Requirements for IP Version 4 Routers

Status of this Memo

PREFACE

   This document is an updated version of RFC 1716, the historical
   Router Requirements document. That RFC preserved the significant
   work that went into the working group, but failed to adequately
   describe current technology for the IESG to consider it a current
   standard.

   The current editor had been asked to bring the document up to date,
   so that it is useful as a procurement specification and a guide to
   implementors. In this, he stands squarely on the shoulders of those
   who have gone before him, and depends largely on expert contributors
   for text. Any credit is theirs; the errors are his.

   The content and form of this document are due, in large part, to the
   working group's chair, and document's original editor and author:
   Philip Almquist. It is also largely due to the efforts of its
   previous editor, Frank Kastenholz. Without their efforts, this
   document would not exist.

Table of Contents

1. INTRODUCTION

   This memo replaces for RFC 1716, "Requirements for Internet Gateways"
   ([INTRO:1]).

   This memo defines and discusses requirements for devices that perform
   the network layer forwarding function of the Internet protocol suite.
   The Internet community usually refers to such devices as IP routers or
   simply routers; The OSI community refers to such devices as
   intermediate systems. Many older Internet documents refer to these
   devices as gateways, a name which more recently has largely passed out
   of favor to avoid confusion with application gateways.

   An IP router can be distinguished from other sorts of packet switching
   devices in that a router examines the IP protocol header as part of
   the switching process. It generally removes the Link Layer header a
   message was received with, modifies the IP header, and replaces the
   Link Layer header for retransmission.

The authors of this memo recognize, as should its readers, that many
routers support more than one protocol. Support for multiple protocol
suites will be required in increasingly large parts of the Internet in
the future. This memo, however, does not attempt to specify Internet
requirements for protocol suites other than TCP/IP.

This document enumerates standard protocols that a router connected to
the Internet must use, and it incorporates by reference the RFCs and
other documents describing the current specifications for these
protocols. It corrects errors in the referenced documents and adds
additional discussion and guidance for an implementor.

For each protocol, this memo also contains an explicit set of
requirements, recommendations, and options. The reader must
understand that the list of requirements in this memo is incomplete by
itself. The complete set of requirements for an Internet protocol
router is primarily defined in the standard protocol specification
documents, with the corrections, amendments, and supplements contained
in this memo.

This memo should be read in conjunction with the Requirements for
Internet Hosts RFCs ([INTRO:2] and [INTRO:3]). Internet hosts and
routers must both be capable of originating IP datagrams and receiving
IP datagrams destined for them. The major distinction between
Internet hosts and routers is that routers implement forwarding
algorithms, while Internet hosts do not require forwarding
capabilities. Any Internet host acting as a router must adhere to the
requirements contained in this memo.

The goal of open system interconnection dictates that routers must
function correctly as Internet hosts when necessary. To achieve this,
this memo provides guidelines for such instances. For simplification
and ease of document updates, this memo tries to avoid overlapping
discussions of host requirements with [INTRO:2] and [INTRO:3] and
incorporates the relevant requirements of those documents by
reference. In some cases the requirements stated in [INTRO:2] and
[INTRO:3] are superseded by this document.

A good-faith implementation of the protocols produced after careful
reading of the RFCs should differ from the requirements of this memo
in only minor ways. Producing such an implementation often requires
some interaction with the Internet technical community, and must
follow good communications software engineering practices. In many
cases, the requirements in this document are already stated or implied
in the standard protocol documents, so that their inclusion here is,
in a sense, redundant. They were included because some past
implementation has made the wrong choice, causing problems of
interoperability, performance, and/or robustness.

This memo includes discussion and explanation of many of the
requirements and recommendations. A simple list of requirements would
be dangerous, because:

o Some required features are more important than others, and some
   features are optional.

o Some features are critical in some applications of routers but
   irrelevant in others.

o There may be valid reasons why particular vendor products that are
   designed for restricted contexts might choose to use different
   specifications.

However, the specifications of this memo must be followed to meet the
general goal of arbitrary router interoperation across the diversity
and complexity of the Internet. Although most current implementations
fail to meet these requirements in various ways, some minor and some
major, this specification is the ideal towards which we need to move.

These requirements are based on the current level of Internet
architecture. This memo will be updated as required to provide
additional clarifications or to include additional information in
those areas in which specifications are still evolving.

1.1 Reading this Document

1.1.1 Organization

This memo emulates the layered organization used by [INTRO:2] and
[INTRO:3]. Thus, Chapter 2 describes the layers found in the Internet
architecture. Chapter 3 covers the Link Layer. Chapters 4 and 5 are
concerned with the Internet Layer protocols and forwarding algorithms.
Chapter 6 covers the Transport Layer. Upper layer protocols are
divided among Chapters 7, 8, and 9. Chapter 7 discusses the protocols
which routers use to exchange routing information with each other.
Chapter 8 discusses network management. Chapter 9 discusses other
upper layer protocols. The final chapter covers operations and
maintenance features. This organization was chosen for simplicity,
clarity, and consistency with the Host Requirements RFCs. Appendices
to this memo include a bibliography, a glossary, and some conjectures
about future directions of router standards.

In describing the requirements, we assume that an implementation
strictly mirrors the layering of the protocols. However, strict
layering is an imperfect model, both for the protocol suite and for
recommended implementation approaches. Protocols in different layers
interact in complex and sometimes subtle ways, and particular

functions often involve multiple layers. There are many design
choices in an implementation, many of which involve creative breaking
of strict layering. Every implementor is urged to read [INTRO:4] and
[INTRO:5].

Each major section of this memo is organized into the following
subsections:

(1) Introduction

(2) Protocol Walk-Through - considers the protocol specification
     documents section-by-section, correcting errors, stating
     requirements that may be ambiguous or ill-defined, and providing
     further clarification or explanation.

(3) Specific Issues - discusses protocol design and implementation
     issues that were not included in the walk-through.

Under many of the individual topics in this memo, there is
parenthetical material labeled DISCUSSION or IMPLEMENTATION. This
material is intended to give a justification, clarification or
explanation to the preceding requirements text. The implementation
material contains suggested approaches that an implementor may want to
consider. The DISCUSSION and IMPLEMENTATION sections are not part of
the standard.

1.1.2 Requirements

In this memo, the words that are used to define the significance of
each particular requirement are capitalized. These words are:

o MUST
   This word means that the item is an absolute requirement of the
   specification. Violation of such a requirement is a fundamental
   error; there is no case where it is justified.

o MUST IMPLEMENT
   This phrase means that this specification requires that the item be
   implemented, but does not require that it be enabled by default.

o MUST NOT
   This phrase means that the item is an absolute prohibition of the
   specification.

o SHOULD
   This word means that there may exist valid reasons in particular
   circumstances to ignore this item, but the full implications should
   be understood and the case carefully weighed before choosing a

      different course.

   o SHOULD IMPLEMENT
     This phrase is similar in meaning to SHOULD, but is used when we
     recommend that a particular feature be provided but does not
     necessarily recommend that it be enabled by default.

   o SHOULD NOT
     This phrase means that there may exist valid reasons in particular
     circumstances when the described behavior is acceptable or even
     useful. Even so, the full implications should be understood and
     the case carefully weighed before implementing any behavior
     described with this label.

   o MAY
     This word means that this item is truly optional. One vendor may
     choose to include the item because a particular marketplace
     requires it or because it enhances the product, for example;
     another vendor may omit the same item.

1.1.3 Compliance

   Some requirements are applicable to all routers. Other requirements
   are applicable only to those which implement particular features or
   protocols. In the following paragraphs, relevant refers to the union
   of the requirements applicable to all routers and the set of
   requirements applicable to a particular router because of the set of
   features and protocols it has implemented.

   Note that not all Relevant requirements are stated directly in this
   memo. Various parts of this memo incorporate by reference sections of
   the Host Requirements specification, [INTRO:2] and [INTRO:3]. For
   purposes of determining compliance with this memo, it does not matter
   whether a Relevant requirement is stated directly in this memo or
   merely incorporated by reference from one of those documents.

   An implementation is said to be conditionally compliant if it
   satisfies all the Relevant MUST, MUST IMPLEMENT, and MUST NOT
   requirements. An implementation is said to be unconditionally
   compliant if it is conditionally compliant and also satisfies all the
   Relevant SHOULD, SHOULD IMPLEMENT, and SHOULD NOT requirements. An
   implementation is not compliant if it is not conditionally compliant
   (i.e., it fails to satisfy one or more of the Relevant MUST, MUST
   IMPLEMENT, or MUST NOT requirements).

   This specification occasionally indicates that an implementation
   SHOULD implement a management variable, and that it SHOULD have a
   certain default value. An unconditionally compliant implementation

implements the default behavior, and if there are other implemented
behaviors implements the variable. A conditionally compliant
implementation clearly documents what the default setting of the
variable is or, in the absence of the implementation of a variable,
may be construed to be. An implementation that both fails to
implement the variable and chooses a different behavior is not
compliant.

For any of the SHOULD and SHOULD NOT requirements, a router may
provide a configuration option that will cause the router to act other
than as specified by the requirement. Having such a configuration
option does not void a router's claim to unconditional compliance if
the option has a default setting, and that setting causes the router
to operate in the required manner.

Likewise, routers may provide, except where explicitly prohibited by
this memo, options which cause them to violate MUST or MUST NOT
requirements. A router that provides such options is compliant
(either fully or conditionally) if and only if each such option has a
default setting that causes the router to conform to the requirements
of this memo. Please note that the authors of this memo, although
aware of market realities, strongly recommend against provision of
such options. Requirements are labeled MUST or MUST NOT because
experts in the field have judged them to be particularly important to
interoperability or proper functioning in the Internet. Vendors
should weigh carefully the customer support costs of providing options
that violate those rules.

Of course, this memo is not a complete specification of an IP router,
but rather is closer to what in the OSI world is called a profile.
For example, this memo requires that a number of protocols be
implemented. Although most of the contents of their protocol
specifications are not repeated in this memo, implementors are
nonetheless required to implement the protocols according to those
specifications.

1.2 Relationships to Other Standards

There are several reference documents of interest in checking the
status of protocol specifications and standardization:

   o INTERNET OFFICIAL PROTOCOL STANDARDS
     This document describes the Internet standards process and lists
     the standards status of the protocols. As of this writing, the
     current version of this document is STD 1, RFC 1780, [ARCH:7].
     This document is periodically re-issued. You should always
     consult an RFC repository and use the latest version of this
     document.

o Assigned Numbers
  This document lists the assigned values of the parameters used in
  the various protocols. For example, it lists IP protocol codes,
  TCP port numbers, Telnet Option Codes, ARP hardware types, and
  Terminal Type names. As of this writing, the current version of
  this document is STD 2, RFC 1700, [INTRO:7]. This document is
  periodically re-issued. You should always consult an RFC
  repository and use the latest version of this document.

o Host Requirements
  This pair of documents reviews the specifications that apply to
  hosts and supplies guidance and clarification for any
  ambiguities. Note that these requirements also apply to routers,
  except where otherwise specified in this memo. As of this
  writing, the current versions of these documents are RFC 1122 and
  RFC 1123 (STD 3), [INTRO:2] and [INTRO:3].

o Router Requirements (formerly Gateway Requirements)
  This memo.

Note that these documents are revised and updated at different times;
in case of differences between these documents, the most recent must
prevail.

These and other Internet protocol documents may be obtained from the:

                        The InterNIC
                      DS.INTERNIC.NET
            InterNIC Directory and Database Service
                    info@internic.net
                     +1-908-668-6587
                URL: http://ds.internic.net/

1.3 General Considerations

   There are several important lessons that vendors of Internet software
   have learned and which a new vendor should consider seriously.

1.3.1 Continuing Internet Evolution

   The enormous growth of the Internet has revealed problems of
   management and scaling in a large datagram based packet communication
   system. These problems are being addressed, and as a result there
   will be continuing evolution of the specifications described in this
   memo. New routing protocols, algorithms, and architectures are
   constantly being developed. New internet layer protocols, and
   modifications to existing protocols, are also constantly being
   devised. Routers play a crucial role in the Internet, and the number

of routers deployed in the Internet is much smaller than the number
of hosts. Vendors should therefore expect that router standards will
continue to evolve much more quickly than host standards. These
changes will be carefully planned and controlled since there is
extensive participation in this planning by the vendors and by the
organizations responsible for operation of the networks.

Development, evolution, and revision are characteristic of computer
network protocols today, and this situation will persist for some
years. A vendor who develops computer communications software for
the Internet protocol suite (or any other protocol suite!) and then
fails to maintain and update that software for changing
specifications is going to leave a trail of unhappy customers. The
Internet is a large communication network, and the users are in
constant contact through it. Experience has shown that knowledge of
deficiencies in vendor software propagates quickly through the
Internet technical community.

1.3.2 Robustness Principle

At every layer of the protocols, there is a general rule (from
[TRANS:2] by Jon Postel) whose application can lead to enormous
benefits in robustness and interoperability:

                    Be conservative in what you do,
              be liberal in what you accept from others.

Software should be written to deal with every conceivable error, no
matter how unlikely. Eventually a packet will come in with that
particular combination of errors and attributes, and unless the
software is prepared, chaos can ensue. It is best to assume that the
network is filled with malevolent entities that will send packets
designed to have the worst possible effect. This assumption will
lead to suitably protective design. The most serious problems in the
Internet have been caused by unforeseen mechanisms triggered by low
probability events; mere human malice would never have taken so
devious a course!

Adaptability to change must be designed into all levels of router
software. As a simple example, consider a protocol specification
that contains an enumeration of values for a particular header field
- e.g., a type field, a port number, or an error code; this
enumeration must be assumed to be incomplete. If the protocol
specification defines four possible error codes, the software must
not break when a fifth code is defined. An undefined code might be
logged, but it must not cause a failure.

The second part of the principal is almost as important: software on
hosts or other routers may contain deficiencies that make it unwise
to exploit legal but obscure protocol features. It is unwise to
stray far from the obvious and simple, lest untoward effects result
elsewhere. A corollary of this is watch out for misbehaving hosts;
router software should be prepared to survive in the presence of
misbehaving hosts. An important function of routers in the Internet
is to limit the amount of disruption such hosts can inflict on the
shared communication facility.

1.3.3 Error Logging

The Internet includes a great variety of systems, each implementing
many protocols and protocol layers, and some of these contain bugs
and misguided features in their Internet protocol software. As a
result of complexity, diversity, and distribution of function, the
diagnosis of problems is often very difficult.

Problem diagnosis will be aided if routers include a carefully
designed facility for logging erroneous or strange events. It is
important to include as much diagnostic information as possible when
an error is logged. In particular, it is often useful to record the
header(s) of a packet that caused an error. However, care must be
taken to ensure that error logging does not consume prohibitive
amounts of resources or otherwise interfere with the operation of the
router.

There is a tendency for abnormal but harmless protocol events to
overflow error logging files; this can be avoided by using a circular
log, or by enabling logging only while diagnosing a known failure.
It may be useful to filter and count duplicate successive messages.
One strategy that seems to work well is to both:

o Always count abnormalities and make such counts accessible through
   the management protocol (see Chapter 8); and
o Allow the logging of a great variety of events to be selectively
   enabled. For example, it might useful to be able to log
   everything or to log everything for host X.

This topic is further discussed in [MGT:5].

1.3.4 Configuration

In an ideal world, routers would be easy to configure, and perhaps
even entirely self-configuring. However, practical experience in the
real world suggests that this is an impossible goal, and that many
attempts by vendors to make configuration easy actually cause
customers more grief than they prevent. As an extreme example, a

router designed to come up and start routing packets without
requiring any configuration information at all would almost certainly
choose some incorrect parameter, possibly causing serious problems on
any networks unfortunate enough to be connected to it.

Often this memo requires that a parameter be a configurable option.
There are several reasons for this. In a few cases there currently
is some uncertainty or disagreement about the best value and it may
be necessary to update the recommended value in the future. In other
cases, the value really depends on external factors - e.g., the
distribution of its communication load, or the speeds and topology of
nearby networks - and self-tuning algorithms are unavailable and may
be insufficient. In some cases, configurability is needed because of
administrative requirements.

Finally, some configuration options are required to communicate with
obsolete or incorrect implementations of the protocols, distributed
without sources, that persist in many parts of the Internet. To make
correct systems coexist with these faulty systems, administrators
must occasionally misconfigure the correct systems. This problem
will correct itself gradually as the faulty systems are retired, but
cannot be ignored by vendors.

When we say that a parameter must be configurable, we do not intend
to require that its value be explicitly read from a configuration
file at every boot time. For many parameters, there is one value
that is appropriate for all but the most unusual situations. In such
cases, it is quite reasonable that the parameter default to that
value if not explicitly set.

This memo requires a particular value for such defaults in some
cases. The choice of default is a sensitive issue when the
configuration item controls accommodation of existing, faulty,
systems. If the Internet is to converge successfully to complete
interoperability, the default values built into implementations must
implement the official protocol, not misconfigurations to accommodate
faulty implementations. Although marketing considerations have led
some vendors to choose misconfiguration defaults, we urge vendors to
choose defaults that will conform to the standard.

Finally, we note that a vendor needs to provide adequate
documentation on all configuration parameters, their limits and
effects.

1.4 Algorithms

   In several places in this memo, specific algorithms that a router
   ought to follow are specified. These algorithms are not, per se,
   required of the router. A router need not implement each algorithm
   as it is written in this document. Rather, an implementation must
   present a behavior to the external world that is the same as a
   strict, literal, implementation of the specified algorithm.

   Algorithms are described in a manner that differs from the way a good
   implementor would implement them. For expository purposes, a style
   that emphasizes conciseness, clarity, and independence from
   implementation details has been chosen. A good implementor will
   choose algorithms and implementation methods that produce the same
   results as these algorithms, but may be more efficient or less
   general.

   We note that the art of efficient router implementation is outside
   the scope of this memo.

2. INTERNET ARCHITECTURE

   This chapter does not contain any requirements. However, it does
   contain useful background information on the general architecture of
   the Internet and of routers.

   General background and discussion on the Internet architecture and
   supporting protocol suite can be found in the DDN Protocol Handbook
   [ARCH:1]; for background see for example [ARCH:2], [ARCH:3], and
   [ARCH:4]. The Internet architecture and protocols are also covered
   in an ever-growing number of textbooks, such as [ARCH:5] and
   [ARCH:6].

2.1 Introduction

   The Internet system consists of a number of interconnected packet
   networks supporting communication among host computers using the
   Internet protocols. These protocols include the Internet Protocol
   (IP), the Internet Control Message Protocol (ICMP), the Internet
   Group Management Protocol (IGMP), and a variety transport and
   application protocols that depend upon them. As was described in
   Section [1.2], the Internet Engineering Steering Group periodically
   releases an Official Protocols memo listing all the Internet
   protocols.

   All Internet protocols use IP as the basic data transport mechanism.
   IP is a datagram, or connectionless, internetwork service and
   includes provision for addressing, type-of-service specification,

fragmentation and reassembly, and security. ICMP and IGMP are
considered integral parts of IP, although they are architecturally
layered upon IP. ICMP provides error reporting, flow control,
first-hop router redirection, and other maintenance and control
functions. IGMP provides the mechanisms by which hosts and routers
can join and leave IP multicast groups.

Reliable data delivery is provided in the Internet protocol suite by
Transport Layer protocols such as the Transmission Control Protocol
(TCP), which provides end-end retransmission, resequencing and
connection control. Transport Layer connectionless service is
provided by the User Datagram Protocol (UDP).

2.2 Elements of the Architecture

2.2.1 Protocol Layering

   To communicate using the Internet system, a host must implement the
   layered set of protocols comprising the Internet protocol suite. A
   host typically must implement at least one protocol from each layer.

   The protocol layers used in the Internet architecture are as follows
   [ARCH:7]:

   o Application Layer
     The Application Layer is the top layer of the Internet protocol
     suite. The Internet suite does not further subdivide the
     Application Layer, although some application layer protocols do
     contain some internal sub-layering. The application layer of the
     Internet suite essentially combines the functions of the top two
     layers - Presentation and Application - of the OSI Reference Model
     [ARCH:8]. The Application Layer in the Internet protocol suite
     also includes some of the function relegated to the Session Layer
     in the OSI Reference Model.

     We distinguish two categories of application layer protocols: user
     protocols that provide service directly to users, and support
     protocols that provide common system functions. The most common
     Internet user protocols are:

     - Telnet (remote login)
     - FTP (file transfer)
     - SMTP (electronic mail delivery)

     There are a number of other standardized user protocols and many
     private user protocols.

Support protocols, used for host name mapping, booting, and
management include SNMP, BOOTP, TFTP, the Domain Name System (DNS)
protocol, and a variety of routing protocols.

Application Layer protocols relevant to routers are discussed in
chapters 7, 8, and 9 of this memo.

o Transport Layer
  The Transport Layer provides end-to-end communication services.
  This layer is roughly equivalent to the Transport Layer in the OSI
  Reference Model, except that it also incorporates some of OSI's
  Session Layer establishment and destruction functions.

  There are two primary Transport Layer protocols at present:

  - Transmission Control Protocol (TCP)
  - User Datagram Protocol (UDP)

  TCP is a reliable connection-oriented transport service that
  provides end-to-end reliability, resequencing, and flow control.
  UDP is a connectionless (datagram) transport service. Other
  transport protocols have been developed by the research community,
  and the set of official Internet transport protocols may be
  expanded in the future.

  Transport Layer protocols relevant to routers are discussed in
  Chapter 6.

o Internet Layer
  All Internet transport protocols use the Internet Protocol (IP) to
  carry data from source host to destination host. IP is a
  connectionless or datagram internetwork service, providing no
  end-to-end delivery guarantees. IP datagrams may arrive at the
  destination host damaged, duplicated, out of order, or not at all.
  The layers above IP are responsible for reliable delivery service
  when it is required. The IP protocol includes provision for
  addressing, type-of-service specification, fragmentation and
  reassembly, and security.

  The datagram or connectionless nature of IP is a fundamental and
  characteristic feature of the Internet architecture.

  The Internet Control Message Protocol (ICMP) is a control protocol
  that is considered to be an integral part of IP, although it is
  architecturally layered upon IP - it uses IP to carry its data
  end-to-end. ICMP provides error reporting, congestion reporting,
  and first-hop router redirection.

The Internet Group Management Protocol (IGMP) is an Internet layer
protocol used for establishing dynamic host groups for IP
multicasting.

The Internet layer protocols IP, ICMP, and IGMP are discussed in
chapter 4.

   o Link Layer
     To communicate on a directly connected network, a host must
     implement the communication protocol used to interface to that
     network. We call this a Link Layer protocol.

     Some older Internet documents refer to this layer as the Network
     Layer, but it is not the same as the Network Layer in the OSI
     Reference Model.

     This layer contains everything below the Internet Layer and above
     the Physical Layer (which is the media connectivity, normally
     electrical or optical, which encodes and transports messages).
     Its responsibility is the correct delivery of messages, among
     which it does not differentiate.

     Protocols in this Layer are generally outside the scope of
     Internet standardization; the Internet (intentionally) uses
     existing standards whenever possible. Thus, Internet Link Layer
     standards usually address only address resolution and rules for
     transmitting IP packets over specific Link Layer protocols.
     Internet Link Layer standards are discussed in chapter 3.

2.2.2 Networks

   The constituent networks of the Internet system are required to
   provide only packet (connectionless) transport. According to the IP
   service specification, datagrams can be delivered out of order, be
   lost or duplicated, and/or contain errors.

   For reasonable performance of the protocols that use IP (e.g., TCP),
   the loss rate of the network should be very low. In networks
   providing connection-oriented service, the extra reliability provided
   by virtual circuits enhances the end-end robustness of the system,
   but is not necessary for Internet operation.

   Constituent networks may generally be divided into two classes:

      o Local-Area Networks (LANs)
        LANs may have a variety of designs. LANs normally cover a small
        geographical area (e.g., a single building or plant site) and
        provide high bandwidth with low delays. LANs may be passive

(similar to Ethernet) or they may be active (such as ATM).

o Wide-Area Networks (WANs)
Geographically dispersed hosts and LANs are interconnected by
wide-area networks, also called long-haul networks. These
networks may have a complex internal structure of lines and
packet-switches, or they may be as simple as point-to-point
lines.

2.2.3 Routers

In the Internet model, constituent networks are connected together by
IP datagram forwarders which are called routers or IP routers. In
this document, every use of the term router is equivalent to IP
router. Many older Internet documents refer to routers as gateways.

Historically, routers have been realized with packet-switching
software executing on a general-purpose CPU. However, as custom
hardware development becomes cheaper and as higher throughput is
required, special purpose hardware is becoming increasingly common.
This specification applies to routers regardless of how they are
implemented.

A router connects to two or more logical interfaces, represented by
IP subnets or unnumbered point to point lines (discussed in section
[2.2.7]). Thus, it has at least one physical interface. Forwarding
an IP datagram generally requires the router to choose the address
and relevant interface of the next-hop router or (for the final hop)
the destination host. This choice, called relaying or forwarding
depends upon a route database within the router. The route database
is also called a routing table or forwarding table. The term
"router" derives from the process of building this route database;
routing protocols and configuration interact in a process called
routing.

The routing database should be maintained dynamically to reflect the
current topology of the Internet system. A router normally
accomplishes this by participating in distributed routing and
reachability algorithms with other routers.

Routers provide datagram transport only, and they seek to minimize
the state information necessary to sustain this service in the
interest of routing flexibility and robustness.

Packet switching devices may also operate at the Link Layer; such
devices are usually called bridges. Network segments that are
connected by bridges share the same IP network prefix forming a
single IP subnet. These other devices are outside the scope of this

document.

2.2.4 Autonomous Systems

An Autonomous System (AS) is a connected segment of a network
topology that consists of a collection of subnetworks (with hosts
attached) interconnected by a set of routes. The subnetworks and the
routers are expected to be under the control of a single operations
and maintenance (O&M) organization. Within an AS routers may use one
or more interior routing protocols, and sometimes several sets of
metrics. An AS is expected to present to other ASs an appearence of
a coherent interior routing plan, and a consistent picture of the
destinations reachable through the AS. An AS is identified by an
Autonomous System number.

The concept of an AS plays an important role in the Internet routing
(see Section 7.1).

2.2.5 Addressing Architecture

An IP datagram carries 32-bit source and destination addresses, each
of which is partitioned into two parts - a constituent network prefix
and a host number on that network. Symbolically:

    IP-address ::= { <Network-prefix>, <Host-number> }

To finally deliver the datagram, the last router in its path must map
the Host-number (or rest) part of an IP address to the host's Link
Layer address.

2.2.5.1 Classical IP Addressing Architecture

Although well documented elsewhere [INTERNET:2], it is useful to
describe the historical use of the network prefix. The language
developed to describe it is used in this and other documents and
permeates the thinking behind many protocols.

The simplest classical network prefix is the Class A, B, C, D, or E
network prefix. These address ranges are discriminated by observing
the values of the most significant bits of the address, and break the
address into simple prefix and host number fields. This is described
in [INTERNET:18]. In short, the classification is:

    0xxx - Class A - general purpose unicast addresses with standard
    8 bit prefix
    10xx - Class B - general purpose unicast addresses with standard
    16 bit prefix

          110x - Class C - general purpose unicast addresses with standard
          24 bit prefix
          1110 - Class D - IP Multicast Addresses - 28 bit prefix, non-
          aggregatable
          1111 - Class E - reserved for experimental use

   This simple notion has been extended by the concept of subnets.
   These were introduced to allow arbitrary complexity of interconnected
   LAN structures within an organization, while insulating the Internet
   system against explosive growth in assigned network prefixes and
   routing complexity. Subnets provide a multi-level hierarchical
   routing structure for the Internet system. The subnet extension,
   described in [INTERNET:2], is a required part of the Internet
   architecture. The basic idea is to partition the <Host-number> field
   into two parts: a subnet number, and a true host number on that
   subnet:

       IP-address ::=
         { <Network-number>, <Subnet-number>, <Host-number> }

   The interconnected physical networks within an organization use the
   same network prefix but different subnet numbers. The distinction
   between the subnets of such a subnetted network is not normally
   visible outside of that network. Thus, routing in the rest of the
   Internet uses only the <Network-prefix> part of the IP destination
   address. Routers outside the network treat <Network-prefix> and
   <Host-number> together as an uninterpreted rest part of the 32-bit IP
   address. Within the subnetted network, the routers use the extended
   network prefix:

       { <Network-number>, <Subnet-number> }

   The bit positions containing this extended network number have
   historically been indicated by a 32-bit mask called the subnet mask.
   The <Subnet-number> bits SHOULD be contiguous and fall between the
   <Network-number> and the <Host-number> fields. More up to date
   protocols do not refer to a subnet mask, but to a prefix length; the
   "prefix" portion of an address is that which would be selected by a
   subnet mask whose most significant bits are all ones and the rest are
   zeroes. The length of the prefix equals the number of ones in the
   subnet mask. This document assumes that all subnet masks are
   expressible as prefix lengths.

   The inventors of the subnet mechanism presumed that each piece of an
   organization's network would have only a single subnet number. In
   practice, it has often proven necessary or useful to have several
   subnets share a single physical cable. For this reason, routers
   should be capable of configuring multiple subnets on the same

physical interfaces, and treat them (from a routing or forwarding
perspective) as though they were distinct physical interfaces.

2.2.5.2 Classless Inter Domain Routing (CIDR)

The explosive growth of the Internet has forced a review of address
assignment policies. The traditional uses of general purpose (Class
A, B, and C) networks have been modified to achieve better use of
IP's 32-bit address space. Classless Inter Domain Routing (CIDR)
[INTERNET:15] is a method currently being deployed in the Internet
backbones to achieve this added efficiency. CIDR depends on
deploying and routing to arbitrarily sized networks. In this model,
hosts and routers make no assumptions about the use of addressing in
the internet. The Class D (IP Multicast) and Class E (Experimental)
address spaces are preserved, although this is primarily an
assignment policy.

By definition, CIDR comprises three elements:

   o topologically significant address assignment,
   o routing protocols that are capable of aggregating network layer
      reachability information, and
   o consistent forwarding algorithm ("longest match").

The use of networks and subnets is now historical, although the
language used to describe them remains in current use. They have
been replaced by the more tractable concept of a network prefix. A
network prefix is, by definition, a contiguous set of bits at the
more significant end of the address that defines a set of systems;
host numbers select among those systems. There is no requirement
that all the internet use network prefixes uniformly. To collapse
routing information, it is useful to divide the internet into
addressing domains. Within such a domain, detailed information is
available about constituent networks; outside it, only the common
network prefix is advertised.

The classical IP addressing architecture used addresses and subnet
masks to discriminate the host number from the network prefix. With
network prefixes, it is sufficient to indicate the number of bits in
the prefix. Both representations are in common use. Architecturally
correct subnet masks are capable of being represented using the
prefix length description. They comprise that subset of all possible
bits patterns that have

   o a contiguous string of ones at the more significant end,
   o a contiguous string of zeros at the less significant end, and
   o no intervening bits.

Routers SHOULD always treat a route as a network prefix, and SHOULD reject configuration and routing information inconsistent with that model.

    IP-address ::= { <Network-prefix>, <Host-number> }

An effect of the use of CIDR is that the set of destinations associated with address prefixes in the routing table may exhibit subset relationship. A route describing a smaller set of destinations (a longer prefix) is said to be more specific than a route describing a larger set of destinations (a shorter prefix); similarly, a route describing a larger set of destinations (a shorter prefix) is said to be less specific than a route describing a smaller set of destinations (a longer prefix). Routers must use the most specific matching route (the longest matching network prefix) when forwarding traffic.

2.2.6 IP Multicasting

IP multicasting is an extension of Link Layer multicast to IP internets. Using IP multicasts, a single datagram can be addressed to multiple hosts without sending it to all. In the extended case, these hosts may reside in different address domains. This collection of hosts is called a multicast group. Each multicast group is represented as a Class D IP address. An IP datagram sent to the group is to be delivered to each group member with the same best-effort delivery as that provided for unicast IP traffic. The sender of the datagram does not itself need to be a member of the destination group.

The semantics of IP multicast group membership are defined in [INTERNET:4]. That document describes how hosts and routers join and leave multicast groups. It also defines a protocol, the Internet Group Management Protocol (IGMP), that monitors IP multicast group membership.

Forwarding of IP multicast datagrams is accomplished either through static routing information or via a multicast routing protocol. Devices that forward IP multicast datagrams are called multicast routers. They may or may not also forward IP unicasts. Multicast datagrams are forwarded on the basis of both their source and destination addresses. Forwarding of IP multicast packets is described in more detail in Section [5.2.1]. Appendix D discusses multicast routing protocols.

2.2.7 Unnumbered Lines and Networks Prefixes

   Traditionally, each network interface on an IP host or router has its
   own IP address. This can cause inefficient use of the scarce IP
   address space, since it forces allocation of an IP network prefix to
   every point-to-point link.

   To solve this problem, a number of people have proposed and
   implemented the concept of unnumbered point to point lines. An
   unnumbered point to point line does not have any network prefix
   associated with it. As a consequence, the network interfaces
   connected to an unnumbered point to point line do not have IP
   addresses.

   Because the IP architecture has traditionally assumed that all
   interfaces had IP addresses, these unnumbered interfaces cause some
   interesting dilemmas. For example, some IP options (e.g., Record
   Route) specify that a router must insert the interface address into
   the option, but an unnumbered interface has no IP address. Even more
   fundamental (as we shall see in chapter 5) is that routes contain the
   IP address of the next hop router. A router expects that this IP
   address will be on an IP (sub)net to which the router is connected.
   That assumption is of course violated if the only connection is an
   unnumbered point to point line.

   To get around these difficulties, two schemes have been conceived.
   The first scheme says that two routers connected by an unnumbered
   point to point line are not really two routers at all, but rather two
   half-routers that together make up a single virtual router. The
   unnumbered point to point line is essentially considered to be an
   internal bus in the virtual router. The two halves of the virtual
   router must coordinate their activities in such a way that they act
   exactly like a single router.

   This scheme fits in well with the IP architecture, but suffers from
   two important drawbacks. The first is that, although it handles the
   common case of a single unnumbered point to point line, it is not
   readily extensible to handle the case of a mesh of routers and
   unnumbered point to point lines. The second drawback is that the
   interactions between the half routers are necessarily complex and are
   not standardized, effectively precluding the connection of equipment
   from different vendors using unnumbered point to point lines.

   Because of these drawbacks, this memo has adopted an alternate
   scheme, which has been invented multiple times but which is probably
   originally attributable to Phil Karn. In this scheme, a router that
   has unnumbered point to point lines also has a special IP address,
   called a router-id in this memo. The router-id is one of the

router's IP addresses (a router is required to have at least one IP
address). This router-id is used as if it is the IP address of all
unnumbered interfaces.

2.2.8 Notable Oddities

2.2.8.1 Embedded Routers

   A router may be a stand-alone computer system, dedicated to its IP
   router functions. Alternatively, it is possible to embed router
   functions within a host operating system that supports connections to
   two or more networks. The best-known example of an operating system
   with embedded router code is the Berkeley BSD system. The embedded
   router feature seems to make building a network easy, but it has a
   number of hidden pitfalls:

   (1) If a host has only a single constituent-network interface, it
        should not act as a router.

        For example, hosts with embedded router code that gratuitously
        forward broadcast packets or datagrams on the same net often
        cause packet avalanches.

   (2) If a (multihomed) host acts as a router, it is subject to the
        requirements for routers contained in this document.

        For example, the routing protocol issues and the router control
        and monitoring problems are as hard and important for embedded
        routers as for stand-alone routers.

        Internet router requirements and specifications may change
        independently of operating system changes. An administration
        that operates an embedded router in the Internet is strongly
        advised to maintain and update the router code. This might
        require router source code.

   (3) When a host executes embedded router code, it becomes part of the
        Internet infrastructure. Thus, errors in software or
        configuration can hinder communication between other hosts. As
        a consequence, the host administrator must lose some autonomy.

