ICS 33.040.30

YD

中华人民共和国通信行业标准

YD/T 1255—2003

具有路由功能的以太网交换机技术要求

Technical Specification for Ethernet Switch with Routing Capability

2003-04-11 发布　　2003-04-11 实施

中华人民共和国信息产业部　发布

目　　次

前　言

本标准是“以太网交换机”系列标准之一。该系列标准预计的结构及名称如下：

1.《以太网交换机设备技术要求》

2.《以太网交换机测试方法》

3. YD/T 1099—2001《千兆以太网交换机设备技术规范》

4. YD/T 1141—2001《千兆以太网交换机测试方法》

5.《具有路由功能的以太网交换机技术要求》

6.《具有路由功能的以太网交换机测试方法》

本标准主要依据 IEEE 标准与 RFC 文档，随着 IP 技术的不断发展，需要对本标准不断补充和完善。

本标准是《具有路由功能的以太网交换机测试方法》的配套标准文件，为其提供技术依据。

本规范附录 A 是规范性附录，附录 B 是资料性附录。

本标准由中国通信标准化协会提出并归口。

本标准起草单位：信息产业部电信传输研究所

华为技术有限公司

深圳市中兴通讯股份有限公司

本标准主要起草人：魏　亮　俞　杰　畅文俊　姚　析　田　辉　袁　琦

具有路由功能的以太网交换机技术要求

1 范围

本标准规定了在公用网使用的具有路由功能的以太网交换机的技术要求，包括功能、指标、通信接口、通信协议环境要求。由于三层交换机的称谓已成为业界共识，下文中所有三层交换机都特指具有路由功能的以太网交换机。本标准列举了三层交换机必须实现的协议与功能。对协议实现的完整要求在 IEEE 标准及 RFC 协议中指出。

本标准适合于公用网使用的具有路由功能的以太网交换机。

在本标准中：

- 必须：表示该条目是本标准必须。违反这样的要求是原则性错误。
- 必须实现：表示该要求必须实现，但不要求缺省使能。
- 不允许（不可以）：标识该条目绝对禁止。
- 应当（建议）：表示在某些特定条件下存在忽视该条目的理由，但是忽视或违反该条目时必须仔细衡量。
- 应当（建议）实现：与应当（建议）类似，实现时不必要缺省使能。
- 不应当（不建议）：表示在某些特定条件下存在所描述行为可接受或有效的理由，但实现该行为时必须仔细衡量。
- 可以：标识该条目确实可选。某些厂商可能出于市场或其他原因实现该选项，另一厂商可能出于类似理由不实现该选项。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件，其随后所有的修改单（不包括勘误的内容）或修订版均不适用于本标准，然而，鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件，其最新版本适用于本标准。

IEEE Std 802.1D（1998）	媒体访问控制（MAC）网桥
IEEE Std 802.1Q（1998）	虚拟桥接局域网
IEEE Std 802.1g（1995）	远程媒体访问控制桥接
IEEE Std 802.2（1998）	逻辑链路控制
IEEE Std 802.3ab（1999）	在 4 对 5 类平衡双绞线上传输千兆比以太网的物理层参数及规定 1000BASE-T
IEEE Std 802.3（2000）	带碰撞检测的载波监听多重访问的访问方式及物理层定义
RFC768	用户数据包协议
RFC791	互联网协议
RFC792	互联网控制消息协议
RFC793	传输控制协议
RFC795	服务映射
RFC796	地址映射
RFC826	以太网地址解释协议（ARP）
RFC854	远程登录（TELNET）协议规范
RFC951	启动捆绑协议（BOOTP）
RFC1058	路由信息协议

RFC1075	距离矢量组播路由协议
RFC1089	以太网上的 SNMP
RFC1108	IP 安全任选域
RFC1112	IP 组播主机扩展
RFC1122	互联网主机要求-通信层
RFC1142	IS-IS 域内路由协议
RFC1157	SNMP 协议
RFC1191	路径 MTU 发现
RFC1195	在 TCP/IP 和双重环境路由中使用 OSI 的 IS-IS
RFC1213（1991）	基于 TCP/IP 的互联网网络管理的管理信息库：MIB-II
RFC1256	ICMP 路由发现消息
RFC1265	BGP 协议分析
RFC1266	BGP 协议的经验
RFC1269	BGP4 MIB
RFC1349	在互联网协议族中服务类型
RFC1350	简单文件传输协议（TFTP）v2
RFC1493（1993）	对网桥管理对象的定义
RFC1584	OSPF v2 组播扩展
RFC1643（1994）	对以太网接口类型管理对象的定义
RFC1657	BGP4 管理对象的定义
RFC1724	RIP v2 MIB 扩展
RFC1757（1995）	远程网络监视管理信息库
RFC1771	边缘网关协议（BGP）v4
RFC1772	边缘网关协议在互联网中的应用
RFC1812	IPv4 路由器技术要求
RFC1850	OSPF v2 管理信息库
RFC1966	BGP 路由反射
RFC1997	BGP 区域（community）属性
RFC2011（1996）	对使用 SMIv2 互联网协议的管理信息库
RFC2012（1996）	对使用 SMIv2 传输控制协议的管理信息库
RFC2013（1996）	对使用 SMIv2 用户数据报协议的管理信息库
RFC2021（1997）	远程网络监视管理信息库版本 2
RFC2074（1997）	远程网络监视 MIB 协议标识符
RFC2082	RIP v2 MD5 认证（Authentication）
RFC2233（1997）	使用 SMIv2 的接口组 MIB
RFC2236	互联网组管理协议 IGMP（版本 2）
RFC2283	BGP4 多协议扩展
RFC2285	对局域网交换设备的测试术语
RFC2328	开放式最短路经优先（版本 2）
RFC2332	下一跳解释协议（NHRP）
RFC2362	协议无关组播-松散模式
RFC2439	BGP4 路由振荡抑制
RFC2453	路由信息协议 RIP（版本 2）
RFC2613（1999）	对交换网络的远程网络监视管理 MIB 扩展 Version 1.0

3 定义术语及缩写

3.1 定义

本标准采用了下列定义。

1）网桥（Bridge）

网桥工作在ISO的OSI 7层参考模型中第二层数据链路层的MAC子层，通过转发MAC帧实现网络互联。网桥的实现应当符合ANSI/IEEE Std802.1D，1998。网桥可以连接同种或不同种MAC技术的网络，利用包含在MAC帧中的目的地址和源地址信息作智能转发决定。在连接以太网时，网桥不但可以扩展物理网络拓扑结构，还可以将端口上的子网隔离成独立的冲突域。

2）以太网交换机（Ethernet Switch）

以太网交换机实质上是支持以太网接口的多端口网桥。交换机通常使用硬件实现过滤、学习和转发数据帧。

交换机必须实现网桥功能中相应功能。

3）具有路由功能的以太网交换机（Ethernet Switch with Routing Capability）

是拥有第三层路由功能的数据包交换机。除实现数据帧转发功能外，能根据收到的数据包中网络层地址以及交换机内部维护的路由表决定输出端口以及下一条交换机地址或主机地址并且重写链路层数据包头。

路由表必须动态维护来反映当前的网络拓扑。具有路由功能的以太网交换机通常通过与其他类似设备/路由器交换路由信息来完成动态维护路由表。

4）三层交换机（Layer3 Lan Switch）

参见具有路由功能的以太网交换机。

5）虚拟局域网（Virtual Local Access Network）

VLAN功能指通过桥接的局域网内活跃拓扑中工作站的划分，各VLAN使用VID（VLAN标识符）区分。各个VLAN是原桥接的局域网的一个子集。

6）远程桥接（Remote MAC Bridgeing）

远程媒体访问控制桥接是指在互连的局域网间使用远程媒体访问控制桥的操作以及远程媒体访问控制桥通过非局域网通信设备按照生成树算法配置被桥接局域网的协议。

7）链路聚合（Link Aggregration）

多链路聚合是指在逻辑上将多条独立的链路作为一条单独链路使用，以此获得灵活的高带宽以及链路冗余。

3.2 缩写

本标准使用下列缩写：

AFC	Asymmetric Flow Control	不对称流量控制
AUI	Attachment Unit Interface	附加单元接口
BPDU	Bridge Protocol Data Unit	桥接协议数据单元
CRC	Cyclic Redundancy Check	循环冗余校验
EAP	Extensible Authentication Protocol	可扩展认证协议
FCS	Frame Check Sequence	帧检验序列
E-ISS	Enhanced Internal Sublayer Service	增强的内部子层服务
EAPOL	EAP Over LANs	局域网上的可扩展认证协议
FID	Filter IDentifier	过滤标识符
GARP	General Attribute Registration Protocol	一般属性注册协议
GARP PDU	GARP Protocol Data Unit	GARP协议数据单元
GID	GARP Information Declaration	GARP信息发布

GIP	GARP Information Propagation	GARP 信息广播
GMII	Gigabit Media Independent Interface	千兆比特媒体无关接口
GMRP	GARP Multicast Registration Protocol	GARP 组播注册协议
GVRP	GARP VLAN Registration Protocol	GARP VLAN 注册协议
IETF	Internet Engineering Task Force	互联网工程任务组
IGMP	Internet Group Management Protocol	互联网组管理协议
ISS	Internal Sublayer Service	内部子层服务
IVL	Independent VLAN Learning	独立的 VLAN 学习
LAN	Local Area Network	局域网
LLC	Logical Link Contral	逻辑链路控制
MAC	Media Access Contral	媒体控制访问
MAU	Medium Attachment Unit	媒体附加接口
MDI	Media Dependent Interface	媒体依赖接口
MIB	Management Information Base	管理信息库
MII	Media Independent Interface	媒体无关接口
MSDU	MAC Service Data Unit	MAC 服务数据单元
NCFI	Non-Canonical Format Indication	非规范的格式标识符
PAE	Port Access Entity	端口访问实体
PCS	Physical Coding Sublayer	物理编码子层
PICS	Protocol Implementation Conformance Statement	协议实现一致性声明
PHY	Physical Layer Device	物理层设备
PLS	Physical Layer Signaling	物理层信令
PDU	Protocol Data Unit	协议数据单元
PMA	Physical Medium Attachment	物理介质接入
PMD	Physical Medium Dependent	物理媒体相关
PVID	Port VID	端口 VID
RADIUS	Remote Authentication Dial In User Service	远程认证拨号用户服务
RIF	Routing Information Field (ISO/IEC8802-5)	路由信息域
STPID	SNAP-encoded Tag Protocol Indentifier	SNAP 编码标记协议标识符
SVL	Shared VLAN Learning	共享 VLAN 学习
TCI	TAG Control Information	标记控制信息
TPID	TAG Protocol Indifier	标记协议信息
VID	Virtual LAN Identifier	虚拟局域网标识符
VLAN	Virtual LAN	虚拟局域网

4 三层交换机功能

三层交换机必须实现：

1）接口功能

三层交换机必须至少拥有一个接口。通常三层交换机拥有以太网接口，也可以拥有 ATM 或 POS 等接口。各种接口必须符合相应规范。

2）逻辑链路层功能

支持以太网接口的三层交换机必须实现一类 LLC 支持类型 1 操作。对 LLC 的实现必须符合 ISO/IEC 8802-2。

3）数据帧转发功能

数据帧转发是指交换机在不同端口所连接的被桥接的链路层实体间交换链路层用户数据帧。交换机必须实现转发数据帧。支持以太网接口的交换机转发数据帧应当实现 IEEE802.1p 中规定的优先级。

4）数据帧过滤功能

过滤是指交换机为防止数据帧重复，对某些端口上数据帧不转发（丢弃）到其他接口的行为。交换机必须实现基本过滤服务。

5）IP 包转发功能

该功能主要负责按照转发表内容在各端口（包括逻辑端口）间转发数据包并且改写链路层数据包头信息。

6）路由信息维护功能

该功能负责运行路由协议，维护路由表。路由协议可包括 RIP、OSPF 等协议。

7）维护决定数据帧转发及过滤的信息

交换机必须实现维护数据帧转发/过滤信息。

8）运行维护功能

交换机必须实现运行维护功能。

9）网络管理功能

交换机必须实现网络管理接口及协议。

10）设备管理和认证

交换机可选实现管理功能包括用户管理认证功能例如 802.1X、业务管理例如组播控制等功能。

5 三层交换机接口

5.1 10/100Mbit/s 以太网接口

三层交换机必须支持 10/100Mbit/s 自适应以太网接口。

10Mbit/s 以太网接口应符合 IEEE802.3，物理层接口上采用曼切斯特编码，用 0.85V 和-0.85V 分别表示“1”和“0”。电缆可采用 10Base-T。

100Mbit/s 以太网接口应符合 IEEE802.3u。100Base-T 技术中可采用 3 类传输介质：100Base-T4、100Base-TX 和 100Base-FX。采用 4B/5B 编码方式。

5.2 千兆比以太网接口

三层交换机必须支持千兆比以太网接口（符合 IEEE802.3z）。

1000Mbit/s 以太网物理接口有 1000Base-SX、1000Base-LX 以及 1000BaseT。1000BaseT 接口应符合 IEEE802.3ab。

