

# 中华人民共和国通信行业标准

YD/T 1698-2007

---

## IPv6 网络设备技术要求 ——具有 IPv6 路由功能的以太网交换机

Technical Specification for IPv6 Network Equipment  
——Ethernet Switch with IPv6 Routing Capability

2007-09-29 发布

2008-01-01 实施

---

中华人民共和国信息产业部 发布

## 目 次

|                         |    |
|-------------------------|----|
| 前 言                     | II |
| 1 范围                    | 1  |
| 2 规范性引用文件               | 1  |
| 3 术语、定义和缩略语             | 3  |
| 4 具有 IPv6 路由功能的以太网交换机功能 | 5  |
| 5 具有 IPv6 路由功能的以太网交换机接口 | 6  |
| 6 链路层功能                 | 7  |
| 7 网络层协议要求               | 20 |
| 8 传输层协议要求               | 28 |
| 9 路由协议                  | 28 |
| 10 MPLS 协议              | 31 |
| 11 QoS 控制机制             | 32 |
| 12 组播协议                 | 32 |
| 13 安全功能要求               | 32 |
| 14 性能指标要求               | 34 |
| 15 运行与维护                | 36 |
| 16 网络管理协议               | 38 |
| 17 环境要求                 | 39 |
| 附录 A（规范性附录） 802.1X      | 40 |

## 前 言

本标准是“IPv6 网络设备”系列标准之一。该系列标准预计的结构及名称如下：

1. YD/T 1452-2006 IPv6 网络设备技术要求——支持 IPv6 的边缘路由器；
2. YD/T 1453-2006 IPv6 网络设备测试方法——支持 IPv6 的边缘路由器；
3. YD/T 1454-2006 IPv6 网络设备技术要求——支持 IPv6 的核心路由器；
4. YD/T 1455-2006 IPv6 网络设备测试方法——支持 IPv6 的核心路由器；
5. YD/T 1698-2007 IPv6 网络设备技术要求——具有 IPv6 路由功能的以太网交换机；
6. IPv6 网络设备测试方法——具有 IPv6 路由功能的以太网交换机。

本标准是《IPv6 网络设备测试方法——具有 IPv6 路由功能的以太网交换机》的配套标准文件，为其提供技术依据。

附录 A 为规范性附录。

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：信息产业部电信研究院

本标准主要起草人：马军锋、魏 亮、赵 锋

# IPv6 网络设备技术要求

## ——具有 IPv6 路由功能的以太网交换机

### 1 范围

本标准规定了在公网中使用的具有 IPv6 路由功能的以太网交换机的技术要求，包括功能要求、通信接口、通信协议、性能指标、安全功能要求以及环境要求等。具有 IPv6 路由功能的以太网交换机是指支持 IPv6 协议、具有路由学习能力和第三层（网络层）包交换能力的 IP 分组交换机。

本标准在指定支持 IPv6 具有路由功能的以太网交换机必须实现的协议时不重复应用协议的内容，只规定协议中必须实现的内容、可选的内容、不需实现的内容，对同类协议作选择。

本标准适用于支持 IPv6 协议的具有路由功能的以太网交换机。

### 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。

|                      |  |
|----------------------|--|
| GB/T 2423.1-2001     | 电工电子产品环境试验 第2部分：试验方法 试验A：低温                      |
| GB/T 2423.2-2001     | 电工电子产品环境试验 第2部分：试验方法 试验B：高温                      |
| GB/T 2423.3-2006     | 电工电子产品环境试验 第2部分：试验方法 试验Cab：恒定湿热试验                |
| GB/T 2423.4          | 电工电子产品环境试验规程 试验Db：交变湿热试验方法                       |
| GB/T 2423.9-2001     | 电工电子产品环境试验 第2部分：试验方法 试验Cb：设备用恒定湿热                |
| YD/T 900-1997        | SDH设备技术要求——时钟                                    |
| YD/T 1061-2003       | 同步数字体系（SDH）上传送 IP 的 LAPS 技术要求                    |
| YD/T1162.1-2005      | 多协议标记交换（MPLS）技术要求                                |
| YD/T 1260-2003       | 基于端口的虚拟局域网（VLAN）技术要求和测试方法                        |
| YD/T 1295-2003       | 支持 IPv6 的路由协议技术要求——开放最短路径优先协议（OSPF）              |
| YD/T 1342-2005       | IPv6 路由协议——支持 IPv6 的边界网关协议（BGP4）                 |
| YDN 099-1998         | 光同步传送网技术体制（暂行规定）                                 |
| ISO/IEC 8802-2       | 信息技术——系统间远程通信和信息交换——局域网和城域网特殊要求<br>第2部分：逻辑链路控制技术 |
| ISO/IEC 15802-3-1998 | 信息技术——系统间远程通信和信息交换——局域和城域网通用规范<br>第3部分：媒体访问控制网桥  |
| ITU-T G.707（1996）    | 同步数字体系（SDH）网络节点接口                                |
| IETF RFC768（1980）    | 用户数据包协议  |
| IETF RFC793（1981）    | 传输控制协议   |
| IETF RFC1075（1988）   | 距离矢量组播路由协议                                       |

|                     |                                   |
|---------------------|-----------------------------------|
| IETF RFC1089 (1989) | 以太网承载 SNMP                        |
| IETF RFC1122 (1989) | 互联网主机要求——通信层                      |
| IETF RFC1142 (1990) | IS-IS 域内路由协议                      |
| IETF RFC1155 (1990) | 基于 TCP/IP 网络的管理信息结构和标识            |
| IETF RFC1195 (1990) | 在 TCP/IP 和双重环境路由中使用 OSI 的 IS-IS   |
| IETF RFC1212 (1991) | 简明管理信息库 (MIB) 定义                  |
| IETF RFC1213 (1991) | 基于 TCP/IP 的互联网网络管理的管理信息库: MIB-II  |
| IETF RFC1224 (1991) | 管理异步产生告警的技术                       |
| IETF RFC1229 (1991) | 通用接口 MIB 扩展                       |
| IETF RFC1418 (1993) | OSI 承载 SNMP                       |
| IETF RFC1493 (1993) | 网桥管理对象定义                          |
| IETF RFC1643 (1994) | 对以太网接口类型管理对象的定义                   |
| IETF RFC1657 (1994) | 使用 SMIv2 定义 BGP4 管理对象             |
| IETF RFC1757 (1995) | 远程网络监控管理信息库                       |
| IETF RFC1771 (1995) | 边界网关路由协议 BGPv4                    |
| IETF RFC1772 (1995) | 边界网关协议 (BGP) 在互联网中的应用             |
| IETF RFC1902 (1996) | SNMPv2 管理信息结构                     |
| IETF RFC1903 (1996) | SNMPv2 文本约定                       |
| IETF RFC1904 (1996) | SNMPv2 一致性声明                      |
| IETF RFC1905 (1996) | SNMPv2 协议操作                       |
| IETF RFC1906 (1996) | SNMPv2 传输映射                       |
| IETF RFC1907 (1996) | SNMPv2 管理信息库                      |
| IETF RFC1966 (1996) | BGP 路由反射                          |
| IETF RFC1981 (1996) | IPv6 路径 MTU 发现                    |
| IETF RFC1997 (1996) | BGP 团体属性                          |
| IETF RFC2080 (1997) | RIPng 路由信息协议                      |
| IETF RFC2283 (1998) | BGP4 多协议扩展                        |
| IETF RFC2439 (1998) | BGP 路由振荡抑制                        |
| IETF RFC2460 (1998) | 互联网 IPv6 协议                       |
| IETF RFC2613 (1999) | 对交换网络的远程网络监视管理 MIB 扩展 Version 1.0 |
| IETF RFC2675 (1999) | IPv6 的超长包                         |
| IETF RFC2710 (1999) | IPv6 组播监听者发现协议                    |
| IETF RFC2711 (1999) | IPv6 路由器告警选项                      |
| IETF RFC2740 (1999) | 支持 IPv6 的 OSPF 路由协议               |
| IETF RFC2858 (2000) | BGP4 多协议扩展                        |
| IETF RFC3414 (2002) | SNMPv3 基于用户的安全模型                  |
| IETF RFC3590 (2003) | 组播监听协议源地址选择                       |

