Multi-Modal Routing and Switch Node Architecture

James P.G. Sterbenz and Rajesh Krishnan
BBN Technologies, Cambridge MA
jpgs@sterbenz.org, *krash@bbn.com*

DARPA / DOE / NASA / NIST / NLM / NSF
Workshop on New Visions for Large-Scale Networks: Research and Applications

12–14 March 2001

ABSTRACT

As the Internet increases in size, heterogeneity, and ubiquity of access, a number of long-held assumptions and characteristics are changing in fundamental ways, including the utility of conventional address-based forwarding and routing. Research in a variety of routing techniques (name, location, characteristics) is underway to better serve the needs of the future Internet. We argue that no one contact mode is capable of efficiently serving the needs of all application flows, and that individual packets may need support for multiple simultaneous modes. Furthermore, substantial performance benefits may be obtained by providing native support for multiple modes in the network layer. We then present the high level architecture of a network switching node capable of supporting multiple modes.

1. Changing Assumptions

As the Internet increases in size, heterogeneity, and ubiquity of access, a number of long-held assumptions and characteristics are changing in fundamental ways. We will focus on two broad categories of these changes: overall network architecture and individual node addressing and identity.

1.1 Network Architecture

Traditional network architectures are relatively static, and have well defined boundaries between end systems and switching nodes. The future Internet will be characterised by very heterogeneous nodes and links under highly dynamic behaviour.

Mobile Wireless Networks

The deployment of wireless networks is rapidly increasing based on the demand for location-independent and mobile network services. This fuels a demand for seamless networking that is eliminating the distinction between the wireless and wired portions of the Internet. One impact is a fundamental need for heterogeneity in network links and node capabilities. Network nodes should support the routing and forwarding mechanisms appropriate to the context in which they reside, as well as the needs of particular packets and flows.

Additionally, mobile wireless networks greatly increase the dynamic behaviour of the network, and challenge current binding mechanisms and routing algorithms that depend on relatively static node location for convergence. This motivates mechanisms that combine the binding, routing, and forwarding process to incrementally determine the routing of packets.

Disconnected operation and weakly connected operation become more likely. As nodes leave one subnetwork to join a different subnetwork a significant tradeoff must be revisited: maintaining an address while being mobile, versus reacquiring a different address within the new subnetwork.

Distinction between end system and switch

One of the consequences of mobile *ad hoc* networks is that nodes may serve the roles of both end system and switch. While it is possible to have distinct switches and endpoints, it may be necessary for nodes to serve both roles to form a multi-hop network among participants while reasonably bounding transmission power and degree of connectivity.

1.2 Node Addressing and Identity

Traditional networks use the address of the endpoint as the means of determining the path that packets traverse. The needs of the future networking environment challenge these mechanisms, at least, particularly due the scale and heterogeneity of the network and user community. At worst, new mechanisms may be needed that do not depend on addresses at all.

Number of End Systems *per* User

Ubiquitous computing research indicates to us that the number of end systems *per* user will increase by at least an order of magnitude. An individual's personal network is likely to consist of a personal computing node [Finn 1998], I/O devices (heads-up display, earphones, microphone, hand-held input device) and sensors (geolocation, environmental, biological). Thus, there will be at least tens of nodes *per* individual. Massive biosensors (*e.g.* smart blood) could increase this by orders of magnitude. Similarly, wearable computational fabrics will contribute to this increase.

The assumption that addressing limits will be fixed in the near future (for example, by IPv6) is increasingly suspect. NATs (network address translators) are only a short-term hack that forestalls the need for more systematic solutions. It is possible that NATs will be in permanent use, and proposals have been made that institutionalise their use [Cheriton 2000].

Sensor networks promise to also dramatically increase the scale of the Internet node count. The majority of these sensors may not be assigned to specific users, which has implications in addition to the scale described subsequently. It is not clear that every device must have a unique disambiguating resource identifier (such as EUI-64) under all situations. Furthermore, unique identifiers are associated with a number of privacy issues. Large numbers of identical mass-produced devices can form networks that may need to interface with the Internet. While they may acquire a unique address in aggregation, *individual* globally unique addresses are not necessary. Although anonymous networks of identical devices cannot elect a leader or unique addressing through purely deterministic algorithms, randomisation and symmetry breaking by tagging a small number of devices provide viable approaches that can be useful in the long-term.