如三层交换机拥有千兆接口，建议实现 1000Base-SX 和 1000Base-LX 两者之一。

5.3 SDH 接口

5.3.1 接口类型

交换机可选支持 SDH STM-1 接口、SDH STM-4 接口和 SDH STM-16 接口。

作为任选，交换机可以支持 SDH STM-64 接口。

STM-1 有光接口和电接口两种。STM-1 电接口适用于局内，干扰信号弱的情况；STM-4、STM-16、STM-64 应采用光接口。

5.3.2 SDH 层要求

— 应符合 YDN 099—1998《SDH 技术体制》和 ITU-T 建议 G.707。

— 应支持以下告警处理功能：LOS，LOF，LAIS，PAIS，LOP，SF，SD。

— 应支持性能监控。

— 应支持 B1，B2，B3 差错计数。

— 应支持本地（内部）或环路定时（从网络恢复时钟），精度要求参照 SDH 标准。

— 应支持保护倒换和本地环回（诊断）和网络环回功能。

5.4 ATM 接口

5.4.1 接口类型

交换机可选支持 ATM 155Mbit/s 接口和 ATM 622Mbit/s 光接口。ATM 155Mbit/s 接口分光接口和电接口两种，电接口适用于局内，干扰信号弱的情况。

ATM 155Mbit/s 接口和 ATM 622Mbit/s 光接口具体要求参见 YDN 067—1998《ATM 交换机设备技术规范》相应的要求。

此外，作为任选，交换机还可以支持 ATM 2.5Gbit/s 光接口，具体要求待定。

5.4.2 ATM 层要求

— 应当支持 PVC，可选支持 SVC。

— 应当支持 AAL5，支持 CBR、UBR 和 VBR 业务，支持业务量整形。

— 应当支持 RFC1483 规定的 AAL5 上的多协议封装。

— 应当支持 LLC/SNAP 和 IP 复用 PVC（路由协议的 LLC 封装）。

— 应当支持 F5 OAM 信元处理。

5.5 WDM 接口

作为任选，交换机可以支持 WDM 接口。

WDM 接口具体要求参见 YDN 120—1999《光波分复用（WDM）系统总体技术要求》。

6 链路层功能

6.1 数据帧的转发及过滤

6.1.1 转发数据帧

交换机转发数据帧必须：

1）符合寻址规定；

2）提供

— 在不提供 48 比特通用管理地址时分配组 MAC 地址来标识网桥协议实体的途径；

— 端口标识符在实现生成树算法及协议时标识交换机每一端口。

交换机转发数据帧可以：

1）提供转发时控制优先级映射的能力；

2）提供多种流量分类；

3）对独立 MAC 地址的转发行为作规定；

4）管理转发帧的优先级。

数据帧的转发可以基于存储转发或直通式转发。

交换机必须支持存储转发。

交换机可以支持直通式转发。实现直通式转发的交换机必须缺省设置为存储转发。

6.1.2 过滤数据帧

交换机过滤数据帧必须符合：

— 实现基本过滤服务，对每个端口至少关联单一流量类（Traffic Class）；

— 对过滤数据库下列参数使用规定的值：

- 过滤数据库大小，过滤数据库所能容纳的最大条目数；
- 静态数据库大小，静态数据库所能容纳的最大条目数。

交换机过滤数据帧可以：

— 提供读取和更新过滤数据库和静态数据库的能力；

— 提供设置过滤数据库更新时间的能力。提供该能力的交换机应当实现本标准指定的所有可选值；

— 对独立 MAC 地址的过滤行为作规定。

6.1.3 支持转发/过滤数据帧的功能

交换机必须实现下列支持数据帧转发/过滤、提供 QoS 的功能：

— 帧接收；
— 丢弃所收到的错误帧；
— 丢弃 frame_type 参数不是 user_data_frame 或 mac_action 参数不是 request_with_no_response 的帧；
— 如果需要，重新产生用户优先级；
— 丢弃过滤信息应用指定需要丢弃的帧；
— 丢弃传输服务用户数据单元大小超过 ISO/IEC 15802-3，6.3.8 中规定的帧；
— 发送所收到得到其他端口的帧；
— 根据过滤信息应用选择流量类；
— 根据流量类对帧排队；
— 丢弃超过最大网桥传输时延的帧；
— 在排队的帧中选择帧传输；
— 选择带外访问优先级（ISO/IEC 15802-3，6.3.9）；
— 如果需要，映射服务数据单元，重新计算帧检验序列；
— 帧发送。

6.1.3.1 帧接收

关联在交换机端口上的 MAC 实体应当检查所连接局域网上所有被传输的帧。

所有正确的帧都被提交到 M_UNITDATA 标识原语，按照下面描述处理。

所有 M_UNITDATA.indication 原语中 frame_type 和 mac_action 参数分别是 user_data_type 和 request_with_no_response 的帧被提交到学习和转发进程。

所有 frame_type 和 mac_action 参数是其他值的帧不应提交到转发进程，但可以提交到学习进程。

所有 frame_type 是 user_data_type，目的地址指向交换机端口的帧应当提交到 LLC。这样的帧的目的地址域中应当携带交换机端口的独立地址或者关联在端口上的组地址。提交给 LLC 的帧同样可以按照上面描述提交给学习进程和/或转发进程。

将交换机端口作为终端的帧和从其他端口转发到端口的帧也应当提交给 LLC。

6.1.3.2 重新生成用户优先级

所收到帧的 user_priority 由包含在帧中的优先级信息和接收端口的用户优先级重新生成表得到。对于每个接收端口，用户优先级重新生成表应当包含 8 个条目，对应于 user_priority 可能的 8 个值（0~7）。每个条目指示给定优先级后，重新生成的用户优先级。

表 1 定义了对所收到的数据中指示的 8 种可能的用户优先级所重新生成的用户优先级的缺省值。

表 1 用户优先级重新生成表

用户优先级	缺省得重新生成用户优先级	范围
0	0	0~7
1	1	0~7
2	2	0~7
3	3	0~7
4	4	0~7
5	5	0~7
6	6	0~7
7	7	0~7

交换机可选支持通过管理功能改变用户优先级重新生成表值。如果支持该功能，交换机应当能对任意接收端口的任意到达优先级独立指定范围中任意值。

6.1.3.3 帧发送

关联在交换机端口上的每个 MAC 实体应当发送由 MAC 中继实体提交的帧。

转发进程提交被中继的帧用于发送。被发送帧所关联 M_UNITDATA.request 原语使用所接收到相应 M_UNITDATA.indication 原语中源和目的地址域。

LLC 协议数据单元由 LLC 作为交换机端口提供的 MAC 服务的使用者提交。用于传输上述协议数据单元所发送的帧，将端口的独立 MAC 地址作为源地址。

每一个帧都发给 MAC 进程，供特定的 IEEE802 局域网技术观察。相应 M_UNITDATA.request 原语中 frame_type 和 mac_action 参数分别应当为 user_data_type 和 request_with_no_response。

由交换机端口所提供的 MAC 服务的 LLC 用户请求的帧发送应当提交给 MAC 中继实体。

6.1.3.4 执行拓扑限制

当且仅当下列条件满足时，交换机的端口才能作为可用的传输端口：

— 接收帧的端口处于转发状态；

— 需要发送帧的端口处于转发状态；

— 发送帧的端口不同于收到该帧的端口；

— 所需发送帧中 mac_service_data_unit 大小不超过发送端口连接的局域网所支持的 mac_service_data_unit 最大尺寸。

6.1.3.5 过滤帧

转发进程应基于下列条件作过滤决定：

— 收到的帧中 MAC 目的地址；

— 过滤数据库中关于 MAC 地址和接收端口的信息；

— 对可用发送端口的缺省组过滤行为。

6.1.3.6 帧排队

转发进程应当为排队的帧提供存储服务，等待时机将帧提交给关联在交换机端口上的 MAC 实体。

交换机可以在端口上提供多个队列。帧基于使用流量类的用户优先级决定所使用的存储队列，上述流量类是关联在每个端口上状态信息的一部分。对每个可能的用户优先级必须对流量类赋值。用户优先级可以由 0~7。队列应当一一对应到流量类。

出于管理考虑，交换机应最多支持 8 个流量类来支持将各个用户优先级的帧独立排队。

表 2 给出用户优先级到流量类映射的建议。

表 2 用户优先级到流量类映射的建议

		可用的流量类数量							
		1	2	3	4	5	6	7	8
用户优先级	0（缺省）	0	0	0	1	1	1	1	2
	1	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	1
	3	0	0	0	1	1	2	2	3
	4	0	1	1	2	2	3	3	4
	5	0	1	1	2	3	4	4	5
	6	0	1	2	3	4	5	5	6
	7	0	1	2	3	4	5	6	7

当帧提交到关联在端口上的 MAC 实体之后，应当从存储的队列中删除。当缓存溢出时，交换机可以在队列中删除帧；交换机可以不按次序发送。

端口离开转发状态时，帧队列应当删除。

在某一特定端口队列中帧删除并不表示删除其他端口帧队列中该帧。

6.2 维护决定数据帧转发/过滤的信息

6.2.1 维护过滤/转发信息

交换机应当实现：

— 计算及配置被桥接的以太网的拓扑；
— 静态配置保留地址；
— 显式配置静态过滤信息；
— 通过察看被桥接局域网的流量中源地址来自动学习对单目的地址的动态过滤信息；
— 对所学到的动态过滤信息设置定时器实现按时间老化；
— 通过 GMRP 协议自动添加/删除动态过滤信息；
— 显式配置关联到交换机端口上的流量类信息；
— 显式配置关联到交换机端口上的端口 VID（PVID）；
— 显式配置关联到交换机端口上的允许接收帧类型参数；
— 显式配置关联到交换机端口上的使能入口过滤参数；
— 通过使用 GVRP 自动配置动态 VLAN 注册实体；
— 显示配置通过静态 VLAN 注册实体方式关联 GVRP 操作的管理控制；
— 通过观察网络流量自动学习关联到 VLAN 的 MAC 地址；
— 显式配置每个端口需要的出口标记。

6.2.2 网桥协议数据单元

交换机应当实现网桥协议数据单元。网桥协议数据单元在 ANSI/IEEE Std 802.1D 1998 中定义。

6.2.2.1 网桥协议数据单元结构

网桥协议数据单元是用于实现生成树算法/协议的网桥协议实体。网桥协议数据单元分如下两种。

1）配置 BPDU

协议标识符	1
	2
协议版本标识符	3
BPDU 类型	4
标志	5
根标识符	6
	7
	8
	9
	10
	11
	12
	13

根路径代价	14
	15
	16
	17
网桥标识符	18
	19
	20
	21
	22
	23
	24
	25
端口标识符	26
	27
消息时限	28
	29
最大时限	30
	31
Hello 时间	32
	33
传递时延	34
	45

图 1　配置 BPDU

在图 1 中配置 BPDU 结构如下：

— 协议标识符位于第 1~2 字节。使用 0000 0000 0000 0000 标识生成树算法/协议。
— 协议版本标识符位于第 3 字节，使用 0000 0000。
— BPDU 类型位于第 4 字节，使用 0000 0000，标识配置 BPDU。
— 拓扑改变确认标志在第 5 字节第 8 比特。
— 拓扑变化标志位于 BPDU 第 5 字节第 1 比特。
— 根标识符位于 BPDU 第 6~13 字节。
— 根路径代价位于 BPDU 第 14~17 字节。
— 网桥标识符位于 BPDU 第 18~25 字节。
— 端口标识符位于 BPDU 第 26、27 字节。
— 消息时限定时器值位于 BPDU 第 28、29 字节。
— 最大时限定时器值位于 BPDU 第 30、31 字节。
— Hello 时间定时器值位于 BPDU 第 32、33 字节。
— 传递时延定时器值位于 BPDU 第 34、35 字节。

消息时限值应小于最大时限值。

2）拓扑变化通知 BPDU

协议标识符	1
	2
协议版本标识符	3
BPDU 类型	4

图 2 拓扑变化通知 BPDU

图 2 BPDU 结构如下：

— 协议标识符位于 BPDU 第 1~2 字节。使用 0000 0000 0000 0000，标识生成树算法/协议。

— 协议版本标识符位于 BPDU 第 3 字节，使用 0000 0000。

— BPDU 类型位于 BPDU 第 4 字节，使用 1000 0000。

6.2.2.2 对 BPDU 的验证

当且仅当 BPDU 长度大于 4 字节且满足下面条件之一时，网桥协议实体应当根据生成树算法/协议中规定处理收到的 BPDU。

— BPDU 类型为配置 BPDU 且 BPDU 包含 35 字节，BPDU 消息时限参数小于最大时限参数。

— BPDU4 字节且类型为拓扑变化通知 BPDU。

6.2.3 GARP 组播注册协议（GMRP）

交换机应当实现 GMRP。GMRP 在 ANSI/IEEE Std 802.1D 1998 中定义。

GMRP 提供一种机制允许交换机和终端系统动态注册组成员信息，该信息在所有支持扩展过滤服务的交换机/网桥间传播。GMRP 利用 GARP 提供服务。

需要通过 GMRP 注册、消除注册或传播的信息有：

— 组成员信息。该信息指示存在一个或多个 GMRP 参与者是某特定组成员，并且携带相关该特定组的组 MAC 地址。对组成员信息的交换可能是因为创建或更新过滤数据库中的组注册实体来指示组成员在哪个端口上被注册。

— 组服务要求信息。该信息只是一个或多个 GMRP 参与者要求将所有组或为注册组应用缺省组过滤行为。

交换机实现 GMRP 的一致性要求

实现 GMRP 的交换机应当符合：

— 实现 GARP 应用及注册状态机以及 LeaveAll generation 机制。

— 按照状态机要求交换 GARP PDU，该 PDU 携带应用特定信息。

— 广播注册信息应符合基本生成树的 GIP 操作。

— GMRP 实现符合 ANSI/IEEE Std 802.1D 1998 中 10.3 字句。

— 对携带 GARP 应用地址的数据帧实行转发、过滤及丢弃。

6.2.4 一般属性注册协议（General Attribute Registration Protocol GARP）

交换机应当实现 GARP。GARP 在 ANSI/IEEE Std 802.1D 1998 中定义。

6.2.4.1 GARP 定义

GARP 提供一种一般属性分发能力，该能力由 GARP 应用在桥接的局域网上与其他 GARP 参与者注册及消除注册属性值。属性的类型、属性的值和属性值相关的语义由 GARP 应用决定。

