

Type of Service in the Internet Protocol Suite

Status of This Memo

This document specifies an IAB standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "IAB Official Protocol Standards" for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Summary

This memo changes and clarifies some aspects of the semantics of the Type of Service octet in the Internet Protocol (IP) header. The handling of IP Type of Service by both hosts and routers is specified in some detail.

This memo defines a new TOS value for requesting that the network minimize the monetary cost of transmitting a datagram. A number of additional new TOS values are reserved for future experimentation and standardization. The ability to request that transmission be optimized along multiple axes (previously accomplished by setting multiple TOS bits simultaneously) is removed. Thus, for example, a single datagram can no longer request that the network simultaneously minimize delay and maximize throughput.

In addition, there is a minor conflict between the Host Requirements (RFC-1122 and RFC-1123) and a number of other standards concerning the sizes of the fields in the Type of Service octet. This memo resolves that conflict.

Table of Contents

1. Introduction	3
2. Goals and Philosophy	3
3. Specification of the Type of Service Octet	4
4. Specification of the TOS Field	5

5.	Use of the TOS Field in the Internet Protocols	6
5.1	Internet Control Message Protocol (ICMP)	6
5.2	Transport Protocols	7
5.3	Application Protocols	7
6.	ICMP and the TOS Facility	8
6.1	Destination Unreachable	8
6.2	Redirect	9
7.	Use of the TOS Field in Routing	9
7.1	Host Routing	10
7.2	Forwarding	12
8.	Other consequences of TOS	13
APPENDIX A.	Updates to Other Specifications	14
A.1	RFC-792 (ICMP)	14
A.2	RFC-1060 (Assigned Numbers)	14
A.3	RFC-1122 and RFC-1123 (Host Requirements)	16
A.4	RFC-1195 (Integrated IS-IS)	16
A.5	RFC-1247 (OSPF) and RFC-1248 (OSPF MIB)	17
APPENDIX B.	Rationale	18
B.1	The Minimize Monetary Cost TOS Value	18
B.2	The Specification of the TOS Field	19
B.3	The Choice of Weak TOS Routing	21
B.4	The Retention of Longest Match Routing	22
B.5	The Use of Destination Unreachable	23
APPENDIX C.	Limitations of the TOS Mechanism	24
C.1	Inherent Limitations	24
C.2	Limitations of this Specification	25
References	27
Acknowledgements	28
Security Considerations	28
Author's Address	28

1. Introduction

Paths through the Internet vary widely in the quality of service they provide. Some paths are more reliable than others. Some impose high call setup or per-packet charges, while others do not do usage-based charging. Throughput and delay also vary widely. Often there are tradeoffs: the path that provides the highest throughput may well not be the one that provides the lowest delay or the lowest monetary cost. Therefore, the "optimal" path for a packet to follow through the Internet may depend on the needs of the application and its user.

Because the Internet itself has no direct knowledge of how to optimize the path for a particular application or user, the IP protocol [11] provides a (rather limited) facility for upper layer protocols to convey hints to the Internet Layer about how the tradeoffs should be made for the particular packet. This facility is the "Type of Service" facility, abbreviated as the "TOS facility" in this memo.

Although the TOS facility has been a part of the IP specification since the beginning, it has been little used in the past. However, the Internet host specification [1,2] now mandates that hosts use the TOS facility. Additionally, routing protocols (including OSPF [10] and Integrated IS-IS [7]) have been developed which can compute routes separately for each type of service. These new routing protocols make it practical for routers to consider the requested type of service when making routing decisions.

This specification defines in detail how hosts and routers use the TOS facility. Section 2 introduces the primary considerations that motivated the design choices in this specification. Sections 3 and 4 describe the Type of Service octet in the IP header and the values which the TOS field of that octet may contain. Section 5 describes how a host (or router) chooses appropriate values to insert into the TOS fields of the IP datagrams it originates. Sections 6 and 7 describe the ICMP Destination Unreachable and Redirect messages and how TOS affects path choice by both hosts and routers. Section 8 describes some additional ways in which TOS may optionally affect packet processing. Appendix A describes how this specification updates a number of existing specifications. Appendices B and C expand on the discussion in Section 2.

2. Goals and Philosophy

The fundamental rule that guided this specification is that a host should never be penalized for using the TOS facility. If a host makes appropriate use of the TOS facility, its network service should be at least as good as (and hopefully better than) it would have been

if the host had not used the facility. This goal was considered particularly important because it is unlikely that any specification which did not meet this goal, no matter how good it might be in other respects, would ever become widely deployed and used. A particular consequence of this goal is that if a network cannot provide the TOS requested in a packet, the network does not discard the packet but instead delivers it the same way it would have been delivered had none of the TOS bits been set.

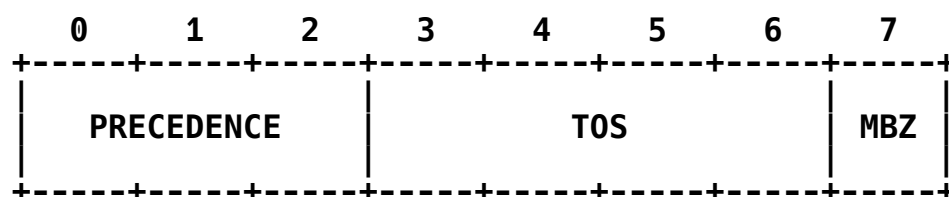
Even though the TOS facility has not been widely used in the past, it is a goal of this memo to be as compatible as possible with existing practice. Primarily this means that existing host implementations should not interact badly with hosts and routers which implement the specifications of this memo, since TOS support is almost non-existent in routers which predate this specification. However, this memo does attempt to be compatible with the treatment of IP TOS in OSPF and Integrated IS-IS.