|                              |  |
|------------------------------|--|
| IETF RFC3810 (2004)          | IPv6 组播监听发现协议 (MLDv2)  |
| IETF RFC4444 (2006)          | ISIS 管理信息库   |
| IETF RFC4541 (2006)          | IGMP 和 MLD 组播侦听  |
| IETF draft-ietf-isis-ipv6-06 | 用于 IPv6 的 ISIS   |
| draft-ietf-ospfv3-mib-10     | OSPFv3 管理信息库   |
| IEEE 802.1D (2004)           | 媒体访问控制 (MAC) 网桥  |
| IEEE 802.1G (1995)           | 远程媒体访问控制桥接   |
| IEEE 802.1Q (2003)           | 虚拟桥接局域网  |
| IEEE 802.3 (2005)            | 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                                      |
| IEEE 802.3ac (2002)          | 局域网和城域网第 3 部分: 带冲突检测载波监听多路访问和物理层规范<br>——万兆以太网媒体访问控制参数、物理层和管理参数 |
| IEEE 802.3x (1997)           | 802.3 全双工操作规范  |

### 3 术语、定义和缩略语

#### 3.1 术语和定义

下列术语和定义适用于本标准。

##### 1) 网桥 (Bridge)

网桥工作在 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 IEEE 802.1D (2004)。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的相应功能。

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

具有第三层路由功能的 IP 数据包交换机。除实现数据帧转发功能外, 能根据接收数据包中的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跳交换机地址或主机地址, 并且重写链路层数据包头。

路由表可以通过静态配置方式维护, 也可以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 (如路由器) 的交换路由信息来完成路由表的动态维护。

注: 如果文中没有特别说明, 那么交换机就特指此类具有路由功能的以太网交换机。

##### 4) 虚拟局域网 (Virtual Local Area Network, VLAN)

是指通过划分不同域来分隔局域网内的广播域。各个 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。

##### 5) 远程桥接 (Remote MAC Bridging)

是指在互联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。

## 6) 链路聚合 (Link Aggregation)

是指在逻辑上将多条独立的链路捆绑成一条单独链路使用，以此获得灵活的高带宽以及链路冗余。

## 7) 流量整形 (Traffic shaping)

是一种用来在接口上平缓通信、避免链路拥塞并满足服务提供商要求的机制。流量整形通过对超过平均速率的分组排队或保存到缓冲区，将突发通信平缓，以满足配置的承诺接入速率。

## 3.2 缩略语

下列缩略语适用于本标准。

|          |  |                |
|----------|--|----------------|
| AFC      | Asymmetric Flow Control                      | 不对称流量控制        |
| AS       | Autonomous System                            | 自治系统           |
| AUI      | Attachment Unit Interface                    | 附加单元接口         |
| BPDU     | Bridge Protocol Data Unit                    | 桥接协议数据单元       |
| BGP      | Border Gateway Protocol                      | 边界网关路由协议       |
| CRC      | Cyclic Redundancy Check                      | 循环冗余校验         |
| DTE      | Data Terminal Equipment                      | 数据终端设备         |
| DES      | Data Encryption Standard                     | 数据加密标准         |
| EAP      | Extensible Authentication Protocol           | 可扩展认证协议        |
| E-ISS    | Enhanced Internal Sublayer Service           | 增强的内部子层服务      |
| EAPOL    | EAP over LANs                                | 在局域网之上的可扩展认证协议 |
| FCS      | Frame Check Sequence                         | 帧检验序列          |
| 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      |
| GMRP     | GARP Multicast Registration Protocol         | GARP 组播注册协议    |
| GVRP     | GARP VLAN Registration Protocol              | GARP VLAN 注册协议 |
| ICMPv6   | Internet Control Messages Protocol Version 6 | 网间控制报文协议版本 6   |
| IS-IS    | Intermediate System-Intermediate System      | 中间系统—中间系统      |
| IPv6     | Internet Protocol Version 6                  | 网际互连协议版本 6     |
| IPsec    | Internet Protocol Security                   | 互联网安全          |
| LAN      | Local Area Network                           | 局域网            |
| LLC      | Logical Link Control                         | 逻辑链路控制         |
| LAIS     | Line Alarm Indication Signal                 | 线警报指示信号        |
| LOF      | Loss of Frame                                | 帧丢失            |
| LOP      | Loss of Pointer                              | 指针丢失           |
| LOS      | Loss of Signal                               | 信号丢失           |
| MAC      | Media Access Control                         | 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     |
| MPLS   | Multi-Protocol Label Switch                  | 多协议标记交换    |
| MLD    | Multicast Listener Discovery                 | 组播监听者发现协议  |
| MSDU   | MAC Service Data Unit                        | MAC 服务数据单元 |
| MSTP   | Multiple Spanning Tree Protocol              | 多生成树协议     |
| MTU    | Maximum Transmission Unit                    | 最大传输单元     |
| NDP    | Neighbor Discovery Protocol                  | 邻居发现协议     |
| NOC    | Network Operation Center                     | 网络运行中心     |
| PHY    | Physical Layer Device                        | 物理层设备      |
| PAIS   | Path Alarm Indication Signal                 | 路径警报指示信号   |
| OSPFv3 | Open Shortest Path First Version 3           | 最短路径优先     |
| 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     |
| QoS    | Quality of Service                           | 服务质量       |
| RADIUS | Remote Authentication Dial In User Service   | 远程认证拨号用户服务 |
| RIF    | Routing Information Field                    | 路由信息域      |
| RIPng  | Routing Information Protocol Next Generation | 下一代路由信息协议  |
| RSTP   | Rapid Spanning Tree Protocol                 | 快速生成树协议    |
| SDH    | Synchronous Digital Hierarchy                | 同步数字系列     |
| SONET  | Synchronous optical network                  | 同步光网络      |
| SNMP   | Simple Network Management Protocol           | 简单网络管理协议   |
| SD     | Signal Degrade                               | 信号劣化       |
| SF     | Signal Fail                                  | 信号失效       |
| STM    | Synchronous Transport Module                 | 同步传送模块     |
| STP    | Spanning Tree Protocol                       | 生成树协议      |
| TCP    | Transfer Control Protocol                    | 传输控制协议     |
| UDP    | User Datagram Protocol                       | 用户数据报协议    |
| VID    | Virtual LAN Identifier                       | 虚拟局域网标识符   |
| VLAN   | Virtual LAN                                  | 虚拟局域网      |

