specific constellation of potential Internet security features associated with the DNS.

This paper does not attempt to argue whether any particular Internet security service is in some sense essential. Rather, the intent is to provide background on the rationale of a particular Internet service whose relatively slow deployment might in some sense be emblematic of a broader issue, to assume *arguendo* that there were some pressing requirement for deployment of that service, and then to pose the question: What impediments to deployment are visible today, and what further impediments might we anticipate in the future? By conducting this thought exercise, we come to a better understanding of the challenges that any feature of this type is likely to encounter.

In this sense, DNS security serves merely as a plausible proxy for any of the Internet-based services that we might have considered.

A. The Domain Name System

The DNS is the primary mechanism whereby names, such as www.fcc.gov, are mapped to Internet addresses, such as 192.104.54.3. The DNS has other mapping or directory functions as well.[34]

A DNS *client,* which might reside in your PC, initiates a DNS request to determine the IP address of www.fcc.gov. The request might be sent to a DNS server maintained by a company or by an ISP, the firm that provides access to the Internet.

The DNS is usually thought of as representing a logical tree structure. The root of that tree is comprised of thirteen groups of DNS servers in the United States, Europe and Asia.[35] Below the root are other groups of servers associated with Top Level Domains (TLDs), which are

34. The DNS is documented in a series of Requests for Comments (RFC) that were developed by the Internet Engineering Task Force (IETF). The primary references are P.V. MOCKAPETRIS, RFC 1034: DOMAIN NAMES - CONCEPTS AND FACILITIES (Internet Engineering Task Force, Nov. 1, 1987), *at* http://www.ietf.org/rfc.html [hereinafter *RFC 1034*] (updated by *RFC 1101*, *RFC 1183*, *RFC 1348*, *RFC 1876*, *RFC 1982*, *RFC 2065*, *RFC 2181*, *RFC 2308*, *RFC 2535*); and P.V. MOCKAPETRIS, RFC 1035: DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION (Internet Engineering Task Force, Nov. 1, 1987), *at* http://www.ietf.org/rfc.html [hereinafter *RFC 1035*] (updated by *RFC 1101*, *RFC 1183*, *RFC 1348*, *RFC 1876*, *RFC 1982*, *RFC 1995*, *RFC 1996*, *RFC 2065*, *RFC 2136*, *RFC 2181*, *RFC 2137*, *RFC 2308*, *RFC 2535*, *RFC 2845*, *RFC 3425*, *RFC 3658*). All RFCs are available at http://www.ietf.org/rfc.html.

35. Some of these root servers are now mirrored in multiple locations.

associated with the rightmost portion of a domain name[36] – for example, .com, .org, or .gov. The servers responsible for .gov provide in turn pointers to the next level down, including servers responsible for .fcc.gov.

This tree structure facilitates delegation of authority.

B. Security Exposures in the DNS

> The opening word was inscribed on the archway all the time! The translation should have been: *Say 'Friend' and enter*. I had only to speak the Elvish word for *friend* and the doors opened. Quite simple. Too simple for a learned loremaster in these suspicious days. Those were happier times.[37]

The DNS was designed in happier times, with little or no regard for security concerns.[38] When a DNS request is transmitted, there is no assurance that the response came from the desired DNS server, nor that the information provided was valid.

If a malefactor (who somehow had the ability to eavesdrop on DNS requests for the address of www.fcc.gov) wished to subvert the FCC's web site, they would not need to hack www.fcc.gov; they could instead create their own bogus site, and respond to DNS requests with the IP address of the bogus site. They might not even have to block legitimate DNS responses; it would be sufficient to respond faster than the legitimate DNS servers. Users accessing the bogus site would presume it to be the real one. There are countless variants on this scenario. Most of them depend on one of several underlying exposures:[39]

---

36. Strictly speaking, we should say the rightmost *customarily visible* portion of the domain name. The rightmost portion is a period denoting the root itself, which is unnamed; however, this is often omitted by convention.

37. J.R.R. TOLKIEN, THE FELLOWSHIP OF THE RING 402 (Ballantine Books 1965).

38. *Cf.* I3P REPORT, *supra* note 21, at iii ("The information infrastructure, taken as a whole, is not an engineered system. . . . Security was not a significant consideration at its inception, and security concerns today do not override market pressures for new uses of technology or innovation, in spite of frequent stories of hackers, criminals, and, increasingly, terrorists and nations using or planning to use the information infrastructure as a weapon to harm the United States.").

39. *Cf.* D. ATKINS & R. AUSTEIN, RFC __: THREAT ANALYSIS OF THE DOMAIN NAME SYSTEM (Internet Engineering Task Force, Feb. 2004), *at* http://www.ietf.org/internet-drafts/draft-ietf-dnsext-dns-threats-07.txt (work in progress: RFC is in preparation). Atkins and Austein primarily characterize threats as (1) packet interception, (2) ID guessing and query prediction, (3) name games, (4) betrayal by trusted server, and (5) denial of service. *Id.* Much work has been done over the years to characterize threats to the DNS, notably including Steven Bellovin, *Using the Domain Name System for System Break-Ins*, USENIX, (Jun. 1995), *at* http://www.usenix.org/publications/library/proceedings/security95/bellovin.html.

- There is no *authentication* of the DNS server, i.e. no assurance that the server is who it purports to be;
- There is no assured *integrity* of the DNS response, i.e. no assurance that the message received is the same as that which was sent;
- There is no assurance that the data maintained by the DNS server was not somehow maliciously modified on the server before being sent. There is in any event no assurance that the data is correct;
- Because the DNS is a logical tree, any compromise potentially impacts everything below that point in the DNS tree.

There is also concern that malefactors might attempt to cripple large portions of the Internet by launching *Distributed Denial of Service* (DDoS) attacks against key DNS servers, preventing users from reaching DNS servers. If users cannot resolve certain domain names, then to all intents and purposes they are unable to use the Internet to access those computers. An attack that was launched on October 21, 2002 received considerable media attention. All indications are that the October 21 attacks had minimal impact; nonetheless, the attacks demonstrated that denial of service is a real threat whose impact should not be underestimated.

C. DNS Security Mechanisms

The Internet community has been aware of these security exposures for many years. A number of responses have been developed within the *Internet Engineering Task Force* (IETF), the relevant standards body. Some of these are potentially more effective than others.