        In many circumstances, a host administrator will need to disable
        router code embedded in the operating system. For this reason,
        it should be straightforward to disable embedded router
        functionality.

        (4) When a host running embedded router code is concurrently used for
            other services, the Operation and Maintenance requirements for
            the two modes of use may conflict.

            For example, router O&M will in many cases be performed remotely
            by an operations center; this may require privileged system
            access that the host administrator would not normally want to
            distribute.

2.2.8.2 Transparent Routers

   There are two basic models for interconnecting local-area networks
   and wide-area (or long-haul) networks in the Internet. In the first,
   the local-area network is assigned a network prefix and all routers
   in the Internet must know how to route to that network. In the
   second, the local-area network shares (a small part of) the address
   space of the wide-area network. Routers that support this second
   model are called address sharing routers or transparent routers. The
   focus of this memo is on routers that support the first model, but
   this is not intended to exclude the use of transparent routers.

   The basic idea of a transparent router is that the hosts on the
   local-area network behind such a router share the address space of
   the wide-area network in front of the router. In certain situations
   this is a very useful approach and the limitations do not present
   significant drawbacks.

   The words in front and behind indicate one of the limitations of this
   approach: this model of interconnection is suitable only for a
   geographically (and topologically) limited stub environment. It
   requires that there be some form of logical addressing in the network
   level addressing of the wide-area network. IP addresses in the local
   environment map to a few (usually one) physical address in the wide-
   area network. This mapping occurs in a way consistent with the { IP
   address <-> network address } mapping used throughout the wide-area
   network.

   Multihoming is possible on one wide-area network, but may present
   routing problems if the interfaces are geographically or
   topologically separated. Multihoming on two (or more) wide-area
   networks is a problem due to the confusion of addresses.

   The behavior that hosts see from other hosts in what is apparently
   the same network may differ if the transparent router cannot fully
   emulate the normal wide-area network service. For example, the
   ARPANET used a Link Layer protocol that provided a Destination Dead
   indication in response to an attempt to send to a host that was off-
   line. However, if there were a transparent router between the

ARPANET and an Ethernet, a host on the ARPANET would not receive a
Destination Dead indication for Ethernet hosts.

2.3 Router Characteristics

   An Internet router performs the following functions:

   (1) Conforms to specific Internet protocols specified in this
        document, including the Internet Protocol (IP), Internet Control
        Message Protocol (ICMP), and others as necessary.

   (2) Interfaces to two or more packet networks. For each connected
        network the router must implement the functions required by that
        network. These functions typically include:

        o Encapsulating and decapsulating the IP datagrams with the
           connected network framing (e.g., an Ethernet header and
           checksum),

        o Sending and receiving IP datagrams up to the maximum size
           supported by that network, this size is the network's Maximum
           Transmission Unit or MTU,

        o Translating the IP destination address into an appropriate
           network-level address for the connected network (e.g., an
           Ethernet hardware address), if needed, and

        o Responding to network flow control and error indications, if
           any.

        See chapter 3 (Link Layer).

   (3) Receives and forwards Internet datagrams. Important issues in
        this process are buffer management, congestion control, and
        fairness.

        o Recognizes error conditions and generates ICMP error and
           information messages as required.

        o Drops datagrams whose time-to-live fields have reached zero.

        o Fragments datagrams when necessary to fit into the MTU of the
           next network.

        See chapter 4 (Internet Layer - Protocols) and chapter 5
        (Internet Layer - Forwarding) for more information.

   (4) Chooses a next-hop destination for each IP datagram, based on the
       information in its routing database. See chapter 5 (Internet
       Layer - Forwarding) for more information.

   (5) (Usually) supports an interior gateway protocol (IGP) to carry
       out distributed routing and reachability algorithms with the
       other routers in the same autonomous system. In addition, some
       routers will need to support an exterior gateway protocol (EGP)
       to exchange topological information with other autonomous
       systems. See chapter 7 (Application Layer - Routing Protocols)
       for more information.

   (6) Provides network management and system support facilities,
       including loading, debugging, status reporting, exception
       reporting and control. See chapter 8 (Application Layer -
       Network Management Protocols) and chapter 10 (Operation and
       Maintenance) for more information.

   A router vendor will have many choices on power, complexity, and
   features for a particular router product. It may be helpful to
   observe that the Internet system is neither homogeneous nor fully
   connected. For reasons of technology and geography it is growing
   into a global interconnect system plus a fringe of LANs around the
   edge. More and more these fringe LANs are becoming richly
   interconnected, thus making them less out on the fringe and more
   demanding on router requirements.

   o The global interconnect system is composed of a number of wide-area
     networks to which are attached routers of several Autonomous
     Systems (AS); there are relatively few hosts connected directly to
     the system.

   o Most hosts are connected to LANs. Many organizations have clusters
     of LANs interconnected by local routers. Each such cluster is
     connected by routers at one or more points into the global
     interconnect system. If it is connected at only one point, a LAN
     is known as a stub network.

   Routers in the global interconnect system generally require:

   o Advanced Routing and Forwarding Algorithms

     These routers need routing algorithms that are highly dynamic,
     impose minimal processing and communication burdens, and offer
     type-of-service routing. Congestion is still not a completely
     resolved issue (see Section [5.3.6]). Improvements in these areas
     are expected, as the research community is actively working on
     these issues.

   o High Availability

     These routers need to be highly reliable, providing 24 hours a
     day, 7 days a week service. Equipment and software faults can
     have a wide-spread (sometimes global) effect. In case of failure,
     they must recover quickly. In any environment, a router must be
     highly robust and able to operate, possibly in a degraded state,
     under conditions of extreme congestion or failure of network
     resources.

   o Advanced O&M Features

     Internet routers normally operate in an unattended mode. They
     will typically be operated remotely from a centralized monitoring
     center. They need to provide sophisticated means for monitoring
     and measuring traffic and other events and for diagnosing faults.

   o High Performance

     Long-haul lines in the Internet today are most frequently full
     duplex 56 KBPS, DS1 (1.544 Mbps), or DS3 (45 Mbps) speeds. LANs,
     which are half duplex multiaccess media, are typically Ethernet
     (10Mbps) and, to a lesser degree, FDDI (100Mbps). However,
     network media technology is constantly advancing and higher speeds
     are likely in the future.

The requirements for routers used in the LAN fringe (e.g., campus
networks) depend greatly on the demands of the local networks. These
may be high or medium-performance devices, probably competitively
procured from several different vendors and operated by an internal
organization (e.g., a campus computing center). The design of these
routers should emphasize low average latency and good burst
performance, together with delay and type-of-service sensitive
resource management. In this environment there may be less formal
O&M but it will not be less important. The need for the routing
mechanism to be highly dynamic will become more important as networks
become more complex and interconnected. Users will demand more out
of their local connections because of the speed of the global
interconnects.

As networks have grown, and as more networks have become old enough
that they are phasing out older equipment, it has become increasingly
imperative that routers interoperate with routers from other vendors.

Even though the Internet system is not fully interconnected, many
parts of the system need to have redundant connectivity. Rich
connectivity allows reliable service despite failures of
communication lines and routers, and it can also improve service by

shortening Internet paths and by providing additional capacity.
Unfortunately, this richer topology can make it much more difficult
to choose the best path to a particular destination.

2.4 Architectural Assumptions

   The current Internet architecture is based on a set of assumptions
   about the communication system. The assumptions most relevant to
   routers are as follows:

   o The Internet is a network of networks.

      Each host is directly connected to some particular network(s); its
      connection to the Internet is only conceptual. Two hosts on the
      same network communicate with each other using the same set of
      protocols that they would use to communicate with hosts on distant
      networks.

   o Routers do not keep connection state information.

      To improve the robustness of the communication system, routers are
      designed to be stateless, forwarding each IP packet independently
      of other packets. As a result, redundant paths can be exploited
      to provide robust service in spite of failures of intervening
      routers and networks.

      All state information required for end-to-end flow control and
      reliability is implemented in the hosts, in the transport layer or
      in application programs. All connection control information is
      thus co-located with the end points of the communication, so it
      will be lost only if an end point fails. Routers control message
      flow only indirectly, by dropping packets or increasing network
      delay.

      Note that future protocol developments may well end up putting
      some more state into routers. This is especially likely for
      multicast routing, resource reservation, and flow based
      forwarding.

   o Routing complexity should be in the routers.

      Routing is a complex and difficult problem, and ought to be
      performed by the routers, not the hosts. An important objective
      is to insulate host software from changes caused by the inevitable
      evolution of the Internet routing architecture.

   o The system must tolerate wide network variation.

     A basic objective of the Internet design is to tolerate a wide
     range of network characteristics - e.g., bandwidth, delay, packet
     loss, packet reordering, and maximum packet size. Another
     objective is robustness against failure of individual networks,
     routers, and hosts, using whatever bandwidth is still available.
     Finally, the goal is full open system interconnection: an Internet
     router must be able to interoperate robustly and effectively with
     any other router or Internet host, across diverse Internet paths.

     Sometimes implementors have designed for less ambitious goals.
     For example, the LAN environment is typically much more benign
     than the Internet as a whole; LANs have low packet loss and delay
     and do not reorder packets. Some vendors have fielded
     implementations that are adequate for a simple LAN environment,
     but work badly for general interoperation. The vendor justifies
     such a product as being economical within the restricted LAN
     market. However, isolated LANs seldom stay isolated for long.
     They are soon connected to each other, to organization-wide
     internets, and eventually to the global Internet system. In the
     end, neither the customer nor the vendor is served by incomplete
     or substandard routers.

     The requirements in this document are designed for a full-function
     router. It is intended that fully compliant routers will be
     usable in almost any part of the Internet.

3. LINK LAYER

   Although [INTRO:1] covers Link Layer standards (IP over various link
   layers, ARP, etc.), this document anticipates that Link-Layer
   material will be covered in a separate Link Layer Requirements
   document. A Link-Layer Requirements document would be applicable to
   both hosts and routers. Thus, this document will not obsolete the
   parts of [INTRO:1] that deal with link-layer issues.

3.1 INTRODUCTION

   Routers have essentially the same Link Layer protocol requirements as
   other sorts of Internet systems. These requirements are given in
   chapter 3 of Requirements for Internet Gateways [INTRO:1]. A router
   MUST comply with its requirements and SHOULD comply with its
   recommendations. Since some of the material in that document has
   become somewhat dated, some additional requirements and explanations
   are included below.

DISCUSSION
    It is expected that the Internet community will produce a
    Requirements for Internet Link Layer standard which will supersede
    both this chapter and the chapter entitled "INTERNET LAYER
    PROTOCOLS" in [INTRO:1].

3.2 LINK/INTERNET LAYER INTERFACE

   This document does not attempt to specify the interface between the
   Link Layer and the upper layers. However, note well that other parts
   of this document, particularly chapter 5, require various sorts of
   information to be passed across this layer boundary.

   This section uses the following definitions:

   o Source physical address

     The source physical address is the Link Layer address of the host
     or router from which the packet was received.

   o Destination physical address

     The destination physical address is the Link Layer address to
     which the packet was sent.

   The information that must pass from the Link Layer to the
   Internetwork Layer for each received packet is:

   (1) The IP packet [5.2.2],

   (2) The length of the data portion (i.e., not including the Link-
        Layer framing) of the Link Layer frame [5.2.2],

   (3) The identity of the physical interface from which the IP packet
        was received [5.2.3], and

   (4) The classification of the packet's destination physical address
        as a Link Layer unicast, broadcast, or multicast [4.3.2],
        [5.3.4].

   In addition, the Link Layer also should provide:

   (5) The source physical address.

   The information that must pass from the Internetwork Layer to the
   Link Layer for each transmitted packet is:

   (1) The IP packet [5.2.1]

   (2) The length of the IP packet [5.2.1]

   (3) The destination physical interface [5.2.1]

   (4) The next hop IP address [5.2.1]

   In addition, the Internetwork Layer also should provide:

   (5) The Link Layer priority value [5.3.3.2]

   The Link Layer must also notify the Internetwork Layer if the packet
   to be transmitted causes a Link Layer precedence-related error
   [5.3.3.3].

3.3 SPECIFIC ISSUES

3.3.1 Trailer Encapsulation

   Routers that can connect to ten megabit Ethernets MAY be able to
   receive and forward Ethernet packets encapsulated using the trailer
   encapsulation described in [LINK:1]. However, a router SHOULD NOT
   originate trailer encapsulated packets. A router MUST NOT originate
   trailer encapsulated packets without first verifying, using the
   mechanism described in [INTRO:2], that the immediate destination of
   the packet is willing and able to accept trailer-encapsulated
   packets. A router SHOULD NOT agree (using these mechanisms) to
   accept trailer-encapsulated packets.

3.3.2 Address Resolution Protocol - ARP

   Routers that implement ARP MUST be compliant and SHOULD be
   unconditionally compliant with the requirements in [INTRO:2].

   The link layer MUST NOT report a Destination Unreachable error to IP
   solely because there is no ARP cache entry for a destination; it
   SHOULD queue up to a small number of datagrams breifly while
   performing the ARP request/reply sequence, and reply that the
   destination is unreachable to one of the queued datagrams only when
   this proves fruitless.

   A router MUST not believe any ARP reply that claims that the Link
   Layer address of another host or router is a broadcast or multicast
   address.

3.3.3 Ethernet and 802.3 Coexistence

   Routers that can connect to ten megabit Ethernets MUST be compliant
   and SHOULD be unconditionally compliant with the Ethernet
   requirements of [INTRO:2].

3.3.4 Maximum Transmission Unit - MTU

   The MTU of each logical interface MUST be configurable within the
   range of legal MTUs for the interface.

   Many Link Layer protocols define a maximum frame size that may be
   sent. In such cases, a router MUST NOT allow an MTU to be set which
   would allow sending of frames larger than those allowed by the Link
   Layer protocol. However, a router SHOULD be willing to receive a
   packet as large as the maximum frame size even if that is larger than
   the MTU.

   DISCUSSION
      Note that this is a stricter requirement than imposed on hosts by
      [INTRO:2], which requires that the MTU of each physical interface
      be configurable.

      If a network is using an MTU smaller than the maximum frame size
      for the Link Layer, a router may receive packets larger than the
      MTU from misconfigured and incompletely initialized hosts. The
      Robustness Principle indicates that the router should successfully
      receive these packets if possible.

3.3.5 Point-to-Point Protocol - PPP

   Contrary to [INTRO:1], the Internet does have a standard point to
   point line protocol: the Point-to-Point Protocol (PPP), defined in
   [LINK:2], [LINK:3], [LINK:4], and [LINK:5].

   A point to point interface is any interface that is designed to send
   data over a point to point line. Such interfaces include telephone,
   leased, dedicated or direct lines (either 2 or 4 wire), and may use
   point to point channels or virtual circuits of multiplexed interfaces
   such as ISDN. They normally use a standardized modem or bit serial
   interface (such as RS-232, RS-449 or V.35), using either synchronous
   or asynchronous clocking. Multiplexed interfaces often have special
   physical interfaces.

   A general purpose serial interface uses the same physical media as a
   point to point line, but supports the use of link layer networks as
   well as point to point connectivity. Link layer networks (such as
   X.25 or Frame Relay) use an alternative IP link layer specification.

Routers that implement point to point or general purpose serial
interfaces MUST IMPLEMENT PPP.

PPP MUST be supported on all general purpose serial interfaces on a
router. The router MAY allow the line to be configured to use point
to point line protocols other than PPP. Point to point interfaces
SHOULD either default to using PPP when enabled or require
configuration of the link layer protocol before being enabled.
General purpose serial interfaces SHOULD require configuration of the
link layer protocol before being enabled.

3.3.5.1 Introduction

This section provides guidelines to router implementors so that they
can ensure interoperability with other routers using PPP over either
synchronous or asynchronous links.

It is critical that an implementor understand the semantics of the
option negotiation mechanism. Options are a means for a local device
to indicate to a remote peer what the local device will accept from
the remote peer, not what it wishes to send. It is up to the remote
peer to decide what is most convenient to send within the confines of
the set of options that the local device has stated that it can
accept. Therefore it is perfectly acceptable and normal for a remote
peer to ACK all the options indicated in an LCP Configuration Request
(CR) even if the remote peer does not support any of those options.
Again, the options are simply a mechanism for either device to
indicate to its peer what it will accept, not necessarily what it
will send.

3.3.5.2 Link Control Protocol (LCP) Options

The PPP Link Control Protocol (LCP) offers a number of options that
may be negotiated. These options include (among others) address and
control field compression, protocol field compression, asynchronous
character map, Maximum Receive Unit (MRU), Link Quality Monitoring
(LQM), magic number (for loopback detection), Password Authentication
Protocol (PAP), Challenge Handshake Authentication Protocol (CHAP),
and the 32-bit Frame Check Sequence (FCS).

A router MAY use address/control field compression on either
synchronous or asynchronous links. A router MAY use protocol field
compression on either synchronous or asynchronous links. A router
that indicates that it can accept these compressions MUST be able to
accept uncompressed PPP header information also.

DISCUSSION
   These options control the appearance of the PPP header. Normally
   the PPP header consists of the address, the control field, and the
   protocol field. The address, on a point to point line, is 0xFF,
   indicating "broadcast". The control field is 0x03, indicating
   "Unnumbered Information." The Protocol Identifier is a two byte
   value indicating the contents of the data area of the frame. If a
   system negotiates address and control field compression it
   indicates to its peer that it will accept PPP frames that have or
   do not have these fields at the front of the header. It does not
   indicate that it will be sending frames with these fields removed.

   Protocol field compression, when negotiated, indicates that the
   system is willing to receive protocol fields compressed to one
   byte when this is legal. There is no requirement that the sender
   do so.

   Use of address/control field compression is inconsistent with the
   use of numbered mode (reliable) PPP.

IMPLEMENTATION
   Some hardware does not deal well with variable length header
   information. In those cases it makes most sense for the remote
   peer to send the full PPP header. Implementations may ensure this
   by not sending the address/control field and protocol field
   compression options to the remote peer. Even if the remote peer
   has indicated an ability to receive compressed headers there is no
   requirement for the local router to send compressed headers.

A router MUST negotiate the Asynchronous Control Character Map (ACCM)
for asynchronous PPP links, but SHOULD NOT negotiate the ACCM for
synchronous links. If a router receives an attempt to negotiate the
ACCM over a synchronous link, it MUST ACKnowledge the option and then
ignore it.

DISCUSSION
   There are implementations that offer both synchronous and
   asynchronous modes of operation and may use the same code to
   implement the option negotiation. In this situation it is
   possible that one end or the other may send the ACCM option on a
   synchronous link.

A router SHOULD properly negotiate the maximum receive unit (MRU).
Even if a system negotiates an MRU smaller than 1,500 bytes, it MUST
be able to receive a 1,500 byte frame.

A router SHOULD negotiate and enable the link quality monitoring
(LQM) option.

DISCUSSION
This memo does not specify a policy for deciding whether the
link's quality is adequate. However, it is important (see Section
[3.3.6]) that a router disable failed links.

A router SHOULD implement and negotiate the magic number option for
loopback detection.

A router MAY support the authentication options (PAP - Password
Authentication Protocol, and/or CHAP - Challenge Handshake
Authentication Protocol).

A router MUST support 16-bit CRC frame check sequence (FCS) and MAY
support the 32-bit CRC.

3.3.5.3 IP Control Protocol (IPCP) Options

A router MAY offer to perform IP address negotiation. A router MUST
accept a refusal (REJect) to perform IP address negotiation from the
peer.

Routers operating at link speeds of 19,200 BPS or less SHOULD
implement and offer to perform Van Jacobson header compression.
Routers that implement VJ compression SHOULD implement an
administrative control enabling or disabling it.

3.3.6 Interface Testing

A router MUST have a mechanism to allow routing software to determine
whether a physical interface is available to send packets or not; on
multiplexed interfaces where permanent virtual circuits are opened
for limited sets of neighbors, the router must also be able to
determine whether the virtual circuits are viable. A router SHOULD
have a mechanism to allow routing software to judge the quality of a
physical interface. A router MUST have a mechanism for informing the
routing software when a physical interface becomes available or
unavailable to send packets because of administrative action. A
router MUST have a mechanism for informing the routing software when
it detects a Link level interface has become available or
unavailable, for any reason.

DISCUSSION
It is crucial that routers have workable mechanisms for
determining that their network connections are functioning
properly. Failure to detect link loss, or failure to take the
proper actions when a problem is detected, can lead to black
holes.

The mechanisms available for detecting problems with network
connections vary considerably, depending on the Link Layer
protocols in use and the interface hardware. The intent is to
maximize the capability to detect failures within the Link-Layer
constraints.

4. INTERNET LAYER - PROTOCOLS

4.1 INTRODUCTION

This chapter and chapter 5 discuss the protocols used at the Internet
Layer: IP, ICMP, and IGMP. Since forwarding is obviously a crucial
topic in a document discussing routers, chapter 5 limits itself to
the aspects of the protocols that directly relate to forwarding. The
current chapter contains the remainder of the discussion of the
Internet Layer protocols.

4.2 INTERNET PROTOCOL - IP

4.2.1 INTRODUCTION

Routers MUST implement the IP protocol, as defined by [INTERNET:1].
They MUST also implement its mandatory extensions: subnets (defined
in [INTERNET:2]), IP broadcast (defined in [INTERNET:3]), and
Classless Inter-Domain Routing (CIDR, defined in [INTERNET:15]).

Router implementors need not consider compliance with the section of
[INTRO:2] entitled "Internet Protocol -- IP," as that section is
entirely duplicated or superseded in this document. A router MUST be
compliant, and SHOULD be unconditionally compliant, with the
requirements of the section entitled "SPECIFIC ISSUES" relating to IP
in [INTRO:2].

In the following, the action specified in certain cases is to
silently discard a received datagram. This means that the datagram
will be discarded without further processing and that the router will
not send any ICMP error message (see Section [4.3]) as a result.
However, for diagnosis of problems a router SHOULD provide the
capability of logging the error (see Section [1.3.3]), including the
contents of the silently discarded datagram, and SHOULD count
datagrams discarded.

4.2.2 PROTOCOL WALK-THROUGH

   RFC 791 [INTERNET:1] is the specification for the Internet Protocol.

4.2.2.1 Options: RFC 791 Section 3.2

   In datagrams received by the router itself, the IP layer MUST
   interpret IP options that it understands and preserve the rest
   unchanged for use by higher layer protocols.

   Higher layer protocols may require the ability to set IP options in
   datagrams they send or examine IP options in datagrams they receive.
   Later sections of this document discuss specific IP option support
   required by higher layer protocols.

      DISCUSSION
         Neither this memo nor [INTRO:2] define the order in which a
         receiver must process multiple options in the same IP header.
         Hosts and routers originating datagrams containing multiple
         options must be aware that this introduces an ambiguity in the
         meaning of certain options when combined with a source-route
         option.

   Here are the requirements for specific IP options:

   (a) Security Option

         Some environments require the Security option in every packet
         originated or received. Routers SHOULD IMPLEMENT the revised
         security option described in [INTERNET:5].

      DISCUSSION
         Note that the security options described in [INTERNET:1] and RFC
         1038 ([INTERNET:16]) are obsolete.

   (b) Stream Identifier Option

         This option is obsolete; routers SHOULD NOT place this option
         in a datagram that the router originates. This option MUST be
         ignored in datagrams received by the router.

   (c) Source Route Options

         A router MUST be able to act as the final destination of a
         source route. If a router receives a packet containing a
         completed source route, the packet has reached its final
         destination. In such an option, the pointer points beyond the
         last field and the destination address in the IP header

addresses the router. The option as received (the recorded
route) MUST be passed up to the transport layer (or to ICMP
message processing).

In the general case, a correct response to a source-routed
datagram traverses the same route. A router MUST provide a
means whereby transport protocols and applications can reverse
the source route in a received datagram. This reversed source
route MUST be inserted into datagrams they originate (see
[INTRO:2] for details) when the router is unaware of policy
constraints. However, if the router is policy aware, it MAY
select another path.

Some applications in the router MAY require that the user be
able to enter a source route.

A router MUST NOT originate a datagram containing multiple
source route options. What a router should do if asked to
forward a packet containing multiple source route options is
described in Section [5.2.4.1].

When a source route option is created (which would happen when
the router is originating a source routed datagram or is
inserting a source route option as a result of a special
filter), it MUST be correctly formed even if it is being
created by reversing a recorded route that erroneously includes
the source host (see case (B) in the discussion below).

DISCUSSION
   Suppose a source routed datagram is to be routed from source S to
   destination D via routers G1, G2, Gn. Source S constructs a
   datagram with G1's IP address as its destination address, and a
   source route option to get the datagram the rest of the way to its
   destination. However, there is an ambiguity in the specification
   over whether the source route option in a datagram sent out by S
   should be (A) or (B):

   (A): {>>G2, G3, ... Gn, D} <--- CORRECT

   (B): {S, >>G2, G3, ... Gn, D} <---- WRONG

   (where >> represents the pointer). If (A) is sent, the datagram
   received at D will contain the option: {G1, G2, ... Gn >>}, with S
   and D as the IP source and destination addresses. If (B) were
   sent, the datagram received at D would again contain S and D as
   the same IP source and destination addresses, but the option would
   be: {S, G1, ...Gn >>}; i.e., the originating host would be the
   first hop in the route.

      (d) Record Route Option

           Routers MAY support the Record Route option in datagrams
           originated by the router.

      (e) Timestamp Option

           Routers MAY support the timestamp option in datagrams
           originated by the router. The following rules apply:

           o When originating a datagram containing a Timestamp Option, a
             router MUST record a timestamp in the option if

             - Its Internet address fields are not pre-specified or
             - Its first pre-specified address is the IP address of the
               logical interface over which the datagram is being sent
               (or the router's router-id if the datagram is being sent
               over an unnumbered interface).

           o If the router itself receives a datagram containing a
             Timestamp Option, the router MUST insert the current time
             into the Timestamp Option (if there is space in the option
             to do so) before passing the option to the transport layer
             or to ICMP for processing. If space is not present, the
             router MUST increment the Overflow Count in the option.

           o A timestamp value MUST follow the rules defined in [INTRO:2].

      IMPLEMENTATION
         To maximize the utility of the timestamps contained in the
         timestamp option, the timestamp inserted should be, as nearly as
         practical, the time at which the packet arrived at the router.
         For datagrams originated by the router, the timestamp inserted
         should be, as nearly as practical, the time at which the datagram
         was passed to the Link Layer for transmission.

         The timestamp option permits the use of a non-standard time clock,
         but the use of a non-synchronized clock limits the utility of the
         time stamp. Therefore, routers are well advised to implement the
         Network Time Protocol for the purpose of synchronizing their
         clocks.

4.2.2.2 Addresses in Options: RFC 791 Section 3.1

   Routers are called upon to insert their address into Record Route,
   Strict Source and Record Route, Loose Source and Record Route, or
   Timestamp Options. When a router inserts its address into such an
   option, it MUST use the IP address of the logical interface on which

the packet is being sent. Where this rule cannot be obeyed because
the output interface has no IP address (i.e., is an unnumbered
interface), the router MUST instead insert its router-id. The
router's router-id is one of the router's IP addresses. The Router
ID may be specified on a system basis or on a per-link basis. Which
of the router's addresses is used as the router-id MUST NOT change
(even across reboots) unless changed by the network manager.
Relevant management changes include reconfiguration of the router
such that the IP address used as the router-id ceases to be one of
the router's IP addresses. Routers with multiple unnumbered
interfaces MAY have multiple router-id's. Each unnumbered interface
MUST be associated with a particular router-id. This association
MUST NOT change (even across reboots) without reconfiguration of the
router.

    DISCUSSION
        This specification does not allow for routers that do not have at
        least one IP address. We do not view this as a serious
        limitation, since a router needs an IP address to meet the
        manageability requirements of Chapter [8] even if the router is
        connected only to point-to-point links.

    IMPLEMENTATION

        One possible method of choosing the router-id that fulfills this
        requirement is to use the numerically smallest (or greatest) IP
        address (treating the address as a 32-bit integer) that is
        assigned to the router.

4.2.2.3 Unused IP Header Bits: RFC 791 Section 3.1

    The IP header contains two reserved bits: one in the Type of Service
    byte and the other in the Flags field. A router MUST NOT set either
    of these bits to one in datagrams originated by the router. A router
    MUST NOT drop (refuse to receive or forward) a packet merely because
    one or more of these reserved bits has a non-zero value; i.e., the
    router MUST NOT check the values of thes bits.

    DISCUSSION
        Future revisions to the IP protocol may make use of these unused
        bits. These rules are intended to ensure that these revisions can
        be deployed without having to simultaneously upgrade all routers
        in the Internet.

4.2.2.4 Type of Service: RFC 791 Section 3.1

   The Type-of-Service byte in the IP header is divided into three
   sections: the Precedence field (high-order 3 bits), a field that is
   customarily called Type of Service or TOS (next 4 bits), and a
   reserved bit (the low order bit).

   Rules governing the reserved bit were described in Section [4.2.2.3].

   A more extensive discussion of the TOS field and its use can be found
   in [ROUTE:11].

   The description of the IP Precedence field is superseded by Section
   [5.3.3]. RFC 795, Service Mappings, is obsolete and SHOULD NOT be
   implemented.

4.2.2.5 Header Checksum: RFC 791 Section 3.1

   As stated in Section [5.2.2], a router MUST verify the IP checksum of
   any packet that is received, and MUST discard messages containing
   invalid checksums. The router MUST NOT provide a means to disable
   this checksum verification.

   A router MAY use incremental IP header checksum updating when the
   only change to the IP header is the time to live. This will reduce
   the possibility of undetected corruption of the IP header by the
   router. See [INTERNET:6] for a discussion of incrementally updating
   the checksum.

      IMPLEMENTATION
         A more extensive description of the IP checksum, including
         extensive implementation hints, can be found in [INTERNET:6] and
         [INTERNET:7].

4.2.2.6 Unrecognized Header Options: RFC 791 Section 3.1

   A router MUST ignore IP options which it does not recognize. A
   corollary of this requirement is that a router MUST implement the End
   of Option List option and the No Operation option, since neither
   contains an explicit length.

      DISCUSSION
         All future IP options will include an explicit length.

4.2.2.7 Fragmentation: RFC 791 Section 3.2

   Fragmentation, as described in [INTERNET:1], MUST be supported by a
   router.

   When a router fragments an IP datagram, it SHOULD minimize the number
   of fragments. When a router fragments an IP datagram, it SHOULD send
   the fragments in order. A fragmentation method that may generate one
   IP fragment that is significantly smaller than the other MAY cause
   the first IP fragment to be the smaller one.

   DISCUSSION
      There are several fragmentation techniques in common use in the
      Internet. One involves splitting the IP datagram into IP
      fragments with the first being MTU sized, and the others being
      approximately the same size, smaller than the MTU. The reason for
      this is twofold. The first IP fragment in the sequence will be
      the effective MTU of the current path between the hosts, and the
      following IP fragments are sized to minimize the further
      fragmentation of the IP datagram. Another technique is to split
      the IP datagram into MTU sized IP fragments, with the last
      fragment being the only one smaller, as described in [INTERNET:1].

      A common trick used by some implementations of TCP/IP is to
      fragment an IP datagram into IP fragments that are no larger than
      576 bytes when the IP datagram is to travel through a router.
      This is intended to allow the resulting IP fragments to pass the
      rest of the path without further fragmentation. This would,
      though, create more of a load on the destination host, since it
      would have a larger number of IP fragments to reassemble into one
      IP datagram. It would also not be efficient on networks where the
      MTU only changes once and stays much larger than 576 bytes.
      Examples include LAN networks such as an IEEE 802.5 network with a
      MTU of 2048 or an Ethernet network with an MTU of 1500).

      One other fragmentation technique discussed was splitting the IP
      datagram into approximately equal sized IP fragments, with the
      size less than or equal to the next hop network's MTU. This is
      intended to minimize the number of fragments that would result
      from additional fragmentation further down the path, and assure
      equal delay for each fragment.

      Routers SHOULD generate the least possible number of IP fragments.

      Work with slow machines leads us to believe that if it is
      necessary to fragment messages, sending the small IP fragment
      first maximizes the chance of a host with a slow interface of
      receiving all the fragments.

4.2.2.8 Reassembly: RFC 791 Section 3.2

   As specified in the corresponding section of [INTRO:2], a router MUST
   support reassembly of datagrams that it delivers to itself.

4.2.2.9 Time to Live: RFC 791 Section 3.2

   Time to Live (TTL) handling for packets originated or received by the
   router is governed by [INTRO:2]; this section changes none of its
   stipulations. However, since the remainder of the IP Protocol
   section of [INTRO:2] is rewritten, this section is as well.

   Note in particular that a router MUST NOT check the TTL of a packet
   except when forwarding it.

   A router MUST NOT originate or forward a datagram with a Time-to-Live
   (TTL) value of zero.

   A router MUST NOT discard a datagram just because it was received
   with TTL equal to zero or one; if it is to the router and otherwise
   valid, the router MUST attempt to receive it.

   On messages the router originates, the IP layer MUST provide a means
   for the transport layer to set the TTL field of every datagram that
   is sent. When a fixed TTL value is used, it MUST be configurable.
   The number SHOULD exceed the typical internet diameter, and current
   wisdom suggests that it should exceed twice the internet diameter to
   allow for growth. Current suggested values are normally posted in
   the Assigned Numbers RFC. The TTL field has two functions: limit the
   lifetime of TCP segments (see RFC 793 [TCP:1], p. 28), and terminate
   Internet routing loops. Although TTL is a time in seconds, it also
   has some attributes of a hop-count, since each router is required to
   reduce the TTL field by at least one.

   TTL expiration is intended to cause datagrams to be discarded by
   routers, but not by the destination host. Hosts that act as routers
   by forwarding datagrams must therefore follow the router's rules for
   TTL.

   A higher-layer protocol may want to set the TTL in order to implement
   an "expanding scope" search for some Internet resource. This is used
   by some diagnostic tools, and is expected to be useful for locating
   the "nearest" server of a given class using IP multicasting, for
   example. A particular transport protocol may also want to specify
   its own TTL bound on maximum datagram lifetime.

   A fixed default value must be at least big enough for the Internet
   "diameter," i.e., the longest possible path. A reasonable value is

about twice the diameter, to allow for continued Internet growth. As
of this writing, messages crossing the United States frequently
traverse 15 to 20 routers; this argues for a default TTL value in
excess of 40, and 64 is a common value.

4.2.2.10 Multi-subnet Broadcasts: RFC 922

   All-subnets broadcasts (called multi-subnet broadcasts in
   [INTERNET:3]) have been deprecated. See Section [5.3.5.3].

4.2.2.11 Addressing: RFC 791 Section 3.2

   As noted in 2.2.5.1, there are now five classes of IP addresses:
   Class A through Class E. Class D addresses are used for IP
   multicasting [INTERNET:4], while Class E addresses are reserved for
   experimental use. The distinction between Class A, B, and C
   addresses is no longer important; they are used as generalized
   unicast network prefixes with only historical interest in their
   class.

   An IP multicast address is a 28-bit logical address that stands for a
   group of hosts, and may be either permanent or transient. Permanent
   multicast addresses are allocated by the Internet Assigned Number
   Authority [INTRO:7], while transient addresses may be allocated
   dynamically to transient groups. Group membership is determined
   dynamically using IGMP [INTERNET:4].

   We now summarize the important special cases for general purpose
   unicast IP addresses, using the following notation for an IP address:

     { <Network-prefix>, <Host-number> }

   and the notation -1 for a field that contains all 1 bits and the
   notation 0 for a field that contains all 0 bits.

   (a) { 0, 0 }

        This host on this network. It MUST NOT be used as a source
        address by routers, except the router MAY use this as a source
        address as part of an initialization procedure (e.g., if the
        router is using BOOTP to load its configuration information).