Even if the address space were significantly larger, there is reason to believe that conventional addressing and identification techniques break down with orders-of-magnitude increase in network scale. In any case, mechanisms that allow flexibility at the node level to deal with a variety of addressing mechanisms will be necessary.

Node Identity and Location

One of the significant barriers to addressing in tera- and peta-node networks is the assignment and administration of addresses. In some cases, it may be more efficient to use a more native way of contacting node, rather than translating to addresses. For example, the problem of administrating addresses to trillions of disposable sensors may exceed the difficulty of forwarding packets using other mechanisms.

Furthermore, nodes may not correspond to users in the conventional sense, nor have administrative owners. In this case, the role of an address is merely an artifact of its need in traditional network layer protocols.

2. Contact Modes

There are a variety of reasons motivating communications, based on who or what an entity is, where an entity is located, or based on its characteristics. We refer to these as *contact modes* to distinguish from the overloaded terms addressing (which refers to a logical network location identifier), and location, which we will use to refer to physical or logical location. Figure 1 shows the routing based on a variety of modes. In each case, the source endpoint wishes to send packets along a path $p(d)$ that will lead to destination endpoint d.

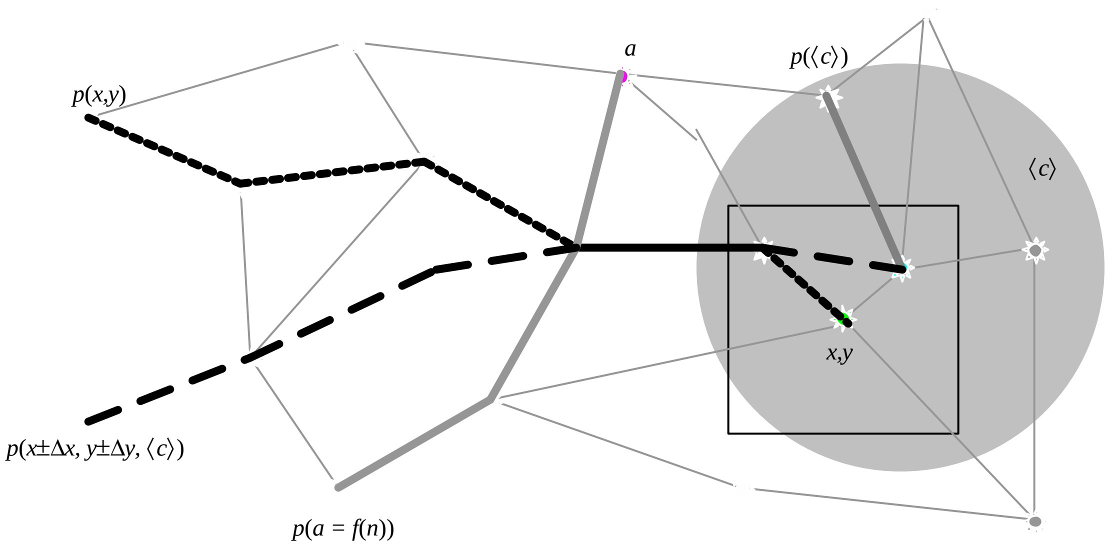

FIGURE 1. CONTACT MODES

2.1 Name Based

On of the most common reason to contact an entity is because of *who* or *what* it is; this can be expressed by a *name*. A name is a globally unique persistent identifier of a person or object in its namespace, such as *Claude Shannon*, *War and Peace*, or *The Sound of Music*.[1] Names may be bound to multiple instances, such as various copies of *War and Peace* (and in different media types). Conversely, names may identify aggregate entities, such as *The Shannon family* and *Acme Transmogrifier Corporation*, or role based, such as *The Father of Information Theory* or CEO *of Acme Transmogrifier Corp*.

