Internet Draft D. Crocker
The Branch Office
R. Hinden
Sun Microsystems
November 11, 1992

IP Address Encapsulation (IPAE): A Mechanism for Introducing a New IP

STATUS OF THIS MEMO

This document is an Internet Draft. Internet Drafts are working documents of the Internet Engineering Task Force (IETF), its Areas, and its Working Groups. Note that other groups may also distribute working documents as Internet Drafts.

Internet Drafts are draft documents valid for a maximum of six months. Internet Drafts may be updated, replaced, or obsoleted by other documents at any time. It is not appropriate to use Internet Drafts as reference material or to cite them other than as a ``working draft" or ``work in progress."

Please check the 1id-abstracts.txt listing contained in the internet-drafts Shadow Directories on nic.ddn.mil, nnsc.nsf.net, nic.nordu.net, ftp.nisc.sri.com, or munnari.oz.auto to learn the current status of any Internet Draft.

SUMMARY

The Internet seeks to increase the amount of IP address space that is available for hosts and to decrease the amount of table storage that is required by routers. Ultimately, the current IP (IP version 4) and any replacement are inherently incompatible and movement to the new version requires very substantial change to the IP installed base. This specification describes an approach which retains as much software, operations and training as possible, and minimizes overall operational disruption by allowing subsets of the Internet to carry the new-format Internet datagrams inside old-style IPv4 datagrams, using a technique called "IP Address Encapsulation" (IPAE). This permits incremental and asynchronous introduction and makes transition entirely optional for portions of the Internet infrastructure. It further permits early reduction to routing table size.

This specification emphasizes extreme ease of transition to a new version of IP that is as close to the existing version as is appropriate. Hence, this specification is notable for attempting to fix only the current and urgent problems and for neither solving nor precluding solution to less well-understood problems. By utilizing a separately-defined protocol for the final deployment, IPAE can be viewed as a transition technology, rather than as a protocol, itself.

ACKNOWLEDGEMENTS

IPAE has had the benefit of on-going contribution by many members of the Internet community, with very substantial changes and enhancements ensuing from the collaborative group engineering process. Bob Hinden originally suggested encapsulating one IP datagram inside another [HIND92b]. Dave Crocker & Bob Hinden brought a modified idea -- to encapsulate new, global addresses inside normal IP datagrams, using the IP address fields for "next hop" carriage [HIND92a] -- to the IPAE working group.

While permitting very smooth transition, this left an end-state for IPAE which was not a very clean, since the IPAE header is a hybrid. Bob Gilligan observed similarities between IPAE's "mini-layer" and Deering's SIP design, resulting in IPAE focus on SIP. Craig Partridge developed a very early code analysis for implementing the original IPAE in the Berkeley kernel, permitting much more concrete understanding of the implementation impact of the encapsulation scheme and the general effects of increasing the IP address size. Recent implementation work by Erik Nordmark and Bob Gilligan has added to the base of empirical data.

The working group has included contributions from: Craig Partridge of BBN, Tom Kessler, Erik Nordmark, and Bob Gilligan of Sun Microsystems, Steve Deering of Xerox PARC, Greg Chesson, Ronald Jacoby,and and Rob Warnock of Silicon Graphics, Hon Wah Chin and Dave Newman of Protocol Engines, Ross Callon, Mike Reilly, Virgil Champlin and Geoff Mulligan of Digital Equipment, Michael Conn of MCI, John Moy of Proteon, John Veizades of Apple, Jeffrey Burgan of NASA, and Vince Fuller of BARRNET.

TABLE OF CONTENTS

1. PROBLEM STATEMENT: IP DEFICIENCIES

The Internet is experiencing explosive growth, usually described as doubling every 12 months. There is no indication that this growth will reduce and development of IP use into mass markets would create an even steeper growth curve. The result is a crisis in IP router table storage and use and in near-term exhaustion of available IP network numbers. A variety of extensions to current IP service also are desired, but they are not part of the current operational crisis.

IP routers which record routes to all networks in the Internet must maintain an entry for each such network, since the IP network address space is flat. That is, neighboring networks do not necessarily have any similarity in the IP network portion of their address. A new address structure is required, probably with a hierarchical scheme, so as to allow route-aggregation in table entries to neighboring networks.

While the 32-bits of the IP address space theoretically can reference about 4 billion nodes, the bits have been partitioned to faciliate assignment and aggregation into networks, thereby reducing the realistic maximum number of nodes and -- more importantly -- networks. While estimates vary considerably, it is possible that IP network number exhaustion will occur within the next few years. For example, IP could begin to make penetrations in mass markets. Since the damage caused by being unable to assign new IP network numbers is so great, it is prudent to pursue an urgent path to increasing the number of bits. Again, there is debate about the total number of bits that is necessary, with proposals ranging from 64-bits to 160-bits (as well as the suggestion that there be no limit.) In any event, the Internet needs to determine the size and format for a new, larger IP address.

Various additional enhancements to IP are being discussed, such as provision for real-time data delivery (i.e., guaranteed throughput and delay), network-level security features, and support for mobile hosts. While there is much discussion about these, and other, issues they represent topics of interest, rather than of concrete and stable experience and agreement. No proposals have achieved concensus in the Internet and none have acquired a substantial operational base. Hence, they must be deemed "research" for current purposes.

2. SOLUTION REQUIREMENTS

This section describes the the characteristics of the required solution that are held to be primary by this specification. While not attempting to exclude additional features and consideration, the specification does place particular emphasis on its concern for the installed base of IP users, while providing relief from the current technical and operational difficulties:

2.1. Address characteristics

The new version of IP must provide a hierarachical address, sufficient to permit distributed assignment and administration and sufficient to permit route-aggregation in IP forwarding devices (routers). A hierarchy which recognizes country boundaries appears to be the most viable, administratively. The address space must permit 10**9 networks.