6.2.4.2 GARP 实现要求

GARP 及相关算法用作在桥接的网络中建立、维护和解析属性注册以及在连接到局域网上的 GARP 参与者间分发注册信息。交换机应当实现 GARP。实现 GARP 的交换机必须满足下面要求。下面要求包括应用及注册要求，失败条件下的错误恢复、性能、可扩展性，与非 GARP 设备反向兼容以及对交换机、终

端和网络的负荷。

1）实现 GARP 必须允许 GARP 参与者连接到桥接的局域网上发布关联于 GARP 应用的属性值声明。

2）实现 GARP 必须允许 GARP 参与者连接到桥接的局域网上撤销关联于 GARP 应用的属性值声明。

3）实现 GARP 必须允许交换机将收到的声明向通过该交换机所连接的局域网能够访问的 GARP 参与者广播。

4）实现 GARP 必须允许 GARP 参与者维护指示当前声明状态和参与设备每个端口属性注册的状态信息。

5）实现 GARP 必须允许 GARP 参与者消除关于部分或所有桥接网络的属性的状态信息。

6）发布/撤销属性注册信息以及属性注册信息在桥接网络中传播的时延应当保持足够小，且按照网络直径线性增长。

7）当一个 GARP 失败时，GARP 应保持可用。

8）丢失一个数据帧时，GARP 应保持可用。

9）GARP 应当能正常运行在：

— 同构网络中：即桥接的网络中所有交换机/网桥都支持基本过滤服务及扩展的过滤服务。

— 异构网络中：即某些交换机/网桥只支持基本过滤服务，某些交换机/网桥支持基本过滤服务和扩展过滤服务。

10）GARP 需要的带宽应当足够小，且与桥接网络上的流量无关。该带宽应当只与局域网上维持成员关系的组的规模有关。

6.2.4.3 GARP 参与者间互操作性要求

1）对每个定义的 GARP 应用必须在协议交换帧的目的地址使用惟一的组 MAC 地址，称为 GARP 应用地址。GARP 应用地址中 01-80-C2-00-00-20 用作 GMRP 地址，01-80-C2-00-00-21 到 01-80-C2-00-00-2F 保留供将来使用。实现已定义 GARP 应用地址的交换机不需要将目的地址是 GARP 地址的帧转发；未实现已定义 GARP 应用地址的交换机需要将目的地址是 GARP 地址的帧向所有活跃拓扑终端口转发。

表 3 GARP 应用地址

分配	值
GMRP 地址	01-80-c2-00-00-20
保留	01-80-c2-00-00-21
保留	01-80-c2-00-00-22
保留	01-80-c2-00-00-23
保留	01-80-c2-00-00-24
保留	01-80-c2-00-00-25
保留	01-80-c2-00-00-26
保留	01-80-c2-00-00-27
保留	01-80-c2-00-00-28
保留	01-80-c2-00-00-29
保留	01-80-c2-00-00-2a
保留	01-80-c2-00-00-2b
保留	01-80-c2-00-00-2c
保留	01-80-c2-00-00-2d
保留	01-80-c2-00-00-2e
保留	01-80-c2-00-00-2f

2）GARP 参与者间 GARP PDU 的收发应采用所考虑 GARP 应用的格式，使用一般 PDU 格式，应通过 LLC 类型 1 过程取得。源和目的 LLC 地址应采用分配给生成树算法的标准 LLC 地址。

3）GARP 参与者收到格式不正确的 PDU 时应丢弃。

4）GARP 参与者的协议行为应符合协议中状态机的描述。

6.2.5 生成树算法及协议（Spanning Tree Algorithm and Protocol）

生成树算法及协议将通过交换机或网桥连接的局域网的拓扑简化成一棵生成树。

交换机应当实现生成树算法及协议。对生成树算法及协议的实现必须符合 ANSI/IEEE Std 802.1D（1998）中第八子句。

生成树算法的实现必须符合下列要求：

— 该算法必须能将一个被桥接的任意拓扑局域网中活跃的拓扑结构配制成一棵生成树。该生成树消除环路，从而在任意两个节点间最多存在一条路径。

— 当由于交换机或网桥失败或瘫痪时，该算法必须能在剩下局域网拓扑中自动重新配置生成树拓扑，达到冗余目的。当存在交换机或网桥端口加入到桥接的局域网中时，该算法必须能自动调节适应新的拓扑结构，避免产生环路。

— 在任意规模桥接的局域网中，构成生成树的活跃的拓扑必须稳定。该算法必须在大多数情况下在可预知的较短间隔内收敛，减少网络不可用时间。

— 该算法的结果应当可预测，可重复。结果应当可以由对算法参数的管理选择，从而通过配置管理与流量分析达到性能管理的目的。

— 该算法应当对终端系统透明，终端系统在使用 MAC 服务时应当无法得知是连接在单一局域网或桥接的局域网上。

— 由交换机或网桥使用用于建立或维护生成树的带宽必须只占链路可用带宽一小部分，并且不随桥接网络的规模增加而增加。

另外为降低交换机及其配置的复杂性，生成树算法的实现必须：

— 该算法关联交换机每端口的内存应当独立于桥接网络规模。

— 交换机在连入桥接网络之前除必要的 MAC 地址外不需作任何附加的配置便可正常工作。

为实现生成树算法，交换机必须：

— 配置一个惟一的 MAC 组地址，由桥接的局域网上所有其他交换机或网桥识别，该地址用于标识连接在局域网上交换机或网桥桥接协议实体。

— 交换机的标识在桥接的局域网范围内惟一。

— 交换机的每个独立端口必须配置单独的端口标识，该标识的分配独立于其他交换机或网桥的端口标识。

交换机必须为上述参数配置，实现分配机制。当交换机使用 48 比特全球统一管理地址时，该惟一的 MAC 地址作为交换机组地址标识桥接协议实体。

另外为使生成树算法结果可配置，交换机必须实现：

— 一种途径用作给予交换机在桥接的局域网中相对其他交换机或网桥的相对优先级。

— 一种途径用作给予某端口在该交换机中相对其他端口的相对优先级。

— 一种途径用作给予端口上路径的代价。

6.2.6 VLAN（虚拟局域网）功能

VLAN 功能指通过桥接的局域网内活跃拓扑中工作站的划分，各 VLAN 使用 VID（VLAN 标识符）区分。各个 VLAN 是原桥接的局域网的一个子集。

交换机应当支持 VLAN 功能。交换机对 VLAN 功能的支持应符合《VLAN 技术规范与测试方法中的相关要求》。

6.2.7 远程桥接

远程媒体访问控制桥接是指在互连的局域网间使用远程媒体访问控制桥的操作以及远程媒体访问控

制桥通过非局域网通信设备按照生成树算法配置被桥接局域网的协议。

如果实现其他接口，可选三层交换机实现远程桥接，对远程桥接的实现必须符合 IEEE Std 802.1g。

6.2.8 多链路聚合

多链路聚合是指在逻辑上将多条独立的链路作为一条单独链路使用，以此获得灵活的高带宽以及链路冗余。

建议三层交换机实现多链路聚合。由于规范链路聚合的协议 IEEE Std 802.1ad 1997 年开始制订，2000 年完成，所以对链路聚合的实现建议符合 IEEE Std 802.3。

6.3 流量控制

当流量超过交换机的最大传输能力时，交换机吞吐量可能下降。本标准规定的交换机不允许因交换能力不足引起拥塞。由于可能存在多个端口向某一端口发流量的情况，因此交换机应当实现流量控制。

流量控制是一种被交换机或拥塞实体用于限制网络访问的机制。流量控制通过对缓存器设置上限、修改发送速率或将发送源关闭一段时间实现。对触发流控机制的参数可以根据流量情况来静态或动态调整。

6.3.1 流量控制性能衡量

流量控制性能衡量需要将不同的流量控制策略进行比较。此外还可以对流量控制策略的参数进行调节优化性能。3 个常用的性能衡量指标是吞吐量、包延时和丢包率。

包吞吐量代表设备处理网络负载的能力。在低负载情况下，吞吐量等于网络负载。在负载较高情况下，由于对资源的竞争吞吐量会下降。

包时延是吞吐量的函数。一种特定的流控可能针对高负载下时延特性优化，对低负载下时延不优化。

丢包率是用来衡量由于重传超时，过多的冲突，缓冲区溢出造成数据包丢失。随着丢包率的增加，吞吐量会降低，时延会增加。

6.3.2 半双工下的流控

6.3.2.1 背压流控

背压流控指交换机拥塞时，某个帧到达输入端口时在帧上加一个强制的冲突，迫使远端 DTE 放弃发送。远端 DTE 经后退间隔后重传。如果此时拥塞尚未解除，可以继续使用背压流控机制。对背压流控端口的选择在本标准范围之外。

背压式流控可以采用如下两种方式：

1）背压流控可以简单地拥塞所有源端口，直到拥塞解除。这样做会影响网络中其他节点的性能。

2）背压流控也可以检查帧中目的地地址，如果属于拥塞端口则产生强制冲突。

交换机可选实现背压流控。实现背压式流控的交换机建议采用方式 2)。

6.3.2.2 载波扩展流控

载波扩展流控是指当交换机拥塞时，交换机产生载波侦听信号。在发送载波侦听信号期间所有的 DTE 被抑制。在拥塞期间，交换机发送载波侦听信号必须发送有效数据位。

交换机可选实现载波扩展流控。对发送载波侦听信号端口的选择在本标准范围之外。

6.3.3 全双工下流控

6.3.3.1 PAUSE 控制

全双工下交换机流控应当采用 IEEE802.3 流量控制（又称 PAUSE 控制）。PAUSE 控制在 IEEE std802.3，1998 版附件 31B 中规定。

PAUSE 控制用作禁止对端在一定时间内发送除 MAC 控制帧之外的帧。是否允许发送 PAUSE 控制在自动协商中决定。IEEE802.3x 定义了两个方向上都支持 PAUSE 功能的 PAUSE 帧格式配置。通过对 PAUSE 位置位，链路两端的设备可以发送并接受 PAUSE 帧。

PAUSE 帧是一个特殊编码的通用 MAC 控制帧。该帧是一个具有最小合法长度的 802.3 以太网帧。该帧具有以下特点：

1）目的地址域是为 MAC 控制 PAUSE 帧单独保留的多目地址（01-80-c2-00-00-01）。

2）源地址域是源/发送站点的48bit地址。

3）两字节的长度/类型域包含16进制值88-08。表示802.3局域网的MAC控制帧。

4）MAC控制操作码使用00-01。

5）MAC控制参数包含2字节的PAUSE定时器值。该定时器值16比特，以LSB在先方式传送。PAUSE时间单位是512位时间+1。

特别指出：

1）当发送器暂停时不禁止发送PAUSE帧。

2）0是在PAUSE帧中传输的有效暂停定时器值。一个被暂停的站点在得到下一个PAUSE帧后可以重载暂停定时器值。

3）由于为PAUSE帧保留的多目地址会被802.1D（生成树）网桥特殊对待，无论802.1D网桥端口状态如何，是否实现MAC控制子层，网桥不传播PAUSE帧。

建议交换机支持PAUSE控制。

6.3.3.2 不对称流量控制（AFC）

PAUSE控制提供了对称的流量控制。交换机可以支持不对称流量控制（AFC）。不对称流量控制是阻止源头的流量，使交换机不需要使用更多的缓冲区。在AFC配置下，交换机具有对端站流控的能力，端站不能对交换机流控。

互联的交换机间不应使用AFC。

交换机对AFC的实现可选。

6.3.4 流量控制策略

启动流量控制通常使用的策略有基于水位的流量控制、基于信用的流量控制和基于速率的流量控制。

对启动流量控制策略的选择在本标准范围之外。

本标准定义了用于流量控制的机制，何时启动流量控制，启动多少时间在本标准范围之外。

6.4 端口镜像

端口镜像通常是指将某一端口上所有输出内容都复制到另一端口。三层交换机可选实现端口镜像。

7 网络层规定

7.1 Internet协议-IP

7.1.1 定义

三层交换机必须实现IP协议，并符合RFC791。交换机必须实现与IP相关的子网（符合RFC950），IP广播（符合RFC922）和无类域间路由选择（符合RFC1519）。

在某些情况下，要求交换机在丢弃数据包时不作任何处理（即不发ICMP差错消息），然而为了诊断故障，交换机可以提供将差错写入日志（包括所丢弃数据包的内容）的能力，以及具有对丢弃数据包进行计数的能力。

7.1.2 协议概述

RFC791对IP协议作描述。

7.1.2.1 选项域

三层交换机对IP协议选项域作如下规定。

1）安全性选项域：可选实现。

2）流标识选项域：不应实现。

3）源路由选项域：可选实现。

4）路由纪录选项域：可选实现。

5）时间戳选项域：应当实现。

7.1.2.2 选项域中的地址

如果实现相应选项域，要求交换机能将地址插入到记录路由、严格源和记录路由、松散源和记录路

由或时间戳选项域中。交换机必须插入发送该数据包的逻辑接口的地址。当发包接口没有 IP 地址时（例如非编号接口），交换机必须插入交换机 ID。交换机 ID 是交换机 IP 地址之一，可以是某接口地址。交换机 ID 必须基于系统或链路指定。除非网络管理员修改，交换机 ID 一般不能改变（即使在重启动之后）。拥有多个非编号接口的交换机可以拥有多个交换机 ID。每个非编号接口必须与一特殊的交换机 ID 相关联。这种关联不能改变，除非重新配置了交换机。