An exhaustive description of these systems is beyond the scope of this paper. The reader who desires more detail should consult the relevant Internet Request for Comments (RFC) documents. I provide a very brief summary here.

1. Domain Name System Security Extensions

The primary response to these security exposures has been the development of a series of specifications for Domain Name Security Extensions,[40] notably *RFC 2535*, that are sometimes termed *DNS Security Extensions* (DNSSEC).[41]

40. DONALD EASTLAKE III, RFC 2535: DOMAIN NAME SYSTEM SECURITY

RFC 2535 provides for the storage of public cryptographic keys as a new DNS resource record. Keys are used both to authenticate the data's origin, and to assure the integrity of an RRset (a set of DNS resource records).

The authentication mechanism depends on the establishment of a *chain of trust*. The chain flows from the root of the DNS system (or from some other point in the DNS tree that is by convention assumed to be trustworthy) down to individual DNS leaf entries. The intent is that DNS servers would intrinsically and reliably be aware of the key for the root zone, and would follow trusted and authenticated entries through each level of the DNS tree in order to reach the correct leaf.[42]

The creators of *RFC 2535* were also concerned about the possible exploitation of negative information in the DNS – responses erroneously claiming that a domain name does *not* exist. Given that the domain name space is sparse, merely signing the entries that are present would not necessarily prove that a domain name did not exist. *RFC 2535* as amended addresses this by providing for an NSEC resource record[43] which points to the next valid domain name in what we can loosely term alphabetical order.

RFC 2535 is currently an IETF Proposed Standard. This means that it "is generally stable, has resolved known design choices, is believed to be well-understood, has received significant community review, and appears to enjoy enough community interest to be considered valuable."[44]

EXTENSIONS (Internet Engineering Task Force, Mar. 1999), *at* http://www.ietf.org/rfc.html (updated by *RFC 2931*, *RFC 3007*, *RFC 3008*, *RFC 3090*, *RFC 3226*, *RFC 3445*, *RFC 3597*, *RFC 3655*, *RFC 3658*) [hereinafter *RFC 2535*]; DONALD EASTLAKE III, RFC 2541: DNS SECURITY OPERATIONAL CONSIDERATIONS (Internet Engineering Task Force, Mar. 1999), *at* http://www.ietf.org/rfc.html [hereinafter *RFC 2541*].

41. To avoid confusion, we use the term "*RFC 2535* DNSSEC" to refer specifically to *RFC 2535* capabilities. Some sources use DNSSEC to refer only to *RFC 2535*, while others use it to encompass additional capabilities, including TSIG, secure dynamic updates (per *RFC 3007*), and the CERT resource record (*RFC 2538*).

42. This seemingly simple assumption masks a world of complexity. For example, the root signature, like all signatures, should be periodically changed in case it has been somehow compromised, and also to minimize the risk of cryptanalysis. If the key is statically configured in every client, how can it reliably be updated? *See RFC 2541, supra* note 40. *See also RFC 2535*, *supra* note 40, at § 6.2.

43. In the original *RFC 2535*, the corresponding RR was referred to an NXT resource record. Based on operational experience, a number of non-backward-compatible changes were made to the DNSSEC protocols, culminating in a renaming of several RRs and renumbering of their code points. See S. WEILER, RFC 3755: LEGACY RESOLVER COMPATIBILITY FOR DELEGATION SIGNER (DS) (Internet Engineering Task Force, May 2004), *at* http://www.ietf.org/rfc.html [hereinafter *RFC 3755*].

44. SCOTT BRADNER, RFC 2026: THE INTERNET STANDARDS PROCESS –REVISION 3, § 4.1.1 (Internet Engineering Task Force, Oct. 1996), *at* http://www.ietf.org/rfc.html [hereinafter *RFC 2026*].

At the same time, early operational tests have raised questions about a number of important protocol details.[45]

RFC 2535 provides for a very comprehensive any-to-any security mechanism, but it is operationally and computationally relatively expensive. There is a natural tendency to focus solely on the incremental cost of hardware and software, but the relevant deployment costs also include training; deployment planning, testing and staging; and ongoing operational complexity and associated incremental expense. Initial generation of public/private key pairs is computationally intensive, as is periodic or episodic re-signing of a DNS zone. Validation of signatures by means of public key cryptography is also computationally intensive – far more so than private key cryptography. The use of *RFC 2535* increases the length of DNS responses, and greatly increases the size of the DNS database.[46] Ultimately, the cost of increased computational power and server storage may be less important than the incremental expense associated with a substantial increase in operational complexity – ensuring the secrecy of the private keys, and effecting re-signing without breaking the chain of trust are just a few examples.[47]

2. Secret Key Transaction Authentication for DNS (TSIG)

A second response has been the use of TSIG to validate, for example, zone transfers[48] (the transfer *en masse* of a possibly large

---

45. For more information on this topic, visit RIPE NCC, DEPLOYMENT OF INTERNET SECURITY INFRASTRUCTURES, *at* http://www.ripe.net/disi/ (last visited May 26, 2004).

46. One source claims that it increases the size of the DNS database by a factor of seven. *See* PAUL ALBITZ & CRICKET LIU, DNS AND BIND 308-74 (4th ed. 2001), *available at* http://www.oreilly.com/catalog/dns4/chapter/ch11.html.

47. *Id.* at 374 ("We realize that DNSSEC is a bit, er, daunting. (We nearly fainted the first time we saw it.)").

48. P. MOCKAPETRIS, RFC 1034: DOMAIN NAMES – CONCEPTS AND FACILITIES § 4.3.5 (Internet Engineering Task Force, Nov. 1987), *at* http://www.ietf.org/rfc.html [hereinafter *RFC 1034*]. *RFC 1034*, describes DNS zone transfers in this way:

> "Part of the job of a zone administrator is to maintain the zones at all of the name servers which are authoritative for the zone. When the inevitable changes are made, they must be distributed to all of the name servers. While this distribution can be accomplished using FTP or some other ad hoc procedure, the preferred method is the zone transfer part of the DNS protocol. The general model of automatic zone transfer or refreshing is that one of the name servers is the master or primary for the zone. Changes are coordinated at the primary, typically by editing a master file for the zone. After editing, the administrator signals the master server to load the new zone. The other non-master or secondary servers for the zone periodically check for changes (at a selectable interval) and obtain new zone copies when changes have been made."