2.2. Timing of benefits

The need for route aggregation is immediate. A specification must permit reductions in route table size in the near term. It is preferred that a router achieve this benefit as soon as the new version of IP is implemented in a router.

2.3. Preservation of Installed Base

A specification for a new IP must consider the current installed base of software and people. IP has an enormous installed base of operational systems, as well as an installed base of people who provide, support and use such systems. Each change to the existing technology carries the expense of changes to people, software and operations. Given the size of the installed base, the aggregate expense of each such change can be quite large, particularly if it causes reduced utility to the Internet.

To this end, a specification must justify each and every change to formats, terminology, services, software and operations, with respect to the current IP Internet. This applies to IP, itself, as well as to related components, such as the addressing framework, internetwork control, subnetwork interfaces and transport-layer interfaces.

2.4. Transition Ease & Independence

Movement from the existing IP to its replacement will be accomplished over an extended period of time. It must not require a highly coordinated change throughout the Internet community, since the diversity of network administrations and operations does not permit such coordination. Further, movement to the new IP must be accomplished in a manner which permits network-layer interoperation between users of the existing and new versions, for as long as is operationally practical. Hence, this specification emphasizes smooth and voluntary transitions, attempting to provide incremental benefit when adopted, and imposes no changes that are not of fundamental and well-understood benefit.

3. IP ADDRESS ENCAPSULATION APPROACH

This specification seeks to solve the technical problems of immediate concern to the Internet and does not attempt to provide other, desired functional and operational enhancements. Neither does it appear to preclude such enhancements and there is some indication that its design will prove entirely adequate for later inclusion of such enhancements.

While a variety of new protocols may satisfy the urgent addressing and routing table concerns for the Internet, this specifications views concerns for protecting the installed base as requiring the new protocol to have as little difference from current IP as possible. Consequently, this specification focusses on Simple IP (SIP), by Steve Deering, as the appropriate choice for the new IP [DEER92a, DEER92b]. While the general IPAE transition scheme, using encapsulation of the new protocol inside the old, can be applied to other candidates, it appears that a number of transition benefits will not accrue with those candidates. In particular, the packet and address similarities to IPv4 greatly facilitate IPv4-SIP translation, which permits combinations of interoperability tht permit an extremely wide range of transition options.

At its simplest, IPAE envisions three phases to the adoption of a new IP:

```
IP   ->   IPAE   ->   SIP
```

Initially, all nodes use only the old IP. The goal is to operate all nodes only with SIP. While it is expected that this end-state will take a very long time to reach for the whole Internet, portions of the Internet may achieve "pure" SIP operation rather quickly. Between the current use of IP and the eventual use of SIP, there will be an extended period of partial adoption and therefore requiring techniques for smooth transition and interoperation. IPAE specifically attends to those requirements.

IPAE permits continued operation of IP-only hosts and IP-only routers. It further supports SIP-knowledgeable hosts and routers. A host may operate SIP-only, without any IP support, and/or may be able to encapsulate SIP within IP datagrams, in an IPAE mode. Routers which support IPAE generally will serve as "border" routers between IP and IPAE or SIP domains. The border routing function can operate at a logical routing level above IP, using IP as a type of subnetwork-layer service. A border router also can take native IP datagrams and turn them into IPAE hybrid datagrams, when the router is acting as an internetwork-layer gateway translation service.

3.1. Encapsulation

This specification defines a technique that is called "IP Address Encapsulation" (IPAE) because it moves the Internet over to a new version IP by allowing hosts and routers to upgrade, to the new addressing and datagram format, in a manner that is transparent to nodes which have not yet made the transition. The technique for accomplishing this is to encapsulate the new protocol inside the old. Hence, users of the enhanced addressing scheme may "tunnel" their datagrams over the existing IP infrastructure. Neighbors which have completed conversion to the new IP can interoperate without encapsulation, whenever that is convenient.

IPAE packets stacks as:

```
+-----------+
| Transport |
+-----------+
|    SIP    |
+-----------+
|    IP     |
+-----------+
|   Link    |
+-----------+
```

3.2. Address Format

Support of interoperation between nodes using the old and new versions of IP is best accomplished by defining the new IP address to contain old IP addresses in its lower 32-bits. This is accomplished by conceptually extending the network field or by adding one or more fields "above" the IP `Network` field. Preserving the definition of the existing 32-bits permits local operations to default the new bits, if they desire, and faciliates translation between old and new IP datagrams.

Further the address, itself, must contain specification of the IP version (old versus new) that is used by the referenced node. This feature eliminates the need for routers at the border of an administrative domain to maintain state information about their user nodes. Such reduced router and operations complexity greatly facilitates transition to SIP.

Any new address format will dictate a change to the Domain Name Service and to any protocol modules which are cognizant of IP addresses. While fundamental to the success of an Internet transition, the details of such changes are beyond the scope of this specification. Further, the details appear to be identical for any IP replacement that is envisioned.

3.3. Interoperation

This specification supports interoperation between hosts supporting only the old IP and those supporting only the new IP, through the use of a translation service. Translation usually is a hand-crafted and operated service, with limited ability to scale to large operations. However, this specification supports general-purpose network-level translation between IP and SIP, as desired. This is possible by ensuring that:

a) Host operating mode (old versus new IP) is detected easily,

b) Address formats map well and algorithmically, and

c) Header field semantics map well and algorithmically.