#### 4 具有 IPv6 路由功能的以太网交换机功能

具有 IPv6 路由功能的以太网交换机必须实现以下功能：

##### 1) 接口功能

交换机必须至少支持下述以太网接口中的一类接口：10/100Mbit/s、1000Mbit/s 或 10Gbit/s；可以支持 PoS 接口，也可以包含 RS232 串口作为管理接口；各种接口必须符合相应的规范。



## 2) 逻辑链路层功能

支持以太网接口的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之间交换链路层用户数据帧。交换机必须实现数据帧的转发。支持以太网接口的交换机转发数据帧应当实现 IEEE802.1D 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或者是按照控制策略，对某些端口上的数据帧不转发（丢弃）到其他端口的行为。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路由转发表内容在各端口（包括逻辑端口）间转发数据包。

## 6) 路由信息维护功能

该功能负责运行路由协议或者提供静态的路由配置模式，生成和维护路由表。路由协议可以包括 Ripng、OSPFv3、IS-ISv6、BGP4+ 以及组播等。

## 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

交换机必须提供网络管理接口并且实现相应的网管协议。

## 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 具有 IPv6 路由功能的以太网交换机接口

## 5.1 10/100Mbit/s 以太网接口

交换机支持 10M/100Mbit/s 自适应以太网接口应符合 IEEE802.3（2005），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 IEEE802.3。100Base-T 技术中可采用三类传输介质：100Base-T4、100Base-TX 和 100Base-FX。采用 4B/5B 编码方式。

## 5.2 千兆以太网接口

交换机支持千兆以太网接口应符合 IEEE802.3。

1000Mbit/s 以太网物理接口类型包括 1000Base-SX、1000Base-LX 以及 1000Base-T。

交换机若支持千兆接口，建议实现 1000Base-SX 和 1000Base-LX 两者之一。

## 5.3 万兆以太网接口

交换机万兆以太网接口应符合 IEEE802.3ae。

10Gbit/s 以太网物理接口类型包括 10GBase-X、10GBase-R、10GBase-W。10GBase-X 使用 8B/10B 编码格式；10GBase-R 使用 64B/66B 编码格式；10GBase-W 是广域网接口，也使用 64B/66B 编码格式，与 SONET OC-192 兼容。

## 5.4 POS 接口（可选）

### 5.4.1 接口类型

交换机应可选支持STM-1、STM-4、STM-16、STM-64PoS接口。

### 5.4.2 SDH 层要求

- 应符合 YDN 099-1998《光同步传送网技术体制（暂行规定）》和 ITU-T G.707；
- 应支持以下告警处理功能：LOS、LOF、LOP、SF、SD；
- 应支持性能监控：支持 B1，B2，B3 差错计数；
- 应支持本地（内部）或从 STM-N 接口提取定时（从网络恢复时钟），精度要求见 YD/T900-1997《SDH 设备技术要求——时钟》。
- 支持本地环回（诊断）和网络环回功能；
- IP over SDH 的帧结构和协议规程应符合 YD/T1061-2003《同步数字体系(SDH)上传送 IP 的 LAPS 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，每个端口至少关联单一流量类。
- 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；
- 丢弃符合过滤信息的帧；
- 丢弃传输服务用户数据单元大小超过 ISO/IEC 15802-3-1998 第 6.3.8 节中规定的帧；
- 发送所收到的到其他端口的帧；
- 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，第 6.3.9 节）；
- 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

接收帧的 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 | 缺省的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的任意值。

### 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。相应地，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

当且仅当下列条件都满足时，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之一作过滤决定：

- 接收帧目的 MAC 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。用户优先级可以为 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<br>优<br>先<br>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流量中的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；（可选）
- 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 IEEE 802.1D 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为配置 BPDU 和拓扑变化通知 BPDU，分别见表 3 和表 4。

表 3 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 |
|          | 35 |

- 表 3 中配置 BPDU 的结构如下：
- 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。

表 4 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 |

表 4 中 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。
- BPDU 4 字节且类型为拓扑变化通知 BPDU。

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

交换机应当实现 GARP。GARP 在 IEEE 802.1D 中定义。

6.2.3.1 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。

6.2.3.2 GARP 实现要求

GARP 及相关算法用作在桥接的网络中建立、维护 and 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, 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。

### 6.2.3.3 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, 见表 5。GARP 应用地址中 01-80-C2-00-00-20 用作 GMRP 地址, 01-80-C2-00-00-22 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端口转发。

表 5 GARP 应用地址

| 分 配     | 值                 |
|---------|-------------------|
| GMRP 地址 | 01-80-c2-00-00-20 |
| GVRP    | 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。



#### 6.2.4 GMRP 组播注册协议

交换机可选实现 GMRP。GMRP 在 IEEE 802.1D 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 的一致性要求如下：

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 IEEE 802.1D 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。

#### 6.2.5 GVRP 注册协议（GVRP – GARP Vlan Registration Protocol）

交换机可选实现 GVRP。GVRP 在 IEEE 802.1Q（2003）中定义

GARP VLAN 注册协议定义了一个提供 VLAN 注册服务的 GARP 应用。GVRP 使用 GARP 信息声明（GID）和 GARP 信息传播（GIP），提供通用的状态机描述和信息传播机制。

GVRP 提供一种机制用于动态维护每个 VLAN 动态注册表项的内容，同时向其他网桥传播该信息。这些信息允许 GVRP 设备动态地建立和更新它们所了解的 VLAN 信息。

GVRP 的操作与 GMRP 非常类似（GMRP 用于注册组成员信息），主要区别如下：

- 1) 协议中携带的属性值是 12 比特的 VID 值，而不是 48 比特 MAC 地址和组服务需求信息；
- 2) VID 的注册和撤销行为影响动态 VLAN 注册表项的内容，而不是组注册表项的内容；
- 3) 在单生成树环境下，每个端口有一个 GVRP 参与者，而不是基于每个端口的每个 VLAN 有一个 GVRP 参与者，GVRP 参与者都运行在一个单一的 GIP 上下文环境；
- 4) 在多生成树环境下，每个端口有一个 GVRP 参与者，但是每个 GVRP 参与者可以运行在多 GIP 上下文环境。

GVRP 允许终端站点和桥接局域网内的网桥发布和撤销有关 VLAN 成员关系的声明。每个 GVRP 参与者接收到该声明都会创建或者是更新在过滤数据库中的一个动态 VLAN 注册表项，该表项表明在该接收端口注册该 VLAN。相应的，如果在一个分段上具有相同 VID 的所有参与者都撤销它们的声明，那么附连到该分段的接口在动态 VLAN 注册表项中被设置成未注册。

GVRP 的参与者由以下几部分构成：

- 1) GVRP 应用；
- 2) GARP 信息发布；

3) GARP 信息声明。

VLAN 注册服务允许 MAC 服务用户指明 MAC 服务的提供者，即它们希望加入的一个 VLAN 集合，服务原语允许服务用户：

- 1) 注册一个 VLAN 成员；
- 2) 注销一个 VLAN 成员。

6.2.5.1 GVRP 应用地址

GVRP 应用地址如下表所示，组 MAC 地址被作为去往 GVRP 参与者的 GARP PDU 的目的地址，它将被 GVRP 地址标识。接收到的 PDU 按照 IEEE 802.1D 12.10 子句中定义的 PDU 格式构造，携带一个与 GVRP 地址（见表 6）相同的目的 MAC 地址，按照下面的方式处理：