Because the Internet community does not have much experience with TOS, it is important that this specification allow easy definition and deployment of new and experimental types of service. This goal has had a significant impact on this specification. In particular, it led to the decision to fix permanently the size of the TOS field and to the decision that hosts and routers should be able to handle a new type of service correctly without having to understand its semantics.

Appendix B of this memo provides a more detailed explanation of the rationale behind particular aspects of this specification.

3. Specification of the Type of Service Octet

The TOS facility is one of the features of the Type of Service octet in the IP datagram header. The Type of Service octet consists of three fields:



The first field, labeled "PRECEDENCE" above, is intended to denote the importance or priority of the datagram. This field is not discussed in detail in this memo.

The second field, labeled "TOS" above, denotes how the network should

make tradeoffs between throughput, delay, reliability, and cost. The TOS field is the primary topic of this memo.

The last field, labeled "MBZ" (for "must be zero") above, is currently unused. The originator of a datagram sets this field to zero (unless participating in an Internet protocol experiment which makes use of that bit). Routers and recipients of datagrams ignore the value of this field. This field is copied on fragmentation.

In the past there has been some confusion about the size of the TOS field. RFC-791 defined it as a three bit field, including bits 3-5 in the figure above. It included bit 6 in the MBZ field. RFC-1122 added bits 6 and 7 to the TOS field, eliminating the MBZ field. This memo redefines the TOS field to be the four bits shown in the figure above. The reasons for choosing to make the TOS field four bits wide can be found in Appendix B.2.

4. Specification of the TOS Field

As was stated just above, this memo redefines the TOS field as a four bit field. Also contrary to RFC-791, this memo defines the TOS field as a single enumerated value rather than as a set of bits (where each bit has its own meaning). This memo defines the semantics of the following TOS field values (expressed as binary numbers):

1000	--	minimize delay
0100	--	maximize throughput
0010	--	maximize reliability
0001	--	minimize monetary cost
0000	--	normal service

The values used in the TOS field are referred to in this memo as "TOS values", and the value of the TOS field of an IP packet is referred to in this memo as the "requested TOS". The TOS field value 0000 is referred to in this memo as the "default TOS."

Because this specification redefines TOS values to be integers rather than sets of bits, computing the logical OR of two TOS values is no longer meaningful. For example, it would be a serious error for a router to choose a low delay path for a packet whose requested TOS was 1110 simply because the router noted that the former "delay bit" was set.

Although the semantics of values other than the five listed above are not defined by this memo, they are perfectly legal TOS values, and hosts and routers must not preclude their use in any way. As will become clear after reading the remainder of this memo, only the default TOS is in any way special. A host or router need not (and

except as described in Section 8 should not) make any distinction between TOS values whose semantics are defined by this memo and those that are not.

It is important to note the use of the words "minimize" and "maximize" in the definitions of values for the TOS field. For example, setting the TOS field to 1000 (minimize delay) does not guarantee that the path taken by the datagram will have a delay that the user considers "low". The network will attempt to choose the lowest delay path available, based on its (often imperfect) information about path delay. The network will not discard the datagram simply because it believes that the delay of the available paths is "too high" (actually, the network manager can override this behavior through creative use of routing metrics, but this is strongly discouraged: setting the TOS field is intended to give better service when it is available, rather than to deny service when it is not).

5. Use of the TOS Field in the Internet Protocols

For the TOS facility to be useful, the TOS fields in IP packets must be filled in with reasonable values. This section discusses how protocols above IP choose appropriate values.

5.1 Internet Control Message Protocol (ICMP)

ICMP [8,9,12] defines a number of messages for performing error reporting and diagnostic functions for the Internet Layer. This section describes how a host or router chooses appropriate TOS values for ICMP messages it originates. The TOS facility also affects the origination and processing of ICMP Redirects and ICMP Destination Unreachables, but that is the topic of Section 6.

For purposes of this discussion, it is useful to divide ICMP messages into three classes:

- o ICMP error messages include ICMP message types 3 (Destination Unreachable), 4 (Source Quench), 5 (Redirect), 11 (Time Exceeded), and 12 (Parameter Problem).
- o ICMP request messages include ICMP message types 8 (Echo), 10 (Router Solicitation), 13 (Timestamp), 15 (Information Request -- now obsolete), and 17 (Address Mask Request).
- o ICMP reply messages include ICMP message types 0 (Echo Reply), 9 (Router Advertisement), 14 (Timestamp Reply), 16 (Information Reply -- also obsolete), and 18 (Address Mask Reply).

An ICMP error message is always sent with the default TOS (0000).

An ICMP request message may be sent with any value in the TOS field. A mechanism to allow the user to specify the TOS value to be used would be a useful feature in many applications that generate ICMP request messages.

An ICMP reply message is sent with the same value in the TOS field as was used in the corresponding ICMP request message.

5.2 Transport Protocols

When sending a datagram, a transport protocol uses the TOS requested by the application. There is no requirement that both ends of a transport connection use the same TOS. For example, the sending side of a bulk data transfer application should request that throughput be maximized, whereas the receiving side might request that delay be minimized (assuming that it is primarily sending small acknowledgement packets). It may be useful for a transport protocol to provide applications with a mechanism for learning the value of the TOS field that accompanied the most recently received data.

It is quite permissible to switch to a different TOS in the middle of a connection if the nature of the traffic being generated changes. An example of this would be SMTP, which spends part of its time doing bulk data transfer and part of its time exchanging short command messages and responses.