3.4. Translation

Translation can be provided by special routers, or by nodes themselves. The latter facility is expected to permit interoperation between hosts operating on the same network, particularly prior to conversion of the interior routers for a network. Hence, nodes operating with the new version of IP may have an implementation which treats the two versions of IP as entirely independent network modules. Alternatively, a node may operate as if all datagrams are coded as new IP datagrams and then treat translation to/from old IP as a special case.

Transmission of new IP datagrams can be direct in their native form, translated into an old IP format, or encapsulated with the new inside the old. The first form is, of course, preferred. But it is expected that few nodes will be able to utilize the new IP initially. Hence, the choice usually will be between translation and encapsulation. If the next SIP hop is the recipient (final) host, and that host supports only old IP, then the SIP datagram must be translated into the format of an old IP datagram. If the next SIP hop is an intermediate router and that router supports only old IP, then the SIP datagram needs to be encapsulated in an IP datagram.

Note: In the following examples, "IPa", "IPb", and “IPc” represent three, different IP headers, generated independently. For example, IPa might be generated by the originating host, but might be discarded when the datagram is translated into a "pure" SIP datagram. Later, a border router might need to have the datagram transit an IP network and would create a fresh IP header (IPb) derived from the SIP header.

3.4.1 Transit Net(s) Only

IPAE can be deployed without awareness of hosts:

```
Orig     Router       Router      Router       Router      Dest
IP       IPAE         IP          IPAE         IP          IP
|          |  |          |  |          |  |         |  |          |
-----------  -----------  -----------  -----------  -----------
  +------+     +------+     +------+     +------+     +------+
  | IPa  |     | SIP  |     | SIP  |     | IPa  |     | IPa  |
  +------+     +------+     +------+     +------+     +------+
               | IPb  |     | IPb  |
               +------+     +------+
```

This permits transit networks to employ IPAE for reduction of routing tables, without waiting for host-level deployment, since routing can be based on the structured SIP address rather than the flat IP address.

3.4.2 IPAE Host to IPAE Host

IPAE packets which move through the Internet between two IPAE hosts may undergro modification, such as:

```
Orig     Router       Router       Router      Router      Dest
IPAE     IP           IP           IPAE        IP          IPAE
|          |  |          |  |          |  |         |  |          |
-----------  -----------  -----------  -----------  -----------
  +------+     +------+     +------+     +------+     +------+
  | SIP  |     | SIP  |     | SIP  |     | SIP  |     | SIP  |
  +------+     +------+     +------+     +------+     +------+
  | IPa  |     | IPa  |     | IPa  |     | IPb  |     | IPb  |
  +------+     +------+     +------+     +------+     +------+
```

In this case, two IPAE hosts go through an Internet in which IPAE routers are rare. This IPAE router discards the IP header it receives (IPa) and generates a new one (IPb) just as it does with media frame headers, when relaying between local area networks.

3.4.3 IP Host to IPAE Host

Common to early stages of deployment, a pure-IP host can talk with an IPAE or pure-SIP host via an IP-accessible router:

```
Orig     Router      Router      Router      Router     Dest
IP       IP          IPAE        IP          IPAE       IPAE
|         |  |         |  |         |  |         |  |         |
-----------  -----------  -----------  -----------  -----------
  +------+     +------+     +------+     +------+     +------+
  | IPa  |     | IPa  |     | SIP  |     | SIP  |     | SIP  |
  +------+     +------+     +------+     +------+     +------+
                            | IPb  |     | IPb  |     | IPc  |
                            +------+     +------+     +------+
```

This is generally identical to the previous example, except that the first IPAE router also serves as an IP-to-IPAE translation gateway, deriving the necessary SIP header from the IP header.

3.4.4 IPAE Host to SIP Host

As localities begin to support "pure" SIP operation, the following scenario will occur:

```
Orig     Router      Router      Router      Router     Dest
IPAE     IP          IP          IPAE        SIP        SIP
|         |  |         |  |         |  |         |  |         |
-----------  -----------  -----------  -----------  -----------
  +------+     +------+     +------+     +------+     +------+
  | SIP  |     | SIP  |     | SIP  |     | SIP  |     | SIP  |
  +------+     +------+     +------+     +------+     +------+
  | IP   |     | IP   |     | IP   |
  +------+     +------+     +------+
```

In this case, the IP header simply is stripped from the datagram and the SIP portion continues on, to a pure-SIP host.

3.4.5 IP Host to SIP Host

Use of IPAE can be helpful even in only one router, to map between to internetworks which employ IP-only and SIP-only:

```
Orig     Router      Router      Router      Dest
IP       IP          IPAE        SIP         SIP
|         |  |         |  |         |  |         |
-----------  -----------  -----------  -----------
  +------+     +------+     +------+     +------+
  | IP   |     | IP   |     | SIP  |     | SIP  |
  +------+     +------+     +------+     +------+
```

A key point, here, is that such pure gatewaying is a straightforward side-effect of the IPAE principles, when coupled with similarity between IP and SIP semantics. The gateway router behaves identically with a configuration of IPAE border router which treats IP encapsulation as a side-effect and which supports translation of IP into SIP.

4. IPAE PROTOCOL COMPONENTS

The primary change in IPAE is to the Internet layer. Other layers must be changed whenever they use Internet addresses. In particular, some change is necessary at the transport and application layers.

The IPAE proposal introduces two new Internet layer packet formats that we refer to as SIP and IPAE. The SIP header is a revised version of the IPv4 header with larger addresses and streamlined functionallity. The IPAE header is composed of a SIP header encapsulated within an IPv4 header. The IPAE packet format is used to transit regions connected with IPv4 routers, while the SIP packet format is used between directly connected IPAE/SIP hosts and routers.

SIP is specified separately [DEER92a, DEER92b], but details are summarized in this specification, to facilitate discussion.