- 1) 支持 GVRP 操作的网桥和终端站点将向与接收端口关联的 GVRP 参与者递交所有的 PDU，并做进一步的处理；
- 2) 不支持 GVRP 操作的网桥，所有的 PDU 都将递交到转发进程。

表 6 GVRP 应用地址

| 赋 值     | 值                 |
|---------|-------------------|
| GVRP 地址 | 01-80-c2-00-00-21 |

6.2.5.2 GVRP 属性类型编码：

GVRP 操作定义了一个单一的 VID 属性类型，该类型在 GARP 协议交换中携带。VID 属性类型用于标识 VLAN 标识符的值。在 GVRP PDU 中携带的组属性类型值应当是 1。

6.2.5.3 GVRP 属性值编码

VID 属性类型的实例值将在 GARP PDU 中编码成 2 字节的属性值，代表一个无符号的二进制数，等于 VLAN 标识符按照 16 进制的编码值。

6.2.5.4 GVRP 的一致性要求：

GVRP 的一致性要求涵盖两种情况，在 MAC 网桥实现 GVRP 和在终端站点实现 GVRP。

6.2.5.4.1 MAC 网桥 GVRP 的一致性要求：

- 1) GARP 应用和注册状态机以及 LeaveAll 生成机制的一致性操作应符合 IEEE 802.1D 中定义。
- 2) 状态机要求交换的 GARP PDU 格式应与 IEEE 802.1D 中描述的通用 PDU 格式一致，能够携带特定的应用信息，使用 GVRP 应用地址。
- 3) 传播注册信息：  
在单生成树网桥环境下，应与 IEEE 802.1D 中定义的基生成树上下文 GIP 操作一致；  
在多生成数网桥环境下，应与 IEEE 802.1Q 11.2.3.4 子句中定义的多生成树上下文 GIP 操作一致。
- 4) GVRP 应用的实现参考 IEEE Std 802.1Q 11.2 中的定义。
- 5) 转发，过滤或者是丢弃携带任何 GARP 应用地址作为目的 MAC 地址的 MAC 帧应与 IEEE 802.1Q 8.14.3 中的要求一致。

6.2.5.4.2 终端站点 GVRP 的一致性要求

- a) 应符合下述操作之一：
  - 1) 在 IEEE 802.1D 12.7.1 子句中定义的申请者状态机；
  - 2) 在 IEEE 802.1D 12.7.5 子句中定义的申请者惟一状态机；

3) 在 IEEE 802.1D 12.7.6 子句中定义的单一申请者状态机。

b) GARP 状态机实现中所要求交换的 GARP PDU 格式应与 IEEE 802.1D 中描述的通用 PDU 格式一致, 能够携带特定的应用信息, 使用 GVRP 应用地址。

c) 支持终端站点注册和注销预定, 依照 IEEE 802.1Q 11.2 子句中的定义。

d) 丢弃携带任意 GARP 应用地址作为目的 MAC 地址的 MAC 帧, 应与 IEEE 802.1Q 8.14.3 子句要求相一致。

对于终端站点 GVRP 的一致性要求以下部分是可选的:

e) 遵循 GARP 注册状态机以及 LeaveAll 生成机制的操作应符合 IEEE 802.1D 中定义;

f) 支持 VLAN 注册和注销预定, 依照 IEEE 802.1Q 11.2 子句中的定义;

g) 过滤目的地为组 MAC 地址的输出帧应与注册 VLAN 成员信息一致, 使用 IEEE 802.1Q 8.7.2 子句描述相一致的转发进程过滤操作和 IEEE 802.1Q 8.8 子句定义的出规则操作。

建议只有那些要求执行源裁减的终端站点遵循申请者状态机的操作。

#### 6.2.6 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 IEEE 802.1D 中第 8 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的配置实现分配机制。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。

### 6.2.7 快速生成树协议（Rapid Spanning Tree Protocol）

快速生成树协议通过网桥将桥接的本地局域网配置成一个具有全连通性，简单对称拓扑结构的局域网。RSTP 协议在实现上应符合下面的要求：

— RSTP 配置每个桥接端口的端口状态，选择其中一些端口用于转发帧，其他的端口则丢弃帧，提供简单全互连的主动拓扑；

— RSTP 应提供容错机制，当局域网部件发生故障时能够实现拓扑的自动重新配置；能够适应网桥或者是网桥端口的增加，在收敛过程中避免形成短暂的数据环路；

— 主动拓扑应当能够在一个短暂的、已知界限的时间间隔内稳定，最小化端节点之间服务中断的时间间隔；

— 主动拓扑应具有可遇见性和可再生性，能够通过算法参数的管理进行拓扑选择，基于流量分析配置管理应用来满足性能管理目标；

— RSTP 协议对于终端节点来说应当是透明的，当使用 MAC 服务时终端节点不用关心它们是连接到一个局域网还是桥接局域网；

— 在任一特定的局域网中，RSTP 占用的通信带宽应是总计可用带宽的一小部分，网络承载的总流量应当是独立的，与网桥或者是局域网的数量无关；

— 每个桥接端口的内存需求与网络中网桥和局域网数量是无关的；

— 网桥在被加入到网络之前不是必须单独配置，而且网桥 MAC 地址的分配应遵循正常的过程；

— 在正常操作中，配置一个构成点到点局域网络主动拓扑的时间与协议计时器值无关。

此外，桥接协议的运行应满足下述要求：

— 连接到局域网络中的网桥应能够识别用于标识网桥生成树协议实体的组 MAC 地址；

— 每个网桥的标识符在桥接的局域网中应当是惟一的；

— 每个网桥端口的标识符在网桥内部应当是惟一的。

这些参数值应当由每个网桥来提供，使用惟一的 MAC 地址作为桥接组地址来标识生成树协议实体。用于主动拓扑管理的参数值分配应符合下面的要求：

— 网络中每个网桥应具有相对优先级；

— 网桥的每个端口应具有相对优先级；

— 每个端口应具有一个端口路径开销。

另外 RSTP 协议在实现上应当能够兼容 STP 协议。

### 6.2.8 多生成树协议（Multiple Spanning Tree Protocol）

多生成树算法和协议为整个互连的本地桥接局域网内分配的 VLAN 帧提供简单的全连通性。MSTP

协议允许分配到不同 VLAN 的数据帧流经不同的路径，每条路径基于由本地局域网或者是多生成树网桥构成的域内相互独立的多生成树实例来形成。这些域和其他的网桥以及本地局域网连接构成一个简单的标准生成树。

MSTP 协议通过一个简单的标准内部生成树（CIST）连接所有的网桥和 LAN。CIST 支持每一个多生成树域的自动决策，选择它的最大可能扩展。CIST 连通性的计算为 CST 提供这些域的互连以及每个域的一个内部生成树。