7.1.2.3 IP 头中的未使用比特

IP 头包含两个未使用比特：一个在服务类型字节中，另一个在标志域中。在交换机产生的数据包中，不允许将上述任何比特置为 1。不允许交换机仅仅因为上述任何比特置为 1 而丢弃（即拒绝接收或拒绝转发）数据包，即交换机不应检查上述比特的值。

7.1.2.4 服务类型

IP 头中服务类型字节分为 3 个部分：优先级字段（最高 3 比特），服务类型字段（TOS）（后续 4 比特）以及保留比特（最后 1 个比特）。

交换机不应实现 RFC795 中规定的服务映射功能。

7.1.2.5 头校验和

IP 必须核实收到的任一数据包的 IP 校验和，必须丢弃包含无效校验和的数据包。交换机不允许提供关闭校验和核实功能的方法。

当 IP 头中仅 TTL 改变时，交换机可以使用递增的 IP 头校验和更新，这可以降低因交换机引起 IP 头毁坏的可能性。

7.1.2.6 不可识别头选项域

交换机必须不理睬不可识别的头选项域，即交换机必须实现选项域列表结尾选项域和无操作选项域，因为上述选项域不包含长度。

7.1.2.7 分段

交换机应当支持分段，分段必须符合 RFC791。

当交换机将 IP 包分段时，它应尽量减少分段数，并按顺序发送。当分段方法可能产生长度明显小于其他分段的一个分段时，则该分段为第 1 个 IP 分段，并首先发送。

7.1.2.8 重组

交换机必须支持将发送给自己的分段 IP 包重组。

重组必须符合标准 RFC1122。

7.1.2.9 生存时间（TTL）

交换机对产生或收到的 IP 包 TTL 的处理必须按照标准 RFC1122。

7.1.2.10 多子网广播

交换机不应支持针对所有子网广播。

7.1.2.11 地址

IP 地址分 5 类，从 A 类到 E 类。D 类用作 IP 组播，E 类保留。A、B、C 类地址一般用作单播网络前缀。

IP 组播地址是逻辑地址，表示一组主机，可以是永久的或临时的。永久组播地址在 RFC1700 中规定。临时地址可以为临时组动态分配，组员由 IGMP（RFC1112）动态决定。

7.1.3 特定规定

7.1.3.1 IP 广播地址

交换机必须符合以下几点。

1）将 255.255.255.255 或 {网络前缀，-1} 作为 IP 广播地址。

2）应悄悄丢弃（即不传递到交换机上层应用）发给 0.0.0.0 或 {网络前缀，0} 的数据包。如果不悄悄丢弃，则这些数据包必须作为 IP 广播包来处理。可以设置一可更改的配置选项域来决定是否接受上述数据包。缺省情况是丢弃上述数据包。

3）在直连网络/子网中发起 IP 广播时，缺省情况，应使用有限广播地址（255.255.255.255）（发送 ICMP 地址掩码应答时除外）。交换机必须接收有限广播。

4）不应始发地址为 0.0.0.0 或 {网络前缀，0} 的数据包。可以设置一可更改的配置选项域决定是否允许产生上述数据包。缺省情况是不产生上述数据包。

7.1.3.2 IP 组播

交换机应满足 RFC1122 中描述的 IP 组播要求。交换机应在所有相连网络上支持本地 IP 组播。当已指定 IP 组播地址到链路层地址映射时，交换机应使用该映射。交换机可以配置成使用链路层广播来替代上述映射。在所有点到点链路和其他端口上，组播被封装成链路层广播。

支持本地 IP 组播可以包括始发组播包、加入组播组、接收组播包、离开组播组等。

7.1.3.3 划分子网

在某些情况下，交换机有必要支持网络划分子网，这些子网可能由不属于该网络的网络互连，上述情况称为支持非连续子网。

交换机必须支持非连续子网。

一个给定网络的地址块可能被划分成不同大小的子块，不同子块的网络前缀可能长度不同，交换机必须在所有接口配置和路由数据库中支持不同长度网络前缀。

7.2 互联网控制消息协议-ICMP

7.2.1 定义

ICMP 是辅助协议，它为 IP 提供路由、诊断和差错处理功能。ICMP 在标准 STD5，RFC792 中描述。交换机必须支持 ICMP。

ICMP 消息分成两类：

1）ICMP 差错消息

2）ICMP 请求消息

7.2.2 一般规定

7.2.2.1 未知消息类型

如果收到未知类型的 ICMP 消息，该消息必须送到 ICMP 用户接口（如果交换机存在用户接口）或者被悄悄丢弃（如果交换机不存在用户接口）。

7.2.2.2 ICMP 消息 TTL

当产生 ICMP 消息时，交换机必须初始化 TTL 值。ICMP 应答的 TTL 值不能从触发该应答的包中得到。

7.2.2.3 初始消息头

ICMP 数据包应在不超过 576 字节的条件下尽可能多包含原始数据包内容。返回的 IP 头（和用户数据）必须与收到的完全一致，除非交换机没有要求恢复所有对 IP 头的改变，这些改变通常是在检查出差错以前转发数据包的例行工作（例如 TTL 减 1，更新选项域）。但对参数问题消息，如果问题在更新的字段中，交换机必须恢复更新。

7.2.2.4 ICMP 消息源地址

除非特别指定，由交换机产生的 ICMP 消息的源地址必须是传输 ICMP 消息的物理端口的 IP 地址。如果该端口没有 IP 地址，则使用交换机 ID 替代。

7.2.2.5 TOS 及优先级

ICMP 差错消息的 TOS 应设置成与触发该 ICMP 差错消息的数据包的 TOS 相同，除非如果设置成该值会导致该差错消息因无法选路到目的地而被立即丢弃，否则 ICMP 差错消息必须将 TOS 设置成 0。ICMP 应答消息的 TOS 必须设置成与引发该应答的 ICMP 请求相同。

7.2.2.6 不发送 ICMP 差错情况

在下面情况下可以不发送 ICMP 差错消息：

— 收到 ICMP 差错消息；或

— IP 头有效性检验失败的数据包（本节中指定允许发送 ICMP 差错消息时除外）；或
— 包的目的地地址是 IP 广播地址或组播地址；或
— 作为链路层广播或组播发送的包；或
— 源地址的网络前缀为 0 或者无效源地址的包；或
— IP 分段后不是第 1 个 IP 分段（IP 头中分段偏移量为非 0）。

另外本标准规定悄悄丢弃包时不允许发送 ICMP 差错消息。

7.2.2.7 速率限制

发送源抑制消息的交换机必须能限制产生该消息的速率。交换机应能限制发送其他类型 ICMP 消息（目的地不可达、重定向、超时、参数问题等）的速率。发送速率限制应是交换机配置工作的一部分。限制的方式（每交换机或每端口）由设备制造者决定。

7.2.3 特殊指定

7.2.3.1 目的地不可达

当交换机因没有到指定目的地的路由（包含无缺省路由）而无法转发包时，交换机必须产生一个“目的地不可达，编码为 0（网络不可达）”的 ICMP 消息。

如果交换机中存在到目的网络的路由，但该路由指定的 TOS 既不是“0000”，也不是需要路由的数据包中的 TOS，交换机必须产生一个“目的地不可达，编码为 11（网络因 TOS 不可达）”的 ICMP 消息（注）。

如果 IP 包需要被转发到直连在交换机上网络中主机（交换机是最后一跳交换机）且交换机确认没有到目标主机的路由，交换机必须产生“目的地不可达，编码 1（主机不可达）”的 ICMP 消息。

如果 IP 包需要被转发到直连在交换机上网络中主机但交换机因为没有到目的地的路由，该路由的 TOS 或者等于“0000”，或者等于需转发数据包的 TOS，不能转发 IP 包时，交换机必须产生一个“目的地不可达，编码为 12（主机因 TOS 不可达）”的 ICMP 消息（注）。

注：本标准暂不要求支持 TOS 应用。

7.2.3.2 重定向

ICMP 重定向消息用作通知本地主机某特定流量需使用不同的下一跳交换机。

与标准 RFC1122 相反，交换机在下面情况下可以不理睬 ICMP 重定向：当交换机运行路由协议，或者交换机正在发送数据包的端口上允许转发时，交换机可以为自身生成的数据包选择路径。

7.2.3.3 源抑制

交换机不应产生 ICMP 源抑制消息。如果交换机产生 ICMP 源抑制消息，则必须能控制该消息产生的速率。

交换机可以不理睬收到的源抑制消息。

7.2.3.4 超时

当交换机转发一 IP 包且 TTL 域减为 0 时，应符合 7. 2.3.8 中的规定。

当交换机重组一个发给该交换机的包时，交换机必须符合 RFC1122。

当交换机收到（即包目的地是该交换机）一个超时消息时，交换机必须符合 RFC1122。

7.2.3.5 参数问题

当交换机遇到其他 ICMP 消息没有覆盖的差错时，应产生参数问题消息。在该 ICMP 消息中，必须包含未经改变的 IP 头及指针指向的参数域。7.2.2 中规定了上述规定的一个例外。

一个参数问题消息的新变量已在标准 RFC1122 规定：

编码 1= 所需选项域丢失。

7.2.3.6 Echo 请求/响应

交换机必须实现 ICMP Echo 服务器功能：接收发给该交换机的 Echo 请求并发送相应的 Echo 响应。交换机必须能够接收，重组及响应一个 ICMP Echo 请求，该请求数据包可能最大到 576 或者所有相连网络的 MTU。

Echo 服务器功能可以选择不响应一个目的地址是 IP 广播或 IP 组播的 IP 包。

交换机应有一个可选项域，能使交换机悄悄丢弃 ICMP Echo 请求，如果提供该选项域，缺省情况下必须设置成回应 Echo 请求。

为了交换机维护目的，交换机必须实现一个用户/应用层接口来发送/接收 Echo 来用作网络诊断。所有的 ICMP Echo 响应必须传送到该接口。

ICMP Echo 响应中的源地址必须与 ICMP Echo 请求中目的地址相同。

在 ICMP Echo 请求中包含的数据必须包含在相应的 Echo 响应中。

如果 ICMP Echo 请求中包含路由记录或/和时间戳选项域，这个/些选项域应被更新后包含在 ICMP Echo 响应消息中。这样，路由记录将得到整个路由轨迹。

如果 ICMP Echo 请求中包含源路由选项域，返回路由必须将 Echo 请求消息中路由记录反向，除非该路由违反交换机的转发策略。

7.2.3.7 信息请求/响应

交换机不应产生或响应这些信息。

7.2.3.8 时间戳及时间戳响应

交换机应实现时间戳及时间戳响应:

— ICMP 时间戳服务器功能必须响应每个收到的时间戳消息。交换机应设计成实现最小的时延变化。

— 对 IP 广播或组播地址的 ICMP 时间戳请求消息可以被悄悄的丢弃。

— ICMP 时间戳响应消息中的 IP 源地址必须与相应 ICMP 时间戳请求消息中目的地址相同。

— 如果 ICMP 时间戳请求中包含源路由选项域，返回路由必须将时间戳请求消息中路由记录反向，除非该路由违反交换机的转发策略。

— 如果交换机提供发送 ICMP 时间戳请求的应用层接口，则收到的 ICMP 时间戳响应必须传送到该用户接口。

— 时间戳值所期望的值（标准值）以 10^{-6}s 从午夜（全球时间）计数。但是很难提供 10^{-6}s 精确率的值，例如一些系统以现行频率更新时钟，50/60 次/s，所以允许标准值必须至少更新 16 次/s，以及标准值的精确性必须接近操作员设值的 CPU 时间。

7.2.3.9 地址掩码请求/响应

交换机必须实现接收 ICMP 地址掩码请求消息并应答回应 ICMP 地址掩码响应，该消息在标准 RFC950 中定义。

交换机应为每个逻辑端口提供是否允许响应该端口上地址掩码请求的选项域，该选项域必须缺省设置成响应 ICMP 地址掩码请求。交换机在知道正确的地址掩码前不允许响应地址掩码请求。

当从一个包含多个逻辑端口的物理端口上得到源地址为 0.0.0.0 的地址掩码请求，且这些逻辑端口的地址掩码不同时，交换机不允许响应该地址掩码请求。

交换机应检查由地址掩码请求得到的掩码是否与交换机了解的掩码匹配。如果 ICMP 地址掩码请求是差错的，交换机应将地址掩码及发送方 IP 地址写入日志。不允许交换机使用 ICMP 地址掩码响应中的内容决定正确的地址掩码。

当交换机宕机时启动的主机可能无法得到地址掩码时，交换机在配置完它的地址掩码之后，可以在每个逻辑端口上广播发送没有必要的 ICMP 地址掩码响应。但是上述特性在变长掩码网络环境中是危险的。所以如果实现该特性，在下面情况下不广播没有必要的地址掩码响应。

— 配置成不广播无必要的地址掩码请求。每一逻辑端口必须拥有可配置的参数来决定是否广播，端口缺省配置必须是不发送无必要的地址掩码响应。

— 共享包含（但不相同）网络前缀和物理接口。

{网络前缀，-1} 格式的 IP 广播地址必须在广播地址掩码中使用。

7.2.3.10 交换机广播和请求

交换机必须在该交换机支持 IP 广播或 IP 组播地址的相连网络上支持 ICMP 交换机发现协议

(RFC1256) 中交换机部分，而且必须包含对交换机特定的、具有缺省值的所有可配置变量。

7.3 互联网组管理协议 (IGMP)

