DATA COMMUNICATIONS MANAGEMENT

Next-Generation Networks, Part 2: The NGN Paradigms

Keith Knightson

INSIDE

Paradigms for NGN; Internet as the NGN; Layer Functionality Reduction and Redistribution; Protocol Architecture; Service Provision: Vertical Aspects; Service/Network Separation: Horizontal Aspects; Network Intelligence Shift; Broadband: Equal Access Service Paradigm; Broadband: Bottlenecked Service Access Paradigm; Mobility Paradigm; Policy Paradigms: Of Carts and Horses

PARADIGMS FOR NGN

To illustrate the dramatic and disruptive shifts inherent in the deployment of next-generation networks (NGNs), a number of graphic illustrations are presented. These shifts stem from the characteristics of an NGN based on Internet-like technology described in Part 1 (50-10-13).

PARADIGM 1: INTERNET AS THE NGN

A major paradigm is the philosophy of the Internet itself and of IP networks in general. The Internet Protocol (IP) acts as the common denominator of the Internet protocol suite. Together with its universal addressing scheme, IP provides the end-to-end network layer[1] connectivity communication mechanism. IP can traverse a multiplicity of underlying (tele)communications technologies, and can convey digital information pertaining to any kind of application. Exhibit 1 shows the "hourglass" paradigm that represents this philosophy.

The Internet paradigm has profound implications for the telecommunications scene. These will be the subject of Part 3. However, at this point we can already see the de-emphasis on the importance of

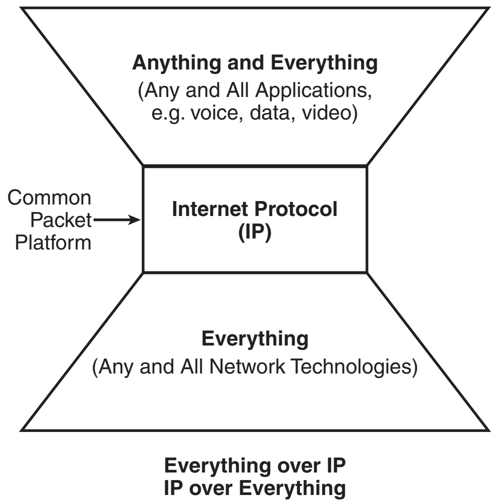
PAYOFF IDEA

This article is the second in a three-part series describing next-generation networks (NGNs), and outlines the significant trends that will shape the NGN, stemming from the characteristics of packet-based networks that were outlined in the first article. The key trends are presented in the form of a set of graphical paradigms that illustrate the dramatic effects that NGNs will have on the communications environment.

The next and third article will identify the effects and impacts that these paradigm shifts will have on users, network service providers, and the communications industry in general.

Auerbach Publications

EXHIBIT 1 — Paradigm 1: Internet Paradigm



telecommunications technologies.[2] Because IP functionality is the only functionality "seen" by the layer(s) above IP, any particular functionality offered by the underlying telecommunications technology is (largely) ignored and not used.

The general philosophy of "everything over IP, and IP over everything" follows the hourglass or wine glass principle, illustrated in Exhibit 1 by the narrowing in the middle and the fan-out above and below.

However, as discussed in Part 1, the Internet as originally envisaged and designed in the 1970s is not adequate for today's requirements. We concluded that the original Internet cannot be considered the NGN as such, but will form the basis of the NGN. Many enhancements have been designed for the Internet to meet NGN requirements for convergence and more can be expected. It is not clear whether a transition to IP version 6 is absolutely necessary for this paradigm to hold true.[3]

The main areas for functional enhancement, as discussed in Part 1, are:

- Quality-of-service (QoS) selection
- Quality-of-service provisioning and management
- Emergency services
- Lawful access/interception
- Accounting
- Fault and performance management
- Mobility

A number of other areas complicate the direction of an architectural design philosophy of the future Internet, including:

- Use of MPLS
- Transition to IP version 6 (IPv6)
- Network address translation (NAT)
- Firewalls

Paradigm 2: Layer Functionality Reduction and Redistribution

The shift from connection-oriented networks to a connectionless IP network represents a simplification in the network layer.[4] IP switches (i.e., routers) do not maintain any information about the relationship between individual packets, and thus are not aware of "conversations" or "calls" between end users.