Id.

volume DNS data).[49] TSIG serves to verify the origin and authenticity of the DNS data.

TSIG dynamically computes a cryptographic hash in response to a specific DNS request, using the well-known HMAC-MD5 algorithm.

TSIG is felt to be a reasonably mature technology. TSIG depends on a cryptographic signature based on *secret keys*, and thus depends on the sender and the receiver possessing a shared secret. As TSIG does not provide a key distribution mechanism, it would become unwieldy[50] if used to mutually authenticate a large number of systems; however, only a small number of systems typically need to perform (for instance) DNS zone transfers to one another for any particular zone, so TSIG works well enough for its intended purpose.

In comparison with *RFC 2535* DNSSEC, TSIG entails far less computational overhead, and does not increase the size of the DNS database. Lewis describes TSIG as less scalable but more efficient than *RFC 2535* DNSSEC.[51] TSIG provides for authentication and integrity of the data transmitted from the point where it leaves the transmitting server, but it does not authenticate the source data (which may have been compromised in the sending server prior to being transmitted) – in other words, TSIG does not provide full *object security*.[52]

D. Deployment of DNS Security Mechanisms

A number of trial deployments of *RFC 2535* DNSSEC have taken place[53], but on the whole the system is not in production deployment.

In a review undertaken by the IETF in December, 2000, Edward Lewis notes that "[i]n 1999 and 2000, more than a half dozen workshops have been held to test the concepts and the earliest versions of implementations. But to date, DNSSEC is not in common use. The current collective wisdom is that DNSSEC is 1) important, 2) a

---

49. PAUL VIXIE ET AL., RFC 2845: SECRET KEY TRANSACTION AUTHENTICATION FOR DNS (TSIG) (Internet Engineering Task Force, May 2000), *at* http://www.ietf.org/rfc.html (updated by *RFC 3645*).

50. In other words, the two systems participating in a TSIG exchange would have to both know the shared secret through some means other than TSIG itself, since TSIG contains no mechanism for distributing the keys. If the keys are to be transmitted through the Internet, by e-mail for example, they must be protected from disclosure to third parties. All of this adds complexity. Since TSIG is normally used for a bounded set of problems where a trust relationship already exists between two systems, the protocol designers have not felt that this extra complexity was warranted.

51. *See generally* EDWARD LEWIS, RFC 3130: NOTES FROM THE STATE-OF-THE-TECHNOLOGY: DNSSEC (Internet Engineering Task Force June 2001), *at* http://www.ietf.org/rfc.html.

52. *See* PAUL VIXIE ET AL., *supra* note 49, at § 6.3; *see also* ATKINS & AUSTEIN, *supra* note 39.

53. *See* LEWIS, *supra* note 51; *see also* RIPE NCC, *supra* note 45.

buzzword, 3) hard, 4) immature." [54] For *RFC 2535* DNSSEC, this is hardly surprising. As previously noted, the true costs of deployment are high.[55]

In addition, *RFC 2535* DNSSEC appears to suffer from many of the characteristics that, as noted in Section I of this paper, potentially complicate deployment. It is not clear that consumers are willing to pay any premium for DNS security;[56] given that implementation costs (largely in the form of operational complexity) are significant, those who must invest to deploy the technology will find it difficult or impossible to craft a clear business case. *RFC 2535* DNSSEC is strongly influenced by network externality effects – *RFC 2535* DNSSEC would be far more valuable to consumers when it is widely deployed than it is today, or even than it would be if it were in modest production deployment. Moreover, because the system depends on a chain of trust, *RFC 2535* DNSSEC is of limited value until those chains are established all the way from the DNS root to the PC on the consumer's desk without breaks.[57] As all of this implicitly requires the cooperation of many independent parties, the economic transaction costs of a comprehensive deployment would tend to be high.[58]

By contrast, indications are that TSIG is deployable today for zone transfers. Per *RFC 3130*, ". . . one component of DNSSEC, TSIG, is more advanced that the others. Use of TSIG to protect zone transfers is already matured to the 'really good idea to do stage' even if other elements of DNSSEC are not."[59]

Based on the discussion of transaction costs earlier in this paper, this is not surprising. The decision to deploy TSIG concerns only a pair (or a small number) of communicating systems, and in most cases a business relationship already exists between the operators of these systems. Thus, transaction costs to deploy are low, and, as we have seen, ongoing costs for computation and storage are also modest.[60]

---

54. LEWIS, *supra* note 51, at § 1.0.

55. *See supra* Section II.C.1.

56. There are also open questions regarding the willingness and ability of consumers to cope with the complexity that DNSSEC implies. Suppose the DNSSEC client software were to notify the consumer that the DNS pointer to a commercial web site such as www.amazon.com had been corrupted. It is not clear what action the consumer should then take, since recovery will generally be beyond the consumer's capabilities. In light of this ambiguity, can the DNSSEC client software provide meaningful and sufficient guidance to the consumer?

57. DNSSEC will be of no use to the average consumer until and unless it is available in the operating system for the consumer's PC – typically Microsoft Windows™.

58. Some have argued for a more piecemeal, selective approach to deployment, but the DNSSEC standards do not currently embrace this approach.

59. LEWIS, *supra* note 51.

60. Unfortunately, the benefits are also modest for the reasons previously noted. The

III. PUBLIC POLICY ALTERNATIVES

To the extent that necessary infrastructure enhancements may not be deployed in the absence of intervention, what is the appropriate role for government?

As we have seen, there is no assurance that industry would deploy a service such as secure DNS based solely on commercial incentives, even assuming the best of intentions on the part of all participants. To the extent that services of this type might be important to the security and robustness of the Internet in the United States, this should be cause for concern.

What role should government play in fostering deployment of Internet capabilities where market forces alone might not suffice? How might government identify and prioritize those capabilities where intervention is warranted (if ever)? For such Internet capabilities as we might deem to be vital, what steps are available to private parties and to the U.S. Government to encourage deployment? Which are likely to be most effective? Which are likely to be least intrusive, and least likely to introduce market distortions?