多协议生成树在实现上应符合下面的要求：

- MSTP 协议为任意指定的 VLAN 配置主动拓扑构成一个简单生成树，在任意两个端节点之间对于具有相同 VLAN 标识的数据存在最多一个数据路径，并且消除数据环路。
- 当网桥故障或者是数据通路中断，MSTP 提供容错机制在受限的可用桥接本地局域网内自动重新配置生成树拓扑，自动适应网络中任意网桥或者是网桥接口的增加，并且不会形成短暂的数据环路；
- 主动拓扑应当能够在一个短暂的、已知界限的时间间隔内稳定，最小化端节点之间服务中断的时间间隔；
- 主动拓扑应具有可遇见性和可再生性，能够通过算法参数的管理进行拓扑选择，基于流量分析配置管理应用来满足性能管理目标；
- MSTP 协议对于终端节点来说应当是透明的，当使用 MAC 服务时终端节点不用关心它们是连接到单个局域网还是桥接局域网；
- 在任一特定的局域网，网桥用于建立和维护生成树所占用的通信带宽应当是总的可用带宽的一小部分，独立于桥接局域网的承载流量，与网桥或者是局域网的数量无关；
- 允许分配到不同 VLAN 的数据帧在已建立的网络管理域内遵循不同的数据通路；
- MSTP 应提供高可能性，即使在出现诸如分配 VLAN 到生成树的管理差错时，也要为数据帧一直提供简单全连通。

另外，算法和协议应实现下面目标，以限定网桥和它们配置的复杂性：

- 每个网桥端口的内存需求应当独立于网桥和桥接局域网内局域网的数量；
  - 除了通过正常的过程分配 MAC 地址，在加入到桥接局域网之前，网桥不需要单独配置。
- 此外，为实现吞吐量、帧丢失和传输时延的性能改进，多生成树协议需要遵循以下要求：
- 在每个可能的 MST 域内，提供一种方式保证 VID 到 MSTIDs 赋值的一致性；
  - 在管理上对用于表示 VID 到 MSTIDs 分配的配置名称和修订等级达成一致；
  - 提供对通信站点集之间潜在流量的分配评估方法；
  - 性能目标选择或者是通信服务质量特性目标的建立应当独立于其他一些因素；
  - 生成树配置参数的选择应提供可用的物理部件。

MSTP 协议在实现上应当能够兼容 STP 和 RSTP 协议。

#### 6.2.9 VLAN（虚拟局域网）功能

VLAN 功能是指通过将桥接的局域网内活跃拓扑中的节点划分到不同域来实现相互隔离，各 VLAN 使用 VID（VLAN 标识符）区分。各个 VLAN 是原桥接局域网的一个子集。

交换机应当支持 VLAN 功能。交换机对 VLAN 功能的支持应符合 YD/T 1260-2003《基于端口的虚拟局域网（VLAN）技术要求和测试方法》中的相关要求。

### 6.2.10 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，可选交换机实现远程桥接，对远程桥接的实现必须符合 IEEE 802.1G。

### 6.2.11 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。

建议交换机实现多链路聚合。对链路聚合的实现建议符合 IEEE 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，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。三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。

丢包率用来衡量由于重传超时、过多的冲突、缓冲区溢出而造成的数据包丢失。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 802.3 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，因此，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的输出内容复制到另一端口。交换机可选实现端口镜像，在镜像时应允许对全部内容或者是部分内容进行复制。此外，为便于网络的安全管理，交换机可选实现向远端监控中心进行镜像的功能。

### 7 网络层协议要求

#### 7.1 Internet 协议-IPv6

##### 7.1.1 定义

具有 IPv6 路由功能的以太网交换机必须实现 IPv6 协议，并符合 IETF RFC2460。

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 ICMPv6 差错消息），但是为了诊断故障，交换机要求能够提供将差错写入日志以及对丢弃数据包进行计数统计的功能。

### 7.1.2 协议概述

IPv6 作为 IPv4 后继的网络层协议，具有更大的地址空间、简化的报文头格式、增强支持选项域和扩展头、支持流标签以及认证和数据保密等特性，该协议具体描述可以参见 IETF RFC2460。

#### 7.1.2.1 跳数限制

交换机应丢弃收到的跳数限制为 0 的数据包，并且向数据源发送 ICMPv6 超时消息。当交换机转发数据包时应将跳数限制域的值减 1。

#### 7.1.2.2 扩展头的分类

交换机必须要能识别并处理以下几种扩展头：

- 逐跳选项头；
- 目的地选项头；
- 路由头；
- 分段头；
- 交换机可以识别并处理以下几种扩展头：
- 认证头；
- 封装安全载荷头。

除了以下两种情况之外，交换机只能检查或处理接收数据包中的逐跳选项头：

- 接收数据包的接口地址与数据包 IPv6 头中的目的地址相同；
- 数据包的 IPv6 头中的目的地址是组播地址，并且该路由器是组播节点组中的一个节点。

交换机必须严格按照扩展头在数据包中出现的顺序来检查或处理接收数据包中的扩展头。交换机不能在数据包中搜索一个特定的扩展头，并且在处理完所有排在它前面的头之前处理它。

如果交换机处理一个头的结果是要进行下一个头的处理，但这个头的“下一个头”域的值不能被交换机所识别，则交换机将丢弃这个数据包并向数据包的源节点发送一个 ICMPv6“参数错误”消息，ICMPv6 代码值为 1（不能识别下一个头的类型），ICMPv6 指针域包含源数据包中不能被识别的域的偏移量。若一个交换机遇到除 IPv6 头外的任一个头的“下一个头”域为 0，则它对这个数据包也应按上面的方法进行处理。

#### 7.1.2.3 扩展头的顺序

当交换机发送的数据包带有多个扩展头时，扩展头应按下面的顺序出现：

- IPv6 头；
- 逐跳选项头；
- 目的地选项头（注 1）；
- 路由头（0 型）；
- 分段头；
- 认证头；
- 封装安全载荷头；



- 目的地选项头（注 2）；
- 上层头。

注 1：这些选项要在 IPv6 目的地址域所列出的第一个目的地进行处理，也要在路由头所列出的后续目的地进行处理。

注 2：这些选项只在数据包的最终目的地进行处理。

交换机必须接收并处理同一个数据包中以任何顺序、任何次数出现的扩展头，只有逐跳选项头才必须严格地接在 IPv6 头之后。

#### 7.1.2.4 选项

交换机在处理一个扩展头时，必须严格按照每个选项在扩展头中出现的顺序来处理它们。

交换机对逐跳选项头或目的地选项头中未知选项的处理必须符合 IETF RFC2460 第 4.1 节中的规定。

交换机可选支持逐跳选项头中的超长包载荷选项。该选项的格式和用法应符合 IETF RFC2675 的规定。

交换机必须支持逐跳选项头中的路由器警告选项。该选项的格式和用法应符合 IETF RFC2711 的规定。

#### 7.1.2.5 扩展头的处理

除了逐跳选项头以外，其他类型的扩展头都不被报文转发路径上的节点处理，只在最终目的地节点做处理。

##### 7.1.2.5.1 路由头

交换机必须支持 0 型路由头。

交换机在处理接收到的数据包时，如果遇到一个路由头包含有不能识别的“路由类型”值，则它应该依据“剩余段”域的值采取措施。具体方法如下所述：

- 如果“剩余段”的值为 0，则交换机忽略这个路由头，继续处理数据包中的下一个头（其类型由路由头的“下一个头”域的值标识）；
- 如果“剩余段”的值不为 0，则交换机必须丢弃这个数据包，并且向数据包的源地址发送一个 ICMPv6 “参数错误”消息（代码值为 0），ICMPv6 指针指向不能识别的“路由类型”。

如果一个交换机在处理完接收数据包的路由头后，决定应将该数据包转发到一条链路 MTU 小于该包长度的链路上，那么该节点必须丢弃此数据包并向该包的源地址发送一个 ICMPv6 “数据包过大”消息。