Another aspect in this area is the concept that networks should be "dumb," and any "intelligence" is best located in the end-user terminals that are in tune with the user's application/service needs.

The advent of more reliable transmission media (e.g., optical networks) makes the need for link-by-link error correction less important. This again simplifies the networks and the need for high functionality layer 2 protocols. Where a layer 2 protocol is present, it is most likely to be the Ethernet protocol, IEEE 802.3. It may be expected that Ethernet will play a significant role in access networks rather than core networks.

The increase in sophistication of optical networks, such as switching, adds functionality to layer 1, a layer that traditionally had little functionality.

It is not entirely clear where all this will lead. It could be postulated, for example, that layer 3 and layer 2 could be merged into a single layer or treated as a single layer.

However, it is clear that an IP network has no need for the rich functions and features of many telecommunications-oriented technologies, such as ATM, Frame Relay, etc.

PARADIGM 3: PROTOCOL ARCHITECTURE

The fundamental Internet paradigm illustrated in Exhibit 2 means that the only major and essential component for end-to-end communication is the Internet Protocol (i.e., IP). Coupled with the functionality shifts identified earlier, and illustrated in Exhibit 2, one could imagine a trend toward a simplification of the basic protocol stack. This is shown in Exhibit 3.

The emergence and dominance of IP as the network layer protocol of choice reduces the need for other network layer protocols and for complex layer 3 switching functionality. Consequently, there is no overwhelming technical requirement for the use of ATM or Frame Relay.

Only legacy interworking requirements would require more complex stacking arrangements. In the absence of legacy requirements, it would appear that other

EXHIBIT 2 — Paradigm 2: Layer Functionality Reduction Paradigm

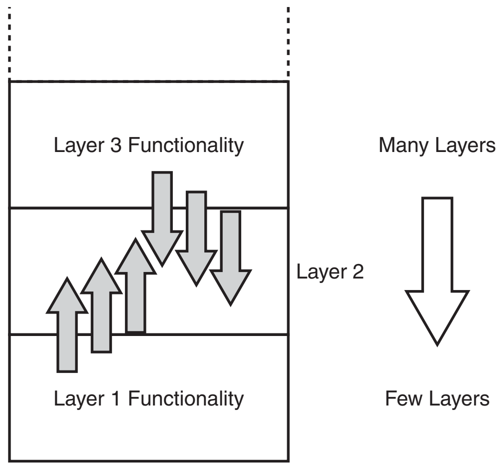
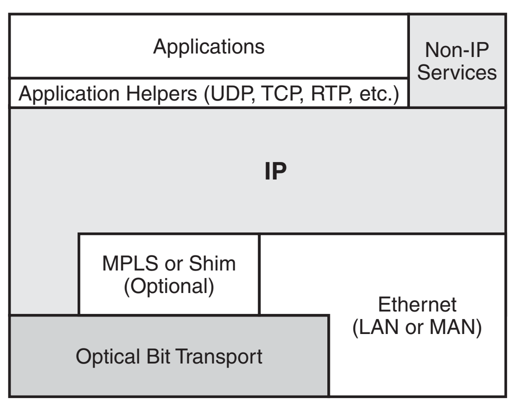


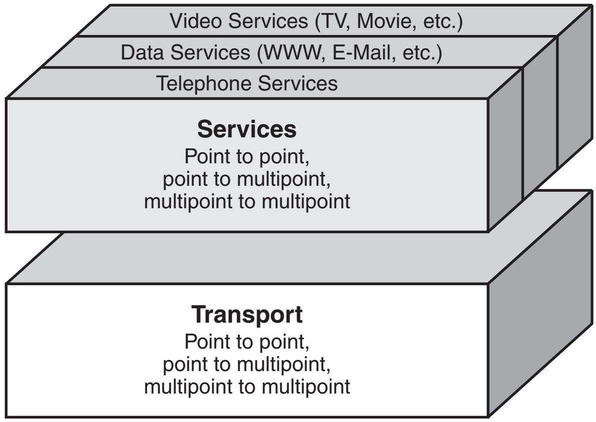
EXHIBIT 3 — Paradigm 3: Protocol Architecture Paradigm



layer 3, or even layer 2 switching functionalities are no longer technically required. The demise of ATM and Frame Relay would appear to be predictable.