IGMP 是用于主机和组播交换机之间的协议，应用于一单个物理网络上以建立特定组播组中的主机成员关系。组播交换机使用该信息和组播路由协议一起支持互联网上的 IP 组播转发。

交换机应支持互联网组管理协议，并符合 RFC1112。

交换机必须实现 IGMP 中的主机部分要求。

7.4 互联网层转发协议

本节描述包转发进程。

7.4.1 转发描述

包转发过程具体规定见互联网层协议 (RFC791，RFC950，RFC922， RFC792，RFC1349)。

7.4.1.1 IP 头确认

三层交换机在处理 IP 包之前，必须作下述的有效性检查来确认该 IP 包头是否有意义。如果 IP 包头在下面任一项检查中失败，该 IP 包被丢弃，差错写入日志。

1) 链路层指示的 IP 包长度必须足够大，能容纳最小的 IP 数据包合法长度 (20 字节)。

2) IP 校验和正确。

3) IP 版本号。

4) IP 头长度必须足够大，能容纳最小的 IP 数据包合法长度 (20Bytes=5Words)。

5) IP 数据包总长度必须足够大能容纳 IP 数据包头，IP 包头的长度在 IP 包头长度字段中指示。

三层交换机中不允许有任何选项域来禁止上述任何一项检查。

7.4.1.2 本地分发决定

三层交换机收到 IP 包时，必须决定该数据包是否发给本三层交换机 (应本地分发) 或另一系统 (应由转发处理)。上述两种情况可能同时发生，某些 IP 广播或 IP 组播可能同时要求本地分发及转发。三层交换机必须使用下面规则决定适用哪一种情况。

1) 未到期的源路由选项域是那些指针值没有超过源路由选项域中最后条目 (entry)。如果 IP 包包含一个未到期的源路由选项域，选项域中的指针应前移，直到指针超过该选项域中最后一个地址，或下一地址已不是该三层交换机的地址。在后者情况下 (即一般情况下)，包被转发 (不是本地分发)。

2) 下列情况，包被本地分发，不考虑转发：

— 数据包的目的地地址完全匹配三层交换机的某个 IP 地址；或者

— 数据包的目的地地址是有限广播地址 ({-1，-1})；或者

— 数据包的目的地地址是从不转发的 IP 组播地址 (例如 224.0.0.1 或 224.0.0.2)，并且收到数据包的物理接口上相关联的逻辑接口至少一个是目标组播组的成员。

1) 下列情况下包被转发及本地分发：

— 数据包的目的地地址是 IP 广播地址，该地址指向至少一个三层交换机的逻辑接口，但是不指向收到该数据包的物理接口上相关联的逻辑接口。

— 数据包的目的地是 IP 组播地址，该地址允许转发 (不像 224.0.0.1 和 224.0.0.2) 和收到数据包的物理接口上相关联的逻辑接口至少一个是目标组播组的成员。

2) 如果数据包的目的地地址是 IP 广播地址 (不是有限广播地址)，该地址至少是收到该数据包物理接口相关联的逻辑接口时，该 IP 包被本地分发，除非收到该 IP 包的链路使用不区分广播与单播封装 (例如使用不同的链路层目标地址) 的 IP 封装，否则，IP 包同时被转发。

3) 在其他所有情况下，IP 包被转发。

7.4.1.3 决定下一跳地址

当三层交换机转发包时，必须决定将包直接发到目的地或者需要发到下一三层交换机。如果是后一种情况，三层交换机需要决定使用哪一个设备作下一三层设备。本节解释如何作决定。

7.4.1.3.1 IP 目的地地址

如果 IP 包头中目的地地址是三层交换机的一个地址，则下一 IP 目标地址是由选项域中指针指向的地址。

如果 IP 头中目的地地址是三层交换机的一个地址，则消息指向分析该消息的系统。

三层交换机决定如何处理数据包时，必须使用 IP 目的地地址。

7.4.1.3.2 本地/远程决定

当决定 IP 包需要按照 7.4.1.3 节指定的规则转发时，必须使用下列算法决定立即目标是否可以直接访问（见 RFC950）。

1）对每个没有 IP 地址的网络地址（上文中描述的非编号线路），对比线路另一端三层交换机 ID 与 IP 目的地地址。如果完全相等，包可以通过该接口传输。

2）如果在第一步中没有选择网络接口，对赋予该三层交换机的每个 IP 地址：

a）隔离该接口使用的网络前缀；

b）将包中 IP 目的地地址相应比特隔离；

c）对比网络前缀。如果相等，该 IP 包经过相应的网络接口传送。

3）如果目的地既不是非编号接口的相邻设备 ID 也不是直连的网络前缀，该 IP 目的地只能通过其他设备访问。对三层交换机及下一跳 IP 地址的选择在 7.4.1.4.3 中描述。对于不作路由器的主机，该选择是配置的缺省路由器。

4）如果选择的“下一跳”可以通过配置 NHRP 的接口访问，则应用下列附加步骤：

a）对比目的地 IP 地址与 NHRP 缓存中的目的地地址。如果该地址在缓存中，将该数据包发送到相应用缓存的链路层地址；

b）如果地址不在缓存中，则建立一个包含目标 IP 地址的 NHRP 请求包。将该请求发送到该接口上配置的 NHRP 服务器。这可能是一个独立的逻辑过程或路由器内一个实体；

c）NHRP 服务器将返回正确的链路层地址，该地址用作传送数据包及后续数据包。当等待 NHRP 服务器响应，系统可以将数据包传送到传统的“下一跳”设备。

7.4.1.3.3 下一跳地址

三层交换机应用前一节中的算法决定 IP 目标地址是否相邻。如果相邻，下一跳地址与目的地 IP 地址相同。否则该包只能通过下一设备转发才能到达立即目标。

如果无法查找到路由，该 IP 包被丢弃，产生一个相应的 ICMP 差错（产生目标主机不可达 ICMP 或者目标网络不可达 ICMP）。

7.4.1.3.4 管理倾向（Administrative Preference）

对厂商策略裁减规则的一个建议机制是使用管理倾向，该机制是简单优先级算法。管理倾向是手工为可选择的路由定义优先级。

每条路由关连一个基于不同属性的倾向性值（为倾向性值指定机制在下面建议）。倾向性值是一个[0...255] 间的整数，0 表示最倾向，254 表示最不倾向，255 是一个特殊值，指该路由不应使用。厂商裁减规则的第一步是丢弃所有最不可能的路由（倾向值为 255 的路由总是首先丢弃）。

由于误用该策略可能导致路由循环，该策略较不安全。

7.4.1.3.5 负荷分担

在下一跳选择进程最后仍可能存在多条路由。这种情况下三层交换机有多种选择。它可以武断地丢弃一些路由。它可以通过比较路由域中路由的度量来减少候选路由。三层交换机也可以保留多个路由，使用负荷分担机制来分摊流量。由于在某个条件下非常有效的负荷分担可能在另一条件下无效，因此应提供允许管理员禁止负荷分担的手段。

7.4.1.4 未使用的 IP 头比特

IP 头中包含几个未使用比特，在 TOS 字段和标志字段中。不允许三层交换机只因为未使用比特非零值而丢弃 IP 包。

三层交换机必须不理睬未使用比特的值，并原封不动转发。如果三层交换机将包分段，必须将上述比特复制到每一个分段的相应位置。

7.4.1.5 分段及重组

三层交换机必须支持 IP 分段。

在转发以前，三层交换机不能重组任何 IP 数据包。

7.4.1.6 互联网控制消息协议-ICMP

本节描述由三层交换机发送的 ICMP 消息。

7.4.1.6.1 目的地不可达

ICMP 路由器不可达消息是三层交换机对因为目的地不可达或服务不可用不能转发的 IP 包的响应。例如消息的目标主机没有开机，不能响应 ARP 请求，消息的目的地网络前缀在三层交换机中没有有效的路由等。

三层交换机必须有能力发送 ICMP 目的地不可达消息并且能选择一个与不可达原因最接近的编码。

编码规则必须符合在 RFC792 和 RFC1122 中规定。

7.4.1.6.2 重定向

ICMP 重定向消息是三层交换机通知本地主机某一类流量应使用另一个不同的下一跳设备。

三层交换机应不产生 RFC792 中规定的网络重定向消息（编码 0），或网络和 TOS 重定向消息(编码 2)。三层交换机必须能产生主机重定向消息（编码 1），应能产生在 RFC792 中指定的 TOS 和主机重定向消息。

三层交换机不能因 RFC792 中指定的网络或 TOS 消息（编码 0，2）发送网络重定向消息。三层交换机必须能产生主机重定向消息（编码 1），应能产生在 RFC792 中规定的 TOS 和主机重定向消息（编码 3)。

当触发重定向的包具有一目的地，其路径需要三层交换机根据所要求的 TOS 来选择时，则三层交换机可以发送编码 3（因主机和 TOS 重定向）消息。

能产生编码 3（主机和 TOS）重定向消息的三层交换机必须设置可配置的选项域（缺省为允许）来允许发送编码 1（主机）重定向来替代编码 3 重定向。如果作相应配置，三层交换机必须发送编码 3 重定向替代编码 1 的重定向。

除非满足下列所有条件，三层交换机不可以产生重定向消息：

1）数据包正在转发到接收该包的同一物理接口，并且

2）包中的源 IP 地址与下一跳 IP 地址在同一 IP 逻辑（子）网上，并且

3）包中没有包含 IP 源路由选项域。

ICMP 重定向消息中源地址必须与目标地址在同一 IP 逻辑（子）网上。

应用路由协议（除静态路由外）的三层交换机不允许在转发包时使用从 ICMP 重定向中学习到的路由。如果三层交换机没有应用路由协议，三层交换机可以设置可配置的选项域允许在转发包时考虑从 ICMP 重定向中学习到的路由。

7.4.1.6.3 超时

当由于 TTL 域超时丢弃一个 IP 包时，三层交换机必须产生一个超时消息（编码 0)。三层交换机可以设置基于每个端口的选项域来禁止产生该消息，但是三层交换机在缺省条件下必须允许产生该消息。

7.4.1.7 互联网组管理协议-IGMP

IGMP 是用于主机和组播三层交换机之间的协议，应用于一单个物理网络上以建立特定组播组中的主机成员关系。组播三层交换机使用该信息和组播路由协议一起支持互联网上的 IP 组播转发。

三层交换机应支持互联网组管理协议，并符合 RFC1112。

三层交换机必须实现 ICMP 中组播路由器部分。

7.4.2 特定规定

7.4.2.1 生存时间（TTL）

IP 头中定义的 TTL 域规定作为限制数据包生存时间的定时器。TTL 是以秒为单位的 8 比特域。即使实际处理时间<1s，处理 IP 包的每个三层交换机（或者其他模块）必须至少将 TTL 减 1。通常情况下，TTL 可以作为衡量数据包在互联网上能传输多少跳的限制。

当三层交换机转发数据包时，必须将 TTL 至少减 1。如果处理包时间>1s，三层交换机可以将 TTL 每秒减 1。

如果 TTL 减到 0 或更少，数据包必须被丢弃，并且，如果目的地不是组播地址，三层交换机必须向 IP 包发送方发送 ICMP 超时消息编码 0。另外，三层交换机不可以因为预测到数据包最终目的地路径上其他三层交换机会丢弃该数据包而预先丢弃该 TTL 不为 0 的单播数据包或广播包。但是三层交换机可以对组播包作预丢弃，以更有效地实现 IP 组播的扩展环搜索算法（见 RFC1112）。

7.4.2.2 服务类型（TOS）

IP 头中 TOS 字节分 3 部分：优先级字段（高 3 比特），TOS 字段（下 4 比特）以及保留比特（最低 1 个比特）。对保留比特的规定见 7.3.1.2.3。

对 TOS 域的规定及使用见 RFC1349。

7.4.2.3 IP 优先级

本小节规定三层交换机中 IP 优先级字段正确处理的要求。优先级机制是基于不同业务流的相对重要性而在网络中分配资源。IP 层规范规定了为不同类型流量而使用不同的特定值。

三层交换机中优先级处理的基本机制是基于资源的优先分配，包括基于优先级顺序的队列服务，基于优先级的拥塞控制，以及链路层优先级属性的选择。三层交换机也为它产生的路由、管理、控制流量分配 IP 优先级。

在本节中讨论的优先级顺序队列服务包括、但并不限于转发过程的排队和输出链路排队。三层交换机支持的优先级应同样包含何时对有限资源（例如包缓存或链路层连接）分配考虑优先级处理。

7.4.2.4 优先级顺序队列服务

三层交换机应实现优先级顺序队列服务。优先级顺序队列服务表示当选择数据包到输出链路时，先选择队列中最高优先级的数据包。实现优先级顺序队列服务的三层交换机必须设置可配置的选项域在互联网层抑制优先级顺序队列服务。

三层交换机可以实现基于其他策略（不同于严格优先级队列）的吞吐量管理规程，但是必须设置可配置的选项域来关闭该吞吐量管理规程。

在拥塞控制中，实现优先级顺序队列服务的三层交换机必须在丢弃高优先级包之前先丢弃低优先级包。预占（包处理或传输的中断）不是互联网层的功能，其他层协议可以提供这种预占特性。