4.1. SIP Address Format

The the basic requirement driving the current effort is to provide enough additional bits to support global administration, with 2 or more levels of hierarchy above the existing IP address portion, and the top level referencing country.

Given the goal of facilitating transition to SIP, IPAE-related operation requires enhancements to the current IP address, beyond the base requirement for a larger address space. These requirements are:

a) A bit indicating whether the referenced host is using old or new IP; this eliminates the need for routers that provide translation gatewaying to maintain state tables about the capabilities of specific hosts; and

b) Preservation of the existing format for 32-bit IP addresses, within the new format, to faciliate local interoperation between old and new IP hosts.

The resulting SIP address specification is:

```
 1    15           32            16
+-+--------+--------+--------+--------+
| |Country+|Site / Subscriber|  Local |
|C|Provider|..........................|
| |/ Metro |         |   IP Address   |
+-+--------+--------+--------+--------+
 C  Region      Attachment      Local

 C (IP Compatibility) bit is 0 for IPAE/SIP destination, 1
   for IP destination

 Multicast is encoded by a special country code.
```

The C-bit is used to indicate the capabilities of the referenced node. If the bit is zero, then the node supports SIP, and possibly IPAE. (In reality, IPAE is viewed as a per-hop encapsulation technique, so that its use depends upon the characteristics of the next-hop node, rather than upon the end-system.)

The Region field is used for global routing and includes the top of the addressing hierarchy which is the country having assignment authority. Subordinate to this is specification either of the service provider for the referenced subscriber or the metropolitan area of the site's attachment to the Internet.

Service providers assign Subscriber IDs. Locations operating with coordinated metropolitan interchanges (MIX) may use Country+Metro code and then assign Site IDs.

The last 4 bytes of an address function as a IPv4 address. However, the actual boundary between IP address and Site/Subscriber information varies. In effect, Site/Subscriber is an extension to the network portion of the IP address.

Until IP network number space is exhausted, 32-bit IP addresses will be assigned according to current practise, so that providers (and MIX operations) will assign only the upper 32 bits of a SIP address. Later, they also will assign those bits currently used to specify the IP Network field.

Hence, the transition model permits an independent network site (an "administrative domain", or AD) to operate internally with 32-bit IP addresses and to pre-pend an additional 32 bits for global routing. If an AD network is multi-homed to the Internet, then the bottom-half of its two SIP addresses will be the same, during the transition period. After IP network number exhaustion, the current specification of SIP means that an AD may have a single "host" field (subnetwork and host) as currently used by IP, but different network addresses for each point of attachment and one more for IP internal exchanges.

4.2. Datagram Formats

IPAE/SIP hosts may transmit packets using any of the three packet formats: IPv4, IPAE, or SIP. IPAE/SIP hosts must be able to receive packets in all three packet formats.

4.2.1 SIP Format

The SIP header is a revision of IP version 4 header format SIP shares all of the datalink layer encapsulations that have been defined for IPv4. A new IP version number differentiates SIP from IP version 4. This is the format of the SIP header:

```
 0 3            15         23         31
+---+------------+---------+---------+ 0
|V=6|          RESERVED              |
+---+------------+---------+---------+ 4
|   DATA LENGTH  | HOP LMT | PAYLOAD |
+----------------+---------+---------+ 8
|                                    |
+      Global IP Source Address      +
|                                    |
+------------------------------------+ 12
|                                    |
+    Global IP Destination Address   +
|                                    |
+------------------------------------+ 16
|             Data                   |
|                                    |
```

4.2.2 IPAE Format

The IPAE packet format uses the SIP header encapsulated within an IPv4 header. This is the format of the IPAE header, with relevant portions of the IP header labeled.

```
             0 3   7       15         23        31
--------- +---+---+--------+---------+---------+  0
          |V=4|            |       Length      |
          +---+------------+-------------------+  4
          |                                    |
 Current  +-------+--------+-------------------+  8
   IP     |  TTL  |PTCL=SIP|                   |
  Header  +-------+--------+-------------------+ 12
          |        Source IP Address           |
          +------------------------------------+ 16
          |       Destination IP Address       |
          +------------------------------------+ 20
          |      IP Options (if present)       |
          |                                    |
--------- +---+------------+---------+---------+ 20+n0
          |V=6|          RESERVED              |
          +---+------------+---------+---------+ 24+n0
          |   DATA LENGTH  | HOP LMT | PAYLOAD |
 Extended +-------+--------+-------------------+ 28+nO
   IP     |                                    |
  Header  +      Global IP Source Address      +
          |                                    |
  (SIP)   +------------------------------------+ 36+nO
          |                                    |
          +    Global IP Destination Address   +
          |                                    |
 -------- +------------------------------------+ 44+nO
          |             Data                   |
          |                                    |

       Note:   nO is size of IP options in octets.
```

When the IPAE format is used, the destination and source IP addresses located in the IPv4 header target the next hop IPAE/SIP system along the path to the destination. In both the SIP and IPAE formats, the global IP source and destination addresses identify the end systems. It is the global source and destination addresses located in the SIP header that are used by UDP and TCP to uniquely identify transport endpoints.

Most of the features of IPv4, including options and fragmentation and reassembly, are available in SIP. Thus it is possible to map an IPv4 packet into an IPAE or SIP packet and vice versa. Options are carried in SIP by using defined `Payload` values and the last SIP Option header contains the actual transport Payload value. When Options are present, the Global destination address references the host that is to process the option.

4.3. ICMP & IGMP

The SIP specification defines any required ICMP and IGMP changes or extensions. Use of IPAE adds no further requirements, except for relaying ICMP messages from IP-only routers to IPAE/SIP source hosts.