However, IP has no inherent QoS functions. It is envisaged that QoS requirements will either be met by the over-provisioning of bandwidth for "IP trunking," or by streaming IP packets over trunks differently provisioned[5] according to the requirements of different applications. In the latter case, technologies such as ATM and Frame Relay may be regarded as "IP helpers" (i.e., technologies that assist IP to achieve QoS differentiation). In this regard, ATM might be useful to

EXHIBIT 4 — Paradigm 4: Service Provision: Vertical Aspects



provide QoS engineering for IP packet streams. However, MPLS also offers such a capability in a simpler manner. Such IP helpers may additionally speed up the routing process by flattening the network and reducing the number of routers to be traversed.

PARADIGM 4: SERVICE PROVISION: VERTICAL ASPECTS

As discussed in Part 1, the separation of services from the network is a key cornerstone of the NGN. The vertical aspects of this separation are shown in Exhibit 4. The separation is represented by two distinct blocks or strata of functionality. There are several points of note to be made. First, there is a set of transport functions that are solely concerned with conveyance of digital information, of any kind, between any two geographically separate points. A complex set of networks may be involved in the transport stratum, constituting layers 1 through 3 of the OSI seven-layer Reference Model.

The transport functions, residing in the lower stratum, provide the connectivity between the service platforms, residing in the upper stratum, which provide the user services (telephone service, Web service, etc.).

The function of the transport network is to provide connectivity. It will provide:

- User-to-user connectivity
- User-to-service platform connectivity
- Service platform-to-service platform connectivity

In general, any and all types of network technologies can be deployed in the transport stratum.

Second, there is a set of application functions related to the service to be invoked. These reside in the services stratum and may comprise voice services (including telephone service), data services (including but not limited to Web-based services), and video services (including but not limited to movies or TV programs), or some combination thereof. The services stratum may involve a complex set of geographically distributed service platforms, or, in the simple case, just the service functions in two end-user sites.

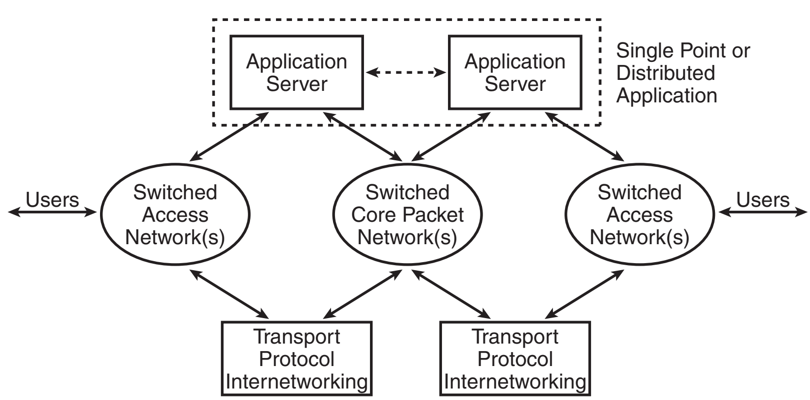
In general, each stratum will have its own set of players and administrative domains. The players and domains involved in application service(s) provision are independent from those involved in transport connectivity provision. Each stratum needs to be treated separately from the technical, commercial, and policy points of view. The horizontal distribution aspects are covered in more detail in the next section.

Some services require more than just point-to-point communication capabilities. Some require point-to-multipoint capabilities,[6] and others may require multipoint-to-multipoint capabilities.[7] Examples of the former include multicasting and broadcasting applications, and an example of the latter is some form of conferencing application. In general, such capabilities can be provided either at the transport level or at the application services level.

PARADIGM 5: SERVICE/NETWORK SEPARATION: HORIZONTAL ASPECTS

In general, one cannot make any assumptions about the geographical location of the "server" (see Exhibit 5). This being so, a different and new interpretation of

EXHIBIT 5 — Paradigm 5: Service Provision: Horizontal Aspects



an access network needs to be made. The network segment used to access a server may include switching/routing components. For example, one might make the following assumptions based on the separation of service from network:

Assumption A: In general, the geographical location of a "server" is not restricted.