Most of what we have to say in this section of the paper is not limited to DNS security, and for that matter is not limited solely to cyber security issues. The challenge of promoting the deployment of public goods that provide benefits to the public, but where deployment may not be warranted based solely by the workings of the marketplace, comes up in a great many contexts.

Among the options worth considering by government as a means of fostering deployment of societally valuable services where market incentives might not otherwise suffice are:

1. Provide leadership.
2. Help industry to forge a consensus.
3. Stimulate standards bodies to focus on relevant problems.
4. Collect relevant statistics.
5. Provide "seed money" for research and for interoperability testing.
6. Support desired functionality in products and services through government's own purchasing preferences.
7. Fund the deployment of desired capabilities.
8. Mandate use of desired services.

---

threats that TSIG guards against are generally irrelevant to the consumer mass market.

An important and overarching consideration is that market intervention should be avoided wherever possible, and kept to a minimum where absolutely necessary. The Communications Act states unambiguously that "[i]t is the policy of the United States . . . to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation."[61] Henry David Thoreau stated it more tersely: "That government is best which governs least."[62]

For a somewhat more expansive comment, we turn to a recent study from the Computer Science and Technology Board ("CSTB") of the National Research Council of the National Academies:

> [A]ppropriate market mechanisms could be more successful than direct regulation in improving the security of the nation's IT infrastructure, even though the market has largely failed to provide sufficient incentives for the private sector to take adequate action with respect to information and network security. The challenge for public policy is to ensure that those appropriate market mechanisms develop. How to deal constructively with prevailing market dynamics has been an enduring challenge for government, which has attempted a variety of programs aimed at stimulating supply and demand but which has yet to arrive at an approach with significant impact. Nevertheless, the committee believes that public policy can have an important influence on the environment in which nongovernment organizations live up to their responsibilities for security.[63]

We now discuss the alternative government options in turn, starting with those that are least intrusive.

A. Provide Leadership

There may be a tendency to overlook the simplest and least intrusive form by which government can seek to foster change: Simply articulating that change is necessary.

It is perhaps counterintuitive that exercise of "the bully pulpit" alone should be sufficient to influence the behavior of industry participants and

---

61. 47 U.S.C. § 230(b)(2) (2000).

62. HENRY DAVID THOREAU, CIVIL DISOBEDIENCE (1849), *available at* http://www.cs.indiana.edu/statecraft/civ.dis.html (quotation is sometimes attributed to Thomas Jefferson).

63. INFORMATION TECHNOLOGY FOR COUNTERTERRORISM: IMMEDIATE ACTIONS AND FUTURE POSSIBILITIES 104 (John L. Hennesy et al. eds., 2003) [hereinafter HENNESY ET AL.].

other private citizens,[64] but there is no question that the simple exercise of government leadership has sometimes driven important change.

Leadership in this sense – sometimes referred to as "jawboning" – is more likely to be most effective where some of the following factors hold:

- Government has succeeded in articulating a clear goal that has broad public support.
- The costs associated with doing as the government requests are small (e.g., within the range of discretionary spending of a senior or chief executive).
- The organization that must act needs to curry the favor of the relevant government agency.

B. Help Industry to Forge a Consensus

The U.S. Government frequently provides fora for discussion in order to help industry to reach consensus. The President's Critical Infrastructure Protection Board (CIPB) did so in meeting with the Internet community in the course of preparing the *National Strategy to Secure Cyberspace*.[65]

Analogously, the FCC encourages the communications industry to work together to enhance overall network robustness through the Network Reliability and Interoperability Council (NRIC). NRIC operates under the Federal Advisory Council Act (FACA). As a FACA, the NRIC provides advice to the FCC; further, NRIC often provides guidance regarding best practices to U.S. industry.

In some instances, this consensus could be expressed as a document or guideline prepared by the participants and embodying industry best practices. FACAs often take this approach.

Adhering to industry best practices, as defined by a body such as the NRIC, may also serve to reduce a firm's legal liability to possible allegations of negligence.[66] This form of government participation is

64. *Cf.* I3P REPORT, *supra* note 21, at 40 ("Currently, the federal government's approach relies on public-private partnerships *and the influence of persuasion*; more rigorous analysis needs to be done on the prospects for success of this approach.") (emphasis added).

65. DRAFT NATIONAL STRATEGY TO SECURE CYBERSPACE, *supra* note 3.

66. Potential tort liability, where a firm might be alleged to have taken less than reasonable care to secure its infrastructure against cyberattacks is an emerging, but still largely undeveloped area of the law. *See* CRITICAL INFORMATION INFRASTRUCTURE PROTECTION AND THE LAW: AN OVERVIEW OF KEY ISSUES (Cynthia A. Patterson & Stewart D. Personick eds., 2003), *available at* http://www7.nationalacademies.org/cstb/pub_ciip.html [hereinafter CRITICAL INFORMATION INFRASTRUCTURE PROTECTION AND THE LAW].

generally viewed as positive by industry and by the broader community. It provides government with the opportunity to offer leadership in a minimally intrusive way.

This form of government participation provides industry with an additional benefit. Companies that routinely compete in the marketplace are understandably uncomfortable meeting to discuss joint action, for fear that their discussions could be misconstrued as being anticompetitive. To the extent that the U.S. Government calls firms together to discuss specific issues in the public interest, antitrust concerns tend to be mitigated.[67]

C. Stimulate Standards Bodies to Focus on Relevant Problems

One form of industry consensus is embodied in the standards process. As described above, government could play a role in helping industry to agree on a standard. If appropriate, government could perhaps reinforce this result by encouraging the relevant standards body or bodies to officially adopt a standard reflecting that consensus.

In general, government would look to industry to develop solutions for the standards process. Government is not well equipped to pick winners and losers.

For some standards bodies, notably including the International Telecommunications Union (ITU), formal U.S. Government advocacy can play a crucial role in achieving adoption of a standard.

The Internet Engineering Task Force (IETF) is the primary standards body for the Internet. By long-standing tradition, the IETF expects standards participants to present their views as an individual expert, rather than those of the organizations that they represent. The U.S. Government thus plays no formal role in the IETF. Even in this case, however, government can when appropriate facilitate the standards process by supporting research and interoperability testing and by identifying problem areas where it appears that the public interest would be well served by a standards-based solution.