TCP [13] should use the same TOS for datagrams containing only TCP control information as it does for datagrams which contain user data. Although it might seem intuitively correct to always request that the network minimize delay for segments containing acknowledgements but no data, doing so could corrupt TCP's round trip time estimates.

5.3 Application Protocols

Applications are responsible for choosing appropriate TOS values for any traffic they originate. The Assigned Numbers document [15] lists the TOS values to be used by a number of common network applications. For other applications, it is the responsibility of the application's designer or programmer to make a suitable choice, based on the nature of the traffic to be originated by the application.

It is essential for many sorts of network diagnostic applications, and desirable for other applications, that the user of the

application be able to override the TOS value(s) which the application would otherwise choose.

The Assigned Numbers document is revised and reissued periodically. Until RFC-1060, the edition current as this is being written, has been superceded, readers should consult Appendix A.2 of this memo.

6. ICMP and the TOS Facility

Routers communicate routing information to hosts using the ICMP protocol [12]. This section describes how support for the TOS facility affects the origination and interpretation of ICMP Redirect messages and certain types of ICMP Destination Unreachable messages. This memo does not define any new extensions to the ICMP protocol.

6.1 Destination Unreachable

The ICMP Destination Unreachable message contains a code which describes the reason that the destination is unreachable. There are four codes [1,12] which are particularly relevant to the topic of this memo:

- 0 -- network unreachable
- 1 -- host unreachable
- 11 -- network unreachable for type of service
- 12 -- host unreachable for type of service

A router generates a code 11 or code 12 Destination Unreachable when an unreachable destination (network or host) would have been reachable had a different TOS value been specified. A router generates a code 0 or code 1 Destination Unreachable in other cases.

A host receiving a Destination Unreachable message containing any of these codes should recognize that it may result from a routing transient. The host should therefore interpret the message as only a hint, not proof, that the specified destination is unreachable.

The use of codes 11 and 12 may seem contrary to the statement in Section 2 that packets should not be discarded simply because the requested TOS cannot be provided. The rationale for having these codes and the limited cases in which they are expected to be used are described in Appendix B.5.

6.2 Redirect

The ICMP Redirect message also includes a code, which specifies the class of datagrams to which the Redirect applies. There are currently four codes defined:

- 0 -- redirect datagrams for the network
- 1 -- redirect datagrams for the host
- 2 -- redirect datagrams for the type of service and network
- 3 -- redirect datagrams for the type of service and host

A router generates a code 3 Redirect when the Redirect applies only to IP packets which request a particular TOS value. A router generates a code 1 Redirect instead when the the optimal next hop on the path to the destination would be the same for any TOS value. In order to minimize the potential for host confusion, routers should refrain from using codes 0 and 2 in Redirects [3,6].

Although the current Internet Host specification [1] only requires hosts to correctly handle code 0 and code 1 Redirects, a host should also correctly handle code 2 and code 3 Redirects, as described in Section 7.1 of this memo. If a host does not, it is better for the host to treat code 2 as equivalent to code 0 and code 3 as equivalent to code 1 than for the host to simply ignore code 2 and code 3 Redirects.

7. Use of the TOS Field in Routing

Both hosts and routers should consider the value of the TOS field of a datagram when choosing an appropriate path to get the datagram to its destination. The mechanisms for doing so are discussed in this section.

Whether a packet's TOS value actually affects the path it takes inside of a particular routing domain is a choice made by the routing domain's network manager. In many routing domains the paths are sufficiently homogeneous in nature that there is no reason for routers to choose different paths based up the TOS field in a datagram. Inside such a routing domain, the network manager may choose to limit the size of the routing database and of routing protocol updates by only defining routes for the default (0000) TOS. Neither hosts nor routers should need to have any explicit knowledge of whether TOS affects routing in the local routing domain.

7.1 Host Routing

When a host (which is not also a router) wishes to send an IP packet to a destination on another network or subnet, it needs to choose an appropriate router to send the packet to. According to the IP Architecture, it does so by maintaining a route cache and a list of default routers. Each entry in the route cache lists a destination (IP address) and the appropriate router to use to reach that destination. The host learns the information stored in its route cache through the ICMP Redirect mechanism. The host learns the list of default routers either from static configuration information or by using the ICMP Router Discovery mechanism [8]. When the host wishes to send an IP packet, it searches its route cache for a route matching the destination address in the packet. If one is found it is used; if not, the packet is sent to one of the default routers. All of this is described in greater detail in section 3.3.1 of RFC-1122 [1].

Adding support for the TOS facility changes the host routing procedure only slightly. In the following, it is assumed that (in accordance with the current Internet Host specification [1]) the host treats code 0 (redirect datagrams for the network) Redirects as if they were code 1 (redirect datagrams for the host) Redirects. Similarly, it is assumed that the host treats code 2 (redirect datagrams for the network and type of service) Redirects as if they were code 3 (redirect datagrams for the host and type of service) Redirects. Readers considering violating these assumptions should be aware that long and careful consideration of the way in which Redirects are treated is necessary to avoid situations where every packet sent to some destination provokes a Redirect. Because these assumptions match the recommendations of Internet Host specification, that careful consideration is beyond the scope of this memo.

As was described in Section 6.2, some ICMP Redirects apply only to IP packets which request a particular TOS. Thus, a host (at least conceptually) needs to store two types of entries in its route cache:

type 1: { destination, TOS, router }

type 2: { destination, *, router }

where type 1 entries result from the receipt of code 3 (or code 1) Redirects and type 2 entries result from the receipt of code 2 (or code 0) Redirects.