        Incoming datagrams with a source address of { 0, 0 } which are
        received for local delivery (see Section [5.2.3]), MUST be
        accepted if the router implements the associated protocol and
        that protocol clearly defines appropriate action to be taken.
        Otherwise, a router MUST silently discard any locally-delivered
        datagram whose source address is { 0, 0 }.

   DISCUSSION
      Some protocols define specific actions to take in response to a
      received datagram whose source address is { 0, 0 }. Two examples
      are BOOTP and ICMP Mask Request. The proper operation of these
      protocols often depends on the ability to receive datagrams whose
      source address is { 0, 0 }. For most protocols, however, it is
      best to ignore datagrams having a source address of { 0, 0 } since
      they were probably generated by a misconfigured host or router.
      Thus, if a router knows how to deal with a given datagram having a
      { 0, 0 } source address, the router MUST accept it. Otherwise,
      the router MUST discard it.

   See also Section [4.2.3.1] for a non-standard use of { 0, 0 }.

   (b) { 0, <Host-number> }

        Specified host on this network. It MUST NOT be sent by routers
        except that the router MAY use this as a source address as part
        of an initialization procedure by which the it learns its own
        IP address.

   (c) { -1, -1 }

        Limited broadcast. It MUST NOT be used as a source address.

        A datagram with this destination address will be received by
        every host and router on the connected physical network, but
        will not be forwarded outside that network.

   (d) { <Network-prefix>, -1 }

        Directed Broadcast - a broadcast directed to the specified
        network prefix. It MUST NOT be used as a source address. A
        router MAY originate Network Directed Broadcast packets. A
        router MUST receive Network Directed Broadcast packets; however
        a router MAY have a configuration option to prevent reception
        of these packets. Such an option MUST default to allowing
        reception.

    (e) { 127, <any> }

        Internal host loopback address. Addresses of this form MUST
        NOT appear outside a host.

   The <Network-prefix> is administratively assigned so that its value
   will be unique in the routing domain to which the device is
   connected.

IP addresses are not permitted to have the value 0 or -1 for the
<Host-number> or <Network-prefix> fields except in the special cases
listed above. This implies that each of these fields will be at
least two bits long.

DISCUSSION
   Previous versions of this document also noted that subnet numbers
   must be neither 0 nor -1, and must be at least two bits in length.
   In a CIDR world, the subnet number is clearly an extension of the
   network prefix and cannot be interpreted without the remainder of
   the prefix. This restriction of subnet numbers is therefore
   meaningless in view of CIDR and may be safely ignored.

For further discussion of broadcast addresses, see Section [4.2.3.1].

When a router originates any datagram, the IP source address MUST be
one of its own IP addresses (but not a broadcast or multicast
address). The only exception is during initialization.

For most purposes, a datagram addressed to a broadcast or multicast
destination is processed as if it had been addressed to one of the
router's IP addresses; that is to say:

o A router MUST receive and process normally any packets with a
   broadcast destination address.

o A router MUST receive and process normally any packets sent to a
   multicast destination address that the router has asked to
   receive.

The term specific-destination address means the equivalent local IP
address of the host. The specific-destination address is defined to
be the destination address in the IP header unless the header
contains a broadcast or multicast address, in which case the
specific-destination is an IP address assigned to the physical
interface on which the datagram arrived.

A router MUST silently discard any received datagram containing an IP
source address that is invalid by the rules of this section. This
validation could be done either by the IP layer or (when appropriate)
by each protocol in the transport layer. As with any datagram a
router discards, the datagram discard SHOULD be counted.

DISCUSSION
   A misaddressed datagram might be caused by a Link Layer broadcast
   of a unicast datagram or by another router or host that is
   confused or misconfigured.

4.2.3 SPECIFIC ISSUES

4.2.3.1 IP Broadcast Addresses

   For historical reasons, there are a number of IP addresses (some
   standard and some not) which are used to indicate that an IP packet
   is an IP broadcast. A router

   (1) MUST treat as IP broadcasts packets addressed to 255.255.255.255
        or { <Network-prefix>, -1 }.

   (2) SHOULD silently discard on receipt (i.e., do not even deliver to
        applications in the router) any packet addressed to 0.0.0.0 or {
        <Network-prefix>, 0 }. If these packets are not silently
        discarded, they MUST be treated as IP broadcasts (see Section
        [5.3.5]). There MAY be a configuration option to allow receipt
        of these packets. This option SHOULD default to discarding
        them.

   (3) SHOULD (by default) use the limited broadcast address
        (255.255.255.255) when originating an IP broadcast destined for
        a connected (sub)network (except when sending an ICMP Address
        Mask Reply, as discussed in Section [4.3.3.9]). A router MUST
        receive limited broadcasts.

   (4) SHOULD NOT originate datagrams addressed to 0.0.0.0 or {
        <Network-prefix>, 0 }. There MAY be a configuration option to
        allow generation of these packets (instead of using the relevant
        1s format broadcast). This option SHOULD default to not
        generating them.

      DISCUSSION
         In the second bullet, the router obviously cannot recognize
         addresses of the form { <Network-prefix>, 0 } if the router has no
         interface to that network prefix. In that case, the rules of the
         second bullet do not apply because, from the point of view of the
         router, the packet is not an IP broadcast packet.

4.2.3.2 IP Multicasting

   An IP router SHOULD satisfy the Host Requirements with respect to IP
   multicasting, as specified in [INTRO:2]. An IP router SHOULD support
   local IP multicasting on all connected networks. When a mapping from
   IP multicast addresses to link-layer addresses has been specified
   (see the various IP-over-xxx specifications), it SHOULD use that
   mapping, and MAY be configurable to use the link layer broadcast
   instead. On point-to-point links and all other interfaces,
   multicasts are encapsulated as link layer broadcasts. Support for

local IP multicasting includes originating multicast datagrams,
joining multicast groups and receiving multicast datagrams, and
leaving multicast groups. This implies support for all of
[INTERNET:4] including IGMP (see Section [4.4]).

DISCUSSION
    Although [INTERNET:4] is entitled Host Extensions for IP
    Multicasting, it applies to all IP systems, both hosts and
    routers. In particular, since routers may join multicast groups,
    it is correct for them to perform the host part of IGMP, reporting
    their group memberships to any multicast routers that may be
    present on their attached networks (whether or not they themselves
    are multicast routers).

    Some router protocols may specifically require support for IP
    multicasting (e.g., OSPF [ROUTE:1]), or may recommend it (e.g.,
    ICMP Router Discovery [INTERNET:13]).

4.2.3.3 Path MTU Discovery

   To eliminate fragmentation or minimize it, it is desirable to know
   what is the path MTU along the path from the source to destination.
   The path MTU is the minimum of the MTUs of each hop in the path.
   [INTERNET:14] describes a technique for dynamically discovering the
   maximum transmission unit (MTU) of an arbitrary internet path. For a
   path that passes through a router that does not support
   [INTERNET:14], this technique might not discover the correct Path
   MTU, but it will always choose a Path MTU as accurate as, and in many
   cases more accurate than, the Path MTU that would be chosen by older
   techniques or the current practice.

   When a router is originating an IP datagram, it SHOULD use the scheme
   described in [INTERNET:14] to limit the datagram's size. If the
   router's route to the datagram's destination was learned from a
   routing protocol that provides Path MTU information, the scheme
   described in [INTERNET:14] is still used, but the Path MTU
   information from the routing protocol SHOULD be used as the initial
   guess as to the Path MTU and also as an upper bound on the Path MTU.

4.2.3.4 Subnetting

   Under certain circumstances, it may be desirable to support subnets
   of a particular network being interconnected only through a path that
   is not part of the subnetted network. This is known as discontiguous
   subnetwork support.

   Routers MUST support discontiguous subnetworks.

IMPLEMENTATION
    In classical IP networks, this was very difficult to achieve; in
    CIDR networks, it is a natural by-product. Therefore, a router
    SHOULD NOT make assumptions about subnet architecture, but SHOULD
    treat each route as a generalized network prefix.

DISCUSSION The Internet has been growing at a tremendous rate of
    late. This has been placing severe strains on the IP addressing
    technology. A major factor in this strain is the strict IP
    Address class boundaries. These make it difficult to efficiently
    size network prefixes to their networks and aggregate several
    network prefixes into a single route advertisement. By
    eliminating the strict class boundaries of the IP address and
    treating each route as a generalized network prefix, these strains
    may be greatly reduced.

    The technology for currently doing this is Classless Inter Domain
    Routing (CIDR) [INTERNET:15].

For similar reasons, an address block associated with a given network
prefix could be subdivided into subblocks of different sizes, so that
the network prefixes associated with the subblocks would have
different length. For example, within a block whose network prefix
is 8 bits long, one subblock may have a 16 bit network prefix,
another may have an 18 bit network prefix, and a third a 14 bit
network prefix.

Routers MUST support variable length network prefixes in both their
interface configurations and their routing databases.

4.3 INTERNET CONTROL MESSAGE PROTOCOL - ICMP

4.3.1 INTRODUCTION

ICMP is an auxiliary protocol, which provides routing, diagnostic and
error functionality for IP. It is described in [INTERNET:8]. A
router MUST support ICMP.

ICMP messages are grouped in two classes that are discussed in the
following sections:

ICMP error messages:

Destination Unreachable   Section 4.3.3.1
Redirect   Section 4.3.3.2
Source Quench   Section 4.3.3.3
Time Exceeded   Section 4.3.3.4
Parameter Problem   Section 4.3.3.5

```
ICMP query messages:
Echo                            Section 4.3.3.6
Information                     Section 4.3.3.7
Timestamp                       Section 4.3.3.8
Address Mask                    Section 4.3.3.9
Router Discovery                Section 4.3.3.10
```

General ICMP requirements and discussion are in the next section.

4.3.2 GENERAL ISSUES

4.3.2.1 Unknown Message Types

   If an ICMP message of unknown type is received, it MUST be passed to
   the ICMP user interface (if the router has one) or silently discarded
   (if the router does not have one).

4.3.2.2 ICMP Message TTL

   When originating an ICMP message, the router MUST initialize the TTL.
   The TTL for ICMP responses must not be taken from the packet that
   triggered the response.

4.3.2.3 Original Message Header

   Historically, every ICMP error message has included the Internet
   header and at least the first 8 data bytes of the datagram that
   triggered the error. This is no longer adequate, due to the use of
   IP-in-IP tunneling and other technologies. Therefore, the ICMP
   datagram SHOULD contain as much of the original datagram as possible
   without the length of the ICMP datagram exceeding 576 bytes. The
   returned IP header (and user data) MUST be identical to that which
   was received, except that the router is not required to undo any
   modifications to the IP header that are normally performed in
   forwarding that were performed before the error was detected (e.g.,
   decrementing the TTL, or updating options). Note that the
   requirements of Section [4.3.3.5] supersede this requirement in some
   cases (i.e., for a Parameter Problem message, if the problem is in a
   modified field, the router must undo the modification). See Section
   [4.3.3.5]).

4.3.2.4 ICMP Message Source Address

   Except where this document specifies otherwise, the IP source address
   in an ICMP message originated by the router MUST be one of the IP
   addresses associated with the physical interface over which the ICMP
   message is transmitted. If the interface has no IP addresses

associated with it, the router's router-id (see Section [5.2.5]) is
used instead.

4.3.2.5 TOS and Precedence

   ICMP error messages SHOULD have their TOS bits set to the same value
   as the TOS bits in the packet that provoked the sending of the ICMP
   error message, unless setting them to that value would cause the ICMP
   error message to be immediately discarded because it could not be
   routed to its destination. Otherwise, ICMP error messages MUST be
   sent with a normal (i.e., zero) TOS. An ICMP reply message SHOULD
   have its TOS bits set to the same value as the TOS bits in the ICMP
   request that provoked the reply.

   ICMP Source Quench error messages, if sent at all, MUST have their IP
   Precedence field set to the same value as the IP Precedence field in
   the packet that provoked the sending of the ICMP Source Quench
   message. All other ICMP error messages (Destination Unreachable,
   Redirect, Time Exceeded, and Parameter Problem) SHOULD have their
   precedence value set to 6 (INTERNETWORK CONTROL) or 7 (NETWORK
   CONTROL). The IP Precedence value for these error messages MAY be
   settable.

   An ICMP reply message MUST have its IP Precedence field set to the
   same value as the IP Precedence field in the ICMP request that
   provoked the reply.

4.3.2.6 Source Route

   If the packet which provokes the sending of an ICMP error message
   contains a source route option, the ICMP error message SHOULD also
   contain a source route option of the same type (strict or loose),
   created by reversing the portion before the pointer of the route
   recorded in the source route option of the original packet UNLESS the
   ICMP error message is an ICMP Parameter Problem complaining about a
   source route option in the original packet, or unless the router is
   aware of policy that would prevent the delivery of the ICMP error
   message.

   DISCUSSION
      In environments which use the U.S. Department of Defense security
      option (defined in [INTERNET:5]), ICMP messages may need to
      include a security option. Detailed information on this topic
      should be available from the Defense Communications Agency.

4.3.2.7 When Not to Send ICMP Errors

   An ICMP error message MUST NOT be sent as the result of receiving:

   o An ICMP error message, or

   o A packet which fails the IP header validation tests described in
      Section [5.2.2] (except where that section specifically permits
      the sending of an ICMP error message), or

   o A packet destined to an IP broadcast or IP multicast address, or

   o A packet sent as a Link Layer broadcast or multicast, or

   o A packet whose source address has a network prefix of zero or is an
      invalid source address (as defined in Section [5.3.7]), or

   o Any fragment of a datagram other then the first fragment (i.e., a
      packet for which the fragment offset in the IP header is nonzero).

   Furthermore, an ICMP error message MUST NOT be sent in any case where
   this memo states that a packet is to be silently discarded.

   NOTE: THESE RESTRICTIONS TAKE PRECEDENCE OVER ANY REQUIREMENT
   ELSEWHERE IN THIS DOCUMENT FOR SENDING ICMP ERROR MESSAGES.

   DISCUSSION
      These rules aim to prevent the broadcast storms that have resulted
      from routers or hosts returning ICMP error messages in response to
      broadcast packets. For example, a broadcast UDP packet to a non-
      existent port could trigger a flood of ICMP Destination
      Unreachable datagrams from all devices that do not have a client
      for that destination port. On a large Ethernet, the resulting
      collisions can render the network useless for a second or more.

      Every packet that is broadcast on the connected network should
      have a valid IP broadcast address as its IP destination (see
      Section [5.3.4] and [INTRO:2]). However, some devices violate
      this rule. To be certain to detect broadcast packets, therefore,
      routers are required to check for a link-layer broadcast as well
      as an IP-layer address.

   IMPLEMENTATION+ This requires that the link layer inform the IP layer
      when a link-layer broadcast packet has been received; see Section
      [3.1].

4.3.2.8 Rate Limiting

   A router which sends ICMP Source Quench messages MUST be able to
   limit the rate at which the messages can be generated. A router
   SHOULD also be able to limit the rate at which it sends other sorts
   of ICMP error messages (Destination Unreachable, Redirect, Time
   Exceeded, Parameter Problem). The rate limit parameters SHOULD be
   settable as part of the configuration of the router. How the limits
   are applied (e.g., per router or per interface) is left to the
   implementor's discretion.

      DISCUSSION
         Two problems for a router sending ICMP error message are:
         (1) The consumption of bandwidth on the reverse path, and
         (2) The use of router resources (e.g., memory, CPU time)

         To help solve these problems a router can limit the frequency with
         which it generates ICMP error messages. For similar reasons, a
         router may limit the frequency at which some other sorts of
         messages, such as ICMP Echo Replies, are generated.

      IMPLEMENTATION
         Various mechanisms have been used or proposed for limiting the
         rate at which ICMP messages are sent:

         (1) Count-based - for example, send an ICMP error message for
              every N dropped packets overall or per given source host.
              This mechanism might be appropriate for ICMP Source Quench,
              if used, but probably not for other types of ICMP messages.

         (2) Timer-based - for example, send an ICMP error message to a
              given source host or overall at most once per T milliseconds.

         (3) Bandwidth-based - for example, limit the rate at which ICMP
              messages are sent over a particular interface to some
              fraction of the attached network's bandwidth.

4.3.3 SPECIFIC ISSUES

4.3.3.1 Destination Unreachable

   If a router cannot forward a packet because it has no routes at all
   (including no default route) to the destination specified in the
   packet, then the router MUST generate a Destination Unreachable, Code
   0 (Network Unreachable) ICMP message. If the router does have routes
   to the destination network specified in the packet but the TOS
   specified for the routes is neither the default TOS (0000) nor the
   TOS of the packet that the router is attempting to route, then the

router MUST generate a Destination Unreachable, Code 11 (Network
Unreachable for TOS) ICMP message.

If a packet is to be forwarded to a host on a network that is
directly connected to the router (i.e., the router is the last-hop
router) and the router has ascertained that there is no path to the
destination host then the router MUST generate a Destination
Unreachable, Code 1 (Host Unreachable) ICMP message. If a packet is
to be forwarded to a host that is on a network that is directly
connected to the router and the router cannot forward the packet
because no route to the destination has a TOS that is either equal to
the TOS requested in the packet or is the default TOS (0000) then the
router MUST generate a Destination Unreachable, Code 12 (Host
Unreachable for TOS) ICMP message.

   DISCUSSION
      The intent is that a router generates the "generic" host/network
      unreachable if it has no path at all (including default routes) to
      the destination. If the router has one or more paths to the
      destination, but none of those paths have an acceptable TOS, then
      the router generates the "unreachable for TOS" message.

4.3.3.2 Redirect

The ICMP Redirect message is generated to inform a local host that it
should use a different next hop router for certain traffic.

Contrary to [INTRO:2], a router MAY ignore ICMP Redirects when
choosing a path for a packet originated by the router if the router
is running a routing protocol or if forwarding is enabled on the
router and on the interface over which the packet is being sent.

4.3.3.3 Source Quench

A router SHOULD NOT originate ICMP Source Quench messages. As
specified in Section [4.3.2], a router that does originate Source
Quench messages MUST be able to limit the rate at which they are
generated.

   DISCUSSION
      Research seems to suggest that Source Quench consumes network
      bandwidth but is an ineffective (and unfair) antidote to
      congestion. See, for example, [INTERNET:9] and [INTERNET:10].
      Section [5.3.6] discusses the current thinking on how routers
      ought to deal with overload and network congestion.

A router MAY ignore any ICMP Source Quench messages it receives.

   DISCUSSION
      A router itself may receive a Source Quench as the result of
      originating a packet sent to another router or host. Such
      datagrams might be, e.g., an EGP update sent to another router, or
      a telnet stream sent to a host. A mechanism has been proposed
      ([INTERNET:11], [INTERNET:12]) to make the IP layer respond
      directly to Source Quench by controlling the rate at which packets
      are sent, however, this proposal is currently experimental and not
      currently recommended.

4.3.3.4 Time Exceeded

   When a router is forwarding a packet and the TTL field of the packet
   is reduced to 0, the requirements of section [5.2.3.8] apply.

   When the router is reassembling a packet that is destined for the
   router, it is acting as an Internet host. [INTRO:2]'s reassembly
   requirements therefore apply.

   When the router receives (i.e., is destined for the router) a Time
   Exceeded message, it MUST comply with [INTRO:2].

4.3.3.5 Parameter Problem

   A router MUST generate a Parameter Problem message for any error not
   specifically covered by another ICMP message. The IP header field or
   IP option including the byte indicated by the pointer field MUST be
   included unchanged in the IP header returned with this ICMP message.
   Section [4.3.2] defines an exception to this requirement.

   A new variant of the Parameter Problem message was defined in
   [INTRO:2]:
        Code 1 = required option is missing.

   DISCUSSION
      This variant is currently in use in the military community for a
      missing security option.

4.3.3.6 Echo Request/Reply

   A router MUST implement an ICMP Echo server function that receives
   Echo Requests sent to the router, and sends corresponding Echo
   Replies. A router MUST be prepared to receive, reassemble and echo
   an ICMP Echo Request datagram at least as the maximum of 576 and the
   MTUs of all the connected networks.

   The Echo server function MAY choose not to respond to ICMP echo
   requests addressed to IP broadcast or IP multicast addresses.

A router SHOULD have a configuration option that, if enabled, causes the router to silently ignore all ICMP echo requests; if provided, this option MUST default to allowing responses.

DISCUSSION
    The neutral provision about responding to broadcast and multicast Echo Requests derives from [INTRO:2]'s "Echo Request/Reply" section.

As stated in Section [10.3.3], a router MUST also implement a user/application-layer interface for sending an Echo Request and receiving an Echo Reply, for diagnostic purposes. All ICMP Echo Reply messages MUST be passed to this interface.

The IP source address in an ICMP Echo Reply MUST be the same as the specific-destination address of the corresponding ICMP Echo Request message.

Data received in an ICMP Echo Request MUST be entirely included in the resulting Echo Reply.

If a Record Route and/or Timestamp option is received in an ICMP Echo Request, this option (these options) SHOULD be updated to include the current router and included in the IP header of the Echo Reply message, without truncation. Thus, the recorded route will be for the entire round trip.

If a Source Route option is received in an ICMP Echo Request, the return route MUST be reversed and used as a Source Route option for the Echo Reply message, unless the router is aware of policy that would prevent the delivery of the message.

4.3.3.7 Information Request/Reply

A router SHOULD NOT originate or respond to these messages.

DISCUSSION
    The Information Request/Reply pair was intended to support self-configuring systems such as diskless workstations, to allow them to discover their IP network prefixes at boot time. However, these messages are now obsolete. The RARP and BOOTP protocols provide better mechanisms for a host to discover its own IP address.

4.3.3.8 Timestamp and Timestamp Reply

A router MAY implement Timestamp and Timestamp Reply. If they are implemented then:

    o The ICMP Timestamp server function MUST return a Timestamp Reply to
      every Timestamp message that is received. It SHOULD be designed
      for minimum variability in delay.

    o An ICMP Timestamp Request message to an IP broadcast or IP
      multicast address MAY be silently discarded.

    o The IP source address in an ICMP Timestamp Reply MUST be the same
      as the specific-destination address of the corresponding Timestamp
      Request message.

    o If a Source Route option is received in an ICMP Timestamp Request,
      the return route MUST be reversed and used as a Source Route
      option for the Timestamp Reply message, unless the router is aware
      of policy that would prevent the delivery of the message.

    o If a Record Route and/or Timestamp option is received in a
      Timestamp Request, this (these) option(s) SHOULD be updated to
      include the current router and included in the IP header of the
      Timestamp Reply message.

    o If the router provides an application-layer interface for sending
      Timestamp Request messages then incoming Timestamp Reply messages
      MUST be passed up to the ICMP user interface.

The preferred form for a timestamp value (the standard value) is
milliseconds since midnight, Universal Time. However, it may be
difficult to provide this value with millisecond resolution. For
example, many systems use clocks that update only at line frequency,
50 or 60 times per second. Therefore, some latitude is allowed in a
standard value:

(a) A standard value MUST be updated at least 16 times per second
      (i.e., at most the six low-order bits of the value may be
      undefined).

(b) The accuracy of a standard value MUST approximate that of
      operator-set CPU clocks, i.e., correct within a few minutes.

   IMPLEMENTATION
      To meet the second condition, a router may need to query some time
      server when the router is booted or restarted. It is recommended
      that the UDP Time Server Protocol be used for this purpose. A
      more advanced implementation would use the Network Time Protocol
      (NTP) to achieve nearly millisecond clock synchronization;
      however, this is not required.

4.3.3.9 Address Mask Request/Reply

   A router MUST implement support for receiving ICMP Address Mask
   Request messages and responding with ICMP Address Mask Reply
   messages. These messages are defined in [INTERNET:2].

   A router SHOULD have a configuration option for each logical
   interface specifying whether the router is allowed to answer Address
   Mask Requests for that interface; this option MUST default to
   allowing responses. A router MUST NOT respond to an Address Mask
   Request before the router knows the correct address mask.

   A router MUST NOT respond to an Address Mask Request that has a
   source address of 0.0.0.0 and which arrives on a physical interface
   that has associated with it multiple logical interfaces and the
   address masks for those interfaces are not all the same.

   A router SHOULD examine all ICMP Address Mask Replies that it
   receives to determine whether the information it contains matches the
   router's knowledge of the address mask. If the ICMP Address Mask
   Reply appears to be in error, the router SHOULD log the address mask
   and the sender's IP address. A router MUST NOT use the contents of
   an ICMP Address Mask Reply to determine the correct address mask.

   Because hosts may not be able to learn the address mask if a router
   is down when the host boots up, a router MAY broadcast a gratuitous
   ICMP Address Mask Reply on each of its logical interfaces after it
   has configured its own address masks. However, this feature can be
   dangerous in environments that use variable length address masks.
   Therefore, if this feature is implemented, gratuitous Address Mask
   Replies MUST NOT be broadcast over any logical interface(s) which
   either:

   o Are not configured to send gratuitous Address Mask Replies. Each
     logical interface MUST have a configuration parameter controlling
     this, and that parameter MUST default to not sending the
     gratuitous Address Mask Replies.

   o Share subsuming (but not identical) network prefixes and physical
     interface.

   The { <Network-prefix>, -1 } form of the IP broadcast address MUST be
   used for broadcast Address Mask Replies.

   DISCUSSION
      The ability to disable sending Address Mask Replies by routers is
      required at a few sites that intentionally lie to their hosts
      about the address mask. The need for this is expected to go away

as more and more hosts become compliant with the Host Requirements standards.

The reason for both the second bullet above and the requirement about which IP broadcast address to use is to prevent problems when multiple IP network prefixes are in use on the same physical network.

4.3.3.10 Router Advertisement and Solicitations

An IP router MUST support the router part of the ICMP Router Discovery Protocol [INTERNET:13] on all connected networks on which the router supports either IP multicast or IP broadcast addressing. The implementation MUST include all the configuration variables specified for routers, with the specified defaults.

DISCUSSION
    Routers are not required to implement the host part of the ICMP Router Discovery Protocol, but might find it useful for operation while IP forwarding is disabled (i.e., when operating as a host).

DISCUSSION We note that it is quite common for hosts to use RIP Version 1 as the router discovery protocol. Such hosts listen to RIP traffic and use and use information extracted from that traffic to discover routers and to make decisions as to which router to use as a first-hop router for a given destination. While this behavior is discouraged, it is still common and implementors should be aware of it.

4.4 INTERNET GROUP MANAGEMENT PROTOCOL - IGMP

IGMP [INTERNET:4] is a protocol used between hosts and multicast routers on a single physical network to establish hosts' membership in particular multicast groups. Multicast routers use this information, in conjunction with a multicast routing protocol, to support IP multicast forwarding across the Internet.

A router SHOULD implement the host part of IGMP.

5. INTERNET LAYER - FORWARDING

5.1 INTRODUCTION

   This section describes the process of forwarding packets.

5.2 FORWARDING WALK-THROUGH

   There is no separate specification of the forwarding function in IP.
   Instead, forwarding is covered by the protocol specifications for the
   internet layer protocols ([INTERNET:1], [INTERNET:2], [INTERNET:3],
   [INTERNET:8], and [ROUTE:11]).

5.2.1 Forwarding Algorithm

   Since none of the primary protocol documents describe the forwarding
   algorithm in any detail, we present it here. This is just a general
   outline, and omits important details, such as handling of congestion,
   that are dealt with in later sections.

   It is not required that an implementation follow exactly the
   algorithms given in sections [5.2.1.1], [5.2.1.2], and [5.2.1.3].
   Much of the challenge of writing router software is to maximize the
   rate at which the router can forward packets while still achieving
   the same effect of the algorithm. Details of how to do that are
   beyond the scope of this document, in part because they are heavily
   dependent on the architecture of the router. Instead, we merely
   point out the order dependencies among the steps:

   (1) A router MUST verify the IP header, as described in section
       [5.2.2], before performing any actions based on the contents of
       the header. This allows the router to detect and discard bad
       packets before the expenditure of other resources.

   (2) Processing of certain IP options requires that the router insert
       its IP address into the option. As noted in Section [5.2.4],
       the address inserted MUST be the address of the logical
       interface on which the packet is sent or the router's router-id
       if the packet is sent over an unnumbered interface. Thus,
       processing of these options cannot be completed until after the
       output interface is chosen.

   (3) The router cannot check and decrement the TTL before checking
       whether the packet should be delivered to the router itself, for
       reasons mentioned in Section [4.2.2.9].

   (4) More generally, when a packet is delivered locally to the router,
       its IP header MUST NOT be modified in any way (except that a

router may be required to insert a timestamp into any Timestamp
options in the IP header). Thus, before the router determines
whether the packet is to be delivered locally to the router, it
cannot update the IP header in any way that it is not prepared
to undo.

5.2.1.1 General

This section covers the general forwarding algorithm. This algorithm
applies to all forms of packets to be forwarded: unicast, multicast,
and broadcast.

(1) The router receives the IP packet (plus additional information
     about it, as described in Section [3.1]) from the Link Layer.

(2) The router validates the IP header, as described in Section
     [5.2.2]. Note that IP reassembly is not done, except on IP
     fragments to be queued for local delivery in step (4).

(3) The router performs most of the processing of any IP options. As
     described in Section [5.2.4], some IP options require additional
     processing after the routing decision has been made.

(4) The router examines the destination IP address of the IP
     datagram, as described in Section [5.2.3], to determine how it
     should continue to process the IP datagram. There are three
     possibilities:

     o The IP datagram is destined for the router, and should be
        queued for local delivery, doing reassembly if needed.

     o The IP datagram is not destined for the router, and should be
        queued for forwarding.

     o The IP datagram should be queued for forwarding, but (a copy)
        must also be queued for local delivery.

5.2.1.2 Unicast

Since the local delivery case is well covered by [INTRO:2], the
following assumes that the IP datagram was queued for forwarding. If
the destination is an IP unicast address:

(5) The forwarder determines the next hop IP address for the packet,
     usually by looking up the packet's destination in the router's
     routing table. This procedure is described in more detail in
     Section [5.2.4]. This procedure also decides which network

interface should be used to send the packet.

(6) The forwarder verifies that forwarding the packet is permitted.
    The source and destination addresses should be valid, as
    described in Section [5.3.7] and Section [5.3.4] If the router
    supports administrative constraints on forwarding, such as those
    described in Section [5.3.9], those constraints must be
    satisfied.

(7) The forwarder decrements (by at least one) and checks the
    packet's TTL, as described in Section [5.3.1].

(8) The forwarder performs any IP option processing that could not be
    completed in step 3.

(9) The forwarder performs any necessary IP fragmentation, as
    described in Section [4.2.2.7]. Since this step occurs after
    outbound interface selection (step 5), all fragments of the same
    datagram will be transmitted out the same interface.

(10) The forwarder determines the Link Layer address of the packet's
     next hop. The mechanisms for doing this are Link Layer-
     dependent (see chapter 3).

(11) The forwarder encapsulates the IP datagram (or each of the
     fragments thereof) in an appropriate Link Layer frame and queues
     it for output on the interface selected in step 5.

(12) The forwarder sends an ICMP redirect if necessary, as described
     in Section [4.3.3.2].

5.2.1.3 Multicast

   If the destination is an IP multicast, the following steps are taken.

   Note that the main differences between the forwarding of IP unicasts
   and the forwarding of IP multicasts are

   o IP multicasts are usually forwarded based on both the datagram's
     source and destination IP addresses,

   o IP multicast uses an expanding ring search,

   o IP multicasts are forwarded as Link Level multicasts, and

   o ICMP errors are never sent in response to IP multicast datagrams.

Note that the forwarding of IP multicasts is still somewhat
experimental. As a result, the algorithm presented below is not
mandatory, and is provided as an example only.

(5a) Based on the IP source and destination addresses found in the
     datagram header, the router determines whether the datagram has
     been received on the proper interface for forwarding. If not,
     the datagram is dropped silently. The method for determining
     the proper receiving interface depends on the multicast routing
     algorithm(s) in use. In one of the simplest algorithms, reverse
     path forwarding (RPF), the proper interface is the one that
     would be used to forward unicasts back to the datagram source.

(6a) Based on the IP source and destination addresses found in the
     datagram header, the router determines the datagram's outgoing
     interfaces. To implement IP multicast's expanding ring search
     (see [INTERNET:4]) a minimum TTL value is specified for each
     outgoing interface. A copy of the multicast datagram is
     forwarded out each outgoing interface whose minimum TTL value is
     less than or equal to the TTL value in the datagram header, by
     separately applying the remaining steps on each such interface.

(7a) The router decrements the packet's TTL by one.

(8a) The forwarder performs any IP option processing that could not
     be completed in step (3).

(9a) The forwarder performs any necessary IP fragmentation, as
     described in Section [4.2.2.7].

(10a) The forwarder determines the Link Layer address to use in the
     Link Level encapsulation. The mechanisms for doing this are
     Link Layer-dependent. On LANs a Link Level multicast or
     broadcast is selected, as an algorithmic translation of the
     datagrams' IP multicast address. See the various IP-over-xxx
     specifications for more details.

(11a) The forwarder encapsulates the packet (or each of the fragments
     thereof) in an appropriate Link Layer frame and queues it for
     output on the appropriate interface.

5.2.2 IP Header Validation

   Before a router can process any IP packet, it MUST perform a the
   following basic validity checks on the packet's IP header to ensure
   that the header is meaningful. If the packet fails any of the
   following tests, it MUST be silently discarded, and the error SHOULD
   be logged.

   (1) The packet length reported by the Link Layer must be large enough
        to hold the minimum length legal IP datagram (20 bytes).

   (2) The IP checksum must be correct.

   (3) The IP version number must be 4. If the version number is not 4
        then the packet may be another version of IP, such as IPng or
        ST-II.

   (4) The IP header length field must be large enough to hold the
        minimum length legal IP datagram (20 bytes = 5 words).

   (5) The IP total length field must be large enough to hold the IP
        datagram header, whose length is specified in the IP header
        length field.

   A router MUST NOT have a configuration option that allows disabling
   any of these tests.

   If the packet passes the second and third tests, the IP header length
   field is at least 4, and both the IP total length field and the
   packet length reported by the Link Layer are at least 16 then,
   despite the above rule, the router MAY respond with an ICMP Parameter
   Problem message, whose pointer points at the IP header length field
   (if it failed the fourth test) or the IP total length field (if it
   failed the fifth test). However, it still MUST discard the packet
   and still SHOULD log the error.

   These rules (and this entire document) apply only to version 4 of the
   Internet Protocol. These rules should not be construed as
   prohibiting routers from supporting other versions of IP.
   Furthermore, if a router can truly classify a packet as being some
   other version of IP then it ought not treat that packet as an error
   packet within the context of this memo.

   IMPLEMENTATION
      It is desirable for purposes of error reporting, though not always
      entirely possible, to determine why a header was invalid. There
      are four possible reasons:

       o The Link Layer truncated the IP header

       o The datagram is using a version of IP other than the standard
          one (version 4).

       o The IP header has been corrupted in transit.

       o The sender generated an illegal IP header.

       It is probably desirable to perform the checks in the order
       listed, since we believe that this ordering is most likely to
       correctly categorize the cause of the error. For purposes of
       error reporting, it may also be desirable to check if a packet
       that fails these tests has an IP version number indicating IPng or
       ST-II; these should be handled according to their respective
       specifications.

   Additionally, the router SHOULD verify that the packet length
   reported by the Link Layer is at least as large as the IP total
   length recorded in the packet's IP header. If it appears that the
   packet has been truncated, the packet MUST be discarded, the error
   SHOULD be logged, and the router SHOULD respond with an ICMP
   Parameter Problem message whose pointer points at the IP total length
   field.

   DISCUSSION
      Because any higher layer protocol that concerns itself with data
      corruption will detect truncation of the packet data when it
      reaches its final destination, it is not absolutely necessary for
      routers to perform the check suggested above to maintain protocol
      correctness. However, by making this check a router can simplify
      considerably the task of determining which hop in the path is
      truncating the packets. It will also reduce the expenditure of
      resources down-stream from the router in that down-stream systems
      will not need to deal with the packet.