Assumption B: In general, users are not restricted to using a single dedicated server.

If either or both of these assumptions are correct, then it seems that servers will have to be addressable, even if they are very local.

So, when we speak about "an access network" do we mean:

1. All the network components used to access a server wherever it may be?
2. Something else (say the last mile(s) passive technology only)?

There are several points to make surrounding this issue.

- If (1) is meant, a different and new interpretation of an "access network" needs to be made.
- If (2) is meant, what exactly is the definition of "access network"? A network in which no switching/routing is possible? This has profound implications in terms of service restriction because it implies that a "bottleneck server" will always be present in an NGN.
- If (2), one could imagine that an access network is just layer 1, bit shifting only. Two sub-cases are then possible:
  a. A server provided by the same organization as that providing the bit shifting
  b. A server provided by a different organization than that providing the bit shifting

However, in both these cases (a) and (b), we end up with a single dedicated bottleneck server.

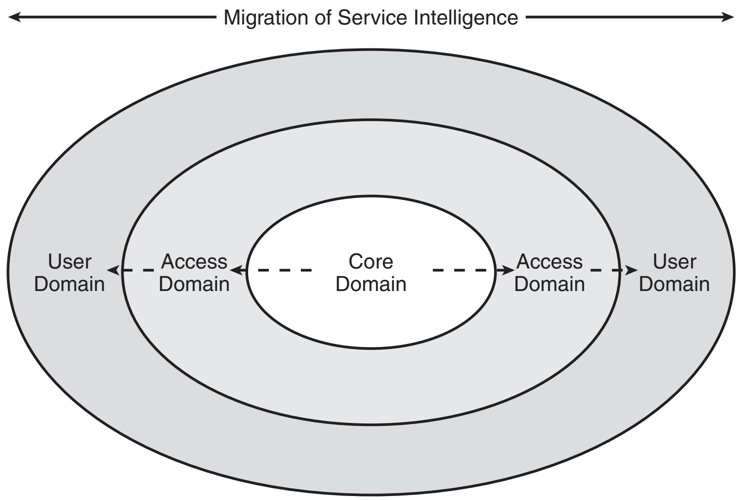
From the NGN services perspective, it would seem that we want Assumptions A and B to hold, forcing us to support definition (1) for an access network (i.e., all the network components used to access a server).

PARADIGM 6: NETWORK INTELLIGENCE SHIFT

Another effect of the basic Internet paradigms is the *migration* of intelligence from the core network to network "edges." So, as far as the geographical distribution is concerned, one can envisage the "dumbing-down" of the core network elements and the migration of network intelligence to the access segments. This is illustrated in Exhibit 6.

It is not clear how much intelligence is necessary in the access or between the access and the core. There is a tendency to imagine that telephone service intelli-

EXHIBIT 6 — Paradigm 6: Network Intelligence Shift Paradigm



gence "belongs" to network. We are used to a centrex access model and a local/long distance hierarchy of telephone exchanges. This model is sometimes transposed to the new environment by taking the existing telecommunications infrastructure and simply swapping the technologies, for example, replacing TDM trunks with IP trunks and telephone exchanges with telephone server platforms. However, it is not clear that this model will be viable, given that intelligence for a telephone service could just as easily reside in customer premise equipment with a little help from Domain Name Servers (DNS) or directory servers. This subject will be discussed further in Part 3.

PARADIGM 7: BROADBAND: EQUAL ACCESS SERVICE PARADIGM

The key to provision of services is gaining access to them. If service innovation is to be achieved, access to services has to be relatively free from impediments. In this respect, users must be able to choose service providers freely and reach them easily, and service providers must be able to deal directly with their customers.

An "equal access" scenario is shown in Exhibit 7. Any user can access any service provider on an equal-access basis. The service providers may be fairly close (local) to a given user or, alternatively, located anywhere on the transport network. In this sense, no distinction is made between an access network and a core network. The interactions between service providers are not shown in Exhibit 7.

It is very important to note that in this scenario, there is independence between service providers and the providers of transport facilities right up to the user's premise. This may be contrasted with Paradigm 8, where no such independence exists.