When a host wants to send a packet, it first searches the route cache for a type 1 entry whose destination matches the destination address of the packet and whose TOS matches the requested TOS in the packet. If it doesn't find one, the host searches its route cache again, this time looking for a type 2 entry whose destination matches the destination address of the packet. If either of these searches finds a matching entry, the packet is sent to the router listed in the matching entry. Otherwise, the packet is sent to one of the routers on the list of default routers.

When a host creates (or updates) a type 2 entry, it must flush from its route cache any type 1 entries which have the same destination. This is necessary for correctness, since the type 1 entry may be obsolete but would continue to be used if it weren't flushed because type 1 entries are always preferred over type 2 entries.

However, the converse is not true: when a host creates a type 1 entry, it should not flush a type 2 entry that has the same destination. In this case, the type 1 entry will properly override the type 2 entry for packets whose destination address and requested TOS match the type 1 entry. Because the type 2 entry may well specify the correct router for some TOS values other than the one specified in the type 1 entry, saving the type 2 entry will likely cut down on the number of Redirects which the host would otherwise receive. This savings can potentially be substantial if one of the Redirects which was avoided would have created a new type 2 entry (thereby causing the new type 1 entry to be flushed). That can happen, for example, if only some of the routers on the local net are part of a routing domain that computes separate routes for each TOS.

As an alternative, a host may treat all Redirects as if they were code 3 (redirect datagrams for hosts and type of service) Redirects. This alternative allows the host to have only type 1 route cache entries, thereby simplifying route lookup and eliminating the need for the rules in the previous two paragraphs. The disadvantage of this approach is that it increases the size of the route cache and the amount of Redirect traffic if the host sends packets with a variety of requested TOS's to a destination for which the host should use the same router regardless of the requested TOS. There is not yet sufficient experience with the TOS facility to know whether that disadvantage would be serious enough in practice to outweigh the simplicity of this approach.

Despite RFC-1122, some hosts acquire their routing information by "wiretapping" a routing protocol instead of by using the

mechanisms described above. Such hosts will need to follow the procedures described in Section 7.2 (except of course that hosts will not send ICMP Destination Unreachables or ICMP Redirects).

7.2 Forwarding

A router in the Internet should be able to consider the value of the TOS field when choosing an appropriate path over which to forward an IP packet. How a router does this is a part of the more general issue of how a router picks appropriate paths. This larger issue can be extremely complex [4], and is beyond the scope of this memo. This discussion should therefore be considered only an overview. Implementors should consult the Router Requirements specification [3] and the the specifications of the routing protocols they implement for details.

A router associates a TOS value with each route in its forwarding table. The value can be any of the possible values of the TOS field in an IP datagram (including those values whose semantics are yet to be defined). Any routes learned using routing protocols which support TOS are assigned appropriate TOS value by those protocols. Routes learned using other routing protocols are always assigned the default TOS value (0000). Static routes have their TOS values assigned by the network manager.

When a router wants to forward a packet, it first looks up the destination address in its forwarding table. This yields a set of candidate routes. The set may be empty (if the destination is unreachable), or it may contain one or more routes to the destination. If the set is not empty, the TOS values of the routes in the set are examined. If the set contains a route whose TOS exactly matches the TOS field of the packet being forwarded then that route is chosen. If not but the set contains a route with the default TOS then that route is chosen.

If no route is found, or if the the chosen route has an infinite metric, the destination is considered to be unreachable. The packet is discarded and an ICMP Destination Unreachable is returned to the source. Normally, the Unreachable uses code 0 (Network unreachable) or 1 (Host unreachable). If, however, a route to the destination exists which has a different TOS value and a non-infinite metric then code 11 (Network unreachable for type of service) or code 12 (Host unreachable for type of service) must be used instead.

8. Other consequences of TOS

The TOS field in a datagram primarily affects the path chosen through the network, but an implementor may choose to have TOS also affect other aspects of how the datagram is handled. For example, a host or router might choose to give preferential queuing on network output queues to datagrams which have requested that delay be minimized. Similarly, a router forced by overload to discard packets might attempt to avoid discarding packets that have requested that reliability be maximized. At least one paper [14] has explored these ideas in some detail, but little is known about how well such special handling would work in practice.

Additionally, some Link Layer protocols have their own quality of service mechanisms. When a router or host transmits an IP packet, it might request from the Link Layer a quality of service as close as possible to the one requested in the TOS field in the IP header. Long ago an attempt (RFC-795) was made to codify how this might be done, but that document describes Link Layer protocols which have since become obsolete and no more recent document on the subject has been written.

APPENDIX A. Updates to Other Specifications

While this memo is primarily an update to the IP protocol specification [11], it also peripherally affects a number of other specifications. This appendix describes those peripheral effects. This information is included in an appendix rather than in the main body of the document because most if not all of these other specifications will be updated in the future. As that happens, the information included in this appendix will become obsolete.

A.1 RFC-792 (ICMP)

RFC-792 [12] defines a set of codes indicating reasons why a destination is unreachable. This memo describes the use of two additional codes:

- 11 -- network unreachable for type of service
- 12 -- host unreachable for type of service

These codes were defined in RFC-1122 [1] but were not included in RFC-792.