   Finally, if the destination address in the IP header is not one of
   the addresses of the router, the router SHOULD verify that the packet
   does not contain a Strict Source and Record Route option. If a
   packet fails this test (if it contains a strict source route option),
   the router SHOULD log the error and SHOULD respond with an ICMP
   Parameter Problem error with the pointer pointing at the offending
   packet's IP destination address.

   DISCUSSION
      Some people might suggest that the router should respond with a
      Bad Source Route message instead of a Parameter Problem message.
      However, when a packet fails this test, it usually indicates a

        protocol error by the previous hop router, whereas Bad Source
        Route would suggest that the source host had requested a
        nonexistent or broken path through the network.

5.2.3 Local Delivery Decision

        When a router receives an IP packet, it must decide whether the
        packet is addressed to the router (and should be delivered locally)
        or the packet is addressed to another system (and should be handled
        by the forwarder). There is also a hybrid case, where certain IP
        broadcasts and IP multicasts are both delivered locally and
        forwarded. A router MUST determine which of the these three cases
        applies using the following rules.

        o An unexpired source route option is one whose pointer value does
          not point past the last entry in the source route. If the packet
          contains an unexpired source route option, the pointer in the
          option is advanced until either the pointer does point past the
          last address in the option or else the next address is not one of
          the router's own addresses. In the latter (normal) case, the
          packet is forwarded (and not delivered locally) regardless of the
          rules below.

        o The packet is delivered locally and not considered for forwarding
          in the following cases:

          - The packet's destination address exactly matches one of the
            router's IP addresses,

          - The packet's destination address is a limited broadcast address
            ({-1, -1}), or

          - The packet's destination is an IP multicast address which is
            never forwarded (such as 224.0.0.1 or 224.0.0.2) and (at least)
            one of the logical interfaces associated with the physical
            interface on which the packet arrived is a member of the
            destination multicast group.

        o The packet is passed to the forwarder AND delivered locally in the
          following cases:

          - The packet's destination address is an IP broadcast address that
            addresses at least one of the router's logical interfaces but
            does not address any of the logical interfaces associated with
            the physical interface on which the packet arrived

      - The packet's destination is an IP multicast address which is
        permitted to be forwarded (unlike 224.0.0.1 and 224.0.0.2) and
        (at least) one of the logical interfaces associated with the
        physical interface on which the packet arrived is a member of
        the destination multicast group.

   o The packet is delivered locally if the packet's destination address
     is an IP broadcast address (other than a limited broadcast
     address) that addresses at least one of the logical interfaces
     associated with the physical interface on which the packet
     arrived. The packet is ALSO passed to the forwarder unless the
     link on which the packet arrived uses an IP encapsulation that
     does not encapsulate broadcasts differently than unicasts (e.g.,
     by using different Link Layer destination addresses).

   o The packet is passed to the forwarder in all other cases.

   DISCUSSION
      The purpose of the requirement in the last sentence of the fourth
      bullet is to deal with a directed broadcast to another network
      prefix on the same physical cable. Normally, this works as
      expected: the sender sends the broadcast to the router as a Link
      Layer unicast. The router notes that it arrived as a unicast, and
      therefore must be destined for a different network prefix than the
      sender sent it on. Therefore, the router can safely send it as a
      Link Layer broadcast out the same (physical) interface over which
      it arrived. However, if the router can't tell whether the packet
      was received as a Link Layer unicast, the sentence ensures that
      the router does the safe but wrong thing rather than the unsafe
      but right thing.

   IMPLEMENTATION
      As described in Section [5.3.4], packets received as Link Layer
      broadcasts are generally not forwarded. It may be advantageous to
      avoid passing to the forwarder packets it would later discard
      because of the rules in that section.

      Some Link Layers (either because of the hardware or because of
      special code in the drivers) can deliver to the router copies of
      all Link Layer broadcasts and multicasts it transmits. Use of
      this feature can simplify the implementation of cases where a
      packet has to both be passed to the forwarder and delivered
      locally, since forwarding the packet will automatically cause the
      router to receive a copy of the packet that it can then deliver
      locally. One must use care in these circumstances to prevent
      treating a received loop-back packet as a normal packet that was
      received (and then being subject to the rules of forwarding,
      etc.).

   Even without such a Link Layer, it is of course hardly necessary
   to make a copy of an entire packet to queue it both for forwarding
   and for local delivery, though care must be taken with fragments,
   since reassembly is performed on locally delivered packets but not
   on forwarded packets. One simple scheme is to associate a flag
   with each packet on the router's output queue that indicates
   whether it should be queued for local delivery after it has been
   sent.

5.2.4 Determining the Next Hop Address

   When a router is going to forward a packet, it must determine whether
   it can send it directly to its destination, or whether it needs to
   pass it through another router. If the latter, it needs to determine
   which router to use. This section explains how these determinations
   are made.

   This section makes use of the following definitions:

   o LSRR - IP Loose Source and Record Route option

   o SSRR - IP Strict Source and Record Route option

   o Source Route Option - an LSRR or an SSRR

   o Ultimate Destination Address - where the packet is being sent to:
     the last address in the source route of a source-routed packet, or
     the destination address in the IP header of a non-source-routed
     packet

   o Adjacent - reachable without going through any IP routers

   o Next Hop Address - the IP address of the adjacent host or router to
     which the packet should be sent next

   o IP Destination Address - the ultimate destination address, except
     in source routed packets, where it is the next address specified
     in the source route

   o Immediate Destination - the node, System, router, end-system, or
     whatever that is addressed by the IP Destination Address.

5.2.4.1 IP Destination Address

   If:

   o the destination address in the IP header is one of the addresses of
     the router,

   o the packet contains a Source Route Option, and

   o the pointer in the Source Route Option does not point past the end
     of the option,

   then the next IP Destination Address is the address pointed at by the
   pointer in that option. If:

   o the destination address in the IP header is one of the addresses of
     the router,

   o the packet contains a Source Route Option, and

   o the pointer in the Source Route Option points past the end of the
     option,

   then the message is addressed to the system analyzing the message.

   A router MUST use the IP Destination Address, not the Ultimate
   Destination Address (the last address in the source route option),
   when determining how to handle a packet.

   It is an error for more than one source route option to appear in a
   datagram. If it receives such a datagram, it SHOULD discard the
   packet and reply with an ICMP Parameter Problem message whose pointer
   points at the beginning of the second source route option.

5.2.4.2 Local/Remote Decision

   After it has been determined that the IP packet needs to be forwarded
   according to the rules specified in Section [5.2.3], the following
   algorithm MUST be used to determine if the Immediate Destination is
   directly accessible (see [INTERNET:2]).

   (1) For each network interface that has not been assigned any IP
       address (the unnumbered lines as described in Section [2.2.7]),
       compare the router-id of the other end of the line to the IP
       Destination Address. If they are exactly equal, the packet can
       be transmitted through this interface.

DISCUSSION
   In other words, the router or host at the remote end of the line
   is the destination of the packet or is the next step in the source
   route of a source routed packet.

(2) If no network interface has been selected in the first step, for
    each IP address assigned to the router:

(a) isolate the network prefix used by the interface.

IMPLEMENTATION
   The result of this operation will usually have been computed and
   saved during initialization.

(b) Isolate the corresponding set of bits from the IP Destination
    Address of the packet.

(c) Compare the resulting network prefixes. If they are equal to
    each other, the packet can be transmitted through the
    corresponding network interface.

(3) If the destination was neither the router-id of a neighbor on an
    unnumbered interface nor a member of a directly connected network
    prefix, the IP Destination is accessible only through some other
    router. The selection of the router and the next hop IP address
    is described in Section [5.2.4.3]. In the case of a host that is
    not also a router, this may be the configured default router.

Ongoing work in the IETF [ARCH:9, NRHP] considers some cases such as
when multiple IP (sub)networks are overlaid on the same link layer
network. Barring policy restrictions, hosts and routers using a
common link layer network can directly communicate even if they are
not in the same IP (sub)network, if there is adequate information
present. The Next Hop Routing Protocol (NHRP) enables IP entities to
determine the "optimal" link layer address to be used to traverse
such a link layer network towards a remote destination.

(4) If the selected "next hop" is reachable through an interface
configured to use NHRP, then the following additional steps apply:

  (a) Compare the IP Destination Address to the destination addresses
      in the NHRP cache. If the address is in the cache, then send
      the datagram to the corresponding cached link layer address.
  (b) If the address is not in the cache, then construct an NHRP
      request packet containing the IP Destination Address. This
      message is sent to the NHRP server configured for that
      interface. This may be a logically separate process or entity
      in the router itself.

         (c) The NHRP server will respond with the proper link layer address
             to use to transmit the datagram and subsequent datagrams to the
             same destination. The system MAY transmit the datagram(s) to
             the traditional "next hop" router while awaiting the NHRP reply.

   5.2.4.3 Next Hop Address

      EDITORS+COMMENTS
         The router applies the algorithm in the previous section to
         determine if the IP Destination Address is adjacent. If so, the
         next hop address is the same as the IP Destination Address.
         Otherwise, the packet must be forwarded through another router to
         reach its Immediate Destination. The selection of this router is
         the topic of this section.

         If the packet contains an SSRR, the router MUST discard the packet
         and reply with an ICMP Bad Source Route error. Otherwise, the
         router looks up the IP Destination Address in its routing table to
         determine an appropriate next hop address.

      DISCUSSION
         Per the IP specification, a Strict Source Route must specify a
         sequence of nodes through which the packet must traverse; the
         packet must go from one node of the source route to the next,
         traversing intermediate networks only. Thus, if the router is not
         adjacent to the next step of the source route, the source route
         can not be fulfilled. Therefore, the router rejects such with an
         ICMP Bad Source Route error.

   The goal of the next-hop selection process is to examine the entries
   in the router's Forwarding Information Base (FIB) and select the best
   route (if there is one) for the packet from those available in the
   FIB.

   Conceptually, any route lookup algorithm starts out with a set of
   candidate routes that consists of the entire contents of the FIB.
   The algorithm consists of a series of steps that discard routes from
   the set. These steps are referred to as Pruning Rules. Normally,
   when the algorithm terminates there is exactly one route remaining in
   the set. If the set ever becomes empty, the packet is discarded
   because the destination is unreachable. It is also possible for the
   algorithm to terminate when more than one route remains in the set.
   In this case, the router may arbitrarily discard all but one of them,
   or may perform "load-splitting" by choosing whichever of the routes
   has been least recently used.

   With the exception of rule 3 (Weak TOS), a router MUST use the
   following Pruning Rules when selecting a next hop for a packet. If a

router does consider TOS when making next-hop decisions, the Rule 3
must be applied in the order indicated below. These rules MUST be
(conceptually) applied to the FIB in the order that they are
presented. (For some historical perspective, additional pruning
rules, and other common algorithms in use, see Appendix E.)

DISCUSSION
   Rule 3 is optional in that Section [5.3.2] says that a router only
   SHOULD consider TOS when making forwarding decisions.

(1) Basic Match
      This rule discards any routes to destinations other than the
      IP Destination Address of the packet. For example, if a
      packet's IP Destination Address is 10.144.2.5, this step
      would discard a route to net 128.12.0.0/16 but would retain
      any routes to the network prefixes 10.0.0.0/8 and
      10.144.0.0/16, and any default routes.

      More precisely, we assume that each route has a destination
      attribute, called route.dest and a corresponding prefix
      length, called route.length, to specify which bits of
      route.dest are significant. The IP Destination Address of
      the packet being forwarded is ip.dest. This rule discards
      all routes from the set of candidates except those for which
      the most significant route.length bits of route.dest and
      ip.dest are equal.

      For example, if a packet's IP Destination Address is
      10.144.2.5 and there are network prefixes 10.144.1.0/24,
      10.144.2.0/24, and 10.144.3.0/24, this rule would keep only
      10.144.2.0/24; it is the only route whose prefix has the same
      value as the corresponding bits in the IP Destination Address
      of the packet.

(2) Longest Match
      Longest Match is a refinement of Basic Match, described
      above. After performing Basic Match pruning, the algorithm
      examines the remaining routes to determine which among them
      have the largest route.length values. All except these are
      discarded.

      For example, if a packet's IP Destination Address is
      10.144.2.5 and there are network prefixes 10.144.2.0/24,
      10.144.0.0/16, and 10.0.0.0/8, then this rule would keep only
      the first (10.144.2.0/24) because its prefix length is
      longest.

   (3) Weak TOS
        Each route has a type of service attribute, called route.tos,
        whose possible values are assumed to be identical to those
        used in the TOS field of the IP header. Routing protocols
        that distribute TOS information fill in route.tos
        appropriately in routes they add to the FIB; routes from
        other routing protocols are treated as if they have the
        default TOS (0000). The TOS field in the IP header of the
        packet being routed is called ip.tos.

        The set of candidate routes is examined to determine if it
        contains any routes for which route.tos = ip.tos. If so, all
        routes except those for which route.tos = ip.tos are
        discarded. If not, all routes except those for which
        route.tos = 0000 are discarded from the set of candidate
        routes.

        Additional discussion of routing based on Weak TOS may be
        found in [ROUTE:11].

   DISCUSSION
      The effect of this rule is to select only those routes that have a
      TOS that matches the TOS requested in the packet. If no such
      routes exist then routes with the default TOS are considered.
      Routes with a non-default TOS that is not the TOS requested in the
      packet are never used, even if such routes are the only available
      routes that go to the packet's destination.

   (4) Best Metric
        Each route has a metric attribute, called route.metric, and a
        routing domain identifier, called route.domain. Each member
        of the set of candidate routes is compared with each other
        member of the set. If route.domain is equal for the two
        routes and route.metric is strictly inferior for one when
        compared with the other, then the one with the inferior metric
        is discarded from the set. The determination of inferior is
        usually by a simple arithmetic comparison, though some
        protocols may have structured metrics requiring more complex
        comparisons.

   (5) Vendor Policy
        Vendor Policy is sort of a catch-all to make up for the fact
        that the previously listed rules are often inadequate to
        choose from the possible routes. Vendor Policy pruning rules
        are extremely vendor-specific. See section [5.2.4.4].

   This algorithm has two distinct disadvantages. Presumably, a
   router implementor might develop techniques to deal with these

disadvantages and make them a part of the Vendor Policy pruning
rule.

(1) IS-IS and OSPF route classes are not directly handled.

(2) Path properties other than type of service (e.g., MTU) are
     ignored.

It is also worth noting a deficiency in the way that TOS is
supported: routing protocols that support TOS are implicitly
preferred when forwarding packets that have non-zero TOS values.

The Basic Match and Longest Match pruning rules generalize the
treatment of a number of particular types of routes. These routes
are selected in the following, decreasing, order of preference:

(1) Host Route: This is a route to a specific end system.

(2) Hierarchical Network Prefix Routes: This is a route to a
     particular network prefix. Note that the FIB may contain
     several routes to network prefixes that subsume each other
     (one prefix is the other prefix with additional bits). These
     are selected in order of decreasing prefix length.

(5) Default Route: This is a route to all networks for which there
     are no explicit routes. It is by definition the route whose
     prefix length is zero.

If, after application of the pruning rules, the set of routes is
empty (i.e., no routes were found), the packet MUST be discarded
and an appropriate ICMP error generated (ICMP Bad Source Route if
the IP Destination Address came from a source route option;
otherwise, whichever of ICMP Destination Host Unreachable or
Destination Network Unreachable is appropriate, as described in
Section [4.3.3.1]).

5.2.4.4 Administrative Preference

One suggested mechanism for the Vendor Policy Pruning Rule is to
use administrative preference, which is a simple prioritization
algorithm. The idea is to manually prioritize the routes that one
might need to select among.

Each route has associated with it a preference value, based on
various attributes of the route (specific mechanisms for assignment
of preference values are suggested below). This preference value
is an integer in the range [0..255], with zero being the most
preferred and 254 being the least preferred. 255 is a special

value that means that the route should never be used. The first
step in the Vendor Policy pruning rule discards all but the most
preferable routes (and always discards routes whose preference
value is 255).

This policy is not safe in that it can easily be misused to create
routing loops. Since no protocol ensures that the preferences
configured for a router is consistent with the preferences
configured in its neighbors, network managers must exercise care in
configuring preferences.

o Address Match
  It is useful to be able to assign a single preference value to
  all routes (learned from the same routing domain) to any of a
  specified set of destinations, where the set of destinations is
  all destinations that match a specified network prefix.

o Route Class
  For routing protocols which maintain the distinction, it is
  useful to be able to assign a single preference value to all
  routes (learned from the same routing domain) which have a
  particular route class (intra-area, inter-area, external with
  internal metrics, or external with external metrics).

o Interface
  It is useful to be able to assign a single preference value to
  all routes (learned from a particular routing domain) that would
  cause packets to be routed out a particular logical interface on
  the router (logical interfaces generally map one-to-one onto the
  router's network interfaces, except that any network interface
  that has multiple IP addresses will have multiple logical
  interfaces associated with it).

o Source router
  It is useful to be able to assign a single preference value to
  all routes (learned from the same routing domain) that were
  learned from any of a set of routers, where the set of routers
  are those whose updates have a source address that match a
  specified network prefix.

o Originating AS
  For routing protocols which provide the information, it is
  useful to be able to assign a single preference value to all
  routes (learned from a particular routing domain) which
  originated in another particular routing domain. For BGP
  routes, the originating AS is the first AS listed in the route's
  AS_PATH attribute. For OSPF external routes, the originating AS
  may be considered to be the low order 16 bits of the route's

external route tag if the tag's Automatic bit is set and the
tag's Path Length is not equal to 3.

o External route tag
  It is useful to be able to assign a single preference value to
  all OSPF external routes (learned from the same routing domain)
  whose external route tags match any of a list of specified
  values. Because the external route tag may contain a structured
  value, it may be useful to provide the ability to match
  particular subfields of the tag.

o AS path
  It may be useful to be able to assign a single preference value
  to all BGP routes (learned from the same routing domain) whose
  AS path "matches" any of a set of specified values. It is not
  yet clear exactly what kinds of matches are most useful. A
  simple option would be to allow matching of all routes for which
  a particular AS number appears (or alternatively, does not
  appear) anywhere in the route's AS_PATH attribute. A more
  general but somewhat more difficult alternative would be to
  allow matching all routes for which the AS path matches a
  specified regular expression.

5.2.4.5 Load Splitting

   At the end of the Next-hop selection process, multiple routes may
   still remain. A router has several options when this occurs. It
   may arbitrarily discard some of the routes. It may reduce the
   number of candidate routes by comparing metrics of routes from
   routing domains that are not considered equivalent. It may retain
   more than one route and employ a load-splitting mechanism to divide
   traffic among them. Perhaps the only thing that can be said about
   the relative merits of the options is that load-splitting is useful
   in some situations but not in others, so a wise implementor who
   implements load-splitting will also provide a way for the network
   manager to disable it.

5.2.5 Unused IP Header Bits: RFC-791 Section 3.1

   The IP header contains several reserved bits, in the Type of
   Service field and in the Flags field. Routers MUST NOT drop
   packets merely because one or more of these reserved bits has a
   non-zero value.

   Routers MUST ignore and MUST pass through unchanged the values of
   these reserved bits. If a router fragments a packet, it MUST copy
   these bits into each fragment.

DISCUSSION
     Future revisions to the IP protocol may make use of these unused
     bits. These rules are intended to ensure that these revisions can
     be deployed without having to simultaneously upgrade all routers
     in the Internet.

5.2.6 Fragmentation and Reassembly: RFC-791 Section 3.2

   As was discussed in Section [4.2.2.7], a router MUST support IP
   fragmentation.

   A router MUST NOT reassemble any datagram before forwarding it.

   DISCUSSION
     A few people have suggested that there might be some topologies
     where reassembly of transit datagrams by routers might improve
     performance. The fact that fragments may take different paths to
     the destination precludes safe use of such a feature.

     Nothing in this section should be construed to control or limit
     fragmentation or reassembly performed as a link layer function by
     the router.

     Similarly, if an IP datagram is encapsulated in another IP
     datagram (e.g., it is tunnelled), that datagram is in turn
     fragmented, the fragments must be reassembled in order to forward
     the original datagram. This section does not preclude this.

5.2.7 Internet Control Message Protocol - ICMP

   General requirements for ICMP were discussed in Section [4.3]. This
   section discusses ICMP messages that are sent only by routers.

5.2.7.1 Destination Unreachable

   The ICMP Destination Unreachable message is sent by a router in
   response to a packet which it cannot forward because the destination
   (or next hop) is unreachable or a service is unavailable. Examples
   of such cases include a message addressed to a host which is not
   there and therefore does not respond to ARP requests, and messages
   addressed to network prefixes for which the router has no valid
   route.

   A router MUST be able to generate ICMP Destination Unreachable
   messages and SHOULD choose a response code that most closely matches
   the reason the message is being generated.

   The following codes are defined in [INTERNET:8] and [INTRO:2]:

   0 = Network Unreachable - generated by a router if a forwarding path
          (route) to the destination network is not available;

   1 = Host Unreachable - generated by a router if a forwarding path
          (route) to the destination host on a directly connected network
          is not available (does not respond to ARP);

   2 = Protocol Unreachable - generated if the transport protocol
          designated in a datagram is not supported in the transport layer
          of the final destination;

   3 = Port Unreachable - generated if the designated transport protocol
          (e.g., UDP) is unable to demultiplex the datagram in the
          transport layer of the final destination but has no protocol
          mechanism to inform the sender;

   4 = Fragmentation Needed and DF Set - generated if a router needs to
          fragment a datagram but cannot since the DF flag is set;

   5 = Source Route Failed - generated if a router cannot forward a
          packet to the next hop in a source route option;

   6 = Destination Network Unknown - This code SHOULD NOT be generated
          since it would imply on the part of the router that the
          destination network does not exist (net unreachable code 0
          SHOULD be used in place of code 6);

   7 = Destination Host Unknown - generated only when a router can
          determine (from link layer advice) that the destination host
          does not exist;

   11 = Network Unreachable For Type Of Service - generated by a router
          if a forwarding path (route) to the destination network with the
          requested or default TOS is not available;

   12 = Host Unreachable For Type Of Service - generated if a router
          cannot forward a packet because its route(s) to the destination
          do not match either the TOS requested in the datagram or the
          default TOS (0).

   The following additional codes are hereby defined:

   13 = Communication Administratively Prohibited - generated if a
          router cannot forward a packet due to administrative filtering;

   14 = Host Precedence Violation. Sent by the first hop router to a
          host to indicate that a requested precedence is not permitted
          for the particular combination of source/destination host or

network, upper layer protocol, and source/destination port;

15 = Precedence cutoff in effect. The network operators have imposed
     a minimum level of precedence required for operation, the
     datagram was sent with a precedence below this level;

NOTE: [INTRO:2] defined Code 8 for source host isolated. Routers
SHOULD NOT generate Code 8; whichever of Codes 0 (Network
Unreachable) and 1 (Host Unreachable) is appropriate SHOULD be used
instead. [INTRO:2] also defined Code 9 for communication with
destination network administratively prohibited and Code 10 for
communication with destination host administratively prohibited.
These codes were intended for use by end-to-end encryption devices
used by U.S military agencies. Routers SHOULD use the newly defined
Code 13 (Communication Administratively Prohibited) if they
administratively filter packets.

Routers MAY have a configuration option that causes Code 13
(Communication Administratively Prohibited) messages not to be
generated. When this option is enabled, no ICMP error message is
sent in response to a packet that is dropped because its forwarding
is administratively prohibited.

Similarly, routers MAY have a configuration option that causes Code
14 (Host Precedence Violation) and Code 15 (Precedence Cutoff in
Effect) messages not to be generated. When this option is enabled,
no ICMP error message is sent in response to a packet that is dropped
because of a precedence violation.

Routers MUST use Host Unreachable or Destination Host Unknown codes
whenever other hosts on the same destination network might be
reachable; otherwise, the source host may erroneously conclude that
all hosts on the network are unreachable, and that may not be the
case.

[INTERNET:14] describes a slight modification the form of Destination
Unreachable messages containing Code 4 (Fragmentation needed and DF
set). A router MUST use this modified form when originating Code 4
Destination Unreachable messages.

5.2.7.2 Redirect

The ICMP Redirect message is generated to inform a local host the it
should use a different next hop router for a certain class of
traffic.

Routers MUST NOT generate the Redirect for Network or Redirect for
Network and Type of Service messages (Codes 0 and 2) specified in

[INTERNET:8]. Routers MUST be able to generate the Redirect for Host
message (Code 1) and SHOULD be able to generate the Redirect for Type
of Service and Host message (Code 3) specified in [INTERNET:8].

DISCUSSION
   If the directly connected network is not subnetted (in the
   classical sense), a router can normally generate a network
   Redirect that applies to all hosts on a specified remote network.
   Using a network rather than a host Redirect may economize slightly
   on network traffic and on host routing table storage. However,
   the savings are not significant, and subnets create an ambiguity
   about the subnet mask to be used to interpret a network Redirect.
   In a CIDR environment, it is difficult to specify precisely the
   cases in which network Redirects can be used. Therefore, routers
   must send only host (or host and type of service) Redirects.

A Code 3 (Redirect for Host and Type of Service) message is generated
when the packet provoking the redirect has a destination for which
the path chosen by the router would depend (in part) on the TOS
requested.

Routers that can generate Code 3 redirects (Host and Type of Service)
MUST have a configuration option (which defaults to on) to enable
Code 1 (Host) redirects to be substituted for Code 3 redirects. A
router MUST send a Code 1 Redirect in place of a Code 3 Redirect if
it has been configured to do so.

If a router is not able to generate Code 3 Redirects then it MUST
generate Code 1 Redirects in situations where a Code 3 Redirect is
called for.

Routers MUST NOT generate a Redirect Message unless all the following
conditions are met:

o The packet is being forwarded out the same physical interface that
   it was received from,

o The IP source address in the packet is on the same Logical IP
   (sub)network as the next-hop IP address, and

o The packet does not contain an IP source route option.

The source address used in the ICMP Redirect MUST belong to the same
logical (sub)net as the destination address.

A router using a routing protocol (other than static routes) MUST NOT
consider paths learned from ICMP Redirects when forwarding a packet.
If a router is not using a routing protocol, a router MAY have a

configuration that, if set, allows the router to consider routes
learned through ICMP Redirects when forwarding packets.

   DISCUSSION
      ICMP Redirect is a mechanism for routers to convey routing
      information to hosts. Routers use other mechanisms to learn
      routing information, and therefore have no reason to obey
      redirects. Believing a redirect which contradicted the router's
      other information would likely create routing loops.

      On the other hand, when a router is not acting as a router, it
      MUST comply with the behavior required of a host.

5.2.7.3 Time Exceeded

   A router MUST generate a Time Exceeded message Code 0 (In Transit)
   when it discards a packet due to an expired TTL field. A router MAY
   have a per-interface option to disable origination of these messages
   on that interface, but that option MUST default to allowing the
   messages to be originated.

5.2.8 INTERNET GROUP MANAGEMENT PROTOCOL - IGMP

   IGMP [INTERNET:4] is a protocol used between hosts and multicast
   routers on a single physical network to establish hosts' membership
   in particular multicast groups. Multicast routers use this
   information, in conjunction with a multicast routing protocol, to
   support IP multicast forwarding across the Internet.

   A router SHOULD implement the multicast router part of IGMP.

5.3 SPECIFIC ISSUES

5.3.1 Time to Live (TTL)

   The Time-to-Live (TTL) field of the IP header is defined to be a
   timer limiting the lifetime of a datagram. It is an 8-bit field and
   the units are seconds. Each router (or other module) that handles a
   packet MUST decrement the TTL by at least one, even if the elapsed
   time was much less than a second. Since this is very often the case,
   the TTL is effectively a hop count limit on how far a datagram can
   propagate through the Internet.

   When a router forwards a packet, it MUST reduce the TTL by at least
   one. If it holds a packet for more than one second, it MAY decrement
   the TTL by one for each second.

   If the TTL is reduced to zero (or less), the packet MUST be
   discarded, and if the destination is not a multicast address the
   router MUST send an ICMP Time Exceeded message, Code 0 (TTL Exceeded
   in Transit) message to the source. Note that a router MUST NOT
   discard an IP unicast or broadcast packet with a non-zero TTL merely
   because it can predict that another router on the path to the
   packet's final destination will decrement the TTL to zero. However,
   a router MAY do so for IP multicasts, in order to more efficiently
   implement IP multicast's expanding ring search algorithm (see
   [INTERNET:4]).

   DISCUSSION
      The IP TTL is used, somewhat schizophrenically, as both a hop
      count limit and a time limit. Its hop count function is critical
      to ensuring that routing problems can't melt down the network by
      causing packets to loop infinitely in the network. The time limit
      function is used by transport protocols such as TCP to ensure
      reliable data transfer. Many current implementations treat TTL as
      a pure hop count, and in parts of the Internet community there is
      a strong sentiment that the time limit function should instead be
      performed by the transport protocols that need it.

      In this specification, we have reluctantly decided to follow the
      strong belief among the router vendors that the time limit
      function should be optional. They argued that implementation of
      the time limit function is difficult enough that it is currently
      not generally done. They further pointed to the lack of
      documented cases where this shortcut has caused TCP to corrupt
      data (of course, we would expect the problems created to be rare
      and difficult to reproduce, so the lack of documented cases
      provides little reassurance that there haven't been a number of
      undocumented cases).

IP multicast notions such as the expanding ring search may not
work as expected unless the TTL is treated as a pure hop count.
The same thing is somewhat true of traceroute.

ICMP Time Exceeded messages are required because the traceroute
diagnostic tool depends on them.

Thus, the tradeoff is between severely crippling, if not
eliminating, two very useful tools and avoiding a very rare and
transient data transport problem that may not occur at all. We
have chosen to preserve the tools.

5.3.2 Type of Service (TOS)

The Type-of-Service byte in the IP header is divided into three
sections: the Precedence field (high-order 3 bits), a field that
is customarily called Type of Service or "TOS (next 4 bits), and a
reserved bit (the low order bit). Rules governing the reserved
bit were described in Section [4.2.2.3]. The Precedence field
will be discussed in Section [5.3.3]. A more extensive discussion
of the TOS field and its use can be found in [ROUTE:11].

A router SHOULD consider the TOS field in a packet's IP header
when deciding how to forward it. The remainder of this section
describes the rules that apply to routers that conform to this
requirement.

A router MUST maintain a TOS value for each route in its routing
table. Routes learned through a routing protocol that does not
support TOS MUST be assigned a TOS of zero (the default TOS).

To choose a route to a destination, a router MUST use an algorithm
equivalent to the following:

(1) The router locates in its routing table all available routes
     to the destination (see Section [5.2.4]).

(2) If there are none, the router drops the packet because the
     destination is unreachable. See section [5.2.4].

(3) If one or more of those routes have a TOS that exactly matches
     the TOS specified in the packet, the router chooses the route
     with the best metric.

(4) Otherwise, the router repeats the above step, except looking
     at routes whose TOS is zero.

(5) If no route was chosen above, the router drops the packet
    because the destination is unreachable. The router returns
    an ICMP Destination Unreachable error specifying the
    appropriate code: either Network Unreachable with Type of
    Service (code 11) or Host Unreachable with Type of Service
    (code 12).

DISCUSSION
   Although TOS has been little used in the past, its use by hosts is
   now mandated by the Requirements for Internet Hosts RFCs
   ([INTRO:2] and [INTRO:3]). Support for TOS in routers may become
   a MUST in the future, but is a SHOULD for now until we get more
   experience with it and can better judge both its benefits and its
   costs.

   Various people have proposed that TOS should affect other aspects
   of the forwarding function. For example:

   (1) A router could place packets that have the Low Delay bit set
       ahead of other packets in its output queues.

   (2) a router is forced to discard packets, it could try to avoid
       discarding those which have the High Reliability bit set.

   These ideas have been explored in more detail in [INTERNET:17] but
   we don't yet have enough experience with such schemes to make
   requirements in this area.

5.3.3 IP Precedence

   This section specifies requirements and guidelines for appropriate
   processing of the IP Precedence field in routers. Precedence is a
   scheme for allocating resources in the network based on the
   relative importance of different traffic flows. The IP
   specification defines specific values to be used in this field for
   various types of traffic.

   The basic mechanisms for precedence processing in a router are
   preferential resource allocation, including both precedence-
   ordered queue service and precedence-based congestion control, and
   selection of Link Layer priority features. The router also
   selects the IP precedence for routing, management and control
   traffic it originates. For a more extensive discussion of IP
   Precedence and its implementation see [FORWARD:6].

   Precedence-ordered queue service, as discussed in this section,
   includes but is not limited to the queue for the forwarding
   process and queues for outgoing links. It is intended that a

router supporting precedence should also use the precedence
indication at whatever points in its processing are concerned with
allocation of finite resources, such as packet buffers or Link
Layer connections. The set of such points is implementation-
dependent.

DISCUSSION
    Although the Precedence field was originally provided for use in
    DOD systems where large traffic surges or major damage to the
    network are viewed as inherent threats, it has useful applications
    for many non-military IP networks. Although the traffic handling
    capacity of networks has grown greatly in recent years, the
    traffic generating ability of the users has also grown, and
    network overload conditions still occur at times. Since IP-based
    routing and management protocols have become more critical to the
    successful operation of the Internet, overloads present two
    additional risks to the network:

    (1) High delays may result in routing protocol packets being lost.
        This may cause the routing protocol to falsely deduce a
        topology change and propagate this false information to other
        routers. Not only can this cause routes to oscillate, but an
        extra processing burden may be placed on other routers.

    (2) High delays may interfere with the use of network management
        tools to analyze and perhaps correct or relieve the problem
        in the network that caused the overload condition to occur.

    Implementation and appropriate use of the Precedence mechanism
    alleviates both of these problems.

5.3.3.1 Precedence-Ordered Queue Service

   Routers SHOULD implement precedence-ordered queue service.
   Precedence-ordered queue service means that when a packet is selected
   for output on a (logical) link, the packet of highest precedence that
   has been queued for that link is sent. Routers that implement
   precedence-ordered queue service MUST also have a configuration
   option to suppress precedence-ordered queue service in the Internet
   Layer.

   Any router MAY implement other policy-based throughput management
   procedures that result in other than strict precedence ordering, but
   it MUST be configurable to suppress them (i.e., use strict ordering).

   As detailed in Section [5.3.6], routers that implement precedence-
   ordered queue service discard low precedence packets before
   discarding high precedence packets for congestion control purposes.

Preemption (interruption of processing or transmission of a packet)
is not envisioned as a function of the Internet Layer. Some
protocols at other layers may provide preemption features.

5.3.3.2 Lower Layer Precedence Mappings

Routers that implement precedence-ordered queuing MUST IMPLEMENT, and
other routers SHOULD IMPLEMENT, Lower Layer Precedence Mapping.

A router that implements Lower Layer Precedence Mapping:

o MUST be able to map IP Precedence to Link Layer priority mechanisms
   for link layers that have such a feature defined.

o MUST have a configuration option to select the Link Layer's default
   priority treatment for all IP traffic

o SHOULD be able to configure specific nonstandard mappings of IP
   precedence values to Link Layer priority values for each
   interface.

DISCUSSION
   Some research questions the workability of the priority features
   of some Link Layer protocols, and some networks may have faulty
   implementations of the link layer priority mechanism. It seems
   prudent to provide an escape mechanism in case such problems show
   up in a network.

   On the other hand, there are proposals to use novel queuing
   strategies to implement special services such as multimedia
   bandwidth reservation or low-delay service. Special services and
   queuing strategies to support them are current research subjects
   and are in the process of standardization.

   Implementors may wish to consider that correct link layer mapping
   of IP precedence is required by DOD policy for TCP/IP systems used
   on DOD networks. Since these requirements are intended to
   encourage (but not force) the use of precedence features in the
   hope of providing better Internet service to all users, routers
   supporting precedence-ordered queue service should default to
   maintaining strict precedence ordering regardless of the type of
   service requested.