---

67. As a somewhat related example, the *National Strategy to Secure Cyberspace* recognizes the importance of establishing mutual assistance agreements to help infrastructure sectors respond to cybersecurity emergencies. *See* NATIONAL STRATEGY TO SECURE CYBERSPACE, *supra* note 4, at 24 (stating that the "[Department of Justice] and the Federal Trade Commission should work with the sectors to address barriers to such cooperation, as appropriate." (emphasis omitted)).

D. Collect Relevant Statistics

In a competitive communications industry, industry participants will have data about their own experiences, but no single industry participant will necessarily have a global view.[68]

Government can collect data where appropriate to identify problems, to determine their magnitude, and to provide a basis on which to evaluate potential solutions.

In determining whether to do so, it would invariably be necessary to balance several conflicting objectives. There may be compelling public interest reasons for gathering certain kinds of information; however, collecting that information represents a regulatory burden on the companies involved. That burden should be avoided where possible, and minimized where the data are truly needed.

Another tension of objectives relates to the sensitivity of data gathered. The public has a right to know information held by the Government, as embodied in the Freedom of Information Act (FOIA) and also by various state "sunshine" acts. At the same time industry participants have a legitimate interest in protecting competitively sensitive information, and in preserving the privacy of their customers. Often, these conflicting demands have been reconciled by having a third party anonymize data before providing it to the Government.[69]

There are specific exemptions from FOIA that address specific needs. One recent report rightly observes that these exemptions provide agencies with substantial ability to shield information of this type from inappropriate disclosure under FOIA;[70] however, that knowledge offers little comfort to industry participants, who must consider not only whether government *can* avoid inappropriate disclosure of their sensitive data, but also whether it *will*.[71]

---

68. *Cf.* NATIONAL STRATEGY TO SECURE CYBERSPACE, *supra* note 4, at 19 ("There is no synoptic or holistic view of cyberspace. Therefore, there is no panoramic vantage point from which we can see attacks coming or spreading.").

69. For example, when industry participants provide incident reports to Information Sharing and Analysis Centers (ISACs) operating under PDD-63, the information might be sanitized or anonymized before being shared with other ISAC participants or with the government.

70. *See* CRITICAL INFORMATION INFRASTRUCTURE PROTECTION AND THE LAW, *supra* note 66, at 25-29.

71. Notably, the Homeland Security Act specifically exempts information about critical infrastructure vulnerabilities provided voluntarily from FOIA obligations. *Cf.* PRESIDENT'S CRITICAL INFRASTRUCTURE PROTECTION BOARD, *supra* note 4, at 25 ("the legislation encourages industry to share information with DHS by ensuring that such voluntarily provided data about threats and vulnerabilities will not be disclosed in a manner that could damage the submitter." This is an area of ongoing concern for the DHS, which is working to ". . .

In those instances where data collection appears warranted in support of some public policy objective, government can work with industry to define the data required, to evaluate necessary safeguards on the dissemination of that information, and then to establish voluntary reporting programs.

Mandatory reporting can be appropriate in some circumstances, but only where the need for the data is compelling, where the data to be collected is well and narrowly defined, and where voluntary reporting for some reason is either inappropriate or unsuccessful.

E. Provide "Seed Money" for Research and for Interoperability Testing

For facilities that may benefit the public interest, but not necessarily individual users or industry participants, it may be that no private funding source is motivated to provide initial "seed" money. Certain security services, for instance, may benefit the public at large rather than any particular individual or company.

Public funding (or funding by public interest sources) may be the only practical way to foster development of such capabilities.

Analogous issues exist with interoperability testing. Many network services are useful only to the extent that they are interoperable with their counterparts in other networks. These counterpart services may be implemented independently and in competing products. Absent testing, there is no assurance that these implementations will interoperate correctly.

The government role in such activities is well established and widely accepted. For an example where this approach worked brilliantly, see the discussion of "Funding for the early Internet – a happier case study" later in this paper. Research[72] and interoperability testing may, in addition, serve to facilitate the standards process. The IETF will not progress a standard to Draft Standard status until interoperability among independent implementations has been rigorously demonstrated.[73]

---

establish uniform procedures for the receipt, care, and storage . . . of critical infrastructure information that is voluntarily submitted to the government.").

72. *See* PRESIDENT'S CRITICAL INFRASTRUCTURE PROTECTION BOARD, *supra* note 4, at 34-35 (explicitly recognizing the importance of prioritizing the Federal research and development agenda and tasking the OSTP with doing so).

73. BRADNER, *supra* note 44.

F. Support Desired Functionality in Products and Services Through Government's Own Purchasing Preferences

To the extent that the U.S. Government is itself a significant user of data networking services, its buying preferences for its own use can serve to influence the evolution of technology.

This represents an interesting proactive lever for change. Industry and the public tend to view this mechanism as legitimate and non-intrusive. It alters the economic incentives of suppliers, but it works *with* the economic system rather than against it.

This form of intervention may be particularly useful as a means of motivating suppliers (e.g., of software) to include desired functionality with the standard distribution versions of their products.

At the same time, it should not be viewed as a panacea. Government purchasing power may not be sufficient to drive widespread adoption (which is still subject to the economic effects of network externalities of the larger market).[74] Consequently, there is always the risk that government will pay a substantial premium in a vain attempt to foster the development and deployment of features and services that, at the end of the day, prove to be of limited utility.

A case in point is the U.S. Government OSI Profile (GOSIP). A massive international standardization effort was in play in the Eighties and into the Nineties on the part of the International Organization for Standardization (ISO) and the Telecommunication Standardization arm of the International Telecommunications Union (ITU-T).[75] They were seeking to develop an entire family of data communications protocols, based on principles of *Open Systems Interconnection* (OSI). The OSI protocols reflected modern concepts of protocol layering, and a full set of applications, including virtual terminal, file transfer, electronic mail, directory, and network management.

It might seem odd in retrospect that the global standards bodies and governments set out to recreate out of whole cloth functionality that already existed. OSI was nominally open to multiple vendors and implementations, but no more so than TCP/IP. Indeed, at the end of

---

74. *Cf.* HENNESSY ET AL., *supra* note 63, at 103 ("the IT sector is one over which the federal government has little leverage. IT sales to the government are a small fraction of the IT sector's overall revenue, and because IT purchasers are generally unwilling to acquire security features at the expense of performance or ease of use, IT vendors have little incentive to include security features at the behest of government alone.").