An IPAE datagram that traverses an IP network may result in having an IP-only interior router generate an ICMP error message. The router will specify a destination for the ICMP datagram which is, in reality, the previous IPAE border router, rather than the originating IPAE (or SIP) host. The border router must function as an ICMP relaying service, when possible, as discussed in the section on Border Routers, below.

4.4. Routing Protocol(s)

IPAE uses SIP and IP routing mechanisms. An administrative domain is free to use any acceptible IP routing protocol, among is interior routers, with an appropriate rule for deriving 32-bit addresses from the larger SIP addresses.

4.5. Transport & Above

IPAE imposes no special requirements on applications or transport, except for pseudo-header checksum calculation. Any development of a larger IP address will directly affect programming interfaces and some application protocols, since they use current IP addresses directly. Primarily, enhancements appears to be straightforward and simply need to handle the larger string, often simply as an uninterpreted string.

A node may implement IPAE (i.e., SIP over IP) as a special network-level interface or may make it equivalent to a subnetwork-/link-level option for SIP. In the former case, the transport layer needs to select IPAE from the set of alternatives, including IP and SIP. In the latter case, the transport layer selects a generic internet layer which, in turn, chooses IP, SIP or IPAE to get to the next (or final) SIP node.

TCP and UDP must be modified somewhat in order to be layered above IPAE and SIP. Changes are needed to the "pseudo-header" used in the checksum algorithm, and to TCP's connection or socket identification algorithm.

4.5.1 Pseudo header checksum

The TCP spec [RFC793] and the UDP spec [RFC768] both define a checksum that covers the data portion of the segment along with a 96-bit "pseudo header" that includes the IP source and destination addresses, protocol ID and length fields from the IP header. Including this pseudo header in the transport checksum protects the transport layer against misrouted segments.

The pseudo header used in the transport checksum when TCP and UDP are layered above IP can viewed logically as this:

```
 0       7        15                 31
+--------+--------+--------+--------+
|          source address           |
+--------+--------+--------+--------+
|        destination address        |
+--------+--------+--------+--------+
|  zero  |protocol|       length    |
+--------+--------+--------+--------+
```

Inclusion of a pseudo header that covers the addresses in the checksum of TCP and UDP layered above IPAE and SIP is even more important since the SIP header is not checksummed.

When communicating with other IPv4/IPAE/SIP hosts, the TCP and UDP pseudo header includes the complete global source and destination addresses. But since the IPAE mechanism allows IPv4 packets to be translated into IPAE or SIP and vice-versa, the IPv4/IPAE/SIP host must implement a compatible pseudo-header checksum when the packets it receives originated from, or the packets it is sending are destined for, an IPv4 host. The IPv4/IPAE/SIP pseudo header checksum algorithm must cover only the low-order 32 bits of the global addresses when it is communicating with an IPv4 host. When transmitting, the IPv4/IPAE/SIP host can use the C-bit of the global destination address to determine whether the peer is an IPv4 host. When receiving, the C-bit of the global source address can be used.

When communicating with an IPv4 host using the IPv4 packet format directly (e.g., an IPv4 host on the same subnet), the 96 bit pseudo header shown above is used.

When sending or receiving packets in the IPAE or SIP format to another IPv4/IPAE/SIP host the following pseudo header is used when the processing node is:

1) Transmitting and C-bit of Global Destination Address is 0, or

2) Receiving and C-bit of Global Source Address is 0

```
 0       7        15                 31
+--------+--------+--------+--------+
|  zero  |protocol|       length    |
+--------+--------+--------+--------+
|                                   |
+    Global source address          +
|                                   |
+--------+--------+--------+--------+
|                                   |
+    Global destination address     +
|                                   |
+--------+--------+--------+--------+
```

IPAE/SIP peer pseudo header

When sending or receiving packets in the IPAE or SIP format to an IPv4 host the following pseudo header is used when the processing node is:

```
1)   Transmitting and C-bit of Global Destination Address is 1, or

2)   Receiving and C-bit of Global Destination Address is 1.
```

```
           0      7       15               31
          +--------+--------+--------+--------+
          |  zero  |protocol|      length     |
          +--------+--------+--------+--------+
          |               zero                |
          +-----------------------------------+
          |   32-LSB of global source addr    |
          +--------+--------+--------+--------+
          |               zero                |
          +-----------------------------------+
          | 32-LSB of global destination addr |
          +--------+--------+--------+--------+

                IPAE/SIP - IPv4 compatible
                     pseudo-header
```

4.5.2 TCP Connection ID

TCP uses the concatenation of local and remote IP address with local and remote port number to uniquely identify a connection. TCP uses the term "socket" to identify one endpoint of a connection. The local socket is identified by the local IP address and local port number, while the remote socket is identified by the remote IP address and remote port number.

In processing received segments and ICMP error messages, TCP must use the destination IP address and port number -- and possibly the source IP address and port number -- from the received segment in a lookup in its list of connection blocks in order to find a matching socket or connection. When communicating with another IPv4/IPAE/SIP host, TCP host must use the full global source and destination addresses to identify connections and sockets. But when communicating with IPv4 hosts using the IPAE or SIP packet formats, TCP must use only the low order 32 bits of global source and destination to identify the connection.