EXHIBIT 7 — Paradigm 7: Broadband: Equal Access Service Paradigm

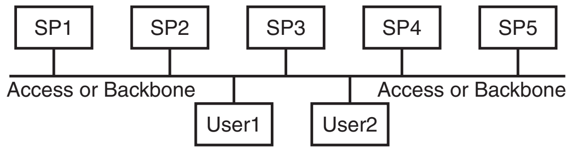
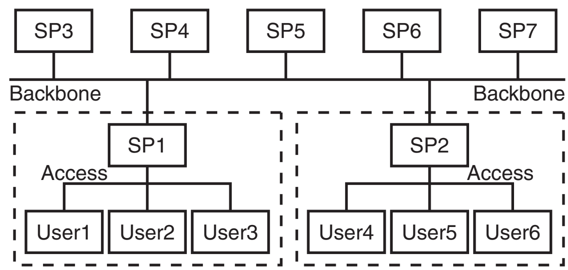


EXHIBIT 8 — Paradigm 8: Broadband: Bottlenecked Service Access



The introduction of Ethernet and/or ATM into the "last mile" infrastructure, over DSL or cable media, make an access network capable, in theory, of providing switching and/or routing capabilities, facilitating equal access scenarios.

PARADIGM 8: BROADBAND: BOTTLENECKED SERVICE ACCESS PARADIGM

Exhibit 8 shows another arrangement for access to services. In this arrangement, the same organization that provides the last-mile medium to the user premise provides services and control access to all other service providers. The service(s) and access transport are bundled together, as indicated in Exhibit 8 by the dotted box. The user can only reach service providers via his "home" network provider.

There are obvious disadvantages to this arrangement, which may or may not be outweighed by economic factors. It is also true that, traditionally, this method allows a distinction to be made between passive access networks which have had no switching and/or routing capabilities, and core networks which have had switching and/or routing capabilities.

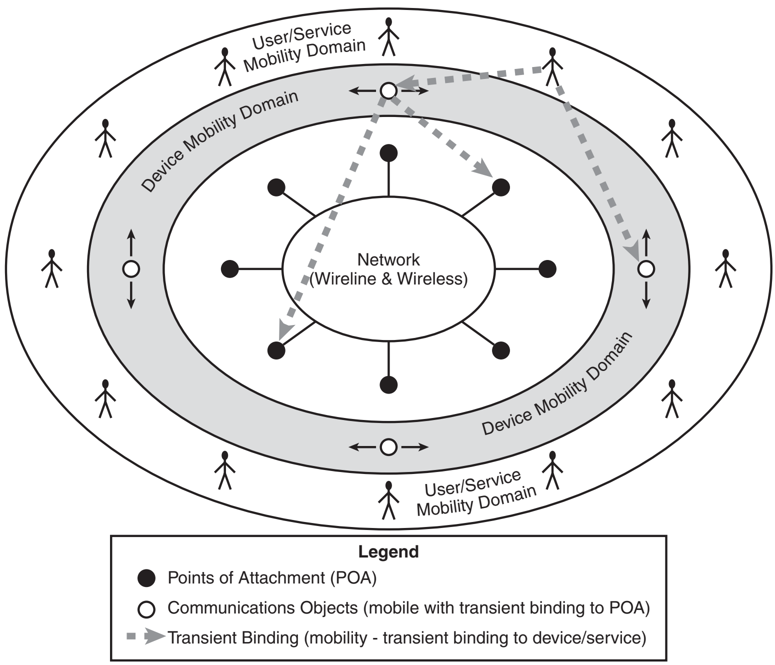
The introduction of Ethernet and/or ATM into the "last-mile" infrastructure makes an access network capable, in theory, of providing switching and/or routing capabilities. Again, the interactions between service providers are not shown in Exhibit 8.

PARADIGM 9: MOBILITY PARADIGM

Mobility has become a major feature in today's world of wireless telephony. Number portability has become commonplace in today's world of wireline telephony. However, these two "mobility" schemes are realized by two different systems, and the scope of mobility is limited according to the particular network. For an NGN to be a single seamless network, these two systems would need to be integrated, and the scope and extent of mobility made equivalent across all types of networks, that is, across both wireless and wireline, via an integrated mobility mechanism.