75. At the time, this was the International Telephone and Telegraph Consultative Committee (CCITT). *See* INTERNATIONAL TELECOMMUNICATIONS UNION, ITU OVERVIEW – HISTORY (Feb. 13, 2002), *at* http://www.itu.int/aboutitu/overview/history.html.

the day, OSI provided no new functionality that users found significant that was not already available under the TCP/IP protocol suite.

Many foreign governments considered TCP/IP to be the creation of the U.S. Department of Defense. Because TCP/IP had not been created by the recognized international standards process, they considered it inappropriate as the basis for a new, global family of communications standards.

The U.S. Government attempted to join a global bandwagon forming in favor of OSI. The National Institutes for Standards and Technology (NIST) published GOSIP Version 1[76] in August 1988, and followed a year later with GOSIP Version 2.[77] A profile was needed because many of the OSI protocols were so specified as to permit a variety of mutually incompatible possible realizations.[78] As of August 1990, Federal agencies were required to acquire OSI products when they required the functionality supplied by the OSI features specified in GOSIP. There was, however, no requirement that Federal agencies procure *only* GOSIP-compliant implementations for these purposes, nor was there an obligation for Federal agencies to *use* the GOSIP-compliant implementations that they had thus procured.

OSI protocols had developed what might have seemed to be an unbreakable momentum in the late Eighties. The ISO and CCITT unequivocally backed the protocols, while the Internet standards groups accepted at least an extended period of coexistence between TCP/IP and OSI protocols.[79] Digital Equipment Corporation (DEC), at the time a leading computer manufacturer, had committed to implementing OSI communications protocols in DECNET Phase V.

Today, however, OSI protocols serve as little more than a historical curiosity, an interesting footnote. Why is it that OSI protocols failed to achieve broad market acceptance?

Some have argued (and sometimes with surprising vehemence) that government support was the kiss of death for OSI protocols. This seems, however, to miss the point. In particular, it fails to explain the

---

76. Approval of Federal Information Processing Standards Publication 146, Government Open Systems Interconnection Profile (GOSIP), 53 Fed. Reg. 32,270, 32,270-02 (Dep't Commerce Aug. 24, 1988).

77. Proposed Revision of Federal Information Processing Standard (FIPS) 146, G3OSIP, 54 Fed. Reg. 29,597, 29,597-602 (Dep't Commerce July 13, 1989).

78. There was no assurance that two independent implementations of, say, the FTAM file transfer and access method would interoperate correctly. This is much less of an issue for TCP/IP protocols, where demonstrated interoperability is a prerequisite to standardization. It would be unusual, for instance, for the FTP support in two different TCP/IP implementations to fail to interoperate correctly.

79. *See* V. CERF & K. MILLS, RFC 1169: EXPLAINING THE ROLE OF GOSIP (Internet Engineering Task Force, Aug. 1990), *at* http://www.ietf.org/rfc.html.

success of TCP/IP protocols, which by all accounts benefited enormously from substantial support from the U.S. Government.

Others have argued that OSI protocols were cumbersome, and evolved slowly, because they were developed by large committees and because the protocol specification effort took place *in advance of* implementation. (Internet protocols, by contrast, would never be standardized until independent implementations had been shown to interoperate.) There probably is some truth to this assertion, and it is moreover plausible in terms of what we know of the economics of transaction costs – the need to obtain concurrence of a great many independent parties invariably exacts costs, one way or another. Nonetheless, it is only a part of the answer.

It must also be noted that OSI protocol implementations tended to be significantly more expensive than TCP/IP protocol implementations, not only in terms of purchase price, but also in terms of memory requirements, processing power requirements, and operational complexity. These were certainly factors, but they may not have been decisive.

A simple and sufficient explanation flows from the economic theory of network externalities. TCP/IP implementations were available on most platforms of interest, and the software was inexpensive or free in many cases, unlike OSI implementations. The deployment of OSI protocols at their peak probably never accounted for more than 1-2% of all traffic on the Internet. Users were motivated to use TCP/IP, because most of the content that they wanted to use or view was available in the TCP/IP world, and not in the OSI world. Content providers and application developers were motivated to use TCP/IP, because the majority of their prospective users were TCP/IP users. (Similar factors may have provided Microsoft Windows with an advantage over the Macintosh and, for that matter, VHS with an advantage over Beta, as noted earlier.)

OSI protocols were starting from a position of zero market share. They could not fully supplant TCP/IP protocols unless they replaced *all* of TCP/IP's functionality; however, TCP/IP began with a huge head start in functionality. Moreover, ongoing investment in new functionality based on the TCP/IP protocols inevitably outstripped that for new OSI functionality by a wide margin. Given that OSI had no compelling inherent advantage over TCP/IP, there was never any means to reverse this trend.

Eventually, the requirement to procure services implementing GOSIP (and its companion standard, the Government Network

Management Profile (GNMP))[80] was lifted. It was presumably recognized that a mandate to procure GOSIP-compliant solutions no longer served a useful purpose. Meanwhile, the U.S. Government had supported the evolution and testing of OSI protocols in many ways, and Federal agencies likely paid more than they otherwise might have to procure functionality that they ultimately did not need and, for the most part, did not use.

G. Fund the Deployment of Desired Capabilities

If deployment of a service is in the public interest, but not in the individual interest of the firms that must deploy it, and if deployment entails significant costs, then those firms have a significant economic disincentive to deploy. In a competitive, deregulated telecommunications marketplace, it is not clear how those firms could recapture their investment.

In those cases, it may be that the only possibility of achieving widespread deployment will be through some combination of subsidizing or funding that deployment as well as any associated incremental operational costs, or possibly by mandating deployment, or both.

The Communications Assistance for Law Enforcement Act (CALEA) is a case in point.[81] CALEA establishes carrier obligations in regard to lawful intercept of communications (e.g. wiretap). No telecommunications customer would wish to pay a premium for the privilege of having his or her own communications amenable to wiretap, nor would any carrier have a business incentive to implement the necessary tools and facilities.