7.4.2.4.1 低层优先级映射

实现优先级顺序排队的三层交换机必须提供低层优先级映射，建议其他三层交换机也提供低层优先级映射。

实现低层优先级映射的三层交换机：

1）必须能将 IP 优先级映射到链路层优先级，如果该链路层具有规定的优先级特性；

2）必须实现一个可配置的选项域来配置对所有 IP 流量选择链路层缺省的优先级处理；

3）建议在每个接口配置 IP 优先级值到链路层优先级值特定的非标准映射。

7.4.2.4.2 所有三层交换机的优先级处理

三层交换机（无论是否使用优先级顺序队列服务）：

1）正常情况下必须接受并处理所有优先级的流量，除非管理员配置要求不这样做。

2）可以实现对特特殊流量源使用管理员规定的优先级过滤器。如果提供上述特性，该过滤器不允许过滤或截断下列类型 ICMP 差错消息：目的地不可达、重定向、超时和参数问题。如果提供该过滤器，该过滤器中同样需要根据地址的包过滤规程。

当数据包被过滤器丢弃时，应发送编码为 14 的 ICMP 目的地不可达消息，除非配置要求禁止发送该 ICMP 消息。

3）可以实现截断功能，该功能允许三层交换机拒绝或丢弃低于某一优先级的流量。该功能可以由管理措施而激活，或者由相关依赖的一些机制来激活，但必须提供一个可配置的选项域禁止非人工干预的激活。当包由截断功能丢弃时，三层交换机必须发送编码 15 的 ICMP 目的地不可达消息，除非配置要求禁止发送。

不允许三层交换机因优先级截断功能拒绝转发优先级为 6（互连网络控制）或 7（网络控制）的数据包。一般情况下，主机流量优先级应该是 5（CRITIC/ECP）或者更低。

4）可以改变不是由本三层交换机产生的数据包的优先级设置。

5）应能为支持的每个路由或管理协议单独配置优先级（某些协议除外，例如 OSPF，该协议指定的优先级值必须使用）。

6）可以不依赖对端地址而配置路由或管理流量优先级。

7）必须能正确响应提供的链路层优先级有关的差错指示。当链路因优先级有关条件不能接收包时，IP 包被丢弃，三层交换机应产生编码 15 的 ICMP 目的地不可达消息，除非配置要求禁止发送。

7.4.2.4.3 链路层广播转发

封装在大多数链路层协议中的 IP 包（PPP 除外）允许接收方只通过检查链路层协议的头（通常是链路层目的地地址）来区别广播、组播或单播包。这一节中提到的链路层广播只适用于可以区分出广播包的链路层协议，同时提到的链路层组播包只是用于可以区分组播包的链路层协议。

三层交换机不允许转发作为链路层广播收到的任何包，除非目的地地址指向一个 IP 组播地址。在后一种情况下假设链路层广播的使用是由于对链路层组播支持的缺乏。

三层交换机不允许转发作为链路层组播收到的包，除非数据包的目的地地址是一个 IP 组播地址。

三层交换机应丢弃经过一链路层广播的帧，该帧既没有规定一个 IP 组播地址，也没有规定一个 IP 广播地址。

当三层交换机作为链路层广播发送包时，IP 目的地地址必须是有效的 IP 组播或者 IP 广播地址。

7.4.2.5 互联网层广播转发

主要有两种 IP 广播地址类型：有限广播与直接广播。直接广播有 3 种子类型：对指定网络前缀的直接广播，向指定子网的直接广播，向指定网络所有子网的广播。对广播进行上述分类根据广播地址和三层交换机对目标网络子网结构的理解。同一广播地址对不同三层交换机可能得到不同的分类。

有限广播地址定义成全 1 地址：{-1，-1} 或者 255.255.255.255。

针对网络前缀广播包含网络前缀以及全 1 的主机地址即 {<网络前缀>，-1}。例如 A 类广播地址：net.255.255.255，B 类广播地址：net.net.255.255，C 类广播地址：net.net.net.255。其中 net 表示网络地址字节。

针对所有子网广播没有在 CIDR 中定义。

当三层交换机遇到一些非标准 IP 广播地址：

1） 0.0.0.0 是废弃的有限广播地址形式。

2） {<网络前缀>，0} 是废弃的针对网络前缀广播地址。

按照本节描述，三层交换机收到上述两种广播包后应丢弃，如果不这样处理，则必须按照未废弃的广播地址处理。这些规则在下面几节中描述。

7.4.2.5.1 有限广播

有限广播不能转发。有限广播不能丢弃。在有限广播有效的情况下，应使用有限广播替代直接广播。

7.4.2.5.2 直接广播

三层交换机必须将网络前缀直接广播作为有效，对远端网络或连接的无子网网络直接广播。

注：在 CIDR 看来像是网络前缀中的主机地址，禁止检查这样的网络前缀后的主机地址。给定一条路由和不相悖的策略，三层交换机必须转发针对网络前缀广播。可以发送针对网络前缀的广播。

三层交换机可以设置选项域来允许接收针对网络前缀广播，可以设置一个选项域允许转发针对网络前缀的广播。这些选项域必须缺省设置成禁止接收且禁止转发针对网络前缀的广播（参见 RFC2644）。

7.4.2.5.3 针对所有子网直接广播

三层交换机不应支持针对所有子网直接广播。

7.4.2.5.4 针对子网直接广播

在 RFC1716 中规定了针对子网广播的算法。在 CIDR 路由域中，针对子网广播与针对网络广播没有区别，所以针对子网广播可以看作针对网络广播。

7.4.2.6 拥塞控制

网络中拥塞可以不严格地定义成一种状态，在这种状态下对资源的需求（通常是带宽或 CPU 时间）超出了能力范围。拥塞避免试图阻止请求过度的能力，拥塞恢复试图恢复运行状态。三层交换机试图混合上述几种机制是可能的。本标准建议使用 RFC1154。

7.4.2.7 组播路由

IP 三层交换机应能支持转发 IP 组播包，转发基于静态组播路由或者由组播路由协议（见 7.10）动态决定的组播路由。转发 IP 组播包的三层交换机称为组播三层交换机。

7.4.2.8 转发控制

对每个物理接口，三层交换机应设置一个可配置的选项域来指定该接口上是否允许转发。当接口上禁止转发时三层交换机应：

1）必须丢弃所有该接口上收到的不是发给该三层交换机的包；

2）除该三层交换机生成的包外，该接口不发送任何其他包；

3）不通过任何路由协议宣布通过该接口的路径的可用性。

7.4.2.9 IP 选项域

一些选项域，例如路由记录和时间戳选项域，包含三层交换机转发路由时需要插入 IP 地址的位置。

然而每一个类似的选项域都有有限个位置以供插入 IP，三层交换机可能发现已没有插入 IP 的位置。在 7.4.1.5 描述如何选择一个 IP 地址插入到选项域中。

7.4.2.9.1 转发中不可识别的选项域和不可识别的 IP 选项域

转发中，包含不可识别的选项域和不可识别的 IP 选项域的包通过后必须毫无改变。

7.4.2.9.2 时间戳选项域

转发包时，三层交换机必须支持时间戳选项域。时间戳的值必须按照 RFC1122 中的规定来处理。

如果标志域=3（时间戳和预指定地址），且下一预指定地址匹配三层交换机中任一地址，三层交换机必须写入时间戳。预指定地址不一定是收到包的接口地址或者转发包使用的接口地址。

三层交换机可以提供一个可设置的选项域，设置该选项域后三层交换机转发包时即使标志域=0（时间戳）或=1（时间戳及注册 IP 地址）三层交换机也不理睬该选项域（转发时不改变）。如果提供上述选项域，该选项域必须关闭（三层交换机不忽略时间戳）。该选项域不影响三层交换机处理收到自身发出的有时间戳选项域的包，在三层交换机收到的数据包时间戳选项域中插入时间戳。

8 传输层协议要求

三层交换机通常应该支持传输控制协议（TCP）和用户数据报协议（UDP）。

8.1 用户数据报协议

用户数据报协议在 RFC768 中规定。

除下面两条外，三层交换机实现的 UDP 必须符合，且无条件符合 RFC1122 的要求：

1）本标准不规定不同协议层间的接口；

2）与 RFC1122 相反，三层交换机应该产生 UDP 校验和。

8.2 传输控制协议-TCP

传输控制协议在标准 RFC793 中规定。

9 路由协议

9.1 概述

互联网路由系统包含两部分：内部路由与外部路由。自治域（AS）允许描述一组三层交换机从内部路由到外部路由的转变。IP 数据包通常要穿过两个或多个 AS 的三层交换机才能到达目的地，AS 系统必须相互提供拓扑信息才能允许这种转发。内部网关协议用作在 AS 内部分发路由信息（即 AS 内部路由）。外部网关协议用作在 AS 间交换路由信息（即 AS 间路由）。

除非特殊路由协议的指定，三层交换机应将携带路由信息流量的 IP 数据包的优先级设置成 6（互联网控制）。

9.2 内部网关协议

9.2.1 定义

内部网关协议（IGP）用作在特定 AS 内部三层交换机间分发路由信息。对特定 IGP 算法的实现相对独立，但必须实现下列功能：

1）应能迅速反映 AS 内部拓扑的改变；

2）提供一种机制使电路振荡时不引起连续的路由更新；

3）提供快速收敛成无环回（loop-free）路由；

4）使用最少的带宽；

5）提供等效路由以便负荷分担；和

6）提供一种认证的路由更新方法。

三层交换机除实现静态路由外，必须实现 RIP v2。ISIS 协议和 OSPF 协议可选实现。

9.2.2 开放最短路径优先-OSPF

基于最短路径优先（SPF）是一类基于链路状态算法的协议，它们基于 Dijistra 的最短路径算法。在基于 SPF 的系统中，每个三层交换机通过称为泛滥（flooding）算法的过程得到完整的拓扑数据库。泛滥过程确保信息可靠传输。每一个运行 SPF 算法的三层交换机在数据路上建立 IP 路由表。

三层交换机可选实现在 RFC2328 中规定的 OSPF v2， 支持可变长子网掩码（VLSM），支持广播网络，支持非广播多接入网络（NBMA），支持虚链路，支持哑（stub）域和 NSSA，支持等开销多路径。

实现 OSPF 的三层交换机必须实现 OSPF MIB RFC1850。

9.2.3 中间系统到中间系统-双重 IS-IS

IS-IS 是基于链路状态（SPF）路由算法，拥有所有该类协议的优点。

三层交换机可选实现双重 IS-IS 三层交换机必须实现双重 IS-IS。

双重 IS-IS 在 RFC1142 和 RFC1195 中规定。

实现双重 IS-IS 的三层交换机必须实现双重 IS-IS MIB “draft-ietf-isis-wg-mib-03”。

9.2.4 路由信息协议

RIP 应用极其广泛，是自治域内路由协议的事实标准之一。

三层交换机应当实现 RIPv2。三层交换机对 RIPv2 的支持必须符合 RFC2453。

实现 RIPv2 的三层交换机必须实现 RIPv2 MIB RFC1724。

9.3 外部网关协议

9.3.1 概述

外部网关协议在自治系统间使用，为特定自治系统内一组网络与相邻自治系统交换可达性信息。

三层交换机可选实现 BGP4。

边缘网关协议（BGP4）是自治域间路由协议，是在 BGP 运行者之间交换网络可达性信息。网络信息包含流量到达某个网络所必须经过的完整 AS 列表。该信息必须确保路径内没有环路。该信息应足够丰富以用作构建 AS 互连图，在 AS 互连图中，必须裁减路由环回，必须被实施 AS 层的策略决定。

BGP4 在 RFC1771 中规定。

实现 BGP4 的三层交换机必须实现 BGP4 MIB RFC1657。

建议实现 BGP 的三层交换机遵从 RFC1772 第 6 章中的规定。

9.4 静态路由

静态路由提供一种途径来显示定义到一个特定目的地的下一跳设备。三层交换机应提供一种途径来定义到特定目的地的静态路由，其中目的地由网络前缀定义。该机制应允许对每一条静态路由指定度量(metric)。一个支持动态路由协议的三层交换机必须允许静态路由定义成任何路由协议使用的有效的度量。三层交换机必须允许用户规定一组静态路由是否通过路由协议扩散。另外如果三层交换机支持使用下列信息的路由协议，应在静态路由中支持这些附加信息。这些信息是：

— TOS；

— 子网掩码；或

— 前缀长度；或

— 对给定路由协议引入静态路由的特定度量。

9.5 路由信息的过滤

网络中每个三层交换机基于转发数据库中包含的信息作转发决定。在一个简单网络中数据库内的信息可以静态配置。当网络变得复杂时，动态更新转发数据库对网络有效运行至关重要。

如果要求通过网络的数据流尽可能地高效，则需要一种机制来控制那些三层交换机用作创建转发数据库的信息的传播。这种控制任务可以通过哪一个路由信息源可以信任，选择相信那一条消息的形式来实现。转发数据库是可用的路由消息经过过滤后的结果。

除有效性之外，控制路由消息传播可以通过阻止不正确或错误的路由信息的扩散而增加转发数据库的稳定性。

在某些情况下，本地策略可能要求不能广泛传播整个路由信息。

这些过滤器要求只用于非 SPF 协议（对不实现距离矢量协议的三层交换机没有影响）。

9.5.1 路由检验

当路由更新宣告中路由违反本标准的规定时，三层交换机应作为差错写入日志，除非接收的更新路由协议使用这些值编码那些特殊路由编码（例如缺省路由）。

9.5.2 基本路由过滤

过滤路由信息允许对三层交换机用作转发包的路径进行控制。三层交换机应可以配置从哪一个路由信息源接收路由消息，哪一条路由可以信任，因此三层交换机必须指定：

— 路由信息可以从哪个逻辑接口接收，从每个逻辑接口上可以接收哪些路由；和

— 在一个逻辑接口上传播所有路由或者只传播缺省路由。

某些路由协议不能将逻辑接口作为路由信息源，在这种情况下，三层交换机必须指定从哪一个相邻三层交换机可以接收路由信息。

10 MPLS 协议

三层交换机体系结构可选支持 MPLS，支持 MPLS LER 和 LSR 功能，支持 MPLS 显式路由 LSP，支持 CR-LDP，能配置备份 LSP，支持负荷分担的多路径 LSP，支持标记压栈，支持基于约束的路径计算，能基于源/目的地址、协议、源/目的端口、选项域、TOS/优先级（Precedence）域、TCP 标志，根据路由表的下一跳等参数将包路由至输出 LSP。