Conventionally, a user who wishes to contact name n will manually map this to an address[2] a that is used to contact the other endpoint of the communication. The goal is for a network name resolution service to perform the binding to an address $a = f(n)$. This may consist of an initial explicit name resolution step, providing an address that is the basis for packet forwarding, but other forms of binding are possible (described subsequently).

---

[1] We are using English language examples here and not considering the possibility that multiple people are named *Claude Shannon*, or even *Claude Elwood Shannon*. A real name for *the* Claude Shannon in this example would have to include a component that guarantees uniqueness, for example *Claude_Shannon_830672224994*.

[2] Such as a an email address or DNS identifier (which is *not* a name, but rather a human readable indirection to an IP address).

2.2 Location Based

In some cases it is *where* an entity is located that is the primary motivation for communication rather than its identity. A coordinate system is used for the location of an entity, such as geographic (*e.g.* sensor at *x*,*y* = N42°23.36′ W071°08.86′), based on a physical coordinate system (*e.g.* the intersection of I-495 and the Massachusetts Turnpike), a network coordinate system (*e.g.* the node at network coordinate 42, 344 in a Manhattan street network), or a relative coordinate system (*e.g.* within a building where absolute GPS-like coordinate systems are less useful). Diffusion routing [Intanagonwiwat 2000] is one technique that is being explored for location based contact.

2.3 Resource Characteristics Based

A final significant reason to communicate is based on the *characteristics* or capabilities of an entity, expressed as metadata $\langle c \rangle$. Examples of these characteristics are device capabilities (such as a color printer) or roles assumed (such as a DARPA program manager). As in the case of name-based contact, there may be a separate resource discovery step (*e.g.* Jini), or routing may be combined with a late binding of the resource characteristics (as in intentional naming [Adjie-Winoto 1998]).

While there has been early work in this area, more research is needed leading to solutions that can perform at high data rates and that scale to a large number of nodes and characteristics. The domain in which characteristics of a given node are known may be limited, as shown by the circular region $\langle c \rangle$ in Figure 1. Furthermore, there are limitations in considering the characteristics purely as an application overlay to be resolved to identifiers from a traditional single addressing, routing, and signaling environment. For example, a mesh of sensors may use characteristics in conjunction with topological coordinates and Cartesian routing rather than DNS resolution, IPv4 addressing, and BGP/OSPF routing. In order to provide the best possible topology and other network characteristics for a given task or application, it is desirable to be able to drive forwarding and routing based on an arbitrary projection on the set of characteristics.

2.4 Multi-Modal Networks

As just described, no single mode is appropriate for all types of communication, and thus at a high level, a network must be capable of supporting name, location, and characteristics contract modes. One way of doing this is to support various modal overlays on a network that supports only a single native mode. It therefore may be desirable to provide this multi-modal support *in the network*, that is in layer 3 protocols and network switching nodes. This is driven by efficiency and scalability concerns that will be expanded on in the next section.

Unless the modes form physically distinct networks, they must be supported simultaneously in network nodes. This is desirable since a combination of modes may be necessary to forward individual packets. For example, contacting a barometric sensor near N42°23.36′ W071°08.86′ requires both location-based and characteristics-based contact modes. Similarly, multiple modes may be used as a performance enhancement, for example contacting the node with characteristics $\langle c \rangle$ which is known to be in the area $x \pm \Delta x$, $y \pm \Delta y$.

We will thus assume that multi-modal routing is desirable as the basis of this white paper, and consider the routing and forwarding mechanisms and node architecture needed.

3. Routing and Forwarding

Before we consider how a node can support multiple modes, it is useful to consider the broad classes of functionality needed: when binding occurs, and whether routing is offline or on-demand.

3.1 Binding Time

The process of resolving a contact mode to a network address may occur initially before the communication begins. This is appropriate when there is a well-defined binding that is relatively stable, for example name resolution to a non-mobile endpoint (with fixed connections to static infrastructure). This is also appropriate when binding based on a set of characteristics that are likely to be needed in the future.