The general paradigm for mobility in an NGN is shown in Exhibit 9. It can be

EXHIBIT 9 — Paradigm 9: Mobility Paradigm



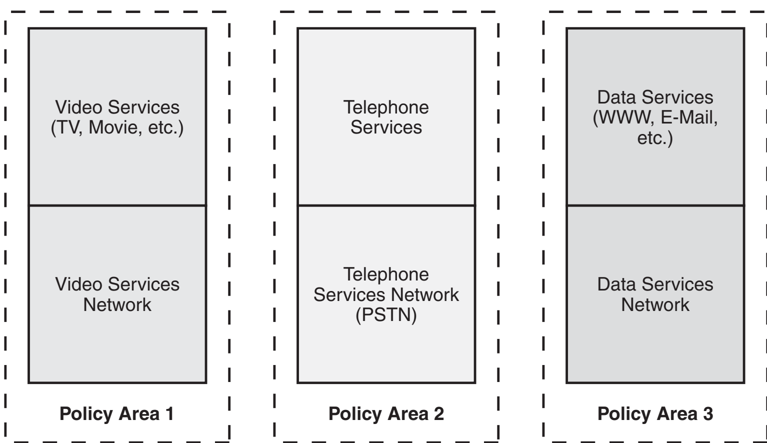
seen from Exhibit 9 that there is no permanent relationship between the identity of an object (e.g., user or device) to be involved in a communication and its location (i.e., the place in which it can be found). In all cases, a transient relationship is established between the communication object and a location. Thus, in general terms, the location/place of a given object can be represented by the physical point of attachment (POA), at which the said object can be reached/found.

Sophisticated directory and/or customized agent-based services allow various identification schemes to be used by both the calling and called parties. These schemes, in general, have no fixed relationship to particular physical locations.

POLICY PARADIGMS: OF CARTS AND HORSES

Current policy is inextricably entangled with the notion of vertically integrated separate networks. In general, separate policies apply to the telephone network, data networks, and broadcast networks, as shown in Exhibit 10.

EXHIBIT 10 — Separate Networks = Separate Policy



In the United States, there is a definition of enhanced service[8] that is used as the basis for policy and regulation (or not, as the case may be). An enhanced service is one in which the processing of user-generated information is deemed to be taking place, and includes any processing of protocols generated by the user.[9] Further, it is based on a model where the PSTN is used to provide dial-up access to the Internet. A basic telecommunications service is where no processing of user information is deemed to be taking place. This has the following consequences:

- The Internet is deemed to be an enhanced service.
- The telephone service is deemed to be an underlying basic service.

It is this model that has resulted in the telephone service being the subject of regulation, and the Internet not being the subject of regulation.

This situation becomes completely inappropriate in the context of a converged network in which the telephone service is carried over the Internet. Which is now the basic service and which is the enhanced service? It is clear that in this case, the policy makers are out of step with technical reality and, as such, the basic/enhanced model represents a seriously flawed approach to future policy regulation unless the technical basis for regulation is changed.

There is confusion regarding the term "telecommunications." This is not simply an academic terminology issue, because legislation could and should be based on a sound technical representational model, including the related technical definitions and technical relationships between architectural components. Similarly, it is not clear what people mean when they speak of the "Internet."

Europe has realized this problem and has completely revamped its regulatory framework. OFTEL has also issued a statement regarding the VoIP Telephone Service.

It is clear that convergence of services in an NGN network follows a particular technical model or models, along the lines of the various paradigms illustrated and discussed above. Our purpose here is not to discuss regulation or specific elements thereof, but to simply point out that any policy or regulation must be based on a set of appropriate technical models. These models must include a technically consistent set of terms and definitions, and architectural components that follow technology in practice.

Catch-up is required by the policy makers. References 3 through 6 provide further insight into this area, which will be discussed in more detail in Part 3.