交换机发送的数据包中如果包含 0 型路由头，则该路由头内一定不能包含组播地址。

交换机只能检查或处理“目的地址”域与接收数据包接口地址相同的 IPv6 数据包中的路由头。

交换机处理 0 型路由头的算法应严格符合 IETF RFC2460 第 4.4 节中的相应规定。

##### 7.1.2.5.2 分段头

如果交换机需要发送一个大于路径 MTU 的数据包到目的节点，则它应将该数据包分段，并将每个分段作为一个独立的数据包传送。交换机在进行数据包分段时应符合 IETF RFC2460 第 4.5 节的相应规定。

交换机必须支持将发送给自己的分段 IPv6 数据包重组。交换机在重组数据包时应符合 IETF RFC2460 第 4.5 节中的重组原则。

如果交换机在接收到第一个到达的分段之后的 60 秒内没有收到该数据包所有要重组的后续分段，则交换机应放弃重组该数据包，并且丢弃所有已接收的分段。在这种情况下，如果第一个分段数据包已接收到，则交换机要向那个分段数据包的源地址发送一个 ICMPv6 “超时——段重组超时”消息。

如果交换机从分段数据包“载荷长度”域中得到的分段长度不是 8 个字节的整数倍，并且这个段的 M 标志位是 1，则交换机必须丢弃这个段，并且要向段的源地址发送一个 ICMPv6 “参数错误”消息（代码为 0），ICMPv6 指针指向分段数据包的“载荷长度”域。

如果交换机收到的一个分段的长度和偏移导致出现这种情况，即由这个分段重组的数据包“载荷长度”超过 65535 个字节，则交换机必须丢弃这个分段，并且向分段的源地址发送一个 ICMPv6 “参数错误”消息（代码为 0），ICMPv6 指针指向分段数据包的“段偏移”域。

#### 7.1.2.6 数据包的长度

交换机任一接口的链路 MTU 均不得小于 1280 字节。如果与某一接口相连的链路不支持 1280 字节的数据包，则交换机必须在网络层以下的一层提供与链路相关的分段和重组功能。

为了发送长度大于路径 MTU 的数据包，交换机必须使用 IPv6 分段头给数据包分段。

#### 7.1.2.7 流标签

对于不支持流标签域功能的交换机来说，当发送一个数据包时，在流标签域填入 0 值；当转发数据包时，对流标签域不作任何改动；当接收数据包时，忽略流标签域。

#### 7.1.2.8 业务等级

支持业务等级域的特殊应用的交换机可以在生成、转发或接收数据包时根据特殊应用的需要改变这一域的值。不具备此能力的交换机应忽略此域并且不能对其进行修改。

### 7.2 邻居发现协议

#### 7.2.1 协议功能

邻居发现协议是 IPv6 协议的一个基本的组成部分，它实现了在 IPv4 中的地址解析协议（ARP）、控制报文协议（ICMP）中的邻居发现部分、重定向协议的所有功能，并具有邻居不可达检测机制。

交换机必须实现邻居发现协议中的邻居和前缀发现、地址解析、下一跳地址确定、重定向、邻居不可达检测、重复地址检测功能，可选实现链路层地址变化、输入负载均衡、泛播地址和代理通告等功能。

#### 7.2.2 消息类型

交换机通过使用五种类型的 IPv6 控制信息报文（ICMPv6）来实现邻居发现功能。交换机必须实现邻居发现协议的五种消息类型：

- 路由器请求（Router Solicitation）：当接口工作时，主机应立即发送路由器请求消息，要求交换机产生路由器通告消息，而不必等到下一个通告周期。
- 路由器通告（Router Advertisement）：交换机周期性地通告其存在以及链路配置参数和网络参数，或者对路由器请求消息作出响应。路由器通告消息包含用于确定本地链路以及配置地址的前缀信息，跳数限制值等。
- 邻居请求（Neighbor Solicitation）：节点发送邻居请求消息来请求邻居的链路层地址，以验证它先前所获得并保存在缓存中的邻居链路层地址的可达性，或者验证它自己的地址在本地链路上是否是惟一的。
- 邻居通告（Neighbor Advertisement）：邻居请求消息的响应；节点也可以发送非请求邻居通告来指示链路层地址的变化；
- 重定向（Redirect）：节点通过重定向消息通知主机，对于特定的目的地址，自己并不是最佳路由，并通知主机到达目的地的最佳下一跳。

### 7.2.3 邻居和前缀发现

交换机必须无条件丢弃不满足有效性检查的路由器请求和路由器通告消息。

邻居发现功能用来标识与给定链路相连的交换机，并获取与地址自动配置相关的前缀和配置参数。

作为对请求消息的响应，交换机应周期地发送组播路由器通告消息，来通告链路上节点的可达性。交换机发出路由器通告消息，指示该发送方是否愿意作为缺省网关。路由器通告还包括前缀信息选项，这些选项列出了一组确认“在连接”IP 地址的前缀。路由器通告消息应包含一些标志位，这些标志位通知主机怎样执行地址的自动配置。另外，路由器通告消息中还应包含网络管理的参数，例如主机产生的数据包中使用的跳数限制参数的缺省值或链路 MTU 值。

当主机向交换机发出路由器请求消息时，交换机应立刻发送路由器通告消息。

### 7.2.4 地址解析

交换机通过邻居请求和邻居通告消息将 IPV6 地址解析成链路层地址，对组播地址不执行地址解析。

交换机通过组播邻居请求消息来激活地址解析过程，邻居请求消息用来请求目标交换机返回它的链路层地址。源节点在邻居请求消息中包含了它的链路层地址，并将邻居请求消息组播到与目标地址相关的请求节点组播地址，目标节点在单播的邻居通告消息中返回它的链路层地址。这一对消息使源和目标节点能够解析出相互的链路层地址。

### 7.2.5 下一跳地址确定

当交换机向目的地发送数据包时，使用目的地缓存、前缀列表、默认网关列表确定合适的下一跳的 IP 地址，然后交换机查询邻居缓存确定邻居的链路层地址。

IPv6 单播地址的下一跳确定操作如下：发送者使用前缀列表中的前缀进行最长前缀匹配，确定数据包的目的地是在直连链路还是非直连链路。如果下一跳是在直连链路，则下一跳地址就和目的地地址相同，否则发送者从默认网关列表中选择下一跳。如果默认网关列表是空，发送者认为目的地是在直连链路上。