A.2 RFC-1060 (Assigned Numbers)

RFC-1060 [15] describes the old interpretation of the TOS field (as three independent bits, with no way to specify that monetary cost should be minimized). Although it is likely obvious how the values in RFC-1060 ought to be interpreted in light of this memo, the information from that RFC is reproduced here. The only actual changes are for ICMP (to conform to Section 5.1 of this memo) and NNTP:

----- Type-of-Service Value -----

Protocol	TOS Value	
TELNET (1)	1000	(minimize delay)
FTP		
Control	1000	(minimize delay)
Data (2)	0100	(maximize throughput)
TFTP	1000	(minimize delay)
SMTP (3)		
Command phase	1000	(minimize delay)
DATA phase	0100	(maximize throughput)

----- Type-of-Service Value -----

Protocol	TOS Value	
Domain Name Service		
UDP Query	1000	(minimize delay)
TCP Query	0000	
Zone Transfer	0100	(maximize throughput)
NNTP	0001	(minimize monetary cost)
ICMP		
Errors	0000	
Requests	0000 (4)	
Responses	<same as request> (4)	
Any IGP	0010	(maximize reliability)
EGP	0000	
SNMP	0010	(maximize reliability)
BOOTP	0000	

Notes:

- (1) Includes all interactive user protocols (e.g., rlogin).
- (2) Includes all bulk data transfer protocols (e.g., rcp).
- (3) If the implementation does not support changing the TOS during the lifetime of the connection, then the recommended TOS on opening the connection is the default TOS (0000).
- (4) Although ICMP request messages are normally sent with the default TOS, there are sometimes good reasons why they would be sent with some other TOS value. An ICMP response always uses the same TOS value as was used in the corresponding ICMP request message. See Section 5.1 of this memo.

An application may (at the request of the user) substitute 0001 (minimize monetary cost) for any of the above values.

This appendix is expected to be obsoleted by the next revision of the Assigned Numbers document.

A.3 RFC-1122 and RFC-1123 (Host Requirements)

The use of the TOS field by hosts is described in detail in RFC-1122 [1] and RFC-1123 [2]. The information provided there is still correct, except that:

- (1) The TOS field is four bits wide rather than five bits wide. The requirements that refer to the TOS field should refer only to the four bits that make up the TOS field.
- (2) An application may set bit 6 of the TOS octet to a non-zero value (but still must not set bit 7 to a non-zero value).

These details will presumably be corrected in the next revision of the Host Requirements specification, at which time this appendix can be considered obsolete.

A.4 RFC-1195 (Integrated IS-IS)

Integrated IS-IS (sometimes known as Dual IS-IS) has multiple metrics for each route. Which of the metrics is used to route a particular IP packet is determined by the TOS field in the packet. This is described in detail in section 3.5 of RFC-1195 [7].

The mapping from the value of the TOS field to an appropriate Integrated IS-IS metric is described by a table in that section. Although the specification in this memo is intended to be substantially compatible with Integrated IS-IS, the extension of the TOS field to four bits and the addition of a TOS value requesting "minimize monetary cost" require minor modifications to that table, as shown here:

The IP TOS octet is mapped onto the four available metrics as follows:

Bits 0-2 (Precedence): (unchanged from RFC-1195)

Bits 3-6 (TOS):

0000	(all normal)	Use default metric
1000	(minimize delay)	Use delay metric
0100	(maximize throughput)	Use default metric
0010	(maximize reliability)	Use reliability metric
0001	(minimize monetary cost)	Use cost metric
other		Use default metric

Bit 7 (MBZ): This bit is ignored by Integrated IS-IS.

It is expected that the next revision of the Integrated IS-IS specification will include this corrected table, at which time this appendix can be considered obsolete.

A.5 RFC-1247 (OSPF) and RFC-1248 (OSPF MIB)

Although the specification in this memo is intended to be substantially compatible with OSPF, the extension of the TOS field to four bits requires minor modifications to the section that describes the encoding of TOS values in Link State Advertisements, described in section 12.3 of RFC-1247 [10]. The encoding is summarized in Table 17 of that memo; what follows is an updated version of table 17. The numbers in the first column are decimal integers, and the numbers in the second column are binary TOS values:

OSPF encoding	TOS	

0	0000	normal service
2	0001	minimize monetary cost
4	0010	maximize reliability
6	0011	
8	0100	maximize throughput
10	0101	
12	0110	
14	0111	
16	1000	minimize delay
18	1001	
20	1010	
22	1011	
24	1100	
26	1101	
28	1110	
30	1111	

The OSPF MIB, described in RFC-1248 [5], is entirely consistent with this memo except for the textual comment which describes the mapping of the old TOS flag bits into TOSType values. TOSType values use the same encoding of TOS values as OSPF's Link State Advertisements do, so the above table also describes the mapping between TOSType values (the first column) and TOS field values (the second column).

If RFC-1247 and RFC-1248 are revised in the future, it is expected that this information will be incorporated into the revised versions. At that time, this appendix may be considered obsolete.

APPENDIX B. Rationale

The main body of this memo has described the details of how TOS facility works. This appendix is for those who wonder why it works that way.

Much of what is in this document can be explained by the simple fact that the goal of this document is to provide a clear and complete specification of the existing TOS facility rather than to design from scratch a new quality of service mechanism for IP. While this memo does amend the facility in some small and carefully considered ways discussed below, the desirability of compatibility with existing specifications and uses of the TOS facility [1,2,7,10,11] was never in doubt. This goal of backwards compatibility determined the broad outlines and many of the details of this specification.

Much of the rest of this specification was determined by two additional goals, which were described more fully in Section 2. The first was that hosts should never be penalized for using the TOS facility, since that would likely ensure that it would never be widely deployed. The second was that the specification should make it easy, or at least possible, to define and deploy new types of service in the future.

The three goals above did not eliminate all need for engineering choices, however, and in a few cases the goals proved to be in conflict with each other. The remainder of this appendix discusses the rationale behind some of these engineering choices.