CONCLUSION

The shift toward Internet technology for a single worldwide next-generation network provokes dramatic changes in the communications environment. The trends are already evidenced by:

- The shift toward the Internet Protocol (IP) for carriage of all types of content (e.g., voice, data, video)
- The simplification of the protocols in the lower layers, at the expense of Frame Relay and ATM networks
- The dumbing-down of core networks in favor of intelligent user systems
- Clean and clear separation between content-related services and systems from transport-related (i.e., delivery) systems
- Increasing importance of maintaining this clean separation between services and transport in broadband access networks
- Increasing emphasis on uniform global mobility and roaming across both fixed and wireless transport systems
- The inappropriateness of the technical assumptions used by policy makers

It is also concluded that the frameworks used by policy makers for vertically integrated networks are not transposable to the horizontally integrated environment of an NGN convergence network. In this regard, it is important to separate the basis of policy from the policy itself. Only in this way can policy reflect the true technical situation to which it is intended to apply.

Part 3 will further address the practical effects and impacts of these trends.

Notes

1. As defined in the ISO Reference Model, IS 7498.
2. It is pertinent to ask, "What is the difference between communications technologies and telecommunications technologies?" One can argue the case for the meanings to be identical. However, for the purposes of this article, we are making the following distinction (for the time being, because this distinction is questionable) in order to separate what have been in the past, two different particular sectors/communities. We use the term "telecommunications" technologies to reflect technologies that traditionally would have been defined in the ITU, and of primary interest to public network operators, usually subject to some form of national regulation. Moreover, a specific industry sector, the telecommunications industry has been traditionally responsible for the manufacture of this type of technology, including companies such as Lucent. On the other hand, we use the term "communications technologies" to reflect those technologies that have arisen from the computer networks community, as for example, developed from the original DARPA network or from other computer network sources such as local area (computer) networks (LANs) such as Ethernet. The specifications for these belong to the Internet Engineering Task Force (IETF) or the Institute of Electrical and Electronic Engineers (IEEE), which are not national or treaty-based organizations. In this case, a different set of manufacturers has been involved, notable examples being Cisco, Juniper, etc.
3. The absolute need for the introduction of IP version 6 depends on exhaustion of the IP version 4 address space. A discussion of this issue is contained in References 1 and 2.
4. As explained in Part 1, for simplicity, this article will continue to use the terms "connection-oriented" (CO) and "connectionless" (CL) to distinguish between two "classical" modes of operation. CO protocols and CO network technologies are generally those in which closely coupled call setup, data transfer, and call tear-down phases are present. In CL protocols/networks, these distinct per user-user conversation phases do not occur. In CL networks, packets, sometimes termed "datagrams," always contain both a source and a destination address. Although religious battles still rage between proponents of CO and CL operation, the distinction between them has become increasingly blurred, as enhancements have been made to each style of operation. As such, all protocols could be regarded as belonging to a continuum of protocols in which CO and CL are two extremes. It is possible to place both CO and CL in the problem space, if one considers that providing "context" for "data" is the key requirement for communications systems. In such a model, call set-up in CO operation is simply a particular method of establishing "associations" with fine granularity of context, and in CL other mechanisms are used to provide "associations" of coarser granularities of context.
5. Known as differentiated services (diff-serv).
6. One sender to many receivers (1:N).
7. Many senders to many receivers (M:N).
8. The Commission defined "enhanced services" as: services, offered over common carrier transmission facilities used in interstate communications, which employ computer processing applications that act on the format, content, code, protocol, or similar aspects of the subscriber's transmitted information; provide the subscriber additional, different, or restructured information; or involve subscriber interaction with stored information.
9. So, in the case of a packet network, network-based bridges and/or routers would be regarded as providing enhanced services. This is not a useful or tenable conclusion.

References

RFC 2101, IPv4 Address Behaviour Today.

RFC 2775, Internet Transparency.

Potential Relevance to the United States or the European Union's Newly Adopted Regulatory Framework for Telecommunications, Federal Communications Commission, Office of Plans and Policy, Paper No. 36, July 2002.

Frequently asked questions on the regulation of Voice over Internet Protocol, OFTEL 2 April 2002.

Green Paper on the Convergence of the Telecommunications, Media and Information Technology Sectors, European Commission, Brussels.
Challenges of Convergence, J. Scott Marcus, Senior Advisor of Internet Technology, FCC.

Keith G. Knightson has been involved in data communications for more than 25 years, and has worked for British Telecom and Nortel Networks. He has operated a consulting company since 1995, concentrating on network architectures and next-generation networks. He can be contacted at kgk@rogers.com.