5.3.3.3 Precedence Handling For All Routers

   A router (whether or not it employs precedence-ordered queue
   service):

   (1) MUST accept and process incoming traffic of all precedence levels
        normally, unless it has been administratively configured to do
        otherwise.

   (2) MAY implement a validation filter to administratively restrict
        the use of precedence levels by particular traffic sources. If
        provided, this filter MUST NOT filter out or cut off the
        following sorts of ICMP error messages: Destination Unreachable,
        Redirect, Time Exceeded, and Parameter Problem. If this filter
        is provided, the procedures required for packet filtering by
        addresses are required for this filter also.

   DISCUSSION
      Precedence filtering should be applicable to specific
      source/destination IP Address pairs, specific protocols, specific
      ports, and so on.

   An ICMP Destination Unreachable message with code 14 SHOULD be sent
   when a packet is dropped by the validation filter, unless this has
   been suppressed by configuration choice.

   (3) MAY implement a cutoff function that allows the router to be set
        to refuse or drop traffic with precedence below a specified
        level. This function may be activated by management actions or
        by some implementation dependent heuristics, but there MUST be a
        configuration option to disable any heuristic mechanism that
        operates without human intervention. An ICMP Destination
        Unreachable message with code 15 SHOULD be sent when a packet is
        dropped by the cutoff function, unless this has been suppressed
        by configuration choice.

        A router MUST NOT refuse to forward datagrams with IP precedence
        of 6 (Internetwork Control) or 7 (Network Control) solely due to
        precedence cutoff. However, other criteria may be used in
        conjunction with precedence cutoff to filter high precedence
        traffic.

   DISCUSSION
      Unrestricted precedence cutoff could result in an unintentional
      cutoff of routing and control traffic. In the general case, host
      traffic should be restricted to a value of 5 (CRITIC/ECP) or
      below; this is not a requirement and may not be correct in certain
      systems.

   (4) MUST NOT change precedence settings on packets it did not
       originate.

   (5) SHOULD be able to configure distinct precedence values to be used
       for each routing or management protocol supported (except for
       those protocols, such as OSPF, which specify which precedence
       value must be used).

   (6) MAY be able to configure routing or management traffic precedence
       values independently for each peer address.

   (7) MUST respond appropriately to Link Layer precedence-related error
       indications where provided. An ICMP Destination Unreachable
       message with code 15 SHOULD be sent when a packet is dropped
       because a link cannot accept it due to a precedence-related
       condition, unless this has been suppressed by configuration
       choice.

   DISCUSSION
      The precedence cutoff mechanism described in (3) is somewhat
      controversial. Depending on the topological location of the area
      affected by the cutoff, transit traffic may be directed by routing
      protocols into the area of the cutoff, where it will be dropped.
      This is only a problem if another path that is unaffected by the
      cutoff exists between the communicating points. Proposed ways of
      avoiding this problem include providing some minimum bandwidth to
      all precedence levels even under overload conditions, or
      propagating cutoff information in routing protocols. In the
      absence of a widely accepted (and implemented) solution to this
      problem, great caution is recommended in activating cutoff
      mechanisms in transit networks.

      A transport layer relay could legitimately provide the function
      prohibited by (4) above. Changing precedence levels may cause
      subtle interactions with TCP and perhaps other protocols; a
      correct design is a non-trivial task.

      The intent of (5) and (6) (and the discussion of IP Precedence in
      ICMP messages in Section [4.3.2]) is that the IP precedence bits
      should be appropriately set, whether or not this router acts upon
      those bits in any other way. We expect that in the future
      specifications for routing protocols and network management
      protocols will specify how the IP Precedence should be set for
      messages sent by those protocols.

      The appropriate response for (7) depends on the link layer
      protocol in use. Typically, the router should stop trying to send
      offensive traffic to that destination for some period of time, and

should return an ICMP Destination Unreachable message with code 15
(service not available for precedence requested) to the traffic
source. It also should not try to reestablish a preempted Link
Layer connection for some time.

5.3.4 Forwarding of Link Layer Broadcasts

The encapsulation of IP packets in most Link Layer protocols (except
PPP) allows a receiver to distinguish broadcasts and multicasts from
unicasts simply by examining the Link Layer protocol headers (most
commonly, the Link Layer destination address). The rules in this
section that refer to Link Layer broadcasts apply only to Link Layer
protocols that allow broadcasts to be distinguished; likewise, the
rules that refer to Link Layer multicasts apply only to Link Layer
protocols that allow multicasts to be distinguished.

A router MUST NOT forward any packet that the router received as a
Link Layer broadcast, unless it is directed to an IP Multicast
address. In this latter case, one would presume that link layer
broadcast was used due to the lack of an effective multicast service.

A router MUST NOT forward any packet which the router received as a
Link Layer multicast unless the packet's destination address is an IP
multicast address.

A router SHOULD silently discard a packet that is received via a Link
Layer broadcast but does not specify an IP multicast or IP broadcast
destination address.

When a router sends a packet as a Link Layer broadcast, the IP
destination address MUST be a legal IP broadcast or IP multicast
address.

5.3.5 Forwarding of Internet Layer Broadcasts

There are two major types of IP broadcast addresses; limited
broadcast and directed broadcast. In addition, there are three
subtypes of directed broadcast: a broadcast directed to a specified
network prefix, a broadcast directed to a specified subnetwork, and a
broadcast directed to all subnets of a specified network.
Classification by a router of a broadcast into one of these
categories depends on the broadcast address and on the router's
understanding (if any) of the subnet structure of the destination
network. The same broadcast will be classified differently by
different routers.

A limited IP broadcast address is defined to be all-ones: { -1, -1 }
or 255.255.255.255.

A network-prefix-directed broadcast is composed of the network prefix
of the IP address with a local part of all-ones or { <Network-
prefix>, -1 }. For example, a Class A net broadcast address is
net.255.255.255, a Class B net broadcast address is net.net.255.255
and a Class C net broadcast address is net.net.net.255 where net is a
byte of the network address.

The all-subnets-directed-broadcast is not well defined in a CIDR
environment, and was deprecated in version 1 of this memo.

As was described in Section [4.2.3.1], a router may encounter certain
non-standard IP broadcast addresses:

o 0.0.0.0 is an obsolete form of the limited broadcast address

o { <Network-prefix>, 0 } is an obsolete form of a network-prefix-
  directed broadcast address.

As was described in that section, packets addressed to any of these
addresses SHOULD be silently discarded, but if they are not, they
MUST be treated according to the same rules that apply to packets
addressed to the non-obsolete forms of the broadcast addresses
described above. These rules are described in the next few sections.

5.3.5.1 Limited Broadcasts

   Limited broadcasts MUST NOT be forwarded. Limited broadcasts MUST
   NOT be discarded. Limited broadcasts MAY be sent and SHOULD be sent
   instead of directed broadcasts where limited broadcasts will suffice.

   DISCUSSION
      Some routers contain UDP servers which function by resending the
      requests (as unicasts or directed broadcasts) to other servers.
      This requirement should not be interpreted as prohibiting such
      servers. Note, however, that such servers can easily cause packet
      looping if misconfigured. Thus, providers of such servers would
      probably be well advised to document their setup carefully and to
      consider carefully the TTL on packets that are sent.

5.3.5.2 Directed Broadcasts

   A router MUST classify as network-prefix-directed broadcasts all
   valid, directed broadcasts destined for a remote network or an
   attached nonsubnetted network. Note that in view of CIDR, such
   appear to be host addresses within the network prefix; we preclude
   inspection of the host part of such network prefixes. Given a route
   and no overriding policy, then, a router MUST forward network-
   prefix-directed broadcasts. Network-Prefix-Directed broadcasts MAY

be sent.

A router MAY have an option to disable receiving network-prefix-
directed broadcasts on an interface and MUST have an option to
disable forwarding network-prefix-directed broadcasts. These options
MUST default to permit receiving and forwarding network-prefix-
directed broadcasts.

DISCUSSION
    There has been some debate about forwarding or not forwarding
    directed broadcasts. In this memo we have made the forwarding
    decision depend on the router's knowledge of the destination
    network prefix. Routers cannot determine that a message is
    unicast or directed broadcast apart from this knowledge. The
    decision to forward or not forward the message is by definition
    only possible in the last hop router.

5.3.5.3 All-subnets-directed Broadcasts

The first version of this memo described an algorithm for
distributing a directed broadcast to all the subnets of a classical
network number. This algorithm was stated to be "broken," and
certain failure cases were specified.

In a CIDR routing domain, wherein classical IP network numbers are
meaningless, the concept of an all-subnets-directed-broadcast is also
meaningless. To the knowledge of the working group, the facility was
never implemented or deployed, and is now relegated to the dustbin of
history.

5.3.5.4 Subnet-directed Broadcasts

The first version of this memo spelled out procedures for dealing
with subnet-directed-broadcasts. In a CIDR routing domain, these are
indistinguishable from net-drected-broadcasts. The two are therefore
treated together in section [5.3.5.2 Directed Broadcasts], and should
be viewed as network-prefix directed broadcasts.

5.3.6 Congestion Control

Congestion in a network is loosely defined as a condition where
demand for resources (usually bandwidth or CPU time) exceeds
capacity. Congestion avoidance tries to prevent demand from
exceeding capacity, while congestion recovery tries to restore an
operative state. It is possible for a router to contribute to both
of these mechanisms. A great deal of effort has been spent studying
the problem. The reader is encouraged to read [FORWARD:2] for a
survey of the work. Important papers on the subject include

[FORWARD:3], [FORWARD:4], [FORWARD:5], [FORWARD:10], [FORWARD:11], [FORWARD:12], [FORWARD:13], [FORWARD:14], and [INTERNET:10], among others.

The amount of storage that router should have available to handle peak instantaneous demand when hosts use reasonable congestion policies, such as described in [FORWARD:5], is a function of the product of the bandwidth of the link times the path delay of the flows using the link, and therefore storage should increase as this Bandwidth*Delay product increases. The exact function relating storage capacity to probability of discard is not known.

When a router receives a packet beyond its storage capacity it must (by definition, not by decree) discard it or some other packet or packets. Which packet to discard is the subject of much study but, unfortunately, little agreement so far. The best wisdom to date suggests discarding a packet from the data stream most heavily using the link. However, a number of additional factors may be relevant, including the precedence of the traffic, active bandwidth reservation, and the complexity associated with selecting that packet.

A router MAY discard the packet it has just received; this is the simplest but not the best policy. Ideally, the router should select a packet from one of the sessions most heavily abusing the link, given that the applicable Quality of Service policy permits this. A recommended policy in datagram environments using FIFO queues is to discard a packet randomly selected from the queue (see [FORWARD:5]). An equivalent algorithm in routers using fair queues is to discard from the longest queue or that using the greatest virtual time (see [FORWARD:13]). A router MAY use these algorithms to determine which packet to discard.

If a router implements a discard policy (such as Random Drop) under which it chooses a packet to discard from a pool of eligible packets:

o If precedence-ordered queue service (described in Section
  [5.3.3.1]) is implemented and enabled, the router MUST NOT discard
  a packet whose IP precedence is higher than that of a packet that
  is not discarded.

o A router MAY protect packets whose IP headers request the maximize
  reliability TOS, except where doing so would be in violation of
  the previous rule.

o A router MAY protect fragmented IP packets, on the theory that
  dropping a fragment of a datagram may increase congestion by
  causing all fragments of the datagram to be retransmitted by the

source.

   o To help prevent routing perturbations or disruption of management
     functions, the router MAY protect packets used for routing
     control, link control, or network management from being discarded.
     Dedicated routers (i.e., routers that are not also general purpose
     hosts, terminal servers, etc.) can achieve an approximation of
     this rule by protecting packets whose source or destination is the
     router itself.

   Advanced methods of congestion control include a notion of fairness,
   so that the 'user' that is penalized by losing a packet is the one
   that contributed the most to the congestion. No matter what
   mechanism is implemented to deal with bandwidth congestion control,
   it is important that the CPU effort expended be sufficiently small
   that the router is not driven into CPU congestion also.

   As described in Section [4.3.3.3], this document recommends that a
   router SHOULD NOT send a Source Quench to the sender of the packet
   that it is discarding. ICMP Source Quench is a very weak mechanism,
   so it is not necessary for a router to send it, and host software
   should not use it exclusively as an indicator of congestion.

5.3.7 Martian Address Filtering

   An IP source address is invalid if it is a special IP address, as
   defined in 4.2.2.11 or 5.3.7, or is not a unicast address.

   An IP destination address is invalid if it is among those defined as
   illegal destinations in 4.2.3.1, or is a Class E address (except
   255.255.255.255).

   A router SHOULD NOT forward any packet that has an invalid IP source
   address or a source address on network 0. A router SHOULD NOT
   forward, except over a loopback interface, any packet that has a
   source address on network 127. A router MAY have a switch that
   allows the network manager to disable these checks. If such a switch
   is provided, it MUST default to performing the checks.

   A router SHOULD NOT forward any packet that has an invalid IP
   destination address or a destination address on network 0. A router
   SHOULD NOT forward, except over a loopback interface, any packet that
   has a destination address on network 127. A router MAY have a switch
   that allows the network manager to disable these checks. If such a
   switch is provided, it MUST default to performing the checks.

   If a router discards a packet because of these rules, it SHOULD log
   at least the IP source address, the IP destination address, and, if

the problem was with the source address, the physical interface on
which the packet was received and the Link Layer address of the host
or router from which the packet was received.

5.3.8 Source Address Validation

A router SHOULD IMPLEMENT the ability to filter traffic based on a
comparison of the source address of a packet and the forwarding table
for a logical interface on which the packet was received. If this
filtering is enabled, the router MUST silently discard a packet if
the interface on which the packet was received is not the interface
on which a packet would be forwarded to reach the address contained
in the source address. In simpler terms, if a router wouldn't route
a packet containing this address through a particular interface, it
shouldn't believe the address if it appears as a source address in a
packet read from this interface.

If this feature is implemented, it MUST be disabled by default.

    DISCUSSION
        This feature can provide useful security improvements in some
        situations, but can erroneously discard valid packets in
        situations where paths are asymmetric.

5.3.9 Packet Filtering and Access Lists

As a means of providing security and/or limiting traffic through
portions of a network a router SHOULD provide the ability to
selectively forward (or filter) packets. If this capability is
provided, filtering of packets SHOULD be configurable either to
forward all packets or to selectively forward them based upon the
source and destination prefixes, and MAY filter on other message
attributes. Each source and destination address SHOULD allow
specification of an arbitrary prefix length.

    DISCUSSION
        This feature can provide a measure of privacy, where systems
        outside a boundary are not permitted to exchange certain protocols
        with systems inside the boundary, or are limited as to which
        systems they may communicate with. It can also help prevent
        certain classes of security breach, wherein a system outside a
        boundary masquerades as a system inside the boundary and mimics a
        session with it.

If supported, a router SHOULD be configurable to allow one of an

    o Include list - specification of a list of message definitions to be
      forwarded, or an

   o Exclude list - specification of a list of message definitions NOT
     to be forwarded.

   A "message definition", in this context, specifies the source and
   destination network prefix, and may include other identifying
   information such as IP Protocol Type or TCP port number.

   A router MAY provide a configuration switch that allows a choice
   between specifying an include or an exclude list, or other equivalent
   controls.

   A value matching any address (e.g., a keyword any, an address with a
   mask of all 0's, or a network prefix whose length is zero) MUST be
   allowed as a source and/or destination address.

   In addition to address pairs, the router MAY allow any combination of
   transport and/or application protocol and source and destination
   ports to be specified.

   The router MUST allow packets to be silently discarded (i.e.,
   discarded without an ICMP error message being sent).

   The router SHOULD allow an appropriate ICMP unreachable message to be
   sent when a packet is discarded. The ICMP message SHOULD specify
   Communication Administratively Prohibited (code 13) as the reason for
   the destination being unreachable.

   The router SHOULD allow the sending of ICMP destination unreachable
   messages (code 13) to be configured for each combination of address
   pairs, protocol types, and ports it allows to be specified.

   The router SHOULD count and SHOULD allow selective logging of packets
   not forwarded.

5.3.10 Multicast Routing

   An IP router SHOULD support forwarding of IP multicast packets, based
   either on static multicast routes or on routes dynamically determined
   by a multicast routing protocol (e.g., DVMRP [ROUTE:9]). A router
   that forwards IP multicast packets is called a multicast router.

5.3.11 Controls on Forwarding

   For each physical interface, a router SHOULD have a configuration
   option that specifies whether forwarding is enabled on that
   interface. When forwarding on an interface is disabled, the router:

   o MUST silently discard any packets which are received on that
     interface but are not addressed to the router

   o MUST NOT send packets out that interface, except for datagrams
     originated by the router

   o MUST NOT announce via any routing protocols the availability of
     paths through the interface

   DISCUSSION
      This feature allows the network manager to essentially turn off an
      interface but leaves it accessible for network management.

      Ideally, this control would apply to logical rather than physical
      interfaces. It cannot, because there is no known way for a router
      to determine which logical interface a packet arrived absent a
      one-to-one correspondence between logical and physical interfaces.

5.3.12 State Changes

   During router operation, interfaces may fail or be manually disabled,
   or may become available for use by the router. Similarly, forwarding
   may be disabled for a particular interface or for the entire router
   or may be (re)enabled. While such transitions are (usually)
   uncommon, it is important that routers handle them correctly.

5.3.12.1 When a Router Ceases Forwarding

   When a router ceases forwarding it MUST stop advertising all routes,
   except for third party routes. It MAY continue to receive and use
   routes from other routers in its routing domains. If the forwarding
   database is retained, the router MUST NOT cease timing the routes in
   the forwarding database. If routes that have been received from
   other routers are remembered, the router MUST NOT cease timing the
   routes that it has remembered. It MUST discard any routes whose
   timers expire while forwarding is disabled, just as it would do if
   forwarding were enabled.

   DISCUSSION
      When a router ceases forwarding, it essentially ceases being a
      router. It is still a host, and must follow all of the
      requirements of Host Requirements [INTRO:2]. The router may still
      be a passive member of one or more routing domains, however. As
      such, it is allowed to maintain its forwarding database by
      listening to other routers in its routing domain. It may not,
      however, advertise any of the routes in its forwarding database,
      since it itself is doing no forwarding. The only exception to
      this rule is when the router is advertising a route that uses only

some other router, but which this router has been asked to
advertise.

A router MAY send ICMP destination unreachable (host unreachable)
messages to the senders of packets that it is unable to forward. It
SHOULD NOT send ICMP redirect messages.

DISCUSSION
    Note that sending an ICMP destination unreachable (host
    unreachable) is a router action. This message should not be sent
    by hosts. This exception to the rules for hosts is allowed so
    that packets may be rerouted in the shortest possible time, and so
    that black holes are avoided.

5.3.12.2 When a Router Starts Forwarding

When a router begins forwarding, it SHOULD expedite the sending of
new routing information to all routers with which it normally
exchanges routing information.

5.3.12.3 When an Interface Fails or is Disabled

If an interface fails or is disabled a router MUST remove and stop
advertising all routes in its forwarding database that make use of
that interface. It MUST disable all static routes that make use of
that interface. If other routes to the same destination and TOS are
learned or remembered by the router, the router MUST choose the best
alternate, and add it to its forwarding database. The router SHOULD
send ICMP destination unreachable or ICMP redirect messages, as
appropriate, in reply to all packets that it is unable to forward due
to the interface being unavailable.

5.3.12.4 When an Interface is Enabled

If an interface that had not been available becomes available, a
router MUST reenable any static routes that use that interface. If
routes that would use that interface are learned by the router, then
these routes MUST be evaluated along with all the other learned
routes, and the router MUST make a decision as to which routes should
be placed in the forwarding database. The implementor is referred to
Chapter [7], Application Layer - Routing Protocols for further
information on how this decision is made.

A router SHOULD expedite the sending of new routing information to
all routers with which it normally exchanges routing information.

5.3.13 IP Options

   Several options, such as Record Route and Timestamp, contain slots
   into which a router inserts its address when forwarding the packet.
   However, each such option has a finite number of slots, and therefore
   a router may find that there is not free slot into which it can
   insert its address. No requirement listed below should be construed
   as requiring a router to insert its address into an option that has
   no remaining slot to insert it into. Section [5.2.5] discusses how a
   router must choose which of its addresses to insert into an option.

5.3.13.1 Unrecognized Options Unrecognized IP options in forwarded
   packets MUST be passed through unchanged.

5.3.13.2 Security Option

   Some environments require the Security option in every packet; such a
   requirement is outside the scope of this document and the IP standard
   specification. Note, however, that the security options described in
   [INTERNET:1] and [INTERNET:16] are obsolete. Routers SHOULD
   IMPLEMENT the revised security option described in [INTERNET:5].

      DISCUSSION
         Routers intended for use in networks with multiple security levels
         should support packet filtering based on IPSO (RFC-1108) labels.
         To implement this support, the router would need to permit the
         router administrator to configure both a lower sensitivity limit
         (e.g. Unclassified) and an upper sensitivity limit (e.g. Secret)
         on each interface. It is commonly but not always the case that
         the two limits are the same (e.g. a single-level interface).
         Packets caught by an IPSO filter as being out of range should be
         silently dropped and a counter should note the number of packets
         dropped because of out of range IPSO labels.

5.3.13.3 Stream Identifier Option

   This option is obsolete. If the Stream Identifier option is present
   in a packet forwarded by the router, the option MUST be ignored and
   passed through unchanged.

5.3.13.4 Source Route Options

   A router MUST implement support for source route options in forwarded
   packets. A router MAY implement a configuration option that, when
   enabled, causes all source-routed packets to be discarded. However,
   such an option MUST NOT be enabled by default.

DISCUSSION
    The ability to source route datagrams through the Internet is
    important to various network diagnostic tools. However, source
    routing may be used to bypass administrative and security controls
    within a network. Specifically, those cases where manipulation of
    routing tables is used to provide administrative separation in
    lieu of other methods such as packet filtering may be vulnerable
    through source routed packets.

EDITORS+COMMENTS
    Packet filtering can be defeated by source routing as well, if it
    is applied in any router except one on the final leg of the source
    routed path. Neither route nor packet filters constitute a
    complete solution for security.

5.3.13.5 Record Route Option

Routers MUST support the Record Route option in forwarded packets.

A router MAY provide a configuration option that, if enabled, will
cause the router to ignore (i.e., pass through unchanged) Record
Route options in forwarded packets. If provided, such an option MUST
default to enabling the record-route. This option should not affect
the processing of Record Route options in datagrams received by the
router itself (in particular, Record Route options in ICMP echo
requests will still be processed according to Section [4.3.3.6]).

DISCUSSION
    There are some people who believe that Record Route is a security
    problem because it discloses information about the topology of the
    network. Thus, this document allows it to be disabled.

5.3.13.6 Timestamp Option

Routers MUST support the timestamp option in forwarded packets. A
timestamp value MUST follow the rules given [INTRO:2].

If the flags field = 3 (timestamp and prespecified address), the
router MUST add its timestamp if the next prespecified address
matches any of the router's IP addresses. It is not necessary that
the prespecified address be either the address of the interface on
which the packet arrived or the address of the interface over which
it will be sent.

IMPLEMENTATION
    To maximize the utility of the timestamps contained in the
    timestamp option, it is suggested that the timestamp inserted be,
    as nearly as practical, the time at which the packet arrived at

the router. For datagrams originated by the router, the timestamp
inserted should be, as nearly as practical, the time at which the
datagram was passed to the network layer for transmission.

A router MAY provide a configuration option which, if enabled, will
cause the router to ignore (i.e., pass through unchanged) Timestamp
options in forwarded datagrams when the flag word is set to zero
(timestamps only) or one (timestamp and registering IP address). If
provided, such an option MUST default to off (that is, the router
does not ignore the timestamp). This option should not affect the
processing of Timestamp options in datagrams received by the router
itself (in particular, a router will insert timestamps into Timestamp
options in datagrams received by the router, and Timestamp options in
ICMP echo requests will still be processed according to Section
[4.3.3.6]).

> DISCUSSION
>   Like the Record Route option, the Timestamp option can reveal
>   information about a network's topology. Some people consider this
>   to be a security concern.

6. TRANSPORT LAYER

A router is not required to implement any Transport Layer protocols
except those required to support Application Layer protocols
supported by the router. In practice, this means that most routers
implement both the Transmission Control Protocol (TCP) and the User
Datagram Protocol (UDP).

6.1 USER DATAGRAM PROTOCOL - UDP

The User Datagram Protocol (UDP) is specified in [TRANS:1].

A router that implements UDP MUST be compliant, and SHOULD be
unconditionally compliant, with the requirements of [INTRO:2], except
that:

o This specification does not specify the interfaces between the
  various protocol layers. Thus, a router's interfaces need not
  comply with [INTRO:2], except where compliance is required for
  proper functioning of Application Layer protocols supported by the
  router.

o Contrary to [INTRO:2], an application SHOULD NOT disable generation
  of UDP checksums.

DISCUSSION
Although a particular application protocol may require that UDP
datagrams it receives must contain a UDP checksum, there is no
general requirement that received UDP datagrams contain UDP
checksums. Of course, if a UDP checksum is present in a received
datagram, the checksum must be verified and the datagram discarded
if the checksum is incorrect.

6.2 TRANSMISSION CONTROL PROTOCOL - TCP

The Transmission Control Protocol (TCP) is specified in [TRANS:2].

A router that implements TCP MUST be compliant, and SHOULD be
unconditionally compliant, with the requirements of [INTRO:2], except
that:

o This specification does not specify the interfaces between the
   various protocol layers. Thus, a router need not comply with the
   following requirements of [INTRO:2] (except of course where
   compliance is required for proper functioning of Application Layer
   protocols supported by the router):

   Use of Push: RFC-793 Section 2.8:
        Passing a received PSH flag to the application layer is now
        OPTIONAL.

   Urgent Pointer: RFC-793 Section 3.1:
        A TCP MUST inform the application layer asynchronously
        whenever it receives an Urgent pointer and there was
        previously no pending urgent data, or whenever the Urgent
        pointer advances in the data stream. There MUST be a way for
        the application to learn how much urgent data remains to be
        read from the connection, or at least to determine whether or
        not more urgent data remains to be read.

   TCP Connection Failures:
        An application MUST be able to set the value for R2 for a
        particular connection. For example, an interactive
        application might set R2 to ``infinity,'' giving the user
        control over when to disconnect.

   TCP Multihoming:
        If an application on a multihomed host does not specify the
        local IP address when actively opening a TCP connection, then
        the TCP MUST ask the IP layer to select a local IP address
        before sending the (first) SYN. See the function
        GET_SRCADDR() in Section 3.4.

        IP Options:
             An application MUST be able to specify a source route when it
             actively opens a TCP connection, and this MUST take
             precedence over a source route received in a datagram.

   o For similar reasons, a router need not comply with any of the
     requirements of [INTRO:2].

   o The requirements concerning the Maximum Segment Size Option in
     [INTRO:2] are amended as follows: a router that implements the
     host portion of MTU discovery (discussed in Section [4.2.3.3] of
     this memo) uses 536 as the default value of SendMSS only if the
     path MTU is unknown; if the path MTU is known, the default value
     for SendMSS is the path MTU - 40.

   o The requirements concerning the Maximum Segment Size Option in
     [INTRO:2] are amended as follows: ICMP Destination Unreachable
     codes 11 and 12 are additional soft error conditions. Therefore,
     these message MUST NOT cause TCP to abort a connection.

   DISCUSSION
        It should particularly be noted that a TCP implementation in a
        router must conform to the following requirements of [INTRO:2]:

        o Providing a configurable TTL. [Time to Live: RFC-793 Section
          3.9]

        o Providing an interface to configure keep-alive behavior, if
           keep-alives are used at all. [TCP Keep-Alives]

        o Providing an error reporting mechanism, and the ability to
           manage it. [Asynchronous Reports]

        o Specifying type of service. [Type-of-Service]

        The general paradigm applied is that if a particular interface is
        visible outside the router, then all requirements for the
        interface must be followed. For example, if a router provides a
        telnet function, then it will be generating traffic, likely to be
        routed in the external networks. Therefore, it must be able to
        set the type of service correctly or else the telnet traffic may
        not get through.

7. APPLICATION LAYER - ROUTING PROTOCOLS

7.1 INTRODUCTION

   For technical, managerial, and sometimes political reasons, the
   Internet routing system consists of two components - interior routing
   and exterior routing. The concept of an Autonomous System (AS), as
   define in Section 2.2.4 of this document, plays a key role in
   separating interior from an exterior routing, as this concept allows
   to deliniate the set of routers where a change from interior to
   exterior routing occurs. An IP datagram may have to traverse the
   routers of two or more Autonomous Systems to reach its destination,
   and the Autonomous Systems must provide each other with topology
   information to allow such forwarding. Interior gateway protocols
   (IGPs) are used to distribute routing information within an AS (i.e.,
   intra-AS routing). Exterior gateway protocols are used to exchange
   routing information among ASs (i.e., inter-AS routing).

7.1.1 Routing Security Considerations

   Routing is one of the few places where the Robustness Principle (be
   liberal in what you accept) does not apply. Routers should be
   relatively suspicious in accepting routing data from other routing
   systems.

   A router SHOULD provide the ability to rank routing information
   sources from most trustworthy to least trustworthy and to accept
   routing information about any particular destination from the most
   trustworthy sources first. This was implicit in the original
   core/stub autonomous system routing model using EGP and various
   interior routing protocols. It is even more important with the
   demise of a central, trusted core.

   A router SHOULD provide a mechanism to filter out obviously invalid
   routes (such as those for net 127).

   Routers MUST NOT by default redistribute routing data they do not
   themselves use, trust or otherwise consider valid. In rare cases, it
   may be necessary to redistribute suspicious information, but this
   should only happen under direct intercession by some human agency.

   Routers must be at least a little paranoid about accepting routing
   data from anyone, and must be especially careful when they distribute
   routing information provided to them by another party. See below for
   specific guidelines.

7.1.2 Precedence

   Except where the specification for a particular routing protocol
   specifies otherwise, a router SHOULD set the IP Precedence value for
   IP datagrams carrying routing traffic it originates to 6
   (INTERNETWORK CONTROL).

   DISCUSSION
      Routing traffic with VERY FEW exceptions should be the highest
      precedence traffic on any network. If a system's routing traffic
      can't get through, chances are nothing else will.

7.1.3 Message Validation

   Peer-to-peer authentication involves several tests. The application
   of message passwords and explicit acceptable neighbor lists has in
   the past improved the robustness of the route database. Routers
   SHOULD IMPLEMENT management controls that enable explicit listing of
   valid routing neighbors. Routers SHOULD IMPLEMENT peer-to-peer
   authentication for those routing protocols that support them.

   Routers SHOULD validate routing neighbors based on their source
   address and the interface a message is received on; neighbors in a
   directly attached subnet SHOULD be restricted to communicate with the
   router via the interface that subnet is posited on or via unnumbered
   interfaces. Messages received on other interfaces SHOULD be silently
   discarded.

   DISCUSSION
      Security breaches and numerous routing problems are avoided by
      this basic testing.

7.2 INTERIOR GATEWAY PROTOCOLS

7.2.1 INTRODUCTION

   An Interior Gateway Protocol (IGP) is used to distribute routing
   information between the various routers in a particular AS.
   Independent of the algorithm used to implement a particular IGP, it
   should perform the following functions:

   (1) Respond quickly to changes in the internal topology of an AS

   (2) Provide a mechanism such that circuit flapping does not cause
        continuous routing updates

   (3) Provide quick convergence to loop-free routing

(4) Utilize minimal bandwidth

(5) Provide equal cost routes to enable load-splitting

(6) Provide a means for authentication of routing updates

Current IGPs used in the internet today are characterized as either being based on a distance-vector or a link-state algorithm.

Several IGPs are detailed in this section, including those most commonly used and some recently developed protocols that may be widely used in the future. Numerous other protocols intended for use in intra-AS routing exist in the Internet community.

A router that implements any routing protocol (other than static routes) MUST IMPLEMENT OSPF (see Section [7.2.2]). A router MAY implement additional IGPs.

7.2.2 OPEN SHORTEST PATH FIRST - OSPF

Shortest Path First (SPF) based routing protocols are a class of link-state algorithms that are based on the shortest-path algorithm of Dijkstra. Although SPF based algorithms have been around since the inception of the ARPANET, it is only recently that they have achieved popularity both inside both the IP and the OSI communities. In an SPF based system, each router obtains the entire topology database through a process known as flooding. Flooding insures a reliable transfer of the information. Each router then runs the SPF algorithm on its database to build the IP routing table. The OSPF routing protocol is an implementation of an SPF algorithm. The current version, OSPF version 2, is specified in [ROUTE:1]. Note that RFC-1131, which describes OSPF version 1, is obsolete.

Note that to comply with Section [8.3] of this memo, a router that implements OSPF MUST implement the OSPF MIB [MGT:14].

7.2.3 INTERMEDIATE SYSTEM TO INTERMEDIATE SYSTEM - DUAL IS-IS

The American National Standards Institute (ANSI) X3S3.3 committee has defined an intra-domain routing protocol. This protocol is titled Intermediate System to Intermediate System Routeing Exchange Protocol.

Its application to an IP network has been defined in [ROUTE:2], and is referred to as Dual IS-IS (or sometimes as Integrated IS-IS). IS-IS is based on a link-state (SPF) routing algorithm and shares all the advantages for this class of protocols.

7.3 EXTERIOR GATEWAY PROTOCOLS

7.3.1 INTRODUCTION

   Exterior Gateway Protocols are utilized for inter-Autonomous System
   routing to exchange reachability information for a set of networks
   internal to a particular autonomous system to a neighboring
   autonomous system.

   The area of inter-AS routing is a current topic of research inside
   the Internet Engineering Task Force. The Exterior Gateway Protocol
   (EGP) described in Section [Appendix F.1] has traditionally been the
   inter-AS protocol of choice, but is now historical. The Border
   Gateway Protocol (BGP) eliminates many of the restrictions and
   limitations of EGP, and is therefore growing rapidly in popularity.
   A router is not required to implement any inter-AS routing protocol.
   However, if a router does implement EGP it also MUST IMPLEMENT BGP.
   Although it was not designed as an exterior gateway protocol, RIP
   (described in Section [7.2.4]) is sometimes used for inter-AS
   routing.

7.3.2 BORDER GATEWAY PROTOCOL - BGP

7.3.2.1 Introduction

   The Border Gateway Protocol (BGP-4) is an inter-AS routing protocol
   that exchanges network reachability information with other BGP
   speakers. The information for a network includes the complete list
   of ASs that traffic must transit to reach that network. This
   information can then be used to insure loop-free paths. This
   information is sufficient to construct a graph of AS connectivity
   from which routing loops may be pruned and some policy decisions at
   the AS level may be enforced.

   BGP is defined by [ROUTE:4]. [ROUTE:5] specifies the proper usage of
   BGP in the Internet, and provides some useful implementation hints
   and guidelines. [ROUTE:12] and [ROUTE:13] provide additional useful
   information.

   To comply with Section [8.3] of this memo, a router that implements
   BGP is required to implement the BGP MIB [MGT:15].

   To characterize the set of policy decisions that can be enforced
   using BGP, one must focus on the rule that an AS advertises to its
   neighbor ASs only those routes that it itself uses. This rule
   reflects the hop-by-hop routing paradigm generally used throughout
   the current Internet. Note that some policies cannot be supported by
   the hop-by-hop routing paradigm and thus require techniques such as

source routing to enforce. For example, BGP does not enable one AS
to send traffic to a neighbor AS intending that traffic take a
different route from that taken by traffic originating in the
neighbor AS. On the other hand, BGP can support any policy
conforming to the hop-by-hop routing paradigm.

Implementors of BGP are strongly encouraged to follow the
recommendations outlined in Section 6 of [ROUTE:5].