As a result, CALEA establishes the Department of Justice Telecommunications Carrier Compliance Fund[82] in an effort to "make the carriers whole." This process has not been painless – carriers have argued that the fund does not adequately reimburse them for costs incurred.[83]

---

80. Approval of Federal Information Processing Standards Publications (FIPS) 146-2, Profiles for Open Systems Internetworking Technologies; and 179-1, Government Network Management Profile, 60 Fed. Reg. 25,888-02 (Nat'l Inst. of Standards and Tech. May 15, 1995), *available at* http://www.itl.nist.gov/fipspubs/fip179-1.htm.

81. Communications Assistance for Law Enforcement Act, Pub. L. No. 103-414, 108 Stat. 4279 (1994) (codified as amended in scattered sections of 18 U.S.C. and 47 U.S.C.) . For a brief background on CALEA, see FCC, CALEA, *at* http://www.fcc.gov/calea/ (last reviewed/updated 6/10/04).

82. Communications Assistance for Law Enforcement Act § 401 (codified as amended at 47 U.S.C. § 1021 (2000)).

83. In practice, the fund reimburses equipment suppliers. There has been to the author's knowledge only one instance where the fund was used to reimburse a service provider. Service providers incur costs for software upgrades to deploy CALEA, and they incur significant additional deployment costs beyond those associated with hardware and software.

Government funding for public goods can take any of a number of forms. It can come from general revenues. It can be a distinct fund, as is the case for CALEA. It can also be a separate fund privately managed on behalf of the government, as is the case for universal service.

H. Mandate Use of Desired Services

If functionality were truly deemed to be essential to the public interest, and if market forces were insufficient to ensure its deployment, then it could in principle be appropriate for government to mandate its deployment and use.

For the Internet, there is no obvious historical example; however, there are many examples in the history of the telephone industry in the United States.

One of these is the previously-noted CALEA. CALEA serves both to oblige telecommunications carriers to provide the technical means of achieving lawful intercept (wiretap) and to provide a mechanism for offsetting their costs in doing so. Lawful intercept is a legitimate societal need, but it does not specifically benefit an individual carrier; consequently, it can only be achieved to the extent that government provides the impetus, in this case by means of an explicit mandate.

Other examples of services that might have been unlikely to deploy absent government action include:

- Disabilities access to telecommunications,[84]
- Provision of 911 services, and
- Local number portability.[85]

This is the most intrusive means the government has of driving deployment. For a number of reasons, it should be used sparingly.[86]

First, as our experience with GOSIP demonstrates, government's ability to prognosticate is limited.[87] If government is to mandate deployment and use, it must be very certain that the functionality in question is truly necessary.

---

84. 47 U.S.C. §§ 225, 255 (2000).

85. *Id.* at § 251.

86. *Cf.* I3P REPORT , *supra* note 21, at 41 ("Aggressive approaches that more fully use the powers of the federal and state governments are also possible, but the costs and benefits are not well understood and the reasons for a general reluctance to regulate are well known. This statement raises the question of who is responsible for security in this information infrastructure 'commons' and who should pay for it.").

87. *Cf.* HENNESSY ET AL., *supra* note 63, at 103-104 ("it is likely that attempts at such regulation will be fought vigorously, or may fail, because of the likely inability of a regulatory process to keep pace with rapid changes in technology.").

Second, mandating a function will generally have a tendency to distort the relevant market. Wherever possible, market mechanisms should be preferred over mandates, especially unfunded mandates.

Finally, there is the risk that a government mandate might lock the industry into the use of a particular technology long after market forces would otherwise have obsoleted it.

I. Adoption of the Metric System – A Sobering Case Study

In considering the prospects for achieving deployment by means of government actions short of an outright mandate, it is helpful to consider historical precedents. We have already discussed GOSIP. Another example, albeit from a different technological domain, is conversion to the metric system.

In 1971, the National Bureau of Standards published a report, *A Metric America*,[88] recommending "[t]hat the Congress, after deciding on a plan for the nation, establish a target date ten years ahead, by which time the U.S. will have become predominantly, though not exclusively, metric. . . ."[89]

The benefits of metric conversion were thought to be manifest. Recognizing this, the U.S. Government has undertaken significant efforts over the years to foster adoption of the metric system,[90] including the passage of the Metric Conversion Act of 1975[91] and the issuance of Executive Order 12770[92] in 1991. Nonetheless, thirty-two years after the publication of *A Metric America*, it can hardly be said that the United States has "become predominantly, though not exclusively, metric".

In *A Metric America*, the National Bureau of Standards report recognized that the United States had become an isolated island in a metric world, and identified the potential costs associated with that isolation. They also attempted to quantify the costs of conversion, and the potential benefits – largely in terms of global trade and simplified

88. NAT'L BUREAU OF STANDARDS, A METRIC AMERICA: A DECISION WHOSE TIME HAS COME, NBS Special Publication 345, July 1971.

89. *Id.* at iii.

90. Interest in the metric system in the U.S. actually began much earlier. John Quincy Adams considered it in his *Report Upon Weights and Measures* in 1821. JOHN QUINCY ADAMS, REPORT ON WEIGHTS AND MEASURES (1821). Beginning in 1866, a series of laws were enacted that legalized the use of metric weights and measures, and directed the Postmaster General to distribute metric postal scales to all post offices exchanging mail with foreign countries. *See* NAT'L BUREAU OF STANDARDS, *supra* note 88. In fact, the U.S. became the first officially metric country by adopting the metric standards in the *Treaty of the Meter* to be the nation's "fundamental standards" of weight and mass in 1889. *Id.* at 14-15.