This requires a change to TCP's connection block lookup for received segments. Assuming that TCP keeps a single list of connection blocks identifying connections to both IPv4 hosts and IPv4/IPAE/SIP hosts, the change can be summarized like this:

```
1)   If the format of the received packet is IPv4,
     Then:
        use the source and destination IP addresses in the
        received packet to compare with the low-order 32 bits of
        the global source and destination address in TCP's
        connection block list,
     Else:
```

```
2)  If the format of the received packet is IPAE or SIP,
    And:

    2a)  If the C-bit of the global source address is 1,
         Then:
            use the low-order 32 bits of the global source and
            destination addresses in the packet to compare the
            low-order 32 bits of the global source and
            destination address in the connection block list,
         Else:

    2b)  If the C-bit of the global source address is 0,
         Then:
            use the full global source and destination addresses
            in the packet to compare with the full global source
            and destination addresses in the connection block
            list.
```

4.6. Subnetwork & Below

IPAE, as SIP-over-IP, employs standard IP-over-media capabilities. SIP-over-media will use the same mechanisms as IP currently uses. Hence, there are no changes to use of ARP, or the like. Note that host references internal to an administrative domain will require 32-bits or less, as they do now, so that the upper 32-bits of SIP address space are not required for subnetwork references.

4.7. Network Management

Management within an AD can continue to operate, without change, since nodes retain their IP addresses. For global management, MIB specifications must be enhanced to accomodate the larger addresses. There is no expectation of additional protocol changes.

5. IPAE HEADER FIELD MAPPINGS

By keeping the semantic details of the new IP and the old IP closely related, it is possible to map between the header fields of either, greatly facilitating support of internetwork-level translation gateways. Three cases need to be supported: Turning an IP datagram into a SIP or IPAE (SIP-over-IP) datagram, generation of an IP header from a SIP header, and general handling of a received IPAE (SIP-over-IP) datagram.

5.1. SIP Derived from IP Datagrams

If a router supports IPAE for hosts that use only IP, then the router is a border IPAE router that also provides internetwork-level translation gatewaying. That is, it turns IP datagrams into SIP datagrams. It may also then send the SIP datagram onward, within an IP encapsulation, but this is a separate function from gateway translation.

On receipt of an IP datagram that is determined to be destined for a SIP host or that is otherwise in need of a SIP header, the gateway module will map IP fields into SIP fields as follows:

`Hop Limit:` The value of the IP `Time To Live` field shall be copied into SIP `Hop Limit` field. (This presumes that the router has already performed its own decrement to the IP `TTL` field.) Though `TTL` has slightly different semantics, there is no need to perform a more complicated translation.

`Payload:` The value of the IP `Protocol` field shall be copied into the SIP `Payload` field.

`Source Address:` The value of the `Source IP Address` shall be mapped to a SIP `Global Source Address`, as appropriate. Usually, this will require pre-pending a constant value to the Source's IP address.

`Destination Address:` The value of the `Source IP Address` shall be mapped to the SIP `Global Destination Address`, as appropriate. Border routers will support a table for such mappings. While IP addresses remain globally unique, this table can be maintained through the Domain Name Service or other coordinated Internet service.

`Options:` If the IP datagram contains options, then each shall be mapped to the appropriate SIP option, as appropriate. Options which do not map are dropped. Note that creation of SIP options alters the value of the SIP `Payload` value, placing the actual value into the `Payload` field of the SIP Option mini-layer.

Other fields: All remaining IP fields are ignored.

5.2. IP Derived from SIP Datagrams

When a border router needs to create an IP header, either for the purpose of IPAE encapsulation to achieve transit through an IP-only domain, or to translate the SIP header for receipt by an IP-only host, the router shall perform the following mappings:

`TTL:` The value of the SIP `Hop Limit` field shall be copied into the IP `Time To Live` field. (This presumes that the router has already performed its own decrement to the SIP `Hop Limit` field.)

`Protocol:` The value of the SIP `Payload` field shall be copied into the IP `Protocol` field.

`Source Address:` The value of the SIP `Global Source Address` shall be mapped to the IP address of the source, if the router has determined that the datagram is traversing its final IPAE hop, that is, if it being delivered to the recipient host, or if it is being fully converted to an IP-only datagram. While IP addresses remain globally unique, the value of the IP `Source Address` field is the value of the lower 32-bits of the SIP Global `Source Address`.

If the IP header is to serve an encapsulation function, with the SIP header being retained, then the `Source Address` shall contain the IP address of the border router that is creating the IP header.

`Destination Address:` The value of the SIP`Global Destination Address` shall be mapped to the IP address of the destination, if the router has determined that the datagram is traversing its final IPAE hop, that is, if it is being delivered to the recipeint host, or if it is being fully converted to an IP-only datagram. While IP addresses remain globally unique, the value of the IP `Destination Address` field is the value of the lower 32-bits of the SIP Global `Destination Address`.

If the IP headers is to serve an encapsulation function, with the SIP header being retained, then the `Destination Address` shall contain the IP address of the next IPAE-knowledgeable hop.

5.3. Receipt of IPAE Datagrams

An IPAE datagram contains a SIP header and an IP header. The IP header is for the purpose of permitting transit of the SIP datagram through IP-only networks. Hence, the IP header should not be viewed as containing the end-to-end information. However, the similarity between SIP/IP relationship and IP/Link relationship has one significant difference: The SIP header needs to reflect changes to the IP header that occur during transit.

In particular, an IPAE router needs to update the SIP header fields in the following manner:

`Hop Limit:` The value of the IP Time to Live field is copied into the SIP `Hop Limit` field. Hence, the SIP `Hop Limit` count must be adequate to count both SIP-knowledgeable and IP-only hops.

Other fields: Values in all other IP fields may be ignored.

6. IPAE NETWORK COMPONENTS

6.1. Hosts

An IPAE host supports IP, SIP and SIP-over-IP. An originating IPAE host must be able to derive an IP header from the SIP header, in order to create the SIP/IP encapsulation.