7.3.2.2 Protocol Walk-through

While BGP provides support for quite complex routing policies (as an
example see Section 4.2 in [ROUTE:5]), it is not required for all BGP
implementors to support such policies. At a minimum, however, a BGP
implementation:

(1) SHOULD allow an AS to control announcements of the BGP learned
    routes to adjacent AS's. Implementations SHOULD support such
    control with at least the granularity of a single network.
    Implementations SHOULD also support such control with the
    granularity of an autonomous system, where the autonomous system
    may be either the autonomous system that originated the route,
    or the autonomous system that advertised the route to the local
    system (adjacent autonomous system).

(2) SHOULD allow an AS to prefer a particular path to a destination
    (when more than one path is available). Such function SHOULD be
    implemented by allowing system administrator to assign weights
    to Autonomous Systems, and making route selection process to
    select a route with the lowest weight (where weight of a route
    is defined as a sum of weights of all AS's in the AS_PATH path
    attribute associated with that route).

(3) SHOULD allow an AS to ignore routes with certain AS's in the
    AS_PATH path attribute. Such function can be implemented by
    using technique outlined in (2), and by assigning infinity as
    weights for such AS's. The route selection process must ignore
    routes that have weight equal to infinity.

7.3.3 INTER-AS ROUTING WITHOUT AN EXTERIOR PROTOCOL

It is possible to exchange routing information between two autonomous
systems or routing domains without using a standard exterior routing
protocol between two separate, standard interior routing protocols.
The most common way of doing this is to run both interior protocols
independently in one of the border routers with an exchange of route
information between the two processes.

As with the exchange of information from an EGP to an IGP, without
appropriate controls these exchanges of routing information between
two IGPs in a single router are subject to creation of routing loops.

7.4 STATIC ROUTING

Static routing provides a means of explicitly defining the next hop
from a router for a particular destination. A router SHOULD provide
a means for defining a static route to a destination, where the
destination is defined by a network prefix. The mechanism SHOULD
also allow for a metric to be specified for each static route.

A router that supports a dynamic routing protocol MUST allow static
routes to be defined with any metric valid for the routing protocol
used. The router MUST provide the ability for the user to specify a
list of static routes that may or may not be propagated through the
routing protocol. In addition, a router SHOULD support the following
additional information if it supports a routing protocol that could
make use of the information. They are:

o TOS,

o Subnet Mask, or

o Prefix Length, or

o A metric specific to a given routing protocol that can import the
  route.

DISCUSSION
    We intend that one needs to support only the things useful to the
    given routing protocol. The need for TOS should not require the
    vendor to implement the other parts if they are not used.

    Whether a router prefers a static route over a dynamic route (or
    vice versa) or whether the associated metrics are used to choose
    between conflicting static and dynamic routes SHOULD be
    configurable for each static route.

    A router MUST allow a metric to be assigned to a static route for
    each routing domain that it supports. Each such metric MUST be
    explicitly assigned to a specific routing domain. For example:

        route 10.0.0.0/8 via 192.0.2.3 rip metric 3

        route 10.21.0.0/16 via 192.0.2.4 ospf inter-area metric 27

        route 10.22.0.0/16 via 192.0.2.5 egp 123 metric 99

DISCUSSION
    It has been suggested that, ideally, static routes should have
    preference values rather than metrics (since metrics can only be
    compared with metrics of other routes in the same routing domain,
    the metric of a static route could only be compared with metrics
    of other static routes). This is contrary to some current
    implementations, where static routes really do have metrics, and
    those metrics are used to determine whether a particular dynamic
    route overrides the static route to the same destination. Thus,
    this document uses the term metric rather than preference.

    This technique essentially makes the static route into a RIP
    route, or an OSPF route (or whatever, depending on the domain of
    the metric). Thus, the route lookup algorithm of that domain
    applies. However, this is NOT route leaking, in that coercing a
    static route into a dynamic routing domain does not authorize the
    router to redistribute the route into the dynamic routing domain.

    For static routes not put into a specific routing domain, the
    route lookup algorithm is:

    (1) Basic match

    (2) Longest match

    (3) Weak TOS (if TOS supported)

    (4) Best metric (where metric are implementation-defined)

    The last step may not be necessary, but it's useful in the case
    where you want to have a primary static route over one interface
    and a secondary static route over an alternate interface, with
    failover to the alternate path if the interface for the primary
    route fails.

7.5 FILTERING OF ROUTING INFORMATION

   Each router within a network makes forwarding decisions based upon
   information contained within its forwarding database. In a simple
   network the contents of the database may be configured statically.
   As the network grows more complex, the need for dynamic updating of
   the forwarding database becomes critical to the efficient operation
   of the network.

   If the data flow through a network is to be as efficient as possible,
   it is necessary to provide a mechanism for controlling the
   propagation of the information a router uses to build its forwarding
   database. This control takes the form of choosing which sources of

routing information should be trusted and selecting which pieces of
the information to believe. The resulting forwarding database is a
filtered version of the available routing information.

In addition to efficiency, controlling the propagation of routing
information can reduce instability by preventing the spread of
incorrect or bad routing information.

In some cases local policy may require that complete routing
information not be widely propagated.

These filtering requirements apply only to non-SPF-based protocols
(and therefore not at all to routers which don't implement any
distance vector protocols).

7.5.1 Route Validation

A router SHOULD log as an error any routing update advertising a
route that violates the specifications of this memo, unless the
routing protocol from which the update was received uses those values
to encode special routes (such as default routes).

7.5.2 Basic Route Filtering

Filtering of routing information allows control of paths used by a
router to forward packets it receives. A router should be selective
in which sources of routing information it listens to and what routes
it believes. Therefore, a router MUST provide the ability to
specify:

o On which logical interfaces routing information will be accepted
   and which routes will be accepted from each logical interface.

o Whether all routes or only a default route is advertised on a
   logical interface.

Some routing protocols do not recognize logical interfaces as a
source of routing information. In such cases the router MUST provide
the ability to specify

o from which other routers routing information will be accepted.

For example, assume a router connecting one or more leaf networks to
the main portion or backbone of a larger network. Since each of the
leaf networks has only one path in and out, the router can simply
send a default route to them. It advertises the leaf networks to the
main network.

7.5.3 Advanced Route Filtering

   As the topology of a network grows more complex, the need for more
   complex route filtering arises. Therefore, a router SHOULD provide
   the ability to specify independently for each routing protocol:

   o Which logical interfaces or routers routing information (routes)
      will be accepted from and which routes will be believed from each
      other router or logical interface,

   o Which routes will be sent via which logical interface(s), and

   o Which routers routing information will be sent to, if this is
      supported by the routing protocol in use.

   In many situations it is desirable to assign a reliability ordering
   to routing information received from another router instead of the
   simple believe or don't believe choice listed in the first bullet
   above. A router MAY provide the ability to specify:

   o A reliability or preference to be assigned to each route received.
      A route with higher reliability will be chosen over one with lower
      reliability regardless of the routing metric associated with each
      route.

   If a router supports assignment of preferences, the router MUST NOT
   propagate any routes it does not prefer as first party information.
   If the routing protocol being used to propagate the routes does not
   support distinguishing between first and third party information, the
   router MUST NOT propagate any routes it does not prefer.

   DISCUSSION
      For example, assume a router receives a route to network C from
      router R and a route to the same network from router S. If router
      R is considered more reliable than router S traffic destined for
      network C will be forwarded to router R regardless of the route
      received from router S.

   Routing information for routes which the router does not use (router
   S in the above example) MUST NOT be passed to any other router.

7.6 INTER-ROUTING-PROTOCOL INFORMATION EXCHANGE

   Routers MUST be able to exchange routing information between separate
   IP interior routing protocols, if independent IP routing processes
   can run in the same router. Routers MUST provide some mechanism for
   avoiding routing loops when routers are configured for bi-directional
   exchange of routing information between two separate interior routing

processes. Routers MUST provide some priority mechanism for choosing
routes from independent routing processes. Routers SHOULD provide
administrative control of IGP-IGP exchange when used across
administrative boundaries.

Routers SHOULD provide some mechanism for translating or transforming
metrics on a per network basis. Routers (or routing protocols) MAY
allow for global preference of exterior routes imported into an IGP.

DISCUSSION
    Different IGPs use different metrics, requiring some translation
    technique when introducing information from one protocol into
    another protocol with a different form of metric. Some IGPs can
    run multiple instances within the same router or set of routers.
    In this case metric information can be preserved exactly or
    translated.

    There are at least two techniques for translation between
    different routing processes. The static (or reachability)
    approach uses the existence of a route advertisement in one IGP to
    generate a route advertisement in the other IGP with a given
    metric. The translation or tabular approach uses the metric in
    one IGP to create a metric in the other IGP through use of either
    a function (such as adding a constant) or a table lookup.

    Bi-directional exchange of routing information is dangerous
    without control mechanisms to limit feedback. This is the same
    problem that distance vector routing protocols must address with
    the split horizon technique and that EGP addresses with the
    third-party rule. Routing loops can be avoided explicitly through
    use of tables or lists of permitted/denied routes or implicitly
    through use of a split horizon rule, a no-third-party rule, or a
    route tagging mechanism. Vendors are encouraged to use implicit
    techniques where possible to make administration easier for
    network operators.

8. APPLICATION LAYER - NETWORK MANAGEMENT PROTOCOLS

   Note that this chapter supersedes any requirements stated under
   "REMOTE MANAGEMENT" in [INTRO:3].

8.1 The Simple Network Management Protocol - SNMP

8.1.1 SNMP Protocol Elements

   Routers MUST be manageable by SNMP [MGT:3]. The SNMP MUST operate
   using UDP/IP as its transport and network protocols. Others MAY be
   supported (e.g., see [MGT:25, MGT:26, MGT:27, and MGT:28]). SNMP

management operations MUST operate as if the SNMP was implemented on
the router itself. Specifically, management operations MUST be
effected by sending SNMP management requests to any of the IP
addresses assigned to any of the router's interfaces. The actual
management operation may be performed either by the router or by a
proxy for the router.

    DISCUSSION
       This wording is intended to allow management either by proxy,
       where the proxy device responds to SNMP packets that have one of
       the router's IP addresses in the packets destination address
       field, or the SNMP is implemented directly in the router itself
       and receives packets and responds to them in the proper manner.

       It is important that management operations can be sent to one of
       the router's IP Addresses. In diagnosing network problems the
       only thing identifying the router that is available may be one of
       the router's IP address; obtained perhaps by looking through
       another router's routing table.

    All SNMP operations (get, get-next, get-response, set, and trap) MUST
    be implemented.

    Routers MUST provide a mechanism for rate-limiting the generation of
    SNMP trap messages. Routers MAY provide this mechanism through the
    algorithms for asynchronous alert management described in [MGT:5].

    DISCUSSION
       Although there is general agreement about the need to rate-limit
       traps, there is not yet consensus on how this is best achieved.
       The reference cited is considered experimental.

8.2 Community Table

    For the purposes of this specification, we assume that there is an
    abstract 'community table' in the router. This table contains
    several entries, each entry for a specific community and containing
    the parameters necessary to completely define the attributes of that
    community. The actual implementation method of the abstract
    community table is, of course, implementation specific.

    A router's community table MUST allow for at least one entry and
    SHOULD allow for at least two entries.

    DISCUSSION
       A community table with zero capacity is useless. It means that
       the router will not recognize any communities and, therefore, all
       SNMP operations will be rejected.

      Therefore, one entry is the minimal useful size of the table.
      Having two entries allows one entry to be limited to read-only
      access while the other would have write capabilities.

   Routers MUST allow the user to manually (i.e., without using SNMP)
   examine, add, delete and change entries in the SNMP community table.
   The user MUST be able to set the community name or construct a MIB
   view. The user MUST be able to configure communities as read-only
   (i.e., they do not allow SETs) or read-write (i.e., they do allow
   SETs).

   The user MUST be able to define at least one IP address to which
   notifications are sent for each community or MIB view, if traps are
   used. These addresses SHOULD be definable on a community or MIB view
   basis. It SHOULD be possible to enable or disable notifications on a
   community or MIB view basis.

   A router SHOULD provide the ability to specify a list of valid
   network managers for any particular community. If enabled, a router
   MUST validate the source address of the SNMP datagram against the
   list and MUST discard the datagram if its address does not appear.
   If the datagram is discarded the router MUST take all actions
   appropriate to an SNMP authentication failure.

   DISCUSSION
      This is a rather limited authentication system, but coupled with
      various forms of packet filtering may provide some small measure
      of increased security.

   The community table MUST be saved in non-volatile storage.

   The initial state of the community table SHOULD contain one entry,
   with the community name string public and read-only access. The
   default state of this entry MUST NOT send traps. If it is
   implemented, then this entry MUST remain in the community table until
   the administrator changes it or deletes it.

   DISCUSSION
      By default, traps are not sent to this community. Trap PDUs are
      sent to unicast IP addresses. This address must be configured
      into the router in some manner. Before the configuration occurs,
      there is no such address, so to whom should the trap be sent?
      Therefore trap sending to the public community defaults to be
      disabled. This can, of course, be changed by an administrative
      operation once the router is operational.

8.3 Standard MIBS

   All MIBS relevant to a router's configuration are to be implemented.
   To wit:

   o The System, Interface, IP, ICMP, and UDP groups of MIB-II [MGT:2]
     MUST be implemented.

   o The Interface Extensions MIB [MGT:18] MUST be implemented.

   o The IP Forwarding Table MIB [MGT:20] MUST be implemented.

   o If the router implements TCP (e.g., for Telnet) then the TCP group
     of MIB-II [MGT:2] MUST be implemented.

   o If the router implements EGP then the EGP group of MIB-II [MGT:2]
     MUST be implemented.

   o If the router supports OSPF then the OSPF MIB [MGT:14] MUST be
     implemented.

   o If the router supports BGP then the BGP MIB [MGT:15] MUST be
     implemented.

   o If the router has Ethernet, 802.3, or StarLan interfaces then the
     Ethernet-Like MIB [MGT:6] MUST be implemented.

   o If the router has 802.4 interfaces then the 802.4 MIB [MGT:7] MUST
     be implemented.

   o If the router has 802.5 interfaces then the 802.5 MIB [MGT:8] MUST
     be implemented.

   o If the router has FDDI interfaces that implement ANSI SMT 7.3 then
     the FDDI MIB [MGT:9] MUST be implemented.

   o If the router has FDDI interfaces that implement ANSI SMT 6.2 then
     the FDDI MIB [MGT:29] MUST be implemented.

   o If the router has interfaces that use V.24 signalling, such as RS-
     232, V.10, V.11, V.35, V.36, or RS-422/423/449, then the RS-232
     [MGT:10] MIB MUST be implemented.

   o If the router has T1/DS1 interfaces then the T1/DS1 MIB [MGT:16]
     MUST be implemented.

   o If the router has T3/DS3 interfaces then the T3/DS3 MIB [MGT:17]
     MUST be implemented.

o If the router has SMDS interfaces then the SMDS Interface Protocol
   MIB [MGT:19] MUST be implemented.

o If the router supports PPP over any of its interfaces then the PPP
   MIBs [MGT:11], [MGT:12], and [MGT:13] MUST be implemented.

o If the router supports RIP Version 2 then the RIP Version 2 MIB
   [MGT:21] MUST be implemented.

o If the router supports X.25 over any of its interfaces then the
   X.25 MIBs [MGT:22, MGT:23 and MGT:24] MUST be implemented.

8.4 Vendor Specific MIBS

   The Internet Standard and Experimental MIBs do not cover the entire
   range of statistical, state, configuration and control information
   that may be available in a network element. This information is,
   nevertheless, extremely useful. Vendors of routers (and other
   network devices) generally have developed MIB extensions that cover
   this information. These MIB extensions are called Vendor Specific
   MIBs.

   The Vendor Specific MIB for the router MUST provide access to all
   statistical, state, configuration, and control information that is
   not available through the Standard and Experimental MIBs that have
   been implemented. This information MUST be available for both
   monitoring and control operations.

   DISCUSSION
      The intent of this requirement is to provide the ability to do
      anything on the router through SNMP that can be done through a
      console, and vice versa. A certain minimal amount of
      configuration is necessary before SNMP can operate (e.g., the
      router must have an IP address). This initial configuration can
      not be done through SNMP. However, once the initial configuration
      is done, full capabilities ought to be available through network
      management.

   The vendor SHOULD make available the specifications for all Vendor
   Specific MIB variables. These specifications MUST conform to the SMI
   [MGT:1] and the descriptions MUST be in the form specified in
   [MGT:4].

   DISCUSSION
      Making the Vendor Specific MIB available to the user is necessary.
      Without this information the users would not be able to configure
      their network management systems to be able to access the Vendor
      Specific parameters. These parameters would then be useless.

      ne 2 The format of the MIB specification is also specified.
      Parsers that read MIB specifications and generate the needed
      tables for the network management station are available. These
      parsers generally understand only the standard MIB specification
      format.

8.5 Saving Changes

   Parameters altered by SNMP MAY be saved to non-volatile storage.

   DISCUSSION
   Reasons why this requirement is a MAY:

      o The exact physical nature of non-volatile storage is not
        specified in this document. Hence, parameters may be saved in
        NVRAM/EEPROM, local floppy or hard disk, or in some TFTP file
        server or BOOTP server, etc. Suppose that this information is
        in a file that is retrieved through TFTP. In that case, a
        change made to a configuration parameter on the router would
        need to be propagated back to the file server holding the
        configuration file. Alternatively, the SNMP operation would
        need to be directed to the file server, and then the change
        somehow propagated to the router. The answer to this problem
        does not seem obvious.

        This also places more requirements on the host holding the
        configuration information than just having an available TFTP
        server, so much more that its probably unsafe for a vendor to
        assume that any potential customer will have a suitable host
        available.

      o The timing of committing changed parameters to non-volatile
        storage is still an issue for debate. Some prefer to commit
        all changes immediately. Others prefer to commit changes to
        non-volatile storage only upon an explicit command.

9. APPLICATION LAYER - MISCELLANEOUS PROTOCOLS

   For all additional application protocols that a router implements,
   the router MUST be compliant and SHOULD be unconditionally compliant
   with the relevant requirements of [INTRO:3].

9.1 BOOTP

9.1.1 Introduction

   The Bootstrap Protocol (BOOTP) is a UDP/IP-based protocol that allows
   a booting host to configure itself dynamically and without user

   supervision. BOOTP provides a means to notify a host of its assigned
   IP address, the IP address of a boot server host, and the name of a
   file to be loaded into memory and executed ([APPL:1]). Other
   configuration information such as the local prefix length or subnet
   mask, the local time offset, the addresses of default routers, and
   the addresses of various Internet servers can also be communicated to
   a host using BOOTP ([APPL:2]).

9.1.2 BOOTP Relay Agents

   In many cases, BOOTP clients and their associated BOOTP server(s) do
   not reside on the same IP (sub)network. In such cases, a third-party
   agent is required to transfer BOOTP messages between clients and
   servers. Such an agent was originally referred to as a BOOTP
   forwarding agent. However, to avoid confusion with the IP forwarding
   function of a router, the name BOOTP relay agent has been adopted
   instead.

   DISCUSSION
      A BOOTP relay agent performs a task that is distinct from a
      router's normal IP forwarding function. While a router normally
      switches IP datagrams between networks more-or-less transparently,
      a BOOTP relay agent may more properly be thought to receive BOOTP
      messages as a final destination and then generate new BOOTP
      messages as a result. One should resist the notion of simply
      forwarding a BOOTP message straight through like a regular packet.

   This relay-agent functionality is most conveniently located in the
   routers that interconnect the clients and servers (although it may
   alternatively be located in a host that is directly connected to the
   client (sub)net).

   A router MAY provide BOOTP relay-agent capability. If it does, it
   MUST conform to the specifications in [APPL:3].

   Section [5.2.3] discussed the circumstances under which a packet is
   delivered locally (to the router). All locally delivered UDP
   messages whose UDP destination port number is BOOTPS (67) are
   considered for special processing by the router's logical BOOTP relay
   agent.

   Sections [4.2.2.11] and [5.3.7] discussed invalid IP source
   addresses. According to these rules, a router must not forward any
   received datagram whose IP source address is 0.0.0.0. However,
   routers that support a BOOTP relay agent MUST accept for local
   delivery to the relay agent BOOTREQUEST messages whose IP source
   address is 0.0.0.0.

10. OPERATIONS AND MAINTENANCE

   This chapter supersedes any requirements of [INTRO:3] relating to
   "Extensions to the IP Module."

   Facilities to support operation and maintenance (O&M) activities form
   an essential part of any router implementation. Although these
   functions do not seem to relate directly to interoperability, they
   are essential to the network manager who must make the router
   interoperate and must track down problems when it doesn't. This
   chapter also includes some discussion of router initialization and of
   facilities to assist network managers in securing and accounting for
   their networks.

10.1 Introduction

   The following kinds of activities are included under router O&M:

   o Diagnosing hardware problems in the router's processor, in its
     network interfaces, or in its connected networks, modems, or
     communication lines.

   o Installing new hardware

   o Installing new software.

   o Restarting or rebooting the router after a crash.

   o Configuring (or reconfiguring) the router.

   o Detecting and diagnosing Internet problems such as congestion,
     routing loops, bad IP addresses, black holes, packet avalanches,
     and misbehaved hosts.

   o Changing network topology, either temporarily (e.g., to bypass a
     communication line problem) or permanently.

   o Monitoring the status and performance of the routers and the
     connected networks.

   o Collecting traffic statistics for use in (Inter-)network planning.

   o Coordinating the above activities with appropriate vendors and
     telecommunications specialists.

   Routers and their connected communication lines are often operated as
   a system by a centralized O&M organization. This organization may
   maintain a (Inter-)network operation center, or NOC, to carry out its

O&M functions. It is essential that routers support remote control
and monitoring from such a NOC through an Internet path, since
routers might not be connected to the same network as their NOC.
Since a network failure may temporarily preclude network access, many
NOCs insist that routers be accessible for network management through
an alternative means, often dial-up modems attached to console ports
on the routers.

Since an IP packet traversing an internet will often use routers
under the control of more than one NOC, Internet problem diagnosis
will often involve cooperation of personnel of more than one NOC. In
some cases, the same router may need to be monitored by more than one
NOC, but only if necessary, because excessive monitoring could impact
a router's performance.

The tools available for monitoring at a NOC may cover a wide range of
sophistication. Current implementations include multi-window,
dynamic displays of the entire router system. The use of AI
techniques for automatic problem diagnosis is proposed for the
future.

Router O&M facilities discussed here are only a part of the large and
difficult problem of Internet management. These problems encompass
not only multiple management organizations, but also multiple
protocol layers. For example, at the current stage of evolution of
the Internet architecture, there is a strong coupling between host
TCP implementations and eventual IP-level congestion in the router
system [OPER:1]. Therefore, diagnosis of congestion problems will
sometimes require the monitoring of TCP statistics in hosts. There
are currently a number of R&D efforts in progress in the area of
Internet management and more specifically router O&M. These R&D
efforts have already produced standards for router O&M. This is also
an area in which vendor creativity can make a significant
contribution.

10.2 Router Initialization

10.2.1 Minimum Router Configuration

There exists a minimum set of conditions that must be satisfied
before a router may forward packets. A router MUST NOT enable
forwarding on any physical interface unless either:

(1) The router knows the IP address and associated subnet mask or
    network prefix length of at least one logical interface
    associated with that physical interface, or

(2) The router knows that the interface is an unnumbered interface
     and knows its router-id.

These parameters MUST be explicitly configured:

o A router MUST NOT use factory-configured default values for its IP
  addresses, prefix lengths, or router-id, and

o A router MUST NOT assume that an unconfigured interface is an
  unnumbered interface.

   DISCUSSION
      There have been instances in which routers have been shipped with
      vendor-installed default addresses for interfaces. In a few
      cases, this has resulted in routers advertising these default
      addresses into active networks.

10.2.2 Address and Prefix Initialization

   A router MUST allow its IP addresses and their address masks or
   prefix lengths to be statically configured and saved in non-volatile
   storage.

   A router MAY obtain its IP addresses and their corresponding address
   masks dynamically as a side effect of the system initialization
   process (see Section 10.2.3]);

   If the dynamic method is provided, the choice of method to be used in
   a particular router MUST be configurable.

   As was described in Section [4.2.2.11], IP addresses are not
   permitted to have the value 0 or -1 in the <Host-number> or
   <Network-prefix> fields. Therefore, a router SHOULD NOT allow an IP
   address or address mask to be set to a value that would make any of
   the these fields above have the value zero or -1.

   DISCUSSION
      It is possible using arbitrary address masks to create situations
      in which routing is ambiguous (i.e., two routes with different but
      equally specific subnet masks match a particular destination
      address). This is one of the strongest arguments for the use of
      network prefixes, and the reason the use of discontiguous subnet
      masks is not permitted.

   A router SHOULD make the following checks on any address mask it
   installs:

  o The mask is neither all ones nor all zeroes (the prefix length is
    neither zero nor 32).

  o The bits which correspond to the network prefix part of the address
    are all set to 1.

  o The bits that correspond to the network prefix are contiguous.

  DISCUSSION
     The masks associated with routes are also sometimes called subnet
     masks, this test should not be applied to them.

10.2.3 Network Booting using BOOTP and TFTP

   There has been much discussion of how routers can and should be
   booted from the network. These discussions have revolved around
   BOOTP and TFTP. Currently, there are routers that boot with TFTP
   from the network. There is no reason that BOOTP could not be used
   for locating the server that the boot image should be loaded from.

   BOOTP is a protocol used to boot end systems, and requires some
   stretching to accommodate its use with routers. If a router is using
   BOOTP to locate the current boot host, it should send a BOOTP Request
   with its hardware address for its first interface, or, if it has been
   previously configured otherwise, with either another interface's
   hardware address, or another number to put in the hardware address
   field of the BOOTP packet. This is to allow routers without hardware
   addresses (like synchronous line only routers) to use BOOTP for
   bootload discovery. TFTP can then be used to retrieve the image
   found in the BOOTP Reply. If there are no configured interfaces or
   numbers to use, a router MAY cycle through the interface hardware
   addresses it has until a match is found by the BOOTP server.

   A router SHOULD IMPLEMENT the ability to store parameters learned
   through BOOTP into local non-volatile storage. A router MAY
   implement the ability to store a system image loaded over the network
   into local stable storage.

   A router MAY have a facility to allow a remote user to request that
   the router get a new boot image. Differentiation should be made
   between getting the new boot image from one of three locations: the
   one included in the request, from the last boot image server, and
   using BOOTP to locate a server.

10.3 Operation and Maintenance

10.3.1 Introduction

   There is a range of possible models for performing O&M functions on a
   router. At one extreme is the local-only model, under which the O&M
   functions can only be executed locally (e.g., from a terminal plugged
   into the router machine). At the other extreme, the fully remote
   model allows only an absolute minimum of functions to be performed
   locally (e.g., forcing a boot), with most O&M being done remotely
   from the NOC. There are intermediate models, such as one in which
   NOC personnel can log into the router as a host, using the Telnet
   protocol, to perform functions that can also be invoked locally. The
   local-only model may be adequate in a few router installations, but
   remote operation from a NOC is normally required, and therefore
   remote O&M provisions are required for most routers.

   Remote O&M functions may be exercised through a control agent
   (program). In the direct approach, the router would support remote
   O&M functions directly from the NOC using standard Internet protocols
   (e.g., SNMP, UDP or TCP); in the indirect approach, the control agent
   would support these protocols and control the router itself using
   proprietary protocols. The direct approach is preferred, although
   either approach is acceptable. The use of specialized host hardware
   and/or software requiring significant additional investment is
   discouraged; nevertheless, some vendors may elect to provide the
   control agent as an integrated part of the network in which the
   routers are a part. If this is the case, it is required that a means
   be available to operate the control agent from a remote site using
   Internet protocols and paths and with equivalent functionality with
   respect to a local agent terminal.

   It is desirable that a control agent and any other NOC software tools
   that a vendor provides operate as user programs in a standard
   operating system. The use of the standard Internet protocols UDP and
   TCP for communicating with the routers should facilitate this.

   Remote router monitoring and (especially) remote router control
   present important access control problems that must be addressed.
   Care must also be taken to ensure control of the use of router
   resources for these functions. It is not desirable to let router
   monitoring take more than some limited fraction of the router CPU
   time, for example. On the other hand, O&M functions must receive
   priority so they can be exercised when the router is congested, since
   often that is when O&M is most needed.

10.3.2 Out Of Band Access

   Routers MUST support Out-Of-Band (OOB) access. OOB access SHOULD
   provide the same functionality as in-band access. This access SHOULD
   implement access controls, to prevent unauthorized access.

   DISCUSSION
      This Out-Of-Band access will allow the NOC a way to access
      isolated routers during times when network access is not
      available.

      Out-Of-Band access is an important management tool for the network
      administrator. It allows the access of equipment independent of
      the network connections. There are many ways to achieve this
      access. Whichever one is used it is important that the access is
      independent of the network connections. An example of Out-Of-Band
      access would be a serial port connected to a modem that provides
      dial up access to the router.

      It is important that the OOB access provides the same
      functionality as in-band access. In-band access, or accessing
      equipment through the existing network connection, is limiting,
      because most of the time, administrators need to reach equipment
      to figure out why it is unreachable. In band access is still very
      important for configuring a router, and for troubleshooting more
      subtle problems.

10.3.2 Router O&M Functions

10.3.2.1 Maintenance - Hardware Diagnosis

   Each router SHOULD operate as a stand-alone device for the purposes
   of local hardware maintenance. Means SHOULD be available to run
   diagnostic programs at the router site using only on-site tools. A
   router SHOULD be able to run diagnostics in case of a fault. For
   suggested hardware and software diagnostics see Section [10.3.3].

10.3.2.2 Control - Dumping and Rebooting

   A router MUST include both in-band and out-of-band mechanisms to
   allow the network manager to reload, stop, and restart the router. A
   router SHOULD also contain a mechanism (such as a watchdog timer)
   which will reboot the router automatically if it hangs due to a
   software or hardware fault.

   A router SHOULD IMPLEMENT a mechanism for dumping the contents of a
   router's memory (and/or other state useful for vendor debugging after
   a crash), and either saving them on a stable storage device local to

the router or saving them on another host via an up-line dump
mechanism such as TFTP (see [OPER:2], [INTRO:3]).

10.3.2.3 Control - Configuring the Router

Every router has configuration parameters that may need to be set.
It SHOULD be possible to update the parameters without rebooting the
router; at worst, a restart MAY be required. There may be cases when
it is not possible to change parameters without rebooting the router
(for instance, changing the IP address of an interface). In these
cases, care should be taken to minimize disruption to the router and
the surrounding network.

There SHOULD be a way to configure the router over the network either
manually or automatically. A router SHOULD be able to upload or
download its parameters from a host or another router. A means
SHOULD be provided, either as an application program or a router
function, to convert between the parameter format and a human-
editable format. A router SHOULD have some sort of stable storage
for its configuration. A router SHOULD NOT believe protocols such as
RARP, ICMP Address Mask Reply, and MAY not believe BOOTP.

   DISCUSSION
      It is necessary to note here that in the future RARP, ICMP Address
      Mask Reply, BOOTP and other mechanisms may be needed to allow a
      router to auto-configure. Although routers may in the future be
      able to configure automatically, the intent here is to discourage
      this practice in a production environment until auto-configuration
      has been tested more thoroughly. The intent is NOT to discourage
      auto-configuration all together. In cases where a router is
      expected to get its configuration automatically it may be wise to
      allow the router to believe these things as it comes up and then
      ignore them after it has gotten its configuration.

10.3.2.4 Net Booting of System Software

A router SHOULD keep its system image in local non-volatile
storage such as PROM, NVRAM, or disk. It MAY also be able to load
its system software over the network from a host or another
router.

A router that can keep its system image in local non-volatile
storage MAY be configurable to boot its system image over the
network. A router that offers this option SHOULD be configurable
to boot the system image in its non-volatile local storage if it
is unable to boot its system image over the network.

DISCUSSION
    It is important that the router be able to come up and run on its
    own. NVRAM may be a particular solution for routers used in large
    networks, since changing PROMs can be quite time consuming for a
    network manager responsible for numerous or geographically
    dispersed routers. It is important to be able to netboot the
    system image because there should be an easy way for a router to
    get a bug fix or new feature more quickly than getting PROMs
    installed. Also if the router has NVRAM instead of PROMs, it will
    netboot the image and then put it in NVRAM.

    Routers SHOULD perform some basic consistency check on any image
    loaded, to detect and perhaps prevent incorrect images.

A router MAY also be able to distinguish between different
configurations based on which software it is running. If
configuration commands change from one software version to another,
it would be helpful if the router could use the configuration that
was compatible with the software.

10.3.2.5 Detecting and responding to misconfiguration

There MUST be mechanisms for detecting and responding to
misconfigurations. If a command is executed incorrectly, the router
SHOULD give an error message. The router SHOULD NOT accept a poorly
formed command as if it were correct.

DISCUSSION
    There are cases where it is not possible to detect errors: the
    command is correctly formed, but incorrect with respect to the
    network. This may be detected by the router, but may not be
    possible.

Another form of misconfiguration is misconfiguration of the network
to which the router is attached. A router MAY detect
misconfigurations in the network. The router MAY log these findings
to a file, either on the router or a host, so that the network
manager will see that there are possible problems on the network.

DISCUSSION
    Examples of such misconfigurations might be another router with
    the same address as the one in question or a router with the wrong
    address mask. If a router detects such problems it is probably
    not the best idea for the router to try to fix the situation.
    That could cause more harm than good.

10.3.2.6 Minimizing Disruption

   Changing the configuration of a router SHOULD have minimal affect on
   the network. Routing tables SHOULD NOT be unnecessarily flushed when
   a simple change is made to the router. If a router is running
   several routing protocols, stopping one routing protocol SHOULD NOT
   disrupt other routing protocols, except in the case where one network
   is learned by more than one routing protocol.

      DISCUSSION
         It is the goal of a network manager to run a network so that users
         of the network get the best connectivity possible. Reloading a
         router for simple configuration changes can cause disruptions in
         routing and ultimately cause disruptions to the network and its
         users. If routing tables are unnecessarily flushed, for instance,
         the default route will be lost as well as specific routes to sites
         within the network. This sort of disruption will cause
         significant downtime for the users. It is the purpose of this
         section to point out that whenever possible, these disruptions
         should be avoided.

10.3.2.7 Control - Troubleshooting Problems

      (1) A router MUST provide in-band network access, but (except as
          required by Section [8.2]) for security considerations this
          access SHOULD be disabled by default. Vendors MUST document
          the default state of any in-band access. This access SHOULD
          implement access controls, to prevent unauthorized access.

   DISCUSSION
      In-band access primarily refers to access through the normal
      network protocols that may or may not affect the permanent
      operational state of the router. This includes, but is not
      limited to Telnet/RLOGIN console access and SNMP operations.

      This was a point of contention between the operational out of the
      box and secure out of The box contingents. Any automagic access
      to the router may introduce insecurities, but it may be more
      important for the customer to have a router that is accessible
      over the network as soon as it is plugged in. At least one vendor
      supplies routers without any external console access and depends
      on being able to access the router through the network to complete
      its configuration.

      It is the vendors call whether in-band access is enabled by
      default; but it is also the vendor's responsibility to make its
      customers aware of possible insecurities.

    (2) A router MUST provide the ability to initiate an ICMP echo.
        The following options SHOULD be implemented:

        o Choice of data patterns

        o Choice of packet size

        o Record route

        and the following additional options MAY be implemented:

        o Loose source route

        o Strict source route

        o Timestamps

    (3) A router SHOULD provide the ability to initiate a traceroute.
        If traceroute is provided, then the 3rd party traceroute
        SHOULD be implemented.

Each of the above three facilities (if implemented) SHOULD have
access restrictions placed on it to prevent its abuse by unauthorized
persons.

10.4 Security Considerations

10.4.1 Auditing and Audit Trails

Auditing and billing are the bane of the network operator, but are
the two features most requested by those in charge of network
security and those who are responsible for paying the bills. In the
context of security, auditing is desirable if it helps you keep your
network working and protects your resources from abuse, without
costing you more than those resources are worth.