B.1 The Minimize Monetary Cost TOS Value

Because the Internet is becoming increasingly commercialized, a number of participants in the IETF's Router Requirements Working Group felt it would be important to have a TOS value which would allow a user to declare that monetary cost was more important than other qualities of the service.

There was considerable debate over what exactly this value should mean. Some felt, for example, that the TOS value should mean "must not cost money". This was rejected for several reasons. Because it would request a particular level of service (cost = 0) rather than merely requesting that some service attribute be minimized or maximized, it would not only philosophically at odds with the other TOS values but would require special code in both hosts and routers. Also, it would not be helpful to users who want their packets to travel via the least-cost path but can accept some level of cost when necessary. Finally, since whether any particular routing domain considers the TOS field when routing

is a choice made by the network manager, a user requiring a free path might not get one if the packet has to pass through a routing domain that does not consider TOS in its routing decisions.

Some proposed a slight variant: a TOS value which would mean "I am willing to pay money to have this packet delivered". This proposal suffers most of the same shortcomings as the previous one and turns out to have an additional interesting quirk: because of the algorithms specified in Section 7.2, any packet which used this TOS value would prefer links that cost money over equally good free links. Thus, such a TOS value would almost be equivalent to a "maximize monetary cost" value!

It seems likely that in the future users may need some mechanism to express the maximum amount they are willing to pay to have a packet delivered. However, an IP option would be a more appropriate mechanism, since there are precedents for having IP options that all routers are required to honor, and an IP option could include parameters such as the maximum amount the user was willing to pay. Thus, the TOS value defined in this memo merely requests that the network "minimize monetary cost".

B.2 The Specification of the TOS Field

There were four goals that guided the decision to have a four bit TOS field and the specification of that field's values:

- (1) To define a new type of service requesting that the network "minimize monetary cost"
- (2) To remain as compatible as possible with existing specifications and uses of the TOS facility
- (3) To allow for the definition and deployment of new types of service in the future
- (4) To permanently fix the size of the TOS field

The last goal may seem surprising, but turns out to be necessary for routing to work correctly when new types of service are deployed. If routers have different ideas about the size of the TOS field they make inconsistent decisions that may lead to routing loops.

At first glance goals (3) and (4) seem to be pretty much mutually exclusive. The IP header currently has only three unused bits, so at most three new type of service bits could be defined without resorting to the impractical step of changing the IP header

format. Since one of them would need to be allocated to meet goal (1), at most two bits could be reserved for new or experimental types of service. Not only is it questionable whether two would be enough, but it is improbable that the IETF and IAB would allow all of the currently unused bits to be permanently reserved for types of service which might or might not ever be defined.

However, some (if not most of) the possible combinations of the individual bits would not be useful. Clearly, setting all of the bits would be equivalent to setting none of the bits, since setting all of the bits would indicate that none of the types of optimization was any more important than any of the others. Although one could perhaps assign reasonable semantics to most pairs of bits, it is unclear that the range of network service provided by various paths could usefully be subdivided in so fine a manner. If some of these non-useful combinations of bits could be assigned to new types of service then it would be possible to meet goal (3) and goal (4) without having to use up all of the remaining reserved bits in the IP header. The obvious way to do that was to change the interpretation of TOS values so that they were integers rather than independently settable bits.

The integers were chosen to be compatible with the bit definitions found in RFC-791. Thus, for example, setting the TOS field to 1000 (minimize delay) sets bit 3 of the Type of Service octet; bit 3 is defined as the Low Delay bit in RFC-791. This memo only defines values which correspond to setting a single one of the RFC-791 bits, since setting multiple TOS bits does not seem to be a common practice. According to [15], none of the common TCP/IP applications currently set multiple TOS bits. However, TOS values corresponding to particular combinations of the RFC-791 bits could be defined if and when they are determined to be useful.

The new TOS value for "minimize monetary cost" needed to be one which would not be too terribly misconstrued by preexisting implementations. This seemed to imply that the value should be one which left all of the RFC-791 bits clear. That would require expanding the TOS field, but would allow old implementations to treat packets which request minimization of monetary cost (TOS 0001) as if they had requested the default TOS. This is not a perfect solution since (as described above) changing the size of the TOS field could cause routing loops if some routers were to route based on a three bit TOS field and others were to route based on a four bit TOS field. Fortunately, this should not be much of a problem in practice because routers which route based on a three bit TOS field are very rare as this is being written and will only become more so once this specification is published.

Because of those considerations, and also in order to allow a reasonable number of TOS values for future definition, it seemed desirable to expand the TOS field. That left the question of how much to expand it. Expanding it to five bits would allow considerable future expansion (27 new TOS values) and would be consistent with Host Requirements, but would reduce to one the number of reserved bits in the IP header. Expanding the TOS field to four bits would restrict future expansion to more modest levels (11 new TOS values), but would leave an additional IP header bit free. The IETF's Router Requirements Working Group concluded that a four bits wide TOS field allow enough values for future use and that consistency with Host Requirements was inadequate justification for unnecessarily increasing the size of the TOS field.

B.3 The Choice of Weak TOS Routing

"Ruminations on the Next Hop" [4] describes three alternative ways of routing based on the TOS field. Briefly, they are:

- (1) Strong TOS --
a route may be used only if its TOS exactly matches the TOS in the datagram being routed. If there is no route with the requested TOS, the packet is discarded.
- (2) Weak TOS --
like Strong TOS, except that a route with the default TOS (0000) is used if there is no route that has the requested TOS. If there is no route with either the requested TOS or the default TOS, the packet is discarded.
- (3) Very Weak TOS --
like Weak TOS, except that a route with the numerically smallest TOS is used if there is no route that has either the requested TOS or the default TOS.