91. Metric Conversion Act, Pub. L. No. 94-168, 89 Stat. 1007 (1975) (codified as amended in 15 U.S.C. § 205 (2000)).

92. Exec. Order No. 12,770, 50 Fed. Reg. 35,801 (July 25, 1991), *available at* http://ts.nist.gov/ts/htdocs/200/202/pub814.htm#president.

education. The Metric Conversion Act of 1975 expressed the advantages in unambiguous bread and butter terms:

> (3) World trade is increasingly geared towards the metric system of measurement.
> (4) Industry in the United States is often at a competitive disadvantage when dealing in international markets because of its nonstandard measurement system, and is sometimes excluded when it is unable to deliver goods which are measured in metric terms.
> (5) The inherent simplicity of the metric system of measurement and standardization of weights and measures has led to major cost savings in certain industries which have converted to that system.
> (6) The Federal Government has a responsibility to develop procedures and techniques to assist industry, especially small business, as it voluntarily converts to the metric system of measurement.
> (7) The metric system of measurement can provide substantial advantages to the Federal Government in its own operations.[93]

An important collective effect of the Metric Conversion Act and of Executive Order 12770 has been to require that each Federal agency ". . . to the extent economically feasible by the end of the fiscal year 1992, use the metric system of measurement in its procurements, grants, and other business-related activities, except to the extent that such use is impractical or is likely to cause significant inefficiencies or loss of markets to United States firms, such as when foreign competitors are producing competing products in non-metric units."

The Metric Conversion Act also attempts to "seek out ways to increase understanding of the metric system of measurement through educational information and guidance and in Government publications." The Act established a United States Metric Board[94] tasked with carrying out "a broad program of planning, coordination, and public education." The Board was to perform extensive public outreach, to "encourage activities of standards organizations," to liaise with foreign governments, to conduct research and surveys, to "collect, analyze, and publish information about the usage of metric measurements," and to "evaluate the costs and benefits of metric usage." Thus, the metric conversion program attempted, to a lesser or greater degree, to employ essentially every tool available to government short of outright deployment funding or an explicit mandate.[95]

93. Metric Conversion Act, 89 Stat. 1007.
94. *Id.*
95. *Id.*

These efforts undoubtedly had effect, but not as great an effect as was intended. Why was this?

> A variety of reasons have been put forward to explain why the metric transition has not made widespread progress in the U.S. in the past. They include lack of national leadership, reluctance to embark on such a change, and *the failure of the voluntary effort that began in 1975*. The many competing national priorities and *the lack of immediate and visible benefit to a transition* clearly were factors. There are political, economic, and social reasons to explain the apparent slow progress and reluctance to make the transition.[96]

It is not the intent of this paper to trivialize or over-simplify what undoubtedly was a very complex process. The key point that the reader should take away from this case study is that, for certain kinds of innovations where economic incentives are not sufficient to motivate their deployment in a free market system, there can be no assurance that government actions short of deployment funding or an explicit mandate will generate substantial deployment.

J. Funding for the Early Internet – A Happier Case Study

In the case of the Internet, by contrast, the historic effects of direct Government funding have in most instances been salutary. The original ARPAnet, the predecessor to the Internet, was funded in the late Sixties by the Advanced Research Projects Agency of the U.S. Department of Defense (DARPA).[97]

In the early Eighties, DARPA funded the University of California at Berkeley to incorporate TCP/IP protocols into Berkeley UNIX®.[98] This effort produced one of the most widely used TCP/IP implementations. Berkeley UNIX was incorporated into an emerging generation of UNIX workstations, thus fostering precisely the network externalities effects that ultimately enabled TCP/IP to prevail in the marketplace.

---

96. DR. GARY P. CARVER, NAT'L INST. OF STANDARDS & TECH., A Metric America: A Decision Whose Time Has Come – For Real, NISTIR 4858 (1992), *available at* http://ts.nist.gov/ts/htdocs/200/202/4858.htm (emphasis added). Dr. Carver was then chief of the Metric Program at the National Institutes of Standards and Technology (NIST).

97. BARRY M. LEINER ET AL, INTERNET SOCIETY, A BRIEF HISTORY OF THE INTERNET (Dec. 10, 2003), *at* http://www.isoc.org/internet/history/brief.shtml#Origins. Note that the Advanced Research Projects Agency (ARPA) changed its name to Defense Advanced Research Projects Agency (DARPA) in 1971, then back to ARPA in 1993, and back to DARPA in 1996.

98. *Id.*

The U.S. National Science Foundation (NSF) provided initial funding for CSNET as a limited-function network for the academic research community. The NSF then invested an estimated $200 million from 1986 to 1995 to build and operate the NSFNET as a general purpose Internet backbone for the research and education community.[99]

Most observers would agree that the modest investments that DARPA and the NSF made in the Internet have collectively been a brilliant success.

IV. CONCLUDING REMARKS

On a hasty reading, this paper might be construed as advocating that government take an intemperate, interventionist approach toward the Internet.

What is called for, in the author's view, is a reasoned and balanced approach. Much has been made of the lack of regulation of the Internet.[100] Yet the very existence of the Internet is a direct result of a succession of government interventions, many of them highly successful. Among these were the initial funding of the ARPAnet, the FCC's Computer Inquiries (simultaneously deregulating services like the Internet while opening up underlying telecommunications facilities for their use), support for CSNET and the NSFNET, and the funding of TCP/IP protocol implementation in Berkeley UNIX.[101] Each of these achieved important and positive results without resorting to a regulatory mandate.

There have also been failures of government intervention. Perhaps the most relevant was the U.S. Government's support of OSI protocols through GOSIP and the GNMP, as described earlier in this paper. That ultimately unsuccessful attempt to use the purchasing power of government to promote global standards that the marketplace had by and large not demanded, likely resulted in significant diversion of attention and waste of resources on the part of both government and industry.

Another example was metric conversion, where the U.S. Government has attempted a combination of practically every conceivable measure short of an outright mandate but has not achieved the widespread deployment that was hoped for.

---

99. *Id.*

100. *See* JASON OXMAN, THE FCC AND THE UNREGULATION OF THE INTERNET (FCC Office of Plans and Policy, Working Paper No. 31, July 1999), *available at* http://ftp.fcc.gov/Bureaus/OPP/working_papers/oppwp31.pdf.

101. LEINER ET AL., *supra* note 97.

Government is neither omniscient nor omnipotent. Government could do too little. Government could also do too much. How to know which is which?

Two principles may be useful going forward:

> BALANCE: Government should recognize both the risks of action and those of inaction, and make cautious and deliberate choices.

> MINIMALISM: Government should choose to err in general on the side of less regulation rather than more. Do not attempt a massive intervention where a less intrusive intervention might suffice. Do not intervene at all unless markets have shown themselves to be unable to deliver a socially important outcome.