On the other hand, if the location of network nodes or endpoints is highly dynamic, or its characteristics are changing, *late binding* is desirable, which binds the contact mode to end node location as late as possible. Similarly, it is impractical to bind the all combinations of the entire set of possible characteristics when the attribute space is very large and only a small subset is of current interest. Late binding combines the resolution of the contact mode with packet traversal, and means that the network switching nodes must implement (or at least participate) in this incremental binding.

In the case of modes that contact endpoint with no network address (such as contacting a geographic location), no explicit binding occurs at all. The switching nodes must examine the packet destination and perform a computation that determines the outgoing link most likely to lead to the destination.

3.2 Forwarding Types

The *forwarding type* refers to the mechanism that determines how a packet is forwarded through a network node. It may or may not be based on IP; while this is an important issue, it orthogonal to the discussion in this white paper. *Routing* refers to the decision that determines the path topology.

In the case of connectionless communication, a packet filter or classifier determines the mode from each packet. In the case of flows with soft state or hard connections, the *per* flow state may contain the mode. Note that the use of state limits the range of forwarding modes that can be applied.

The extremes in forwarding types are offline (routing and forwarding asynchronous with on another) and on-demand (*per* packet routing decisions).

Offline routing

Traditional networks use *offline routing*, in which a routing algorithm run asynchronously to forwarding, and populates the forwarding tables that are used to either forward packets (connectionless) or to perform virtual connection path establishment. Multiple modes using offline routing requires switching between algorithms and forwarding tables [Partridge 2000]. Any contact mode that can be resolved to a network layer address is a candidate for offline routing.

On-demand routing

Contact modes that do not have a distinct initial binding to a network address require on-demand routing. A route computation is performed *per* packet that is used to determine how the packet will be forwarded through the switching node.

Candidates for on-demand routing include highly dynamic environments where late binding is desirable, as well as nodes with no addresses where binding to an address is not possible. On-demand routing may range from static algorithms permanently in the network node to active processing where algorithms can by dynamically installed and modified.

4. Node Architecture

The architecture of a switching node that supports multi-modal routing is shown in Figure 2, and consists of a set of input and output processing modules interconnected by a switch fabric. Each input processing module contains a packet classifier that is used to determine the mode.