(1) Configuration Changes

    Router SHOULD provide a method for auditing a configuration
    change of a router, even if it's something as simple as
    recording the operator's initials and time of change.

DISCUSSION
    Configuration change logging (who made a configuration change,
    what was changed, and when) is very useful, especially when
    traffic is suddenly routed through Alaska on its way across town.
    So is the ability to revert to a previous configuration.

(2) Packet Accounting

   Vendors should strongly consider providing a system for
   tracking traffic levels between pairs of hosts or networks.
   A mechanism for limiting the collection of this information
   to specific pairs of hosts or networks is also strongly
   encouraged.

DISCUSSION
   A host traffic matrix as described above can give the network
   operator a glimpse of traffic trends not apparent from other
   statistics. It can also identify hosts or networks that are
   probing the structure of the attached networks - e.g., a single
   external host that tries to send packets to every IP address in
   the network address range for a connected network.

(3) Security Auditing

   Routers MUST provide a method for auditing security related
   failures or violations to include:

   o Authorization Failures: bad passwords, invalid SNMP
     communities, invalid authorization tokens,

   o Violations of Policy Controls: Prohibited Source Routes,
     Filtered Destinations, and

   o Authorization Approvals: good passwords - Telnet in-band
     access, console access.

   Routers MUST provide a method of limiting or disabling such
   auditing but auditing SHOULD be on by default. Possible
   methods for auditing include listing violations to a console
   if present, logging or counting them internally, or logging
   them to a remote security server through the SNMP trap
   mechanism or the Unix logging mechanism as appropriate. A
   router MUST implement at least one of these reporting
   mechanisms - it MAY implement more than one.

10.4.2 Configuration Control

   A vendor has a responsibility to use good configuration control
   practices in the creation of the software/firmware loads for their
   routers. In particular, if a vendor makes updates and loads
   available for retrieval over the Internet, the vendor should also
   provide a way for the customer to confirm the load is a valid one,
   perhaps by the verification of a checksum over the load.

   DISCUSSION
      Many vendors currently provide short notice updates of their
      software products through the Internet. This a good trend and
      should be encouraged, but provides a point of vulnerability in the
      configuration control process.

   If a vendor provides the ability for the customer to change the
   configuration parameters of a router remotely, for example through a
   Telnet session, the ability to do so SHOULD be configurable and
   SHOULD default to off. The router SHOULD require valid
   authentication before permitting remote reconfiguration. This
   authentication procedure SHOULD NOT transmit the authentication
   secret over the network. For example, if telnet is implemented, the
   vendor SHOULD IMPLEMENT Kerberos, S-Key, or a similar authentication
   procedure.

   DISCUSSION
      Allowing your properly identified network operator to twiddle with
      your routers is necessary; allowing anyone else to do so is
      foolhardy.

   A router MUST NOT have undocumented back door access and master
   passwords. A vendor MUST ensure any such access added for purposes
   of debugging or product development are deleted before the product is
   distributed to its customers.

   DISCUSSION
      A vendor has a responsibility to its customers to ensure they are
      aware of the vulnerabilities present in its code by intention -
      e.g., in-band access. Trap doors, back doors and master passwords
      intentional or unintentional can turn a relatively secure router
      into a major problem on an operational network. The supposed
      operational benefits are not matched by the potential problems.

11. REFERENCES

   Implementors should be aware that Internet protocol standards are
   occasionally updated. These references are current as of this
   writing, but a cautious implementor will always check a recent
   version of the RFC index to ensure that an RFC has not been updated
   or superseded by another, more recent RFC. Reference [INTRO:6]
   explains various ways to obtain a current RFC index.

   APPL:1.
        Croft, B., and J. Gilmore, "Bootstrap Protocol (BOOTP)", RFC
        951, Stanford University, Sun Microsystems, September 1985.

     APPL:2.
          Alexander, S., and R. Droms, "DHCP Options and BOOTP Vendor
          Extensions", RFC 1533, Lachman Technology, Inc., Bucknell
          University, October 1993.

     APPL:3.
          Wimer, W., "Clarifications and Extensions for the Bootstrap
          Protocol", RFC 1542, Carnegie Mellon University, October 1993.

     ARCH:1.
          DDN Protocol Handbook, NIC-50004, NIC-50005, NIC-50006 (three
          volumes), DDN Network Information Center, SRI International,
          Menlo Park, California, USA, December 1985.

     ARCH:2.
          V. Cerf and R. Kahn, "A Protocol for Packet Network
          Intercommunication", IEEE Transactions on Communication, May
          1974. Also included in [ARCH:1].

     ARCH:3.
          J. Postel, C. Sunshine, and D. Cohen, "The ARPA Internet
          Protocol", Computer Networks, volume 5, number 4, July 1981.
          Also included in [ARCH:1].

     ARCH:4.
          B. Leiner, J. Postel, R. Cole, and D. Mills, :The DARPA
          Internet Protocol Suite", Proceedings of INFOCOM '85, IEEE,
          Washington, DC, March 1985. Also in: IEEE Communications
          Magazine, March 1985. Also available from the Information
          Sciences Institute, University of Southern California as
          Technical Report ISI-RS-85-153.

     ARCH:5.
          D. Comer, "Internetworking With TCP/IP Volume 1: Principles,
          Protocols, and Architecture", Prentice Hall, Englewood Cliffs,
          NJ, 1991.

     ARCH:6.
          W. Stallings, "Handbook of Computer-Communications Standards
          Volume 3: The TCP/IP Protocol Suite", Macmillan, New York, NY,
          1990.

     ARCH:7.
          Postel, J., "Internet Official Protocol Standards", STD 1, RFC
          1780, Internet Architecture Board, March 1995.

ARCH:8.
     Information processing systems - Open Systems Interconnection -
     Basic Reference Model, ISO 7489, International Standards
     Organization, 1984.

ARCH:9
     R.   Braden, J.   Postel, Y.   Rekhter, "Internet Architecture
     Extensions for Shared Media", 05/20/1994

FORWARD:1.
     IETF CIP Working Group (C. Topolcic, Editor), "Experimental
     Internet Stream Protocol", Version 2 (ST-II), RFC 1190, October
     1990.

FORWARD:2.
     Mankin, A., and K.   Ramakrishnan, Editors, "Gateway Congestion
     Control Survey", RFC 1254, MITRE, Digital Equipment Corporation,
     August 1991.

FORWARD:3.
     J.   Nagle, "On Packet Switches with Infinite Storage", IEEE
     Transactions on Communications, volume COM-35, number 4, April
     1987.

FORWARD:4.
     R.   Jain, K.   Ramakrishnan, and D.   Chiu, "Congestion Avoidance
     in Computer Networks With a Connectionless Network Layer",
     Technical Report DEC-TR-506, Digital Equipment Corporation.

FORWARD:5.
     V.   Jacobson, "Congestion Avoidance and Control", Proceedings of
     SIGCOMM '88, Association for Computing Machinery, August 1988.

FORWARD:6.
     W.   Barns, "Precedence and Priority Access Implementation for
     Department of Defense Data Networks", Technical Report MTR-
     91W00029, The Mitre Corporation, McLean, Virginia, USA, July
     1991.

FORWARD:7
     Fang, Chen, Hutchins, "Simulation Results of TCP Performance
     over ATM with and without Flow Control", presentation to the ATM
     Forum, November 15, 1993.

FORWARD:8
     V. Paxson, S.   Floyd "Wide Area Traffic: the Failure of Poisson
     Modeling", short version in SIGCOMM '94.

FORWARD:9
     Leland, Taqqu, Willinger and Wilson, "On the Self-Similar Nature
     of Ethernet Traffic", Proceedings of SIGCOMM '93, September,
     1993.

FORWARD:10
     S.   Keshav "A Control Theoretic Approach to Flow Control",
     SIGCOMM 91, pages 3-16

FORWARD:11
     K.K.   Ramakrishnan and R.   Jain, "A Binary Feedback Scheme for
     Congestion Avoidance in Computer Networks", ACM Transactions of
     Computer Systems, volume 8, number 2, 1980.

FORWARD:12
     H.   Kanakia, P.   Mishara, and A.   Reibman].   "An adaptive
     congestion control scheme for real-time packet video transport",
     In Proceedings of ACM SIGCOMM 1994, pages 20-31, San Francisco,
     California, September 1993.

FORWARD:13
     A.   Demers, S.   Keshav, S.   Shenker, "Analysis and Simulation of
     a Fair Queuing Algorithm",
      93 pages 1-12

FORWARD:14
     Clark, D., Shenker, S., and L.   Zhang, "Supporting Real-Time
     Applications in an Integrated Services Packet Network:
     Architecture and Mechanism", 92 pages 14-26

INTERNET:1.
     Postel, J., "Internet Protocol", STD 5, RFC 791, USC/Information
     Sciences Institute, September 1981.

INTERNET:2.
     Mogul, J., and J.   Postel, "Internet Standard Subnetting
     Procedure", STD 5, RFC 950, Stanford, USC/Information Sciences
     Institute, August 1985.

INTERNET:3.
     Mogul, J., "Broadcasting Internet Datagrams in the Presence of
     Subnets", STD 5, RFC 922, Stanford University, October 1984.

INTERNET:4.
     Deering, S., "Host Extensions for IP Multicasting", STD 5, RFC
     1112, Stanford University, August 1989.

     INTERNET:5.
          Kent, S., "U.S. Department of Defense Security Options for the
          Internet Protocol", RFC 1108, BBN Communications, November 1991.

     INTERNET:6.
          Braden, R., Borman, D., and C. Partridge, "Computing the
          Internet Checksum", RFC 1071, USC/Information Sciences
          Institute, Cray Research, BBN Communications, September 1988.

     INTERNET:7.
          Mallory T., and A. Kullberg, "Incremental Updating of the
          Internet Checksum", RFC 1141, BBN Communications, January 1990.

     INTERNET:8.
          Postel, J., "Internet Control Message Protocol", STD 5, RFC 792,
          USC/Information Sciences Institute, September 1981.

     INTERNET:9.
          A. Mankin, G. Hollingsworth, G. Reichlen, K. Thompson, R.
          Wilder, and R. Zahavi, "Evaluation of Internet Performance -
          FY89", Technical Report MTR-89W00216, MITRE Corporation,
          February, 1990.

     INTERNET:10.
          G. Finn, A "Connectionless Congestion Control Algorithm",
          Computer Communications Review, volume 19, number 5, Association
          for Computing Machinery, October 1989.

     INTERNET:11.
          Prue, W., and J. Postel, "The Source Quench Introduced Delay
          (SQuID)", RFC 1016, USC/Information Sciences Institute, August
          1987.

     INTERNET:12.
          McKenzie, A., "Some comments on SQuID", RFC 1018, BBN Labs,
          August 1987.

     INTERNET:13.
          Deering, S., "ICMP Router Discovery Messages", RFC 1256, Xerox
          PARC, September 1991.

     INTERNET:14.
          Mogul J., and S. Deering, "Path MTU Discovery", RFC 1191,
          DECWRL, Stanford University, November 1990.

INTERNET:15
     Fuller, V., Li, T., Yu, J., and K. Varadhan, "Classless Inter-
     Domain Routing (CIDR): an Address Assignment and Aggregation
     Strategy" RFC 1519, BARRNet, cisco, Merit, OARnet, September
     1993.

INTERNET:16
     St. Johns, M., "Draft Revised IP Security Option", RFC 1038,
     IETF, January 1988.

INTERNET:17
     Prue, W., and J. Postel, "Queuing Algorithm to Provide Type-
     of-service For IP Links", RFC 1046, USC/Information Sciences
     Institute, February 1988.

INTERNET:18
     Postel, J., "Address Mappings", RFC 796, USC/Information
     Sciences Institute, September 1981.

INTRO:1.
     Braden, R., and J. Postel, "Requirements for Internet
     Gateways", STD 4, RFC 1009, USC/Information Sciences Institute,
     June 1987.

INTRO:2.
     Internet Engineering Task Force (R. Braden, Editor),
     "Requirements for Internet Hosts - Communication Layers", STD 3,
     RFC 1122, USC/Information Sciences Institute, October 1989.

INTRO:3.
     Internet Engineering Task Force (R. Braden, Editor),
     "Requirements for Internet Hosts - Application and Support", STD
     3, RFC 1123, USC/Information Sciences Institute, October 1989.

INTRO:4.
     Clark, D., "Modularity and Efficiency in Protocol
     Implementations", RFC 817, MIT Laboratory for Computer Science,
     July 1982.

INTRO:5.
     Clark, D., "The Structuring of Systems Using Upcalls",
     Proceedings of 10th ACM SOSP, December 1985.

INTRO:6.
     Jacobsen, O., and J. Postel, "Protocol Document Order
     Information", RFC 980, SRI, USC/Information Sciences Institute,
     March 1986.

INTRO:7.
     Reynolds, J., and J. Postel, "Assigned Numbers", STD 2, RFC
     1700, USC/Information Sciences Institute, October 1994. This
     document is periodically updated and reissued with a new number.
     It is wise to verify occasionally that the version you have is
     still current.

INTRO:8.
     DoD Trusted Computer System Evaluation Criteria, DoD publication
     5200.28-STD, U.S. Department of Defense, December 1985.

INTRO:9
     Malkin, G., and T. LaQuey Parker, Editors, "Internet Users'
     Glossary", FYI 18, RFC 1392, Xylogics, Inc., UTexas, January
     1993.

LINK:1.
     Leffler, S., and M. Karels, "Trailer Encapsulations", RFC 893,
     University of California at Berkeley, April 1984.

LINK:2
     Simpson, W., "The Point-to-Point Protocol (PPP)", STD 51, RFC
     1661, Daydreamer July 1994.

LINK:3
     McGregor, G., "The PPP Internet Protocol Control Protocol
     (IPCP)", RFC 1332, Merit May 1992.

LINK:4
     Lloyd, B., and W. Simpson, "PPP Authentication Protocols", RFC
     1334, L&A, Daydreamer, May 1992.

LINK:5
     Simpson, W., "PPP Link Quality Monitoring", RFC 1333,
     Daydreamer, May 1992.

MGT:1.
     Rose, M., and K. McCloghrie, "Structure and Identification of
     Management Information of TCP/IP-based Internets", STD 16, RFC
     1155, Performance Systems International, Hughes LAN Systems, May
     1990.

MGT:2.
     McCloghrie, K., and M. Rose (Editors), "Management Information
     Base of TCP/IP-Based Internets: MIB-II", STD 16, RFC 1213,
     Hughes LAN Systems, Inc., Performance Systems International,
     March 1991.

MGT:3.
     Case, J., Fedor, M., Schoffstall, M., and J. Davin, "Simple
     Network Management Protocol", STD 15, RFC 1157, SNMP Research,
     Performance Systems International, MIT Laboratory for Computer
     Science, May 1990.

MGT:4.
     Rose, M., and K. McCloghrie (Editors), "Towards Concise MIB
     Definitions", STD 16, RFC 1212, Performance Systems
     International, Hughes LAN Systems, March 1991.

MGT:5.
     Steinberg, L., "Techniques for Managing Asynchronously Generated
     Alerts", RFC 1224, IBM Corporation, May 1991.

MGT:6.
     Kastenholz, F., "Definitions of Managed Objects for the
     Ethernet-like Interface Types", RFC 1398, FTP Software, Inc.,
     January 1993.

MGT:7.
     McCloghrie, K., and R. Fox "IEEE 802.4 Token Bus MIB", RFC 1230,
     Hughes LAN Systems, Inc., Synoptics, Inc., May 1991.

MGT:8.
     McCloghrie, K., Fox R., and E. Decker, "IEEE 802.5 Token Ring
     MIB", RFC 1231, Hughes LAN Systems, Inc., Synoptics, Inc., cisco
     Systems, Inc., February 1993.

MGT:9.
     Case, J., and A. Rijsinghani, "FDDI Management Information
     Base", RFC 1512, The University of Tennesse and SNMP Research,
     Digital Equipment Corporation, September 1993.

MGT:10.
     Stewart, B., Editor "Definitions of Managed Objects for RS-232-
     like Hardware Devices", RFC 1317, Xyplex, Inc., April 1992.

MGT:11.
     Kastenholz, F., "Definitions of Managed Objects for the Link
     Control Protocol of the Point-to-Point Protocol", RFC 1471, FTP
     Software, Inc., June 1992.

MGT:12.
     Kastenholz, F., "The Definitions of Managed Objects for the
     Security Protocols of the Point-to-Point Protocol", RFC 1472,
     FTP Software, Inc., June 1992.

MGT:13.
     Kastenholz, F., "The Definitions of Managed Objects for the IP
     Network Control Protocol of the Point-to-Point Protocol", RFC
     1473, FTP Software, Inc., June 1992.

MGT:14.
     Baker, F., and R. Coltun, "OSPF Version 2 Management
     Information Base", RFC 1253, ACC, Computer Science Center,
     August 1991.

MGT:15.
     Willis, S., and J. Burruss, "Definitions of Managed Objects for
     the Border Gateway Protocol (Version 3)", RFC 1269, Wellfleet
     Communications Inc., October 1991.

MGT:16.
     Baker, F., and J. Watt, "Definitions of Managed Objects for the
     DS1 and E1 Interface Types", RFC 1406, Advanced Computer
     Communications, Newbridge Networks Corporation, January 1993.

MGT:17.
     Cox, T., and K. Tesink, Editors "Definitions of Managed Objects
     for the DS3/E3 Interface Types", RFC 1407, Bell Communications
     Research, January 1993.

MGT:18.
     McCloghrie, K., "Extensions to the Generic-Interface MIB", RFC
     1229, Hughes LAN Systems, August 1992.

MGT:19.
     Cox, T., and K. Tesink, "Definitions of Managed Objects for the
     SIP Interface Type", RFC 1304, Bell Communications Research,
     February 1992.

MGT:20
     Baker, F., "IP Forwarding Table MIB", RFC 1354, ACC, July 1992.

MGT:21.
     Malkin, G., and F. Baker, "RIP Version 2 MIB Extension", RFC
     1724, Xylogics, Inc., Cisco Systems, November 1994

MGT:22.
     Throop, D., "SNMP MIB Extension for the X.25 Packet Layer", RFC
     1382, Data General Corporation, November 1992.

MGT:23.
     Throop, D., and F. Baker, "SNMP MIB Extension for X.25 LAPB",
     RFC 1381, Data General Corporation, ACC, November 1992.

MGT:24.
     Throop, D., and F. Baker, "SNMP MIB Extension for MultiProtocol
     Interconnect over X.25", RFC 1461, Data General Corporation, May
     1993.

MGT:25.
     Rose, M., "SNMP over OSI", RFC 1418, Dover Beach Consulting,
     Inc., March 1993.

MGT:26.
     Minshall, G., and M. Ritter, "SNMP over AppleTalk", RFC 1419,
     Novell, Inc., Apple Computer, Inc., March 1993.

MGT:27.
     Bostock, S., "SNMP over IPX", RFC 1420, Novell, Inc., March
     1993.

MGT:28.
     Schoffstall, M., Davin, C., Fedor, M., and J. Case, "SNMP over
     Ethernet", RFC 1089, Rensselaer Polytechnic Institute, MIT
     Laboratory for Computer Science, NYSERNet, Inc., University of
     Tennessee at Knoxville, February 1989.

MGT:29.
     Case, J., "FDDI Management Information Base", RFC 1285, SNMP
     Research, Incorporated, January 1992.

OPER:1.
     Nagle, J., "Congestion Control in IP/TCP Internetworks", RFC
     896, FACC, January 1984.

OPER:2.
     Sollins, K., "TFTP Protocol (revision 2)", RFC 1350, MIT, July
     1992.

ROUTE:1.
     Moy, J., "OSPF Version 2", RFC 1583, Proteon, March 1994.

ROUTE:2.
     Callon, R., "Use of OSI IS-IS for Routing in TCP/IP and Dual
     Environments", RFC 1195, DEC, December 1990.

ROUTE:3.
     Hedrick, C., "Routing Information Protocol", RFC 1058, Rutgers
     University, June 1988.

ROUTE:4.
     Lougheed, K., and Y. Rekhter, "A Border Gateway Protocol 3
     (BGP-3)", RFC 1267, cisco, T.J. Watson Research Center, IBM
     Corp., October 1991.

ROUTE:5.
     Gross, P, and Y. Rekhter, "Application of the Border Gateway
     Protocol in the Internet", RFC 1772, T.J. Watson Research
     Center, IBM Corp., MCI, March 1995.

ROUTE:6.
     Mills, D., "Exterior Gateway Protocol Formal Specification", RFC
     904, UDEL, April 1984.

ROUTE:7.
     Rosen, E., "Exterior Gateway Protocol (EGP)", RFC 827, BBN,
     October 1982.

ROUTE:8.
     Seamonson, L, and E. Rosen, "STUB" "Exterior Gateway Protocol",
     RFC 888, BBN, January 1984.

ROUTE:9.
     Waitzman, D., Partridge, C., and S. Deering, "Distance Vector
     Multicast Routing Protocol", RFC 1075, BBN, Stanford, November
     1988.

ROUTE:10.
     Deering, S., Multicast Routing in Internetworks and Extended
     LANs, Proceedings of '88, Association for Computing Machinery,
     August 1988.

ROUTE:11.
     Almquist, P., "Type of Service in the Internet Protocol Suite",
     RFC 1349, Consultant, July 1992.

ROUTE:12.
     Rekhter, Y., "Experience with the BGP Protocol", RFC 1266, T.J.
     Watson Research Center, IBM Corp., October 1991.

ROUTE:13.
     Rekhter, Y., "BGP Protocol Analysis", RFC 1265, T.J. Watson
     Research Center, IBM Corp., October 1991.

TRANS:1.
      Postel, J., "User Datagram Protocol", STD 6, RFC 768,
      USC/Information Sciences Institute, August 1980.

TRANS:2.
      Postel, J., "Transmission Control Protocol", STD 7, RFC 793,
      USC/Information Sciences Institute, September 1981.

APPENDIX A. REQUIREMENTS FOR SOURCE-ROUTING HOSTS

   Subject to restrictions given below, a host MAY be able to act as an
   intermediate hop in a source route, forwarding a source-routed
   datagram to the next specified hop.

   However, in performing this router-like function, the host MUST obey
   all the relevant rules for a router forwarding source-routed
   datagrams [INTRO:2]. This includes the following specific
   provisions:

   (A) TTL
       The TTL field MUST be decremented and the datagram perhaps
       discarded as specified for a router in [INTRO:2].

   (B) ICMP Destination Unreachable
       A host MUST be able to generate Destination Unreachable messages
       with the following codes:
       4 (Fragmentation Required but DF Set) when a source-routed
         datagram cannot be fragmented to fit into the target network;
       5 (Source Route Failed) when a source-routed datagram cannot be
         forwarded, e.g., because of a routing problem or because the
         next hop of a strict source route is not on a connected
         network.

   (C) IP Source Address
       A source-routed datagram being forwarded MAY (and normally will)
       have a source address that is not one of the IP addresses of the
       forwarding host.

   (D) Record Route Option
       A host that is forwarding a source-routed datagram containing a
       Record Route option MUST update that option, if it has room.

   (E) Timestamp Option
       A host that is forwarding a source-routed datagram containing a
       Timestamp Option MUST add the current timestamp to that option,
       according to the rules for this option.

   To define the rules restricting host forwarding of source-routed
   datagrams, we use the term local source-routing if the next hop will
   be through the same physical interface through which the datagram
   arrived; otherwise, it is non-local source-routing.

   A host is permitted to perform local source-routing without
   restriction.

A host that supports non-local source-routing MUST have a
configurable switch to disable forwarding, and this switch MUST
default to disabled.

The host MUST satisfy all router requirements for configurable policy
filters [INTRO:2] restricting non-local forwarding.

If a host receives a datagram with an incomplete source route but
does not forward it for some reason, the host SHOULD return an ICMP
Destination Unreachable (code 5, Source Route Failed) message, unless
the datagram was itself an ICMP error message.

APPENDIX B. GLOSSARY

This Appendix defines specific terms used in this memo. It also
defines some general purpose terms that may be of interest. See also
[INTRO:9] for a more general set of definitions.

Autonomous System (AS)
     An Autonomous System (AS) is a connected segment of a network
     topology that consists of a collection of subnetworks (with
     hosts attached) interconnected by a set of routes. The
     subnetworks and the routers are expected to be under the control
     of a single operations and maintenance (O&M) organization.
     Within an AS routers may use one or more interior routing
     protocols, and sometimes several sets of metrics. An AS is
     expected to present to other ASs an appearence of a coherent
     interior routing plan, and a consistent picture of the
     destinations reachable through the AS. An AS is identified by
     an Autonomous System number.
Connected Network
     A network prefix to which a router is interfaced is often known
     as a local network or the subnetwork of that router. However,
     these terms can cause confusion, and therefore we use the term
     Connected Network in this memo.

Connected (Sub)Network
     A Connected (Sub)Network is an IP subnetwork to which a router
     is interfaced, or a connected network if the connected network
     is not subnetted. See also Connected Network.

Datagram
     The unit transmitted between a pair of internet modules. Data,
     called datagrams, from sources to destinations. The Internet
     Protocol does not provide a reliable communication facility.
     There are no acknowledgments either end-to-end or hop-by-hop.
     There is no error no retransmissions. There is no flow control.
     See IP.

Default Route
     A routing table entry that is used to direct any data addressed
     to any network prefixes not explicitly listed in the routing
     table.

Dense Mode
     In multicast forwarding, two paradigms are possible: in Dense
     Mode forwarding, a network multicast is forwarded as a data link
     layer multicast to all interfaces except that on which it was
     received, unless and until the router is instructed not to by a
     multicast routing neighbor. See Sparse Mode.

EGP

     Exterior Gateway Protocol A protocol that distributes routing
     information to the gateways (routers) which connect autonomous
     systems. See IGP.

EGP-2
     Exterior Gateway Protocol version 2 This is an EGP routing
     protocol developed to handle traffic between Autonomous Systems
     in the Internet.

Forwarder
     The logical entity within a router that is responsible for
     switching packets among the router's interfaces. The Forwarder
     also makes the decisions to queue a packet for local delivery,
     to queue a packet for transmission out another interface, or
     both.

Forwarding
     Forwarding is the process a router goes through for each packet
     received by the router. The packet may be consumed by the
     router, it may be output on one or more interfaces of the
     router, or both. Forwarding includes the process of deciding
     what to do with the packet as well as queuing it up for
     (possible) output or internal consumption.

Forwarding Information Base (FIB)
     The table containing the information necessary to forward IP
     Datagrams, in this document, is called the Forwarding
     Information Base. At minimum, this contains the interface
     identifier and next hop information for each reachable
     destination network prefix.

Fragment
     An IP datagram that represents a portion of a higher layer's
     packet that was too large to be sent in its entirety over the
     output network.

General Purpose Serial Interface
     A physical medium capable of connecting exactly two systems, and
     therefore configurable as a point to point line, but also
     configurable to support link layer networking using protocols
     such as X.25 or Frame Relay. A link layer network connects
     another system to a switch, and a higher communication layer
     multiplexes virtual circuits on the connection. See Point to
     Point Line.

IGP
     Interior Gateway Protocol A protocol that distributes routing
     information with an Autonomous System (AS). See EGP.

Interface IP Address
     The IP Address and network prefix length that is assigned to a
     specific interface of a router.

Internet Address
     An assigned number that identifies a host in an internet. It
     has two parts: an IP address and a prefix length. The prefix
     length indicates how many of the most specific bits of the
     address constitute the network prefix.

IP
     Internet Protocol The network layer protocol for the Internet.
     It is a packet switching, datagram protocol defined in RFC 791.
     IP does not provide a reliable communications facility; that is,
     there are no end-to-end of hop-by-hop acknowledgments.

IP Datagram
     An IP Datagram is the unit of end-to-end transmission in the
     Internet Protocol. An IP Datagram consists of an IP header
     followed by all of higher-layer data (such as TCP, UDP, ICMP,
     and the like). An IP Datagram is an IP header followed by a
     message.

     An IP Datagram is a complete IP end-to-end transmission unit.
     An IP Datagram is composed of one or more IP Fragments.

     In this memo, the unqualified term Datagram should be understood
     to refer to an IP Datagram.

IP Fragment
     An IP Fragment is a component of an IP Datagram. An IP Fragment
     consists of an IP header followed by all or part of the higher-
     layer of the original IP Datagram.

     One or more IP Fragments comprises a single IP Datagram.

In this memo, the unqualified term Fragment should be understood
to refer to an IP Fragment.

IP Packet
     An IP Datagram or an IP Fragment.

     In this memo, the unqualified term Packet should generally be
     understood to refer to an IP Packet.

Logical [network] interface
     We define a logical [network] interface to be a logical path,
     distinguished by a unique IP address, to a connected network.

Martian Filtering
     A packet that contains an invalid source or destination address
     is considered to be martian and discarded.

MTU (Maximum Transmission Unit)
     The size of the largest packet that can be transmitted or
     received through a logical interface. This size includes the IP
     header but does not include the size of any Link Layer headers
     or framing.

Multicast
     A packet that is destined for multiple hosts. See broadcast.

Multicast Address
     A special type of address that is recognizable by multiple
     hosts.

     A Multicast Address is sometimes known as a Functional Address
     or a Group Address.

Network Prefix
     The portion of an IP Address that signifies a set of systems.
     It is selected from the IP Address by logically ANDing a subnet
     mask with the address, or (equivalently) setting the bits of the
     address not among the most significant <prefix-length> bits of
     the address to zero.

Originate
     Packets can be transmitted by a router for one of two reasons:
     1) the packet was received and is being forwarded or 2) the
     router itself created the packet for transmission (such as route
     advertisements). Packets that the router creates for
     transmission are said to originate at the router.

Packet
      A packet is the unit of data passed across the interface between
      the Internet Layer and the Link Layer. It includes an IP header
      and data. A packet may be a complete IP datagram or a fragment
      of an IP datagram.

Path
      The sequence of routers and (sub-)networks that a packet
      traverses from a particular router to a particular destination
      host. Note that a path is uni-directional; it is not unusual to
      have different paths in the two directions between a given host
      pair.

Physical Network
      A Physical Network is a network (or a piece of an internet)
      which is contiguous at the Link Layer. Its internal structure
      (if any) is transparent to the Internet Layer.

      In this memo, several media components that are connected using
      devices such as bridges or repeaters are considered to be a
      single Physical Network since such devices are transparent to
      the IP.

Physical Network Interface
      This is a physical interface to a Connected Network and has a
      (possibly unique) Link-Layer address. Multiple Physical Network
      Interfaces on a single router may share the same Link-Layer
      address, but the address must be unique for different routers on
      the same Physical Network.

Point to Point Line
      A physical medium capable of connecting exactly two systems. In
      this document, it is only used to refer to such a line when used
      to connect IP entities. See General Purpose Serial Interface.

router
      A special-purpose dedicated computer that connects several
      networks. Routers switch packets between these networks in a
      process known as forwarding. This process may be repeated
      several times on a single packet by multiple routers until the
      packet can be delivered to the final destination - switching the
      packet from router to router to router... until the packet gets
      to its destination.

RPF
      Reverse Path Forwarding - A method used to deduce the next hops
      for broadcast and multicast packets.

Silently Discard
      This memo specifies several cases where a router is to Silently
      Discard a received packet (or datagram). This means that the
      router should discard the packet without further processing, and
      that the router will not send any ICMP error message (see
      Section [4.3.2]) as a result. However, for diagnosis of
      problems, the router should provide the capability of logging
      the error (see Section [1.3.3]), including the contents of the
      silently discarded packet, and should record the event in a
      statistics counter.

Silently Ignore
      A router is said to Silently Ignore an error or condition if it
      takes no action other than possibly generating an error report
      in an error log or through some network management protocol, and
      discarding, or ignoring, the source of the error. In
      particular, the router does NOT generate an ICMP error message.

Sparse Mode
      In multicast forwarding, two paradigms are possible: in Sparse
      Mode forwarding, a network layer multicast datagram is forwarded
      as a data link layer multicast frame to routers and hosts that
      have asked for it. The initial forwarding state is the inverse
      of dense-mode in that it assumes no part of the network wants
      the data. See Dense Mode.

Specific-destination address
      This is defined to be the destination address in the IP header
      unless the header contains an IP broadcast or IP multicast
      address, in which case the specific-destination is an IP address
      assigned to the physical interface on which the packet arrived.

subnet
      A portion of a network, which may be a physically independent
      network, which shares a network address with other portions of
      the network and is distinguished by a subnet number. A subnet
      is to a network what a network is to an internet.

subnet number
      A part of the internet address that designates a subnet. It is
      ignored for the purposes internet routing, but is used for
      intranet routing.

TOS
      Type Of Service A field in the IP header that represents the
      degree of reliability expected from the network layer by the
      transport layer or application.

     TTL
          Time To Live A field in the IP header that represents how long a
          packet is considered valid. It is a combination hop count and
          timer value.

APPENDIX C. FUTURE DIRECTIONS

     This appendix lists work that future revisions of this document may
     wish to address.

     In the preparation of Router Requirements, we stumbled across several
     other architectural issues. Each of these is dealt with somewhat in
     the document, but still ought to be classified as an open issue in
     the IP architecture.

     Most of the he topics presented here generally indicate areas where
     the technology is still relatively new and it is not appropriate to
     develop specific requirements since the community is still gaining
     operational experience.

     Other topics represent areas of ongoing research and indicate areas
     that the prudent developer would closely monitor.

     (1) SNMP Version 2

     (2) Additional SNMP MIBs

     (7) More detailed requirements for leaking routes between routing
          protocols

     (8) Router system security

     (9) Routing protocol security

     (10) Internetwork Protocol layer security. There has been extensive
          work refining the security of IP since the original work writing
          this document. This security work should be included in here.

     (12) Load Splitting

     (13) Sending fragments along different paths

     (15) Multiple logical (sub)nets on the same wire. Router
          Requirements does not require support for this. We made some
          attempt to identify pieces of the architecture (e.g., forwarding
          of directed broadcasts and issuing of Redirects) where the
          wording of the rules has to be done carefully to make the right

thing happen, and tried to clearly distinguish logical
interfaces from physical interfaces. However, we did not study
this issue in detail, and we are not at all confident that all
the rules in the document are correct in the presence of
multiple logical (sub)nets on the same wire.

(15) Congestion control and resource management. On the advice of
the IETF's experts (Mankin and Ramakrishnan) we deprecated
(SHOULD NOT) Source Quench and said little else concrete
(Section 5.3.6).

(16) Developing a Link-Layer requirements document that would be
common for both routers and hosts.

(17) Developing a common PPP LQM algorithm.

(18) Investigate of other information (above and beyond section
[3.2]) that passes between the layers, such as physical network
MTU, mappings of IP precedence to Link Layer priority values,
etc.

(19) Should the Link Layer notify IP if address resolution failed
(just like it notifies IP when there is a Link Layer priority
value problem)?

(20) Should all routers be required to implement a DNS resolver?

(21) Should a human user be able to use a host name anywhere you can
use an IP address when configuring the router? Even in ping and
traceroute?

(22) Almquist's draft ruminations on the next hop and ruminations on
route leaking need to be reviewed, brought up to date, and
published.

(23) Investigation is needed to determine if a redirect message for
precedence is needed or not. If not, are the type-of-service
redirects acceptable?

(24) RIPv2 and RIP+CIDR and variable length network prefixes.

(25) BGP-4 CIDR is going to be important, and everyone is betting on
BGP-4. We can't avoid mentioning it. Probably need to describe
the differences between BGP-3 and BGP-4, and explore upgrade
issues...

(26) Loose Source Route Mobile IP and some multicasting may require
this. Perhaps it should be elevated to a SHOULD (per Fred

Baker's Suggestion).

APPENDIX D. Multicast Routing Protocols

   Multicasting is a relatively new technology within the Internet
   Protocol family. It is not widely deployed or commonly in use yet.
   Its importance, however, is expected to grow over the coming years.