This specification has adopted Weak TOS.

Strong TOS was quickly rejected. Because it requires that each router a packet traverses have a route with the requested TOS, packets which requested non-zero TOS values would have (at least until the TOS facility becomes widely used) a high probability of being discarded as undeliverable. This violates the principle (described in Section 2) that hosts should not be penalized for choosing non-zero TOS values.

The choice between Weak TOS and Very Weak TOS was not as straightforward. Weak TOS was chosen because it is slightly

simpler to implement and because it is consistent with the OSPF and Integrated IS-IS specifications. In addition, many dislike Very Weak TOS because its algorithm for choosing a route when none of the available routes have either the requested or the default TOS cannot be justified by intuition (there is no reason to believe that having a numerically smaller TOS makes a route better). Since a router would need to understand the semantics of all of the TOS values to make a more intelligent choice, there seems to be no reasonable way to fix this particular deficiency of Very Weak TOS.

In practice it is expected that the choice between Weak TOS and Very Weak TOS will make little practical difference, since (except where the network manager has intentionally set things up otherwise) there will be a route with the default TOS to any destination for which there is a route with any other TOS.

B.4 The Retention of Longest Match Routing

An interesting issue is how early in the route choice process TOS should be considered. There seem to be two obvious possibilities:

- (1) Find the set of routes that best match the destination address of the packet. From among those, choose the route which best matches the requested TOS.
- (2) Find the set of routes that best match the requested TOS. From among those, choose the route which best matches the destination address of the packet.

The two approaches are believed to support an identical set of routing policies. Which of the two allows the simpler configuration and minimizes the amount of routing information that needs to be passed around seems to depend on the topology, though some believe that the second option has a slight edge in this regard.

Under the first option, if the network manager neglects some pieces of the configuration the likely consequence is that some packets which would benefit from TOS-specific routes will be routed as if they had requested the default TOS. Under the second option, however, a network manager can easily (accidentally) configure things in such a way that packets which request a certain TOS and should be delivered locally will instead follow a default route for that TOS and be dumped into the Internet. Thus, the first option would seem to have a slight edge with regard to robustness in the face of errors by the network manager.

It has been also been suggested that the first option provides the additional benefit of allowing loop-free routing in routing domains which contain both routers that consider TOS in their routing decisions and routers that do not. Whether that is true in all cases is unknown. It is certainly the case, however, that under the second option it would not work to mix routers that consider TOS and routers which do not in the same routing domain.

All in all, there were no truly compelling arguments for choosing one way or the other, but it was nonetheless necessary to make a choice: if different routers were to make the choice differently, chaos (in the form of routing loops) would result. The mechanisms specified in this memo reflect the first option because that will probably be more intuitive to most network managers. Internet routing has traditionally chosen the route which best matches the destination address, with other mechanisms serving merely as tie-breakers. The first option is consistent with that tradition.

B.5 The Use of Destination Unreachable

Perhaps the most contentious and least defensible part of this specification is that a packet can be discarded because the destination is considered to be unreachable even though a packet to the same destination but requesting a different TOS would have been deliverable. This would seem to fall perilously close to violating the principle that hosts should never be penalized for requesting non-default TOS values in packets they originate.

This can happen in only three, somewhat unusual, cases:

- (1) There is a route to the packet's destination which has the TOS value requested in the packet, but the route has an infinite metric.
- (2) The only routes to the packet's destination have TOS values other than the one requested in the packet. One of them has the default TOS, but it has an infinite metric.
- (3) The only routes to the packet's destination have TOS values other than the one requested in the packet. None of them have the default TOS.

It is commonly accepted that a router which has a default route should nonetheless discard a packet if the router has a more specific route to the destination in its forwarding table but that route has an infinite metric. The first two cases seem to be analogous to that rule.

In addition, it is worth noting that, except perhaps during brief transients resulting from topology changes, routes with infinite metrics occur only as the result of deliberate action (or serious error) on the part of the network manager. Thus, packets are unlikely to be discarded unless the network manager has taken deliberate action to cause them to be. Some people believe that this is an important feature of the specification, allowing the network to (for example) keep packets which have requested that cost be minimized off of a link that is so expensive that the network manager feels confident that the users would want their packets to be dropped. Others (including the author of this memo) believe that this "feature" will prove not to be useful, and that other mechanisms may be required for access controls on links, but couldn't justify changing this specification in the ways necessary to eliminate the "feature".

Case (3) above is more problematic. It could have been avoided by using Very Weak TOS, but that idea was rejected for the reasons discussed in Appendix B.3. Some suggested that case (3) could be fixed by relaxing longest match routing (described in Appendix B.4), but that idea was rejected because it would add complexity to routers without necessarily making their routing choices particularly more intuitive. It is also worth noting that this is another case that a network manager has to try rather hard to create: since OSPF and Integrated IS-IS both enforce the constraint that there must be a route with the default TOS to any destination for which there is a route with a non-zero TOS, a network manager would have to await the development of a new routing protocol or create the problem with static routes. The eventual conclusion was that any fix to case (3) was worse than the problem.

APPENDIX C. Limitations of the TOS Mechanism

It is important to note that the TOS facility has some limitations. Some are consequences of engineering choices made in this specification. Others, referred to as "inherent limitations" below, could probably not have been avoided without either replacing the TOS facility defined in RFC-791 or accepting that things wouldn't work right until all routers in the Internet supported the TOS facility.