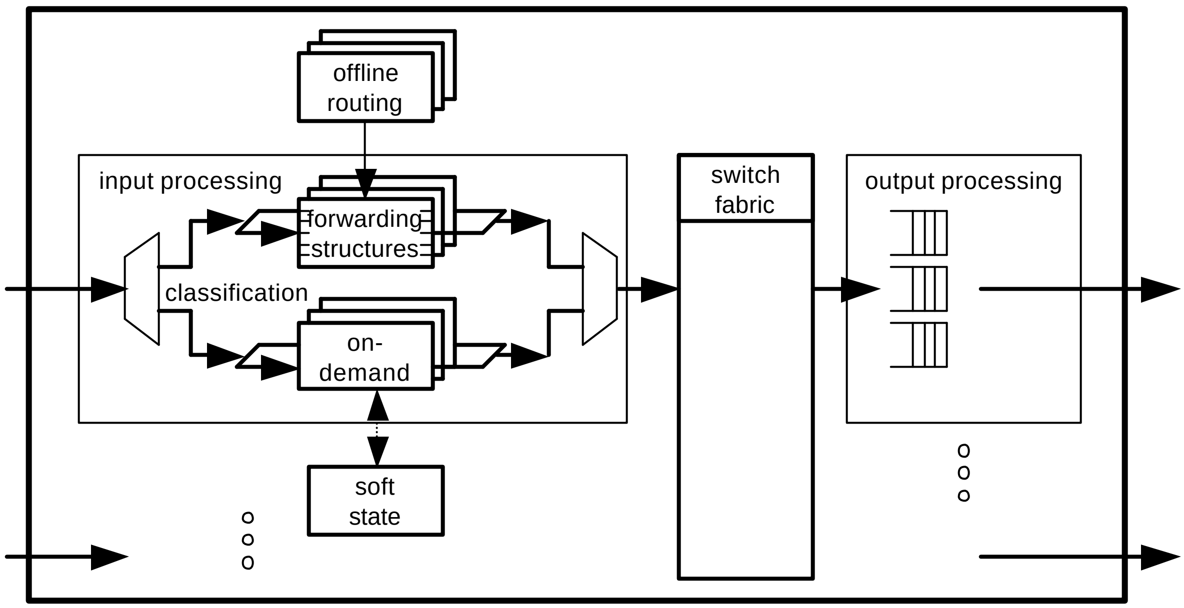

FIGURE 2. MULTI-MODAL NODE.

Based on the mode, a forwarding decision is made that determines the egress port that directs the packet toward its destination. In the case of a mode that uses on offline routing algorithm, the appropriate forwarding structure is used to look up the outgoing link. An individual node can contain multiple forwarding structures and routing algorithms.

In the case of on-demand routing, the packet is matched to the appropriate on-demand routing algorithm that uses information in the packet to compute the egress link. In the extreme case, this is done *per* packet. In some cases, aggregation *per* flow may be used to populate soft state that can accelerate the forwarding decision for subsequent packets.

The advances in processing capabilities that are enabling *per* class and *per* flow queueing, as well as sophisticated multidimensional packet classification, are the enablers that allow us to consider the implementation of a switch node that supports multi-modal addressing.

5. Conclusions

As the Internet increases in size, heterogeneity, and ubiquity of access, a number of long-held assumptions and characteristics are changing in fundamental ways. These include a dramatic increase in dynamic behaviour due to mobile nodes and the need to rethink the current address-based model of routing and forwarding. A variety of contact modes are the basis for communications, based on name, location, or characteristics. No one mode is capable of efficiently serving the needs of all application flows, and individual flows and packets may use multiple simultaneous contact modes. Substantial performance benefits may be obtained by providing native support in layer 3 protocols and switching nodes for multiple simultaneous modes.

While initial resolution of contact mode to endpoint address is adequate in some cases, late binding is indicated for dynamic networks and with large attribute spaces. This motivates direct support for on-demand routing in addition to the traditional separation of offline routing from forwarding. We propose a high-level node architecture capable of supporting multiple contact modes, and capable of both on-demand and offline routing. Continual increases in processing capability allows us to consider switching nodes that can perform the requisite packet classification and control-plane active processing at line rate, providing an enabler for further research

References

[Adjie-Winoto 1998] William Adjie-Winoto, Elliott Schwartz, Hari Balakrishnan, and Jeremy Lilley, "The Design and Implementation of an Intentional Naming System", *Proceedings Symposium on Operating Systems Principles (SOSP) '99*, *ACM Operating Systems Review*, vol.34 #5, December 1999, pp. 186–201, Dec. 1999.

[Cheriton 2000] David R Cheriton, and Mark Gritter, "TRIAD: A New Next-Generation Internet Architecture", Stanford Computer Science Techical Report, July 2000, available from `http://www.dsg.stanford.edu/triad`.

[Finn1998] Gregory G. Finn and Joseph D. Touch, "The Personal Node", *Proceedings Usenix Workshop on Embedded Systems*, (Cambridge MA US), Mar. 1999, Usenix, Berkely CA US, also ISI Research Report, ISI/RR-98-461, July, 1998.

[Intanagonwiwat 2000] Chalermek Intanagonwiwat, Ramesh Govindan, and Deborah Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks", *Proceedings of MobiCom 2000* (Boston MA US), pp. 56–67.

[Partridge 2000] Craig Partridge, Alex C. Snoeren, W. Timothy Strayer, Beverly Schwartz, Matthew Condell and Isidro Castineyra, "FIRE: Flexible Intra-AS Routing Environment", *Proceedings of SIGCOMM 2000* (Stockholm SE), *ACM SIGCOMM CCR*, vol. 30, #4, pp. 191–203.