   This Appendix describes some of the technologies being investigated
   for routing multicasts through the Internet.

   A diligent implementor will keep abreast of developments in this area
   to properly develop multicast facilities.

   This Appendix does not specify any standards or requirements.

D.1 Introduction

   Multicast routing protocols enable the forwarding of IP multicast
   datagrams throughout a TCP/IP internet. Generally these algorithms
   forward the datagram based on its source and destination addresses.
   Additionally, the datagram may need to be forwarded to several
   multicast group members, at times requiring the datagram to be
   replicated and sent out multiple interfaces.

   The state of multicast routing protocols is less developed than the
   protocols available for the forwarding of IP unicasts. Three
   experimental multicast routing protocols have been documented for
   TCP/IP. Each uses the IGMP protocol (discussed in Section [4.4]) to
   monitor multicast group membership.

D.2 Distance Vector Multicast Routing Protocol - DVMRP

   DVMRP, documented in [ROUTE:9], is based on Distance Vector or
   Bellman-Ford technology. It routes multicast datagrams only, and
   does so within a single Autonomous System. DVMRP is an
   implementation of the Truncated Reverse Path Broadcasting algorithm
   described in [ROUTE:10]. In addition, it specifies the tunneling of
   IP multicasts through non-multicast-routing-capable IP domains.

D.3 Multicast Extensions to OSPF - MOSPF

   MOSPF, currently under development, is a backward-compatible addition
   to OSPF that allows the forwarding of both IP multicasts and unicasts
   within an Autonomous System. MOSPF routers can be mixed with OSPF
   routers within a routing domain, and they will interoperate in the
   forwarding of unicasts. OSPF is a link-state or SPF-based protocol.

By adding link state advertisements that pinpoint group membership,
MOSPF routers can calculate the path of a multicast datagram as a
tree rooted at the datagram source. Those branches that do not
contain group members can then be discarded, eliminating unnecessary
datagram forwarding hops.

D.4 Protocol Independent Multicast - PIM

PIM, currently under development, is a multicast routing protocol
that runs over an existing unicast infrastructure. PIM provides for
both dense and sparse group membership. It is different from other
protocols, since it uses an explicit join model for sparse groups.
Joining occurs on a shared tree and can switch to a per-source tree.
Where bandwidth is plentiful and group membership is dense, overhead
can be reduced by flooding data out all links and later pruning
exception cases where there are no group members.

APPENDIX E Additional Next-Hop Selection Algorithms

Section [5.2.4.3] specifies an algorithm that routers ought to use
when selecting a next-hop for a packet.

This appendix provides historical perspective for the next-hop
selection problem. It also presents several additional pruning rules
and next-hop selection algorithms that might be found in the
Internet.

This appendix presents material drawn from an earlier, unpublished,
work by Philip Almquist; Ruminations on the Next Hop.

This Appendix does not specify any standards or requirements.

E.1. Some Historical Perspective

It is useful to briefly review the history of the topic, beginning
with what is sometimes called the "classic model" of how a router
makes routing decisions. This model predates IP. In this model, a
router speaks some single routing protocol such as RIP. The protocol
completely determines the contents of the router's Forwarding
Information Base (FIB). The route lookup algorithm is trivial: the
router looks in the FIB for a route whose destination attribute
exactly matches the network prefix portion of the destination address
in the packet. If one is found, it is used; if none is found, the
destination is unreachable. Because the routing protocol keeps at
most one route to each destination, the problem of what to do when
there are multiple routes that match the same destination cannot
arise.

Over the years, this classic model has been augmented in small ways.
With the deployment of default routes, subnets, and host routes, it
became possible to have more than one routing table entry which in
some sense matched the destination. This was easily resolved by a
consensus that there was a hierarchy of routes: host routes should be
preferred over subnet routes, subnet routes over net routes, and net
routes over default routes.

With the deployment of technologies supporting variable length subnet
masks (variable length network prefixes), the general approach
remained the same although its description became a little more
complicated; network prefixes were introduced as a conscious
simplification and regularization of the architecture. We now say
that each route to a network prefix route has a prefix length
associated with it. This prefix length indicates the number of bits
in the prefix. This may also be represented using the classical
subnet mask. A route cannot be used to route a packet unless each
significant bit in the route's network prefix matches the
corresponding bit in the packet's destination address. Routes with
more bits set in their masks are preferred over routes that have
fewer bits set in their masks. This is simply a generalization of
the hierarchy of routes described above, and will be referred to for
the rest of this memo as choosing a route by preferring longest
match.

Another way the classic model has been augmented is through a small
amount of relaxation of the notion that a routing protocol has
complete control over the contents of the routing table. First,
static routes were introduced. For the first time, it was possible
to simultaneously have two routes (one dynamic and one static) to the
same destination. When this happened, a router had to have a policy
(in some cases configurable, and in other cases chosen by the author
of the router's software) which determined whether the static route
or the dynamic route was preferred. However, this policy was only
used as a tie-breaker when longest match didn't uniquely determine
which route to use. Thus, for example, a static default route would
never be preferred over a dynamic net route even if the policy
preferred static routes over dynamic routes.

The classic model had to be further augmented when inter-domain
routing protocols were invented. Traditional routing protocols came
to be called "interior gateway protocols" (IGPs), and at each
Internet site there was a strange new beast called an "exterior
gateway", a router that spoke EGP to several "BBN Core Gateways" (the
routers that made up the Internet backbone at the time) at the same
time as it spoke its IGP to the other routers at its site. Both
protocols wanted to determine the contents of the router's routing
table. Theoretically, this could result in a router having three

routes (EGP, IGP, and static) to the same destination. Because of
the Internet topology at the time, it was resolved with little debate
that routers would be best served by a policy of preferring IGP
routes over EGP routes. However, the sanctity of longest match
remained unquestioned: a default route learned from the IGP would
never be preferred over a net route from learned EGP.

Although the Internet topology, and consequently routing in the
Internet, have evolved considerably since then, this slightly
augmented version of the classic model has survived intact to this
day in the Internet (except that BGP has replaced EGP). Conceptually
(and often in implementation) each router has a routing table and one
or more routing protocol processes. Each of these processes can add
any entry that it pleases, and can delete or modify any entry that it
has created. When routing a packet, the router picks the best route
using longest match, augmented with a policy mechanism to break ties.
Although this augmented classic model has served us well, it has a
number of shortcomings:

o It ignores (although it could be augmented to consider) path
  characteristics such as quality of service and MTU.

o It doesn't support routing protocols (such as OSPF and Integrated
  IS-IS) that require route lookup algorithms different than pure
  longest match.

o There has not been a firm consensus on what the tie-breaking
  mechanism ought to be. Tie-breaking mechanisms have often been
  found to be difficult if not impossible to configure in such a way
  that the router will always pick what the network manger considers
  to be the "correct" route.

E.2. Additional Pruning Rules

   Section [5.2.4.3] defined several pruning rules to use to select
   routes from the FIB. There are other rules that could also be
   used.

   o OSPF Route Class
     Routing protocols that have areas or make a distinction between
     internal and external routes divide their routes into classes
     by the type of information used to calculate the route. A
     route is always chosen from the most preferred class unless
     none is available, in which case one is chosen from the second
     most preferred class, and so on. In OSPF, the classes (in
     order from most preferred to least preferred) are intra-area,
     inter-area, type 1 external (external routes with internal
     metrics), and type 2 external. As an additional wrinkle, a

          router is configured to know what addresses ought to be
          accessible using intra-area routes, and will not use inter-
          area or external routes to reach these destinations even when
          no intra-area route is available.

          More precisely, we assume that each route has a class
          attribute, called route.class, which is assigned by the routing
          protocol. The set of candidate routes is examined to determine
          if it contains any for which route.class = intra-area. If so,
          all routes except those for which route.class = intra-area are
          discarded. Otherwise, router checks whether the packet's
          destination falls within the address ranges configured for the
          local area. If so, the entire set of candidate routes is
          deleted. Otherwise, the set of candidate routes is examined to
          determine if it contains any for which route.class = inter-
          area. If so, all routes except those for which route.class =
          inter-area are discarded. Otherwise, the set of candidate
          routes is examined to determine if it contains any for which
          route.class = type 1 external. If so, all routes except those
          for which route.class = type 1 external are discarded.

     o IS-IS Route Class
       IS-IS route classes work identically to OSPF's. However, the
       set of classes defined by Integrated IS-IS is different, such
       that there isn't a one-to-one mapping between IS-IS route
       classes and OSPF route classes. The route classes used by
       Integrated IS-IS are (in order from most preferred to least
       preferred) intra-area, inter-area, and external.

       The Integrated IS-IS internal class is equivalent to the OSPF
       internal class. Likewise, the Integrated IS-IS external class
       is equivalent to OSPF's type 2 external class. However,
       Integrated IS-IS does not make a distinction between inter-area
       routes and external routes with internal metrics - both are
       considered to be inter-area routes. Thus, OSPF prefers true
       inter-area routes over external routes with internal metrics,
       whereas Integrated IS-IS gives the two types of routes equal
       preference.

     o IDPR Policy
       A specific case of Policy. The IETF's Inter-domain Policy
       Routing Working Group is devising a routing protocol called
       Inter-Domain Policy Routing (IDPR) to support true policy-based
       routing in the Internet. Packets with certain combinations of
       header attributes (such as specific combinations of source and
       destination addresses or special IDPR source route options) are
       required to use routes provided by the IDPR protocol. Thus,
       unlike other Policy pruning rules, IDPR Policy would have to be

applied before any other pruning rules except Basic Match.

Specifically, IDPR Policy examines the packet being forwarded
to ascertain if its attributes require that it be forwarded
using policy-based routes. If so, IDPR Policy deletes all
routes not provided by the IDPR protocol.

E.3 Some Route Lookup Algorithms

This section examines several route lookup algorithms that are in
use or have been proposed. Each is described by giving the
sequence of pruning rules it uses. The strengths and weaknesses
of each algorithm are presented

E.3.1 The Revised Classic Algorithm

The Revised Classic Algorithm is the form of the traditional
algorithm that was discussed in Section [E.1]. The steps of this
algorithm are:

1. Basic match
2. Longest match
3. Best metric
4. Policy

Some implementations omit the Policy step, since it is needed only
when routes may have metrics that are not comparable (because they
were learned from different routing domains).

The advantages of this algorithm are:

(1) It is widely implemented.

(2) Except for the Policy step (which an implementor can choose to
     make arbitrarily complex) the algorithm is simple both to
     understand and to implement.

Its disadvantages are:

(1) It does not handle IS-IS or OSPF route classes, and therefore
     cannot be used for Integrated IS-IS or OSPF.

(2) It does not handle TOS or other path attributes.

(3) The policy mechanisms are not standardized in any way, and are
     therefore are often implementation-specific. This causes
     extra work for implementors (who must invent appropriate
     policy mechanisms) and for users (who must learn how to use

          the mechanisms. This lack of a standardized mechanism also
          makes it difficult to build consistent configurations for
          routers from different vendors. This presents a significant
          practical deterrent to multi-vendor interoperability.

     (4) The proprietary policy mechanisms currently provided by
          vendors are often inadequate in complex parts of the
          Internet.

     (5) The algorithm has not been written down in any generally
          available document or standard. It is, in effect, a part of
          the Internet Folklore.

E.3.2 The Variant Router Requirements Algorithm

     Some Router Requirements Working Group members have proposed a
     slight variant of the algorithm described in the Section
     [5.2.4.3]. In this variant, matching the type of service
     requested is considered to be more important, rather than less
     important, than matching as much of the destination address as
     possible. For example, this algorithm would prefer a default
     route that had the correct type of service over a network route
     that had the default type of service, whereas the algorithm in
     [5.2.4.3] would make the opposite choice.

     The steps of the algorithm are:

     1. Basic match
     2. Weak TOS
     3. Longest match
     4. Best metric
     5. Policy

     Debate between the proponents of this algorithm and the regular
     Router Requirements Algorithm suggests that each side can show
     cases where its algorithm leads to simpler, more intuitive routing
     than the other's algorithm does. This variant has the same set of
     advantages and disadvantages that the algorithm specified in
     [5.2.4.3] does, except that pruning on Weak TOS before pruning on
     Longest Match makes this algorithm less compatible with OSPF and
     Integrated IS-IS than the standard Router Requirements Algorithm.

E.3.3 The OSPF Algorithm

     OSPF uses an algorithm that is virtually identical to the Router
     Requirements Algorithm except for one crucial difference: OSPF
     considers OSPF route classes.

The algorithm is:

1. Basic match
2. OSPF route class
3. Longest match
4. Weak TOS
5. Best metric
6. Policy

Type of service support is not always present. If it is not present then, of course, the fourth step would be omitted

This algorithm has some advantages over the Revised Classic Algorithm:

(1) It supports type of service routing.

(2) Its rules are written down, rather than merely being a part of the Internet folklore.

(3) It (obviously) works with OSPF.

However, this algorithm also retains some of the disadvantages of the Revised Classic Algorithm:

(1) Path properties other than type of service (e.g., MTU) are ignored.

(2) As in the Revised Classic Algorithm, the details (or even the existence) of the Policy step are left to the discretion of the implementor.

The OSPF Algorithm also has a further disadvantage (which is not shared by the Revised Classic Algorithm). OSPF internal (intra-area or inter-area) routes are always considered to be superior to routes learned from other routing protocols, even in cases where the OSPF route matches fewer bits of the destination address. This is a policy decision that is inappropriate in some networks.

Finally, it is worth noting that the OSPF Algorithm's TOS support suffers from a deficiency in that routing protocols that support TOS are implicitly preferred when forwarding packets that have non-zero TOS values. This may not be appropriate in some cases.

E.3.4 The Integrated IS-IS Algorithm

   Integrated IS-IS uses an algorithm that is similar to but not quite
   identical to the OSPF Algorithm. Integrated IS-IS uses a different
   set of route classes, and differs slightly in its handling of type of
   service. The algorithm is:

   1. Basic Match
   2. IS-IS Route Classes
   3. Longest Match
   4. Weak TOS
   5. Best Metric
   6. Policy

   Although Integrated IS-IS uses Weak TOS, the protocol is only capable
   of carrying routes for a small specific subset of the possible values
   for the TOS field in the IP header. Packets containing other values
   in the TOS field are routed using the default TOS.

   Type of service support is optional; if disabled, the fourth step
   would be omitted. As in OSPF, the specification does not include the
   Policy step.

   This algorithm has some advantages over the Revised Classic
   Algorithm:

   (1) It supports type of service routing.
   (2) Its rules are written down, rather than merely being a part of
       the Internet folklore.
   (3) It (obviously) works with Integrated IS-IS.

   However, this algorithm also retains some of the disadvantages of the
   Revised Classic Algorithm:

   (1) Path properties other than type of service (e.g., MTU) are
       ignored.
   (2) As in the Revised Classic Algorithm, the details (or even the
       existence) of the Policy step are left to the discretion of the
       implementor.
   (3) It doesn't work with OSPF because of the differences between IS-
       IS route classes and OSPF route classes. Also, because IS-IS
       supports only a subset of the possible TOS values, some obvious
       implementations of the Integrated IS-IS algorithm would not
       support OSPF's interpretation of TOS.

   The Integrated IS-IS Algorithm also has a further disadvantage (which
   is not shared by the Revised Classic Algorithm): IS-IS internal
   (intra-area or inter-area) routes are always considered to be

superior to routes learned from other routing protocols, even in
cases where the IS-IS route matches fewer bits of the destination
address and doesn't provide the requested type of service. This is a
policy decision that may not be appropriate in all cases.

Finally, it is worth noting that the Integrated IS-IS Algorithm's TOS
support suffers from the same deficiency noted for the OSPF
Algorithm.

Security Considerations

Although the focus of this document is interoperability rather than
security, there are obviously many sections of this document that
have some ramifications on network security.

Security means different things to different people. Security from a
router's point of view is anything that helps to keep its own
networks operational and in addition helps to keep the Internet as a
whole healthy. For the purposes of this document, the security
services we are concerned with are denial of service, integrity, and
authentication as it applies to the first two. Privacy as a security
service is important, but only peripherally a concern of a router -
at least as of the date of this document.

In several places in this document there are sections entitled ...
Security Considerations. These sections discuss specific
considerations that apply to the general topic under discussion.

Rarely does this document say do this and your router/network will be
secure. More likely, it says this is a good idea and if you do it,
it *may* improve the security of the Internet and your local system
in general.

Unfortunately, this is the state-of-the-art AT THIS TIME. Few if any
of the network protocols a router is concerned with have reasonable,
built-in security features. Industry and the protocol designers have
been and are continuing to struggle with these issues. There is
progress, but only small baby steps such as the peer-to-peer
authentication available in the BGP and OSPF routing protocols.

In particular, this document notes the current research into
developing and enhancing network security. Specific areas of
research, development, and engineering that are underway as of this
writing (December 1993) are in IP Security, SNMP Security, and common
authentication technologies.

Notwithstanding all the above, there are things both vendors and
users can do to improve the security of their router. Vendors should

get a copy of Trusted Computer System Interpretation [INTRO:8]. Even
if a vendor decides not to submit their device for formal
verification under these guidelines, the publication provides
excellent guidance on general security design and practices for
computing devices.

APPENDIX F: HISTORICAL ROUTING PROTOCOLS

   Certain routing protocols are common in the Internet, but the authors
   of this document cannot in good conscience recommend their use. This
   is not because they do not work correctly, but because the
   characteristics of the Internet assumed in their design (simple
   routing, no policy, a single "core router" network under common
   administration, limited complexity, or limited network diameter) are
   not attributes of today's Internet. Those parts of the Internet that
   still use them are generally limited "fringe" domains with limited
   complexity.

   As a matter of good faith, collected wisdom concerning their
   implementation is recorded in this section.

F.1 EXTERIOR GATEWAY PROTOCOL - EGP

F.1.1 Introduction

   The Exterior Gateway Protocol (EGP) specifies an EGP that is used to
   exchange reachability information between routers of the same or
   differing autonomous systems. EGP is not considered a routing
   protocol since there is no standard interpretation (i.e. metric) for
   the distance fields in the EGP update message, so distances are
   comparable only among routers of the same AS. It is however designed
   to provide high-quality reachability information, both about neighbor
   routers and about routes to non-neighbor routers.

   EGP is defined by [ROUTE:6]. An implementor almost certainly wants
   to read [ROUTE:7] and [ROUTE:8] as well, for they contain useful
   explanations and background material.

   DISCUSSION
      The present EGP specification has serious limitations, most
      importantly a restriction that limits routers to advertising only
      those networks that are reachable from within the router's
      autonomous system. This restriction against propagating third
      party EGP information is to prevent long-lived routing loops.
      This effectively limits EGP to a two-level hierarchy.

      RFC-975 is not a part of the EGP specification, and should be
      ignored.

F.1.2 Protocol Walk-through

     Indirect Neighbors: RFC-888, page 26

          An implementation of EGP MUST include indirect neighbor
          support.

     Polling Intervals: RFC-904, page 10

          The interval between Hello command retransmissions and the
          interval between Poll retransmissions SHOULD be configurable
          but there MUST be a minimum value defined.

          The interval at which an implementation will respond to Hello
          commands and Poll commands SHOULD be configurable but there
          MUST be a minimum value defined.

     Network Reachability: RFC-904, page 15

An implementation MUST default to not providing the external list of
routers in other autonomous systems; only the internal list of
routers together with the nets that are reachable through those
routers should be included in an Update Response/Indication packet.
However, an implementation MAY elect to provide a configuration
option enabling the external list to be provided. An implementation
MUST NOT include in the external list routers that were learned
through the external list provided by a router in another autonomous
system. An implementation MUST NOT send a network back to the
autonomous system from which it is learned, i.e. it MUST do split-
horizon on an autonomous system level.

If more than 255 internal or 255 external routers need to be
specified in a Network Reachability update, the networks reachable
from routers that can not be listed MUST be merged into the list for
one of the listed routers. Which of the listed routers is chosen for
this purpose SHOULD be user configurable, but SHOULD default to the
source address of the EGP update being generated.

An EGP update contains a series of blocks of network numbers, where
each block contains a list of network numbers reachable at a
particular distance through a particular router. If more than 255
networks are reachable at a particular distance through a particular
router, they are split into multiple blocks (all of which have the
same distance). Similarly, if more than 255 blocks are required to
list the networks reachable through a particular router, the router's
address is listed as many times as necessary to include all the
blocks in the update.

Unsolicited Updates: RFC-904, page 16

   If a network is shared with the peer, an implementation MUST send an
   unsolicited update upon entry to the Up state if the source network
   is the shared network.

Neighbor Reachability: RFC-904, page 6, 13-15

   The table on page 6 that describes the values of j and k (the
   neighbor up and down thresholds) is incorrect. It is reproduced
   correctly here:

      Name   Active Passive Description
      ----
       j   3   1   neighbor-up threshold
       k   1   0   neighbor-down threshold

   The value for k in passive mode also specified incorrectly in RFC-
   904, page 14 The values in parenthesis should read:

      (j = 1, k = 0, and T3/T1 = 4)

   As an optimization, an implementation can refrain from sending a
   Hello command when a Poll is due. If an implementation does so, it
   SHOULD provide a user configurable option to disable this
   optimization.

Abort timer: RFC-904, pages 6, 12, 13

   An EGP implementation MUST include support for the abort timer (as
   documented in section 4.1.4 of RFC-904). An implementation SHOULD
   use the abort timer in the Idle state to automatically issue a Start
   event to restart the protocol machine. Recommended values are P4 for
   a critical error (Administratively prohibited, Protocol Violation and
   Parameter Problem) and P5 for all others. The abort timer SHOULD NOT
   be started when a Stop event was manually initiated (such as through
   a network management protocol).

Cease command received in Idle state: RFC-904, page 13

   When the EGP state machine is in the Idle state, it MUST reply to
   Cease commands with a Cease-ack response.

Hello Polling Mode: RFC-904, page 11

   An EGP implementation MUST include support for both active and
   passive polling modes.

Neighbor Acquisition Messages: RFC-904, page 18

   As noted the Hello and Poll Intervals should only be present in
   Request and Confirm messages. Therefore the length of an EGP
   Neighbor Acquisition Message is 14 bytes for a Request or Confirm
   message and 10 bytes for a Refuse, Cease or Cease-ack message.
   Implementations MUST NOT send 14 bytes for Refuse, Cease or Cease-ack
   messages but MUST allow for implementations that send 14 bytes for
   these messages.

Sequence Numbers: RFC-904, page 10

   Response or indication packets received with a sequence number not
   equal to S MUST be discarded. The send sequence number S MUST be
   incremented just before the time a Poll command is sent and at no
   other times.

F.2 ROUTING INFORMATION PROTOCOL - RIP

F.2.1 Introduction

   RIP is specified in [ROUTE:3]. Although RIP is still quite important
   in the Internet, it is being replaced in sophisticated applications
   by more modern IGPs such as the ones described above. A router
   implementing RIP SHOULD implement RIP Version 2 [ROUTE:?], as it
   supports CIDR routes. If occasional access networking is in use, a
   router implementing RIP SHOULD implement Demand RIP [ROUTE:?].

   Another common use for RIP is as a router discovery protocol.
   Section [4.3.3.10] briefly touches upon this subject.

F.2.2 Protocol Walk-Through

   Dealing with changes in topology: [ROUTE:3], page 11

      An implementation of RIP MUST provide a means for timing out
      routes. Since messages are occasionally lost, implementations
      MUST NOT invalidate a route based on a single missed update.

      Implementations MUST by default wait six times the update
      interval before invalidating a route. A router MAY have
      configuration options to alter this value.

   DISCUSSION
      It is important to routing stability that all routers in a RIP
      autonomous system use similar timeout value for invalidating
      routes, and therefore it is important that an implementation
      default to the timeout value specified in the RIP specification.

However, that timeout value is too conservative in environments
where packet loss is reasonably rare. In such an environment, a
network manager may wish to be able to decrease the timeout period
to promote faster recovery from failures.

   IMPLEMENTATION
      There is a very simple mechanism that a router may use to meet the
      requirement to invalidate routes promptly after they time out.
      Whenever the router scans the routing table to see if any routes
      have timed out, it also notes the age of the least recently
      updated route that has not yet timed out. Subtracting this age
      from the timeout period gives the amount of time until the router
      again needs to scan the table for timed out routes.

Split Horizon: [ROUTE:3], page 14-15

   An implementation of RIP MUST implement split horizon, a scheme used
   for avoiding problems caused by including routes in updates sent to
   the router from which they were learned.

   An implementation of RIP SHOULD implement Split horizon with poisoned
   reverse, a variant of split horizon that includes routes learned from
   a router sent to that router, but sets their metric to infinity.
   Because of the routing overhead that may be incurred by implementing
   split horizon with poisoned reverse, implementations MAY include an
   option to select whether poisoned reverse is in effect. An
   implementation SHOULD limit the time in which it sends reverse routes
   at an infinite metric.

   IMPLEMENTATION
      Each of the following algorithms can be used to limit the time for
      which poisoned reverse is applied to a route. The first algorithm
      is more complex but does a more thorough job of limiting poisoned
      reverse to only those cases where it is necessary.

      The goal of both algorithms is to ensure that poison reverse is
      done for any destination whose route has changed in the last Route
      Lifetime (typically 180 seconds), unless it can be sure that the
      previous route used the same output interface. The Route Lifetime
      is used because that is the amount of time RIP will keep around an
      old route before declaring it stale.

      The time intervals (and derived variables) used in the following
      algorithms are as follows:

      Tu The Update Timer; the number of seconds between RIP updates.
           This typically defaults to 30 seconds.

Rl The Route Lifetime, in seconds. This is the amount of time
   that a route is presumed to be good, without requiring an
   update. This typically defaults to 180 seconds.

Ul The Update Loss; the number of consecutive updates that have to
   be lost or fail to mention a route before RIP deletes the
   route. Ul is calculated to be (Rl/Tu)+1. The +1 is to
   account for the fact that the first time the ifcounter is
   decremented will be less than Tu seconds after it is
   initialized. Typically, Ul will be 7: (180/30)+1.

In The value to set ifcounter to when a destination is newly
   learned. This value is Ul-4, where the 4 is RIP's garbage
   collection timer/30

The first algorithm is:

- Associated with each destination is a counter, called the
  ifcounter below. Poison reverse is done for any route whose
  destination's ifcounter is greater than zero.

- After a regular (not triggered or in response to a request)
  update is sent, all the non-zero ifcounters are decremented by
  one.

- When a route to a destination is created, its ifcounter is set
  as follows:

  - If the new route is superseding a valid route, and the old
    route used a different (logical) output interface, then the
    ifcounter is set to Ul.

  - If the new route is superseding a stale route, and the old
    route used a different (logical) output interface, then the
    ifcounter is set to MAX(0, Ul - INT(seconds that the route
    has been stale/Ut).

  - If there was no previous route to the destination, the
    ifcounter is set to In.

  - Otherwise, the ifcounter is set to zero

- RIP also maintains a timer, called the resettimer below. Poison
  reverse is done on all routes whenever resettimer has not
  expired (regardless of the ifcounter values).

- When RIP is started, restarted, reset, or otherwise has its
  routing table cleared, it sets the resettimer to go off in Rl
  seconds.

The second algorithm is identical to the first except that:

- The rules which set the ifcounter to non-zero values are changed
  to always set it to Rl/Tu, and

- The resettimer is eliminated.

Triggered updates: [ROUTE:3], page 15-16; page 29

Triggered updates (also called flash updates) are a mechanism for
immediately notifying a router's neighbors when the router adds or
deletes routes or changes their metrics. A router MUST send a
triggered update when routes are deleted or their metrics are
increased. A router MAY send a triggered update when routes are
added or their metrics decreased.

Since triggered updates can cause excessive routing overhead,
implementations MUST use the following mechanism to limit the
frequency of triggered updates:

(1) When a router sends a triggered update, it sets a timer to a
    random time between one and five seconds in the future. The
    router must not generate additional triggered updates before
    this timer expires.

(2) If the router would generate a triggered update during this
    interval it sets a flag indicating that a triggered update is
    desired. The router also logs the desired triggered update.

(3) When the triggered update timer expires, the router checks the
    triggered update flag. If the flag is set then the router
    sends a single triggered update which includes all the
    changes that were logged. The router then clears the flag
    and, since a triggered update was sent, restarts this
    algorithm.

(4) The flag is also cleared whenever a regular update is sent.

Triggered updates SHOULD include all routes that have changed
since the most recent regular (non-triggered) update. Triggered
updates MUST NOT include routes that have not changed since the
most recent regular update.

   DISCUSSION
      Sending all routes, whether they have changed recently or not, is
      unacceptable in triggered updates because the tremendous size of
      many Internet routing tables could otherwise result in
      considerable bandwidth being wasted on triggered updates.

Use of UDP: [ROUTE:3], page 18-19.

   RIP packets sent to an IP broadcast address SHOULD have their initial
   TTL set to one.

   Note that to comply with Section [6.1] of this memo, a router SHOULD
   use UDP checksums in RIP packets that it originates, MUST discard RIP
   packets received with invalid UDP checksums, but MUST NOT discard
   received RIP packets simply because they do not contain UDP
   checksums.

Addressing Considerations: [ROUTE:3], page 22

   A RIP implementation SHOULD support host routes. If it does not, it
   MUST (as described on page 27 of [ROUTE:3]) ignore host routes in
   received updates. A router MAY log ignored hosts routes.

   The special address 0.0.0.0 is used to describe a default route. A
   default route is used as the route of last resort (i.e., when a route
   to the specific net does not exist in the routing table). The router
   MUST be able to create a RIP entry for the address 0.0.0.0.

Input Processing - Response: [ROUTE:3], page 26

   When processing an update, the following validity checks MUST be
   performed:

   o The response MUST be from UDP port 520.

   o The source address MUST be on a directly connected subnet (or on a
      directly connected, non-subnetted network) to be considered valid.

   o The source address MUST NOT be one of the router's addresses.

   DISCUSSION
      Some networks, media, and interfaces allow a sending node to
      receive packets that it broadcasts. A router must not accept its
      own packets as valid routing updates and process them. The last
      requirement prevents a router from accepting its own routing
      updates and processing them (on the assumption that they were sent
      by some other router on the network).

An implementation MUST NOT replace an existing route if the metric
received is equal to the existing metric except in accordance with
the following heuristic.

An implementation MAY choose to implement the following heuristic to
deal with the above situation. Normally, it is useless to change the
route to a network from one router to another if both are advertised
at the same metric. However, the route being advertised by one of
the routers may be in the process of timing out. Instead of waiting
for the route to timeout, the new route can be used after a specified
amount of time has elapsed. If this heuristic is implemented, it
MUST wait at least halfway to the expiration point before the new
route is installed.

F.2.3 Specific Issues

RIP Shutdown

    An implementation of RIP SHOULD provide for a graceful shutdown
    using the following steps:

    (1) Input processing is terminated,

    (2) Four updates are generated at random intervals of between two
        and four seconds, These updates contain all routes that were
        previously announced, but with some metric changes. Routes
        that were being announced at a metric of infinity should
        continue to use this metric. Routes that had been announced
        with a non-infinite metric should be announced with a metric
        of 15 (infinity - 1).

    DISCUSSION
      The metric used for the above really ought to be 16 (infinity);
      setting it to 15 is a kludge to avoid breaking certain old hosts
      that wiretap the RIP protocol. Such a host will (erroneously)
      abort a TCP connection if it tries to send a datagram on the
      connection while the host has no route to the destination (even if
      the period when the host has no route lasts only a few seconds
      while RIP chooses an alternate path to the destination).

RIP Split Horizon and Static Routes

    Split horizon SHOULD be applied to static routes by default. An
    implementation SHOULD provide a way to specify, per static route,
    that split horizon should not be applied to this route.

F.3 GATEWAY TO GATEWAY PROTOCOL - GGP

   The Gateway to Gateway protocol is considered obsolete and SHOULD NOT
   be implemented.

Acknowledgments

   O that we now had here
   But one ten thousand of those men in England
   That do no work to-day!

   What's he that wishes so?
   My cousin Westmoreland? No, my fair cousin:
   If we are mark'd to die, we are enow
   To do our country loss; and if to live,
   The fewer men, the greater share of honour.
   God's will! I pray thee, wish not one man more.
   By Jove, I am not covetous for gold,
   Nor care I who doth feed upon my cost;
   It yearns me not if men my garments wear;
   Such outward things dwell not in my desires:
   But if it be a sin to covet honour,
   I am the most offending soul alive.
   No, faith, my coz, wish not a man from England:
   God's peace! I would not lose so great an honour
   As one man more, methinks, would share from me
   For the best hope I have. O, do not wish one more!
   Rather proclaim it, Westmoreland, through my host,
   That he which hath no stomach to this fight,
   Let him depart; his passport shall be made
   And crowns for convoy put into his purse:
   We would not die in that man's company
   That fears his fellowship to die with us.
   This day is called the feast of Crispian:
   He that outlives this day, and comes safe home,
   Will stand a tip-toe when the day is named,
   And rouse him at the name of Crispian.
   He that shall live this day, and see old age,
   Will yearly on the vigil feast his neighbours,
   And say 'To-morrow is Saint Crispian:'
   Then will he strip his sleeve and show his scars.
   And say 'These wounds I had on Crispin's day.'
   Old men forget: yet all shall be forgot,
   But he'll remember with advantages
   What feats he did that day: then shall our names.
   Familiar in his mouth as household words
   Harry the king, Bedford and Exeter,
   Warwick and Talbot, Salisbury and Gloucester,

Be in their flowing cups freshly remember'd.
This story shall the good man teach his son;
And Crispin Crispian shall ne'er go by,
From this day to the ending of the world,
But we in it shall be remember'd;
We few, we happy few, we band of brothers;
For he to-day that sheds his blood with me
Shall be my brother; be he ne'er so vile,
This day shall gentle his condition:
And gentlemen in England now a-bed
Shall think themselves accursed they were not here,
And hold their manhoods cheap whiles any speaks
That fought with us upon Saint Crispin's day.

                         -- William Shakespeare

This memo is a product of the IETF's Router Requirements Working
Group. A memo such as this one is of necessity the work of many more
people than could be listed here. A wide variety of vendors, network
managers, and other experts from the Internet community graciously
contributed their time and wisdom to improve the quality of this
memo. The editor wishes to extend sincere thanks to all of them.

The current editor also wishes to single out and extend his heartfelt
gratitude and appreciation to the original editor of this document;
Philip Almquist. Without Philip's work, both as the original editor
and as the Chair of the working group, this document would not have
been produced. He also wishes to express deep and heartfelt
gratitude to the previous editor, Frank Kastenholz. Frank changed
the original document from a collection of information to a useful
description of IP technology - in his words, a "snapshot" of the
technology in 1991. One can only hope that this snapshot, of the
technology in 1994, is as clear.

Philip Almquist, Jeffrey Burgan, Frank Kastenholz, and Cathy
Wittbrodt each wrote major chapters of this memo. Others who made
major contributions to the document included Bill Barns, Steve
Deering, Kent England, Jim Forster, Martin Gross, Jeff Honig, Steve
Knowles, Yoni Malachi, Michael Reilly, and Walt Wimer.

Additional text came from Andy Malis, Paul Traina, Art Berggreen,
John Cavanaugh, Ross Callon, John Lekashman, Brian Lloyd, Gary
Malkin, Milo Medin, John Moy, Craig Partridge, Stephanie Price, Yakov
Rekhter, Steve Senum, Richard Smith, Frank Solensky, Rich Woundy, and
others who have been inadvertently overlooked.

Some of the text in this memo has been (shamelessly) plagiarized from
earlier documents, most notably RFC-1122 by Bob Braden and the Host

Editor's Address

   The address of the current editor of this document is

      Fred Baker
      Cisco Systems
      519 Lado Drive
      Santa Barbara, California 93111
      USA

      Phone:+1 805-681-0115

      EMail: fred@cisco.com