下一跳确定的信息存储在目的地缓存中，下一个包可以使用这些信息。当交换机发送包时，首先检查目的地缓存，如果目的地缓存没有相关信息存在，就激活下一跳确定过程。

在学习到下一跳交换机的 IPv6 地址后，发送者检查邻居缓存以决定链路层地址。如果没有下一跳 IPv6 地址的表项存在，交换机的工作如下：

- 创建一个新表项，并设置其状态为不完全；
- 开始进行地址解析；
- 对传送的包进行排队。

当地址解析结束时，获得链路层地址，存储在邻居缓存中。此时表项到达新的可达状态，排队的包能够转发。

对于组播包，下一跳总是认为在直连链路上。确定组播 IPV6 地址的链路层地址取决于链路类型。

当邻居缓存开始转发单播包时，发送者根据邻居不可达检测算法检测相关的可达性信息，验证邻居的可达性。

当邻居不可达时，再次执行下一跳确定，验证到达目的地的另一条路径是否是可达的。

### 7.2.6 重定向功能

当数据包发送到一个非直连的目的地时，需要选择中间转发交换机。当选择的交换机作为报文转发

的下一跳并不是最佳时，交换机应当产生重定向消息，通知源节点到达目的地存在一个更佳的下一跳交换机。

交换机必须能够确定每个相邻交换机的本地链路地址，以保证重定向消息里的目标地址根据本地链路地址来识别相邻交换机。

在源端没有正确应答重定向消息，或者源端选择忽略没有被验证的重定向消息的情况下，为了节省带宽和处理的费用，交换机必须限定发送重定向消息的速率。

在收到重定向消息时，交换机不能更新路由表。

#### 7.2.7 邻居不可达检测

交换机应当进行邻居不可达性检测，以检测邻居或邻居前向路径发生的故障。

如果交换机新近从邻居节点接收到对发送到它的数据包的确认，那么该邻居就是可达的。邻居不可达检测使用两种方法进行确认：一种是从上层协议来的提示，提供“连接正在处理”的确认；另一种是交换机发送单播邻居请求消息，等待接收应答的邻居通告消息。为了减少不必要的网络流量，探测消息仅发送到邻居。

邻居不可达性检测与向邻居发送数据包同时进行。在邻居可达性确认期间，交换机继续向缓存链路层地址的邻居发送数据包。如果没有数据包发向邻居，则不发送检测。

#### 7.3 路径 MTU 发现协议

为了有效利用网络带宽资源并尽量减少 IPv6 报文分段的发生，有必要发现端到端的 MTU。在 IETF RFC1981 中描述了发现路径 MTU 的机制。交换机可选支持路径 MTU 发现协议。如果交换机不支持该协议，则在转发数据包时应以缺省的 IPv6 最小链路 MTU（1280 字节）作为最大包长。

交换机在发送 IPv6 数据包时，应以发送下一跳链路的 MTU 作为路径 MTU 的初始预测值并根据 IETF RFC1981 中描述的方法修改此预测值。

当交换机路径 MTU 的预测值小于或等于实际的路径 MTU 时，交换机应终止发现路径 MTU 的处理过程。

当数据包的目的地址是组播地址时，交换机应选择所有组播路径的路径 MTU 的最小值作为转发数据包时的路径 MTU。

#### 7.4 互联网控制消息协议-ICMPv6

##### 7.4.1.1 定义

在交换机中 ICMPv6 用来报告处理报文过程中遇到的错误，以及实现一些网络层功能，如诊断（ICMPv6“Ping”、“Traceroute”）。ICMPv6 是 IPv6 的一个必要组成部分，交换机必须实现 ICMPv6 协议。

##### 7.4.1.2 消息类型

交换机必须实现两类 ICMPv6 消息：差错消息和信息消息。差错消息的消息类型字段高位比特为 0，所以差错消息的消息类型代码为 0~127，信息消息的消息类型代码为 128~255。

ICMPv6 差错消息：

- 目的不可达（类型 1）；
- 包长超长（类型 2）；
- 超时（类型 3）；
- 参数错误（类型 4）。

ICMPv6 信息消息：

- 回显请求（类型 128）；
- 回显应答（类型 129）。

#### 7.4.1.3 消息源地址的确定

发送 ICMPv6 消息的交换机必须在对 IPv6 报头计算校验和之前确定消息源和目的的 IPv6 地址。如果该交换机具有多个单播地址，那么它需要依据以下原则选择消息的源地址：

- 如果要发送的消息是对发送到这个交换机某个单播地址的报文响应，那么源地址必须与该单播地址相同。
- 如果要发送的消息是对发送到这个交换机所属组的组播或泛播组地址的报文响应，那么响应消息的源地址必须是接收组播包的接口的一个单播地址。
- 如果一个发送到非本交换机地址的报文发生了错误，交换机应该发送一个 ICMPv6 消息作为响应，这个 ICMPv6 消息的源地址应该是本交换机的一个单播地址，并且这个地址应该与错误报文的目的地具有最大的相关性。例如，如果待发送的 ICMPv6 消息是对一个数据包无法成功传送情况的响应，那么 ICMPv6 消息的源地址应该是该数据包转发失败的那个接口的单播地址。

• 另外，传送消息时应根据其目的地地址检查交换机的路由表，以确定转发接口，并且消息的源地址应该是这个接口上的一个单播地址。

#### 7.4.1.4 消息校验和计算

交换机对校验和的计算是将 ICMPv6 消息分成 16 比特长的段，每段计算其二进制补码，然后对其求和，这里的 ICMPv6 消息从 ICMPv6 消息类型字段开始，它由 IPv6 报头中的“伪报头”指明，在伪报头中的“下一报头”值为 58。

在计算校验和之前，校验和字段要先置为 0。

#### 7.4.1.5 消息处理规则

交换机对 ICMPv6 消息的处理应符合以下规则：

[1] 如果接收到带有未知类型的 ICMPv6 差错消息，必须将其交至上层。

[2] 如果接收到带有未知类型的信息消息，必须将其丢弃。

[3] 每个 ICMPv6 差错消息（类型值<128）的消息体中将包含导致错误的那个 IPv6 数据包，但要保证最后的差错消息长度不超过 IPv6 最小 MTU 的限制。

[4] 当网络层协议要求将 ICMPv6 差错消息传送给上层进程时，从原始数据包中取出（包含在 ICMPv6 差错消息的消息体中）上层协议的类型，根据协议的类型选择适当的上层进程来处理差错。如果原始数据包含有过多的扩展报头，由于要满足 IPv6 最小 MTU 的要求，在 ICMPv6 消息中可能不包含上层协议类型。这种情况下，这个差错消息将在 IPv6 层处理后丢弃。

[5] 当接收到如下报文时，不能发送 ICMPv6 差错消息：

- 一个 ICMPv6 差错消息；
- 一个发往某 IPv6 组播地址的报文（对这个规则有两种例外情况：1）包长过大消息—使用 IPv6 组播的路径 MTU 发现机制进行工作；2）代码为 2 的参数错误消息—表明有一个未知的 IPv6 选项，且选项类型的最高位的两个比特为 10）；
- 一个链路层组播包；

- 链路层广播包；
- 数据包的源地址不是 IPv6 单播地址，也就是说，其源地址是一个非 IPv6 定义的地址，一个 IPv6 组播地址或是一个 ICMP 消息发送者已知的 IPv6 “泛播”地址。

最后，为了限制发送 ICMPv6 差错消息所用的带宽，减少开销，每个 IPv6 交换机都必须限制 ICMPv6 差错消息的发送速率。有几种方法可以实现速率限制的功能，例如：

- 基于定时器：例如，限制将差错消息发往指定信源或任何信源的速率，最多每 T 微秒发送一次。
- 基于带宽：例如限制在某个接口上发送差错消息的速率为与其相连的链路带宽的某一比例 (F)。