6.2. Interior Routers

For IPAE operation, it is presumed that the interior routers of an administrative domain are not IPAE-aware and hence need no modification, since they will transmit the IPAE datagram on the basis of the IP encapsulating datagram. However, administrative domains are expected to convert their routers over to use of IPAE and/or SIP, eventually, in which case their behavior for doing IP/SIP translation, address mapping, and the like will be similar to that of a border router.

6.3. Border Routers

An IPAE/SIP border router performs the following basic functions:

a) Typical router store-and-forward relaying, for SIP datagrams,

b) Participation in local IP infrastuctures, sufficiently to appear to other IP routers as if the border router is also an IP router,

c) IP-base encapsulation/decapsulation of SIP datagrams, and

d) Translation gateway creation of a SIP header based on the contents of an IP header, and vice-versa.

SIP routing functions are discussed in [DEER92b].

Encapsulation/decapsulation functions are to be performed using the derivation rules specified in the previous section.

It is expected that conversion of IP datagrams to SIP or IPAE datagrams will permit the converting border router in an AD to maintain smaller routing tables, immediately. For IP, router tables include a source information base with a copy of the information about all Internet nets, for each of the router's neighbors. For SIP, the table still needs to maintain a copy of the information received from each neighbor, but it can be reduced to a portion of the global hierarchy, such as all countries, as well as all "local" networks of interest.

Currently for IPv4, the real-time information base used to make direct datagram forwarding decisions needs to have an entry for each network on the Internet. For SIP, it needs to have only the relevant portions of the hierarchy, so that the table information about Internet neighbors can share the same entry.

7. IPAE ADDRESSING EXAMPLE

The relationship between IP and SIP addressing, particularly when both are present in an IPAE datagram, appears to be the greatest source of confusion concerning IPAE dynamics. This section provides an extended example of the end-to-end handling of both types of addresses as a datagram moves through the Internet.

Notation: IP addresses are represented in the usual, four decimal fields, separated by periods. SIP addresses also use a decimal dotted notation, showing Country.Metro/Provider.Site/Subscriber, followed by the site's IP address. The SIP portion is separated from the IP portion by a slash. Reference to IP or SIP networks, without referencing specific nodes on those networks, uses only the network portion of the address. For example, a Class C IP network would be shown as the first 3 octets of its address, such as 192.3.4. A SIP network address would show only the first 4 octets of the 64-bit address, indicating only the Region and Metro/Provider.

IPAE forms a set of logical networks "above" the set of IP networks, so that one or more IP networks are the equivalent of IPAE subnetworks. Hence, IPAE network boundaries occur only at IPAE routers. In this example, note that the last system of the first line (Router IPAEx) is the same as the first system on the second line.

This shows one intermediate IPAE hop, with an intermediate IP hop between the originating host and the IPAE router, and another between the IPAE router and the destination host.

The transit sequence for the datagram is:

```
-----------------------------------------------------------------

SIP Net                         <- 1.213.77. ->              ||
IP Net                     <- 192.3.4. |   <- 128.1. ->      ||
                                       |                     ||
Node:                H(IP-O)          R(IPa)           R(IPAEx)...
                      |1:213:77:5/     |  |  1:213:77:5/|
                      |192.3.4.1       |  |    128.1.1.2|
                      |                |  |             |
                      |      192.3.4.2 |  |128.1.1.1    |
                      |                |  |             |
                      -------->--------   ------>--------

Datagram:             SIP                 SIP
                      SRC:1:213:77:5:     SRC:1:213:77:5:
                          192.3.4.1           192.3.4.1
                      DST:1:2:2:8:        DST:1:2:2:8:
                          201.5.7.250         201.5.7.250
                      -----------------   -----------------
                      IP                  IP
                      SRC: 192.3.4.1      SRC:192.3.4.1
                      DST: 128.1.1.2      DST:128.1.1.2

-----------------------------------------------------------------
```

```
-----------------------------------------------------------------
SIP Net               ||         <- 1.002.2. ->
IP Net                ||        <- 10.  | 201.5.7. ->
                      ||                |
Node:            ...R(IPAEx)        R(IPb)         H(IPAE-D)
                      |1.2.2.75:       |  |     1.2.2.8:|
                      |10.0.0.3        |  |  201.5.7.250|
                      |                |  |             |
                      |       10.1.2.3 |  |201.5.7.9    |
                      |                |  |             |
                      -------->--------   ------->-------

Datagram:             SIP                 SIP
                      SRC:1:213:77:5:     SRC:1:213:77:5:
                          192.3.4.1           192.3.4.1
                      DST:1:2:2:8:        DST:1:2:2:8:
                          201.5.7.250         201.5.7.250
                      ----------------    ----------------
                      IP                  IP
                      SRC:10.0.0.3        SRC:10.0.0.3
                      DST:201.5.7.250     DST:201.5.7.250
-----------------------------------------------------------------
```

The pattern of manipulation to observe is that the SIP source and destination address fiels are unchanged. The IP source and destination addresses show the addresses of the IPAE nodes for the specific SIP hop and are unchanged when passing through an IP-only router.

8. TRANSITION SEQUENCE

The transition plan for IPAE is based around a timeline which has a number of milestones. The timeline is as follows:

```
|                                                          |
|----------|---------|----------|-----------|--------------|
|                                                          |

TODAY  ->  I         H          R           S    ->   FUTURE
```

Milestone I is the initial deployment of IPAE in selected Border Routers. Milestone H is large scale deployment of IPAE in Hosts. Milestone R is when the Internet runs out of the current 32-bit IP addresses. Milestone S indicates large scale conversion of administrative domains to pure SIP operations. Each milestone in the transition plan is discussed in the following sections.

The first objective of the transition plan is to install IPAE in critical router locations. This will reduce the size of routing tables and the load of routing computation in the Internet. The authors of this proposal believe this is very important to keep the Internet growing. It will be done without any changes to hosts.