有关 MPLS 协议具体参见行标《多协议标记交换（MPLS）总体技术要求》。

11 排队策略和拥塞控制

11.1 排队策略

1）可选支持公平排队算法（Fair Queuing 或 Round Robin）。

2）可选支持加权公平排队算法（WFQ）。

11.2 拥塞控制

1）可选支持 WFQ、随机早期探测 RED、加权的随机早期探测 WRED 等拥塞控制机制。

2）在有可能存在输出队列争用的交换环境中，必须提供有效的方法消除头部拥塞。

12 组播协议

三层交换机建议实现组播协议，可选支持互联网组管理协议 IGMP v2（RFC2236），可选支持 IGMP Snooping，可选支持和协议无关组播协议--分散模式（PIM-SM），可选实现距离矢量组播路由协议 DVMRP（RFC1075）和组播源发现协议 MSDP。在实现 PIM-SM 协议时应考虑与距离矢量组播路由协议 DVMRP 的互通。域间组播实现可选 MBGP。

有关各组播路由协议具体参见行标《组播路由协议技术规范》。

13 性能指标要求

13.1 端口数量

设备拥有的端口的数量。

要求端口数量应当≥2，如 ATM 端口则端口数量应当≥1。

13.2 设备吞吐量

设备吞吐量指设备所有端口同时收发数据速率能力的总和。

本标准建议二层交换时吞吐量=∑端口速率（半双工），吞吐量=∑端口速率x2（全双工）。

本标准对三层交换的吞吐量本标准不作规范。

13.3 突发长度

突发（burst）指以最小合法帧间隔发送的一组帧，突发长度（burst size）指一定数量的突发帧。突发长度可以从 1 到无限。在全双工接口上无论是单向/双向流量，理论上突发长度没有限制。在半双工接口上双向流量的突发长度有限，因为接口发送序列可能被接收帧中断。

建议全双工端口上突发长度无限。半双工端口上单向流量的突发长度无限。半双工端口上双向流量突发长度不作规范。

13.4 突发间隔

突发间隔指突发之间的时间间隔。

建议突发间隔=最小帧间隔。

13.5 过负荷

过负荷指试图使被测设备以超出端口媒体限制的速度传输。过负荷可以通过缓存或拥塞控制方式实现。

本标准建议实现过负荷。

13.6 转发速率

在一定负荷下，被测设备可以观察到正确转发帧的速率。

本标准建议设备端口线速转发数据帧。

13.7 拥塞控制

任何用作为避免帧丢失，请求外部数据源降低发送速率以免拥塞端口的机制。

本标准对拥塞控制不作规范。

13.8 队头阻塞

队头阻塞是指由于输入端口试图向某一拥塞端口发送数据帧而导致该输入端口上目的地为不拥塞端口帧的丢失或附加时延。

本标准不强制实现避免队头阻塞，但建议生产厂家实现。

13.9 地址缓存能力

每个端口/模块/设备上能够缓存的 MAC 地址的能力。由于缓存的 MAC 地址才能使到达的帧不被丢

弃或广播。

本标准建议端口平均 MAC 地址缓存能力不低于 1024 个。

13.10 地址学习能力

交换机可以学习新的 MAC 地址（不用广播或丢弃数据帧）的速度。该指标用作衡量网络重启后地址表建立速度。

本标准建议端口地址学习能力>1000 个/s。

13.11 时延

对于存储转发设备，时延为被测设备收到最后一比特到发出第一比特的时间间隔。对于按比特转发设备，时延为被测设备收到第一比特到发出第一比特的时间间隔。本标准定义的时延为测试设备发出带时戳的测试帧到收到该帧的时间间隔。

本标准建议 64Byte 长的数据帧时延不超过 1ms。

13.12 时延抖动

时延抖动指时延变化。

由于交换机时延较小，本标准对时延抖动不作规范。

13.13 丢包率

丢包率是指交换机因资源不足引起的包丢失率。3 层交换机丢包率分二层交换丢包率和三层交换丢包率。

本标准建议交换机三层交换丢包率为 0（暂定）。

本标准建议交换机二层交换丢包率为 0（测试时间由测试规范定义）。

13.14 乱序

乱序指设备入口处有顺序的数据报序列在设备出口处的顺序情况。乱序的衡量方式待定。

本标准对乱序结果不作规范，只作为重要指标供比较。

13.15 错帧过滤

指设备收到错帧/非正常帧时正确处理的能力。

本标准建议系统实现错帧过滤能力。

13.16 路由表容量

路由表容量指交换机运行中可以容纳的路由数量。由于交换机设计使用在不同目的和应用环境，对路由表容量作范围限制没有意义。

本标准对路由表容量不作规范，只作为重要的性能指标供比较。

13.17 可靠性

此处可靠性指系统无故障运行时间。

本标准只对用于核心网的三层交换机作规范。建议系统的无故障工作时间 MTBF>17 520h。建议系统故障恢复时间<1h。建议主要部件热备份冗余。

13.18 VLAN 数量

VLAN 数量指系统支持的 VLAN 个数。

本标准建议三层交换机所支持的 VLAN 数量≥交换机端口数量。

14 运行与维护

14.1 定义

下面行为应包含在交换机的 O&M 中：

— 提供设备资源利用率；

— 提供网络接口带宽利用率；

— 诊断交换机的处理器、网络接口、相连的网络、调制解调器的硬件问题；

— 安装新硬件；

— 安装新软件；
— 在崩溃后重新启动或重新引导交换机；
— 配置（重新配置）交换机；
— 发现及诊断互联网问题，例如拥塞、环路、错误 MAC 地址等错误行为；
— 改变网络拓扑，暂时（例如绕过有问题的通信链路）或者永久；
— 监视交换机及相连网络的状态及性能；
— 为网络设计收集流量统计；
— 在恰当的厂商及电信规范中协调上述行为。

14.2 交换机初始化

14.2.1 最少交换机配置

交换机应当能在不配置任何参数条件下正确转发数据帧。

14.2.2 地址及前缀初始化

交换机可以允许静态配置 IP 地址，地址掩码或前缀长度，并存储在非-不稳定存储器中，用作管理。

如上文所述，交换机的 IP 地址中主机地址部分和网络前缀部分不允许是 0 或-1。所以交换机应当不允许将 IP 地址设置成上述形式。

交换机应当对设置的掩码实施下述检查：
— 掩码既不是全 0，也不是全 1（前缀长度不为 0 或 32）；
— 相应与地址网络前缀部分的比特为全 1；
— 对应于网络前缀部分的比特是连续的。

14.3 运行和维护具体规定

14.3.1 定义

在交换机上实施 O&M 功能有多个可用的模型：一个是仅在本地模型，该模型要求 O&M 功能只能在本地执行（例如，接在交换机上的终端）；一个是完全远程管理，在本地只允许作最少的操作（例如，强迫引导），大多数 O&M 从远端由 NOC 执行；另一个是中间模型，例如 NOC 人员可以登录到交换机上作为一个主机，使用 Telenet 协议执行本地也能执行的功能。本地模型一般在交换机安装时使用，交换机通常需要由 NOC 远端操作，所以交换机应实现远端操作。

远端 O&M 功能可以通过控制代理（程序）实现。在直接应用中，O&M 功能直接由 NOC 通过标准互联网协议实现（例如 SNMP、UDP、TCP）。在间接应用中，控制代理支持这些协议并控制交换机使用恰当的协议。建议使用直接应用的方式。

厂商应提供这样一种环境：用户使用控制代理或其他 NOC 软件应像在标准操作系统中编程一样。

交换机远程监视和远程控制存在重要的访问控制问题：一方面应确保应用这些功能时交换机资源的有效控制，例如交换机监视时必须不过分占用 CPU 资源；另一方面，O&M 功能必须具有相对高的优先级，因为交换机拥塞通常是最需要 O&M 操作的时候。

14.3.2 带外访问

交换机应当提供带外（Out-Of-Band OOB）访问。OOB 访问应当提供所有带内访问的功能。带外访问应当实现访问控制，防止非授权访问。

14.3.3 交换机 O&M 功能

1）维护-硬件诊断

在本地硬件维护时，每个交换机应当像一个独立设备一样被操作，在交换机端应当提供运行诊断程序需要的方法。交换机应当能在出错时运行诊断程序。

2）控制-配置交换机

每台交换机都可能有需要配置的参数。交换机参数更新后应当不需要重新启动；最坏情况下需要重新运行。可能存在某些情况，改变参数后必须重启交换机（例如改变 IP 地址）。这些情况下，必须小心将对交换机和周边网络带来的影响减少到最小。

3）对错误配置的检查与反应

必须实现一种机制检查错误配置并做出反应。如果命令不正确运行，交换机应当给出错误消息。交换机不应接受错误格式的命令，即使该命令本身是正确的。

另一种错误是对交换机连接网络的错误配置。交换机可以实现检查网络的误配置。交换机可以将发现的错误记录到日志或者网络上其他交换机或主机，管理员能看到可能存在的问题。

14.4 安全性考虑

14.4.1 审计与审计记录

1）配置改变

交换机应当提供一种方法来记录配置的改变，指示记录操作人员改变配置的时间。

2）安全性审计

交换机必须提供一种机制审计与安全性相关的失败与冲突。

— 授权失败：错误口令，无效的 SNMP 通信，非法的授权令牌。

— 对控制策略的违反：被过滤掉的目的地。

— 授权通过：正确口令，远程登录带内访问，配置口访问。

交换机必须提供一种机制来限制或禁止这样的审计，但缺省情况下审计应当存在。审计可能的方法包括在如果存在的控制口列出冲突，计数或者写入日志，通过 SNMP trap 机制送到远程安全服务器，或者使用 UNIX 的日志机制。交换机必须实现至少一种上述方法，可以实现多种方法。

14.4.2 配置控制

在为交换机生产软件/固件时，厂商应负责良好的配置控制。

如果厂商提供用户远程改变交换机配置的能力，例如通过远程登录；这种能力应当是可配置的，缺省情况应当是不允许远程配置。在允许远程配置前，交换机应当要求有效的授权。这种授权不应当在网络上传输明文。例如：如果实现远程登录，厂商应当实现 Kerberos，S-Key，或者其他类似授权机制。

交换机不允许存在未记载于文档的访问后门，或通用密码。厂商必须确保这样的用于调试或者开发的访问途径在产品分销到客户手中之前已删除。

15 网络管理协议

15.1 简单网络管理协议-SNMP

15.1.1 SNMP 协议元素

交换机必须支持 RFC1902 至 RFC1906 中规定的 SNMP v2。

SNMP 必须使用 UDP/IP 作为传输层/网络层协议。也可以使用其他协议（例如，RFC1418，RFC1089）。

SNMP 管理请求向交换机任何一个接口发出时，该操作必须生效。实际的管理动作应由交换机或交换机的代理完成。

支持 SNMP v2 协议的交换机必须实现 SNMP v2 MIB RFC1907。

交换机必须实现所有的 SNMP 操作。

交换机必须提供一种机制来限制 SNMP 陷阱（trap）消息的产生速率。交换机可以通过 RFC1224 中描述的异步告警管理算法来实现上述机制。

15.1.2 团体表格

为本标准描述方便，假设交换机中存在一个抽象的团体表格。该表格包含多个条目，每个条目给一个特定区域，包含完全定义该区域属性需要的参数。对抽象团体表格的实现方法在本标准范围之外，由实现者决定。

交换机的团体表格必须至少包含一个条目，建议至少包含两个条目。

交换机必须允许用户手工（即不使用 SNMP）检查、增、删、改 SNMP 团体表格中的条目。用户必须能够设置区域名，或者构造 MIB 视图。用户必须能以只读（即不允许 SET）或者读写（允许 SET）的方式配置区域。

用户必须能定义至少一个 IP 地址，当使用 trap 时，对每个区域或 MIB 视图的通知将送到该 IP 地址。这些 IP 地址应当被定义在区域或 MIB 视图库内。允许或不允许在区域或 MIB 视图库上发通知应当是可配置的。

交换机应当提供为特定区域提供有效管理员列表的能力。如果提供上述列表，交换机必须验证 SNMP 数据报源地址的有效性；如果该地址没有在上述列表中出现则必须丢弃该数据报。如果数据报被丢弃，交换机必须作 SNMP 授权失败时相应动作。

团体表必须存储在非-不稳定的存储器内。

团体表的初始状态应当包含一个条目，其中区域名串为 Public，访问权限为只读。该条目的缺省状态不允许发送 trap。如果实现，该条目必须保存在团体表中，直到管理员改变或者删除。

15.1.3 标准 MIBS

所有关于交换机配置的 MIB 都应实现：

- MIB-Ⅱ RFC1213 中的系统、接口、IP、ICMP 和 UDP 组必须实现；
- 接口扩展 MIB RFC1229 必须实现；
- 如果交换机实现 TCP（例如，远程登录），MIB-Ⅱ，RFC1213 中的 TCP 组必须实现；
- 以太网-链路 MIB RFC1643 必须实现；
- 网桥 MIB RFC1493 必须实现；
- 远程网络监视 MIB RFC1757 必须实现接口、IP、ICMP 和 UDP 组；
- 远程网络监视 MIB 对交换网络的扩展 V1，RFC2613 可选支持。