C.1 Inherent Limitations

The most important of the inherent limitations is that the TOS facility is strictly an advisory mechanism. It is not an appropriate mechanism for requesting service guarantees. There are two reasons why this is so:

- (1) Not all networks will consider the value of the TOS field when deciding how to handle and route packets. Partly this is a transition issue: there will be a (probably lengthy) period when some networks will use equipment that predates this specification. Even long term, however, many networks will not be able to provide better service by considering the value of the TOS field. For example, the best path through a network composed of a homogeneous collection of interconnected LANs is probably the same for any possible TOS value. Inside such a network, it would make little sense to require routers and routing protocols to do the extra work needed to consider the value of the TOS field when forwarding packets.
- (2) The TOS mechanism is not powerful enough to allow an application to quantify the level of service it desires. For example, an application may use the TOS field to request that the network choose a path which maximizes throughput, but cannot use that mechanism to say that it needs or wants a particular number of kilobytes or megabytes per second. Because the network cannot know what the application requires, it would be inappropriate for the network to decide to discard a packet which requested maximal throughput because no "high throughput" path was available.

The inability to provide resource guarantees is a serious drawback for certain kinds of network applications. For example, a system using packetized voice simply creates network congestion when the available bandwidth is inadequate to deliver intelligible speech. Likewise, the network oughtn't even bother to deliver a voice packet that has suffered more delay in the network than the application can tolerate. Unfortunately, resource guarantees are problematic in connectionless networks. Internet researchers are actively studying this problem, and are optimistic that they will be able to invent ways in which the Internet Architecture can evolve to support resource guarantees while preserving the advantages of connectionless networking.

C.2 Limitations of this Specification

There are a couple of additional limitations of the TOS facility which are not inherent limitations but instead are consequences of engineering choices made in this specification:

- (1) Routing is not really optimal for some TOS values. This is because optimal routing for those TOS values would require that routing protocols be cognizant of the semantics of the TOS values and use special algorithms to compute routes for

them. For example, routing protocols traditionally compute the metric for a path by summing the costs of the individual links that make up the path. However, to maximize reliability, a routing protocol would instead have to compute a metric which was the product of the probabilities of successful delivery over each of the individual links in the path. While this limitation is in some sense a limitation of current routing protocols rather than of this specification, this specification contributes to the problem by specifying that there are a number of legal TOS values that have no currently defined semantics.

- (2) This specification assumes that network managers will do "the right thing". If a routing domain uses TOS, the network manager must configure the routers in such a way that a reasonable path is chosen for each TOS. While this ought not to be terribly difficult, a network manager could accidentally or intentionally violate our rule that using the TOS facility should provide service at least as good as not using it.

References

- [1] Internet Engineering Task Force (R. Braden, Editor), "Requirements for Internet Hosts -- Communication Layers", RFC 1122, USC/Information Sciences Institute, October 1989.
- [2] Internet Engineering Task Force (R. Braden, Editor), "Requirements for Internet Hosts -- Application and Support", RFC 1123, USC/Information Sciences Institute, October 1989.
- [3] Almquist, P., "Requirements for IP Routers", Work in progress.
- [4] Almquist, P., "Ruminations on the Next Hop", Work in progress.
- [5] Baker, F. and R. Coltun, "OSPF Version 2 Management Information Base", RFC 1248, ACC, Computer Science Center, August 1991.
- [6] Braden, R. and J. Postel, "Requirements for Internet Gateways", RFC 1009, USC/Information Sciences Institute, June 1987.
- [7] Callon, R., "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments", RFC 1195, Digital Equipment Corporation, December 1990.
- [8] Deering, S., "ICMP Router Discovery Messages", RFC 1256, Xerox PARC, September 1991.
- [9] Mogul, J. and J. Postel, "Internet Standard Subnetting Procedure", RFC 950, USC/Information Sciences Institute, August 1985.
- [10] Moy, J., "OSPF Version 2", RFC 1247, Proteon, Inc., July 1991.
- [11] Postel, J., "Internet Protocol", RFC 791, DARPA, September 1981.
- [12] Postel, J., "Internet Control Message Protocol", RFC 792, DARPA, September 1981.
- [13] Postel, J., "Transmission Control Protocol", RFC 793, DARPA, September 1981.
- [14] Prue, W. and J. Postel, "A Queuing Algorithm to Provide Type-of-Service for IP Links", RFC 1046, USC/Information Sciences Institute, February 1988.
- [15] Reynolds, J. and J. Postel, "Assigned Numbers", RFC 1060, USC/Information Sciences Institute, March 1990.

Acknowledgements

Some of the ideas presented in this memo are based on discussions held by the IETF's Router Requirements Working Group. Much of the specification of the treatment of Type of Service by hosts is merely a restatement of the ideas of the IETF's former Host Requirements Working Group, as captured in RFC-1122 and RFC-1123. The author is indebted to John Moy and Ross Callon for their assistance and cooperation in achieving consistency among the OSPF specification, the Integrated IS-IS specification, and this memo.

This memo has been substantially improved as the result of thoughtful comments from a number of reviewers, including Dave Borman, Bob Braden, Ross Callon, Vint Cerf, Noel Chiappa, Deborah Estrin, Phill Gross, Bob Hinden, Steve Huston, Jon Postel, Greg Vaudreuil, John Wobus, and the Router Requirements Working Group.

The initial work on this memo was done while its author was an employee of BARRNet. Their support is gratefully acknowledged.

Security Considerations

This memo does not explicitly discuss security issues. The author does not believe that the specifications in this memo either weaken or enhance the security of the IP Protocol or of the other protocols mentioned herein.

Author's Address

Philip Almquist
214 Cole Street, Suite 2
San Francisco, CA 94117-1916

Phone: 415-752-2427

Email: almquist@Jessica.Stanford.EDU