The second objective is to install IPAE in hosts before the current 32-bit IP network addresses are all allocated. This will permit an orderly transition to a global routing and addressing scheme with a minimum of disruption to the existing internet.

As an administrative domain comes to use IPAE universally, then it may decide to permit pure SIP exchanges, without IPAE encapsulation.

8.1. Initial Deployment of IPAE (Milestone I)

The first step of IPAE deployment is to define the first SIP administrative domains. Initially these should be large areas of the Internet, such as portions of a single continent. The goal is to reduce the amount of routing traffic and the size of routing tables by an order of magnitude or more.

The first deployment of IPAE will occur in border routers. The border routers will create the first SIP administrative domains. However, at that time the current 32-bit IP addresses will still be globally unique. This will allow the border routers to look up the global IP addresses from the 32-bit IP addresses and build packets with extended IP headers for the hosts. The hosts will not have to implement IPAE in the first stage of deployment. The benefit of this stage of the deployment will be to reduce the size of routing tables in border routers and to reduce the routing computation load on these routers. The functions performed by the border routers are primarily of inserting/removing SIP headers and looking up Global IP addresses based 32-bit addresses.

Several approaches are possible for the procedure to lookup the Destination Global IP Address. One approach is a static table kept in the border routers. The mappings in this table would be based on the number of networks in the Internet. While this would be a large table towards the time the existing 32-bit IP Address space runs out, even then it would be possible to maintain this table within the small set of border routers.

Another approach is to use the Internet's Domain Name System to maintain this mapping table. The DNS is well-suited to this task and it would be a straightforward extension to add the information required. This approach has the border routers periodically performa a DNS lookup to obtain recent additions to the routing table. (That is, the DNS would provide a means of maintained a distributed copy of the Internet address mappings.) This would require careful engineering of the border router/DNS interactions to prevent their becoming a bottleneck. It is expected that the size of the resulting cache in border routers will be much smaller in size than with current operations because most Internet traffic tends to stay in its own IPAC and only a small percentage of total networks will pass through a border router.

It should be noted that any requirements for address mapping, whether static or via the DNS, are in no way specific to the IPAE proposal. Any scheme which wishes to provide communication among old-style hosts and new-style hosts having a different address format will require some type of address translation. IPAE simply makes the mapping process easier than would be required if the new address format were entirely unrelated to old-style IP addresses.

8.2. IPAE Deployment in Hosts (Milestone H)

The second phase of IPAE deployment is deployment to Internet hosts. Hosts will need to be able to support both IPv4 Headers format and IPAE/SIP Header format. They will use IPv4 Headers when communicating within their administrative domain and will use SIP Headers when communicating with hosts in different domains.

The goal for this phase of IPAE deployment it to have the majority of Internet hosts implement IPAE before the Internet runs out of 32-bit IP addresses. While it is not possible to have every host implement IPAE, there is sufficient time before Milestone R occurs that it can be implemented by the vast majority, since the amount of new software, documentation and training required will be quite small. This will reduce the need for address translation support in the next phase of deployment.

8.3. Internet Runs Out of 32-Bit IPv4 Network Numbers (Milestone R)

At the third phase of the IPAE deployment, hosts which do not implement IPAE will be able to communicate directly only with hosts in their own administrative domain, via IP. There are several possible methods to extend their connectvity if it is necessary to provide these hosts with global Internet service. One straightforward approach is to stage their communication through a host which supports IPAE. This would permit services such as electronic mail and news to operate transparently. Even in today's Internet, mail service often operates in this fashion. Other services such as Telnet and FTP would require an application-level forwarding agent or double login.

Another approach is to perform automatic address mapping. Various schemes are possible to support this and IPAE imposes no special constraints on the choices.

8.4. Administrative Domains Fully Convert to SIP (Milestone S)

The final stage of IPAE deployment is the demise of IPAE usage within individual administrative domains. As neighboring domains support pure SIP, then the border routers between them can be configured to omit the IPAE SIP-over-IP encapsulation for inter-domain exchanges. It is important to note, however, that decisions to permit pure SIP operation are entirely at the discretion of the local administrations, rather than requiring larger, Internet coordination.

9. REFERENCES

[BRAD89a] Braden, R.T., RFC 1127: Perspective on the Host Requirements RFCs. 1989 October.

[BRAD89b] Braden, R.T., ed., RFC 1122: Requirements for Internet hosts - communication layers. 1989 October.

[BRAD89c] Braden, R.T.,ed., RFC 1123: Requirements for Internet hosts - application and support. 1989 October.

[DEER92a] Deering, S. Simple Internet Protocol (SIP) Specification. 1992 November.

[DEER92b] Deering, S. Simple Internet Protocol (SIP) Addressing and Routing. 1992 November.

[E.163] CCITT, Numbering Plan for the International Telephone Services.

[HIND92a] Hinden, B., "New Scheme for Internet Routing and Addressing (ENCAPS)", Email message to Big-Internet mailing list, March 16, 1992

[HIND92b] A Proposal for IP Address Encapsulation (IPAE): A Compatible Version of IP with Large Addresses; (draft, June 1992)

[WOOD92] Woodburn, R & D. Mills, "A Scheme for an Internet Encapsulation Protocol: Version 1", RFC 1241. July 1991.

10. CONTACTS

David H. Crocker <dcrocker@mordor.stanford.edu>
The Branch Office
675 Spruce Dr.
Sunnyvale, CA 94086
+1 408 246 8253

Robert Hinden <hinden@eng.sun.com>
Sun Microsystems, Inc.
MS MTV-44
2550 Garcia Avenue
Mountain View, CA 94043-1100
+1 415 336 2082