15.1.4 厂商指定的 MIBS

互联网标准和根据实验的 MIB 不能完全覆盖网络元素统计、状态、配置和控制信息。交换机（或其他网络设备）厂商通常自己开发覆盖上述信息的 MIB 扩展。这些 MIB 扩展称为厂商特定的 MIB。

交换机上厂商特定的 MIB 必须提供一种方法来存取所有实现的统计、状态、配置和控制信息，这些信息不能由标准或实验得到的 MIB 得到。这些信息必须能被控制和监视操作使用。

厂商应当使所有厂商特定的 MIB 变量可用。这些指定必须符合 RFC1155，并且必须以 RFC1212 指定的方式描述。

16 环境要求

16.1 环境要求

16.1.1 交换机正常工作的温度、湿度条件：

a）长期工作条件：温度保持 15~30℃、相对湿度保持 40%~65%

b）短期工作条件：温度保持 0~40℃、相对湿度保持 20%~90%

注：

1）交换机的正常工作的温度和相对湿度的测量点指在地板以上 2m 和交换机前方 0.4m 处测量值。

2）短期工作条件是指连续不超过 48h 和每年累计不超过 15 天。

3）相对湿度低于 20%的环境应采用抗静电地面。

16.1.2 交换机的防尘要求

机房内灰尘粒子应是非导电，非导磁和非腐蚀性的。

16.2 防电磁干扰要求

交换机产生的电磁干扰应满足以下要求。

a）由交换机射出的无线电电磁干扰应符合表 5 的规定。

表 5 无线电电磁干扰要求

频率，MHz	电磁强度，dB（μV/m）	频率，MHz	电磁强度，dB（μV/m）
0.01~0.024	148.6–60lg$d$	47.7/d~88	59.1–20lg$d$
0.024~0.8	116.2–60lg$d$–20lg$f$	88~216	63.6–20lg$d$
0.8~1.59	118.2–60lg$d$	2160~10000	66.6–20lg$d$
1.59~47.7/d	120.2–60lg$d$–40lg$f$		

注：
1 d 为测试天线与靠近被测物间水平距离，单位为 m，d 限于 30m 内。
2 f 为频率，以 MHz 为单位。
3 dB（μV/m）表示微伏（μV/m）为参考单元的分贝数。

b）由交换机进入交流馈电线的无线电电磁干扰应符合表 6 的规定。

表 6 进入交流馈电线的无线电电磁干扰

频率，MHz	最大线路电流，dB（μA）
0.000061~0.001	I–20lg$f$–84.4
0.001~0.01	（124.4–I）lg$f$+348.8–2I
0.01~0.8	–21.05lg$f$+57.9
0.8~100	60

注：
1 f 为频率，以 MHz 为单位。
2 I 为接入到交流电源处的输入线路电流电平。
3 dB（μV/m）表示微伏（μV/m）为参考单元的分贝数。

c）由交换机进入直流馈线和信号线的无线电电磁干扰应符合表 7 的规定。

表 7 进入直流馈线和信号线的无线电电磁干扰

频率，MHz	最大线路电流，dB（μA）
0.01~0.8	–21.05lg$f$+57.9
0.8~100	60

16.3 交换机抗电磁干扰的能力

交换机在受到 0.01~1000MHz 频率范围内电场强度为 140dB（μV/m）的外界电磁干扰时应不出现故障和性能下降。

在直流或交流电源线受到表 8 所示，0.01~100MHz 频率范围的外界电磁干扰电流时应不出现故障和性能下降。

表 8

频率，MHz	最大线路电流，dB（μA）
0.01~0.8	–21.05lg$f$+67.9
0.8~100	70

16.4 交换机防雷击能力

交换机设备防雷击能力应当符合 GB3483—83《电子设备雷击试验导则》。

17 电源与接地

17.1 电源

a）直流电压及其波动范围要求

额定电压：为-48V 的直流电源。

电压波动范围：在直流输入端子处测量-48V 电压允许变动范围为-57~-40V。交换机在此范围内应工作正常。

b）杂音电压指标

在直流配电盘输出端子处测量的限值如下：

300~3400Hz，杂音电压≤2mV；

0~300Hz，峰峰值杂音电压≤400mV；

3.4~15kHz，宽带杂音电压≤100mV 有效值；

150kHz~30MHz，宽带杂音电压≤30mV 有效值。

c）离散频率杂音电压

3.4~15kHz，≤5mV 有效值；

150~200kHz，≤3mV 有效值；

200~500kHz，≤2mV 有效值；

500kHz~2MHz，≤1mV 有效值。

d）交流电压及其波动范围要求

单相 220V±10%，频率 50Hz±5%。

线电压波形畸变率<5%。

17.2 交换机接地要求

a）接地方式应符合工作地、保护地和建筑防雷接地公用一组接地体的联合接地方式。

b）接地线截面积

接地线截面积根据可能通过的最大电流负荷确定。应采用良导体导线，不能使用裸导线布放。

接地电阻值：联合接地的电阻值应<5Ω。

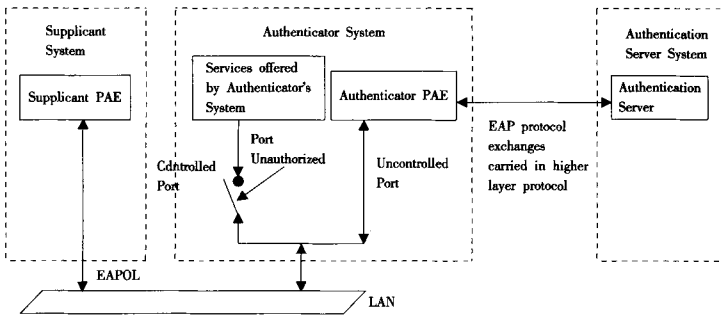
附录 A
(规范性附录)
802.1X

A.1 802.1X 定义

802.1X 定义了基于端口的网络接入控制协议，该协议适用于用户接入设备与接入端口间点到点的连接方式，实现对局域网用户接入的认证与服务管理。 其中端口可以是物理端口，也可以是逻辑端口。

A.2 802.1X 实现要求

A.2.1 802.1X 体系结构



802.1X 的体系结构中包括 3 个部分：Supplicant System，用户接入设备；Authenticator System，接入控制单元；Authentication Sever System，认证服务器。

在交换机中需要实现 802.1X 的接入控制单元部分，即 Authenticator System；用户接入设备实现在用户终端，如用户 PC 中；认证服务器是需要具有 AAA 功能的服务器。接入控制单元根据用户接入设备的认证状态控制物理接入，在用户接入设备没有认证通过时，接入控制单元限制用户接入设备对网络的使用，用户接入设备发起认证，接入控制单元把认证信息发送到认证服务器，认证服务器对用户接入设备进行认证并把用户认证的结果送回接入控制设备，决定是否允许用户认证设备对网络的使用。

Supplicant 与 Authenticator 间运行 EAPOL 协议；Authenticator 与 Authentication Sever 间协议应该支持两种方式：一是运行 EAP 协议，EAP 帧中封装认证数据，该协议承载在其他高层次协议中，如 Radius；二是直接在 Authenticator 终结 EAP 协议，采用 Radius 等高层协议传送认证数据。

A.2.2 802.1X 的端口概念

Authenticator 内部有受控端口（Controlled Port）和非受控端口（Uncontrolled Port）。非受控端口始终处于双向连通状态，主要用来传递 EAPOL 协议帧，可保证 Supplicant 始终可以发出或接受认证。受控端口只有在认证通过的状态下才打开，用于传递网络资源和服务。受控端口可配置为双向受控、仅输入受控两种方式，以适应不同的应用环境。

A.2.3 受控端口类型

受控端口是一个逻辑上的概念，实现方式有下列 3 种：

- 物理端口：一个逻辑端口对应一个交换机的物理端口。
- 物理端口和源 MAC 的组合：将交换机物理端口和用户终端的源地址组合对应一个逻辑端口，当有客户端发送 EAPOL-Start 请求认证报文时交换机提取客户端源 MAC 地址，和物理端口组合做为受控端口的标识。客户端注销时对应的受控端口资源被释放。
- 物理端口和 VLANID、源 MAC 的组合：将交换机物理端口和用户终端的 VLANID、源地址组合对应一个逻辑端口，当有客户端发送 EAPOL-Start 请求认证报文时交换机提取客户端源 MAC 地址/VLANID 和物理端口组合做为受控端口的标识。客户端注销时对应的受控端口资源被释放。

A.2.4 用户接入设备状态维护

为了维护用户接入设备的状态，接入控制设备需要定时同用户接入设备握手：利用客户端的 MAC 地址作为 EAPoL 报文的单播目的地址，接入控制设备定时发送格式为 EAPoL-Request/Identity 的握手报文，接收客户端 EAPoL-Response/Identity 的握手响应报文，重新进行一次认证过程，可以避免客户终端异常关机造成的用户上网时间统计不准确和客户端仿冒的问题。定时发送握手报文的时间间隔应<15s。

附录 B
（资料性附录）
受控组播

B.1 定义

组播业务的顺利开展依赖于有效的业务管理、监控及计费。受控组播解决组播业务运营必须的用户管理、业务管理和计费等方面的问题。受控组播（(Managable Multicast）可管理可运营的组播）是建立在组播技术基础上的可管理可运营的组播。它的主要功能是对组播业务接收者实现组播认证和组播计费，并能抑制非法组播流进网络，防止用户非法开设组播业务。

受控组播系统由用户侧主机、业务节点（局侧交换机）、业务控制平台等 3 个功能模块组成。用户主机发起加入组过程；业务节点的组播协议处理模块维护组播转发项；业务节点的认证或认证代理模块完成组播认证代理和发起组播计费；业务控制平台则完成对用户的认证鉴权与用户计费的功能。

B.2 技术要求

B.2.1 组播认证

对于组播认证，业务节点（局侧交换机）应首先具备标记用户接入数据流端口的能力。即业务节点（局侧交换机）能够检查用户与端口的对应关系。其次，业务节点（局侧交换机）应能对用户接入进行认证或代理认证。即业务节点（局侧交换机）能够在用户接入时对用户授权，允许用户访问网络。

当用户欲加入组时，它首先发送 IGMP 报告。局侧交换机收到 IGMP 报告后，并不立即将收到报告的端口添加到转发出端口列表中，而是先根据用户信息对该报告进行认证。然后，局侧交换机根据认证的结果决定是否将收到报告的端口添加到出端口列表中。如果认证成功，则局侧交换机将端口添加到出端口列表中；如果失败，则默默丢弃该 IGMP 报告或作记录。

B.2.2 组播计费

当组播认证成功后即可开始计费。当接收者发送 IGMP 离开消息离开组时，业务节点（局侧交换机）将端口从出端口列表中删除，并停止计费。如果接收者默默离开或者业务节点（局侧交换机）长时间没有收到接收者的 IGMP 报告，业务节点（局侧交换机）也将用户端口从出端口列表中删除，并停止计费。对于预付费的用户，业务节点（局侧交换机）应能在费用结束或不足时，强制拆除连接，将用户端口从出端口列表中删除。

B.2.2.1 组播认证与计费流程

组播认证与计费需要经过以下步骤。

组播认证需要经过以下过程：

用户访问网络并被授权允许接入；

用户欲访问组，发送 IGMP 组报告报文；

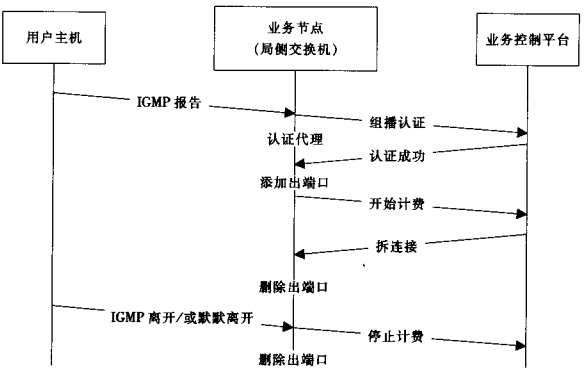
业务节点（局侧交换机）对 IGMP 报告作认证代理，向用户业务控制平台发送组播认证请求；

用户业务控制平台返回认证结果。

如果认证成功，业务节点（局侧交换机）添加收到 IGMP 报告的端口为出端口；如果认证失败，业务节点（局侧交换机）默默丢弃 IGMP 报告或作记录。

组播计费需要经过以下过程：

业务节点（局侧交换机）添加出端口成功，则向用户业务控制平台发送开始计费请求。



组播用户离开需要以下过程：

用户向业务节点（局侧交换机）发送 IGMP 离开组；

业务节点（局侧交换机）删除出端口，并向用户业务控制平台发送停止计费请求。

如果用户默默离开或业务节点（局侧交换机）长时间收不到用户的 IGMP 报告，业务节点（局侧交换机）删除出端口，并向用户业务控制平台发送停止计费请求。

如果预付费用户费用不足或用完时，用户业务控制平台通知业务节点（局侧交换机）强制拆连接。

业务节点（局侧交换机）删除出端口。

B.2.2.2 组播源管理

由于目前的组播源软件无需授权就能向网络发送组播流（就是目的 IP 地址在 224.0.1.0–239.255.255.255 之间的 IP 报文）。因此，对非法组播流的抑制可以在交换机或路由器上设置过滤规则来完成。初始情况下，交换机/路由器不接收任何组播流；然后针对特定的源允许特定的组播流进入网络。对特定源的特定组播流的授权可以采用手工配置的方式，也可以动态认证授权的方式。