在交换机上限制参数（上边例子中的 T 和 F）应该配置一个保守的缺省值（例如：T=1s，而不是 0；或 F=2%，而不是 100%）。

#### 7.4.1.6 ICMPv6 差错消息

##### 7.4.1.6.1 目的不可达消息

目的不可达消息应该由交换机的网络层产生，作为对数据包由于非阻塞的原因无法送达目的端的响应（如果数据包由于网络阻塞而被丢弃，则不能发送 ICMPv6 消息）。

如果传送失败的原因是在传送交换机上没有匹配的路由表项，则代码值应置为 0。（注意：这种错误只能当交换机上没有“缺省路由”时才会发生）

如果传送失败是由于网络管理上的原因，如存在“防火墙”，则代码值应置为 1。

如果传送失败是由于其他原因引起，例如无法将 IPv6 地址解析为响应的链路层地址，或由于链路层的某些原因，则代码值应置为 3。

当目的交换机上的传输层协议（如 UDP）对数据包并没有接收者，并且传输层协议本身也没有措施去通知发送端，则接收端应该发送一个代码为 4 的目的不可达消息。

当交换机接收到以自身为目的地 ICMPv6 目的不可达消息之后必须通知高层进程。

##### 7.4.1.6.2 包长超长消息

当数据包大于出口链路的 MTU 而无法传送时，交换机应该发送一个包长超长消息作为响应。消息中的信息用来作为路径 MTU 发现过程的一部分。

在如下情况下交换机需要发送包长超长响应消息：接收到一个目的地址为 IPv6 组播地址的报文，或者一个链路层组播报文，或者一个链路层广播报文。

当交换机接收到以自身为目的地的包长超长消息时，必须送交上层处理。

##### 7.4.1.6.3 超时消息

如果交换机接收到跳数限制为 0 的包，或交换机将数据包的跳数限制减至 0，则必须丢弃这个数据包，并向信源发送代码为 0 的 ICMPv6 超时消息，这个消息指明出现了路由环路或跳数初始值过小。

当交换机接收到以自身为目的地的超时消息时，必须送交上层处理。

##### 7.4.1.6.4 参数错误消息

如果交换机在处理报文时，发现报文的头部或扩展头中发生错误，以至于无法进行处理，则该交换机必须丢弃这个报文，并应该向信源发送一个 ICMPv6 参数错误消息，以指明错误的类型和位置。

指针指示了原始数据包中发生错误的位置（以字节为标志）。例如：类型为 4，代码为 1，指针域为 40 的 ICMPv6 消息表明在原始数据包中紧跟着 IPv6 报头的扩展报头中存在一个未知的“下一报头”值。

当交换机接收到以自身为目的地的参数错误消息时，必须送交上层处理。

#### 7.4.1.7 ICMPv6 信息消息

##### 7.4.1.7.1 回显请求消息

交换机必须实现 ICMPv6 回显响应功能，当收到回显请求时能够发送相应的回显应答。同时，交换机应实现应用层的接口，以发送回显请求，接收回显应答，作为一种诊断的手段。

回显请求消息可以送交上层进行 ICMP 消息的处理。

##### 7.4.1.7.2 回显应答消息

交换机必须实现 ICMPv6 的回显功能，以接收回显请求消息并发送相应的回显应答。同时交换机应该实现应用层接口，用来发送回显请求并接收回显应答，以达到诊断的目的。

与发往单播地址的回显请求消息相对应的回显应答消息的源地址必须与该回显请求消息中的目的地址相同。

对于发往 IPv6 组播地址的回显请求消息，也应该发送回显应答作为响应。回显应答必须使用接收到该回显请求的接口的一个单播地址作为源地址。

回显应答消息必须交付给交换机上产生回显请求的进程进行处理，它也可能交付给那些没有产生回显请求的进程。

#### 7.5 组播监听者发现协议（MLD）

MLD 是用于主机和组播路由器之间的协议，该协议是非对称的协议，分别定义了组播监听者和组播路由器的行为。MLD 应用于单个物理网络，旨在建立特定组播组中的主机成员关系。MLD 使用 ICMPv6 的消息类型。组播路由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。

交换机可选支持组播监听者发现协议，并符合 IETF RFC2710、IETF RFC3590 和 IETF RFC3810。实现组播监听者发现协议的交换机应该实现 MLD 中的主机部分要求。

#### 7.6 组播监听者嗅探协议（MLD Snooping）

交换机应可选实现 MLD Snooping 功能，其实现参见 IETF RFC4541。

### 8 传输层协议要求

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）。传输层对于 IPv6 超大包的处理参见 IETF RFC2675。

#### 8.1 用户数据报协议

用户数据报协议参见 IETF RFC768。

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

- 1) 本标准不规定不同协议层间的接口；
- 2) 与 IETF RFC1122 相反，具有 IPv6 路由功能的以太网交换机应该产生 UDP 校验和。

#### 8.2 传输控制协议——TCP

传输控制协议参见 IETF RFC793。

### 9 路由协议

#### 9.1 概述

互联网路由系统可以划分为自治系统域内路由和自治系统域间路由两大类。自治域是一个运行相同控制协议和管理策略的逻辑区域，它可以描述一组交换机从域内到域间的路由路径。在一个大规模的网

络拓扑中，IP 数据包通常需要穿越两个甚至多个自治系统域内的交换机才能够到达目的地，因此要实现报文的正确转发，自治系统必须能够相互提供拓扑信息，这样才能通过运用路由算法计算出正确的转发路径。内部网关协议用作在自治系统域内分发路由信息（即 AS 内部路由），外部网关协议用作在自治系统域间交换路由信息（即 AS 域间路由）。

除非特殊路由协议的指定，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）用作在自治系统内部交换机之间分发路由信息。对特定 IGP 算法的实现相对独立，但必须实现下列功能：

- 1) 应能够迅速反映自治系统内部的拓扑变化；
- 2) 提供一种机制，确保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路（loop-free）路由的机制；
- 4) 使用最少的带宽资源。

交换机除必须实现静态路由外，根据设备在组网中的实际需求选择实现 RIPng、IS-ISv6 和 OSPFv3 三种动态内部网关路由协议中的一种。

### 9.2.2 路由信息协议 RIPng

RIPng 路由协议适用于在中等规模的 IPv6 网络中（网络直径小于等于 15 跳）交换网络可达信息，该协议是距离向量路由协议，使用 Bellman-Ford 算法计算到特定目的网络的最佳路径。

交换机必须实现 RIPng 协议。交换机对 RIPng 协议的支持必须符合 IETF RFC2080。

### 9.2.3 开放最短路径优先——OSPFv3

OSPFv3 路由协议是基于链路状态的路由协议，该类协议的一个重要特征就是在整个路由区域内所有路由设备都维护一个完全相同的拓扑数据库，该数据库描述了路由区域内相应的链路状态信息。每一个运行 OSPFv3 的路由设备根据这个数据库利用 Dijkstra 最短路径算法计算出到网络其他节点的单源最短路径。为了实现拓扑数据库的同步，OSPFv3 协议定义了 Hello 协议、交换协议、邻居有限状态机、接口有限状态机以及扩散算法等。

交换机必须支持可变长子网掩码（VLSM），支持广播网络、非广播多接入网络（NBMA），支持虚链路（可选），支持 Stub 域和 NSSA 域，支持等价多路径，具体的规定见 YD/T1295-2003《支持 IPv6 的路由协议——开放最短路径优先协议（OSPF）》。

交换机对 OSPFv3 协议的实现必须符合 IETF RFC2740，并且参考 IETF draft-ietf-ospfv3-mib-10 实现 OSPFv3 的 MIB 库。

### 9.2.4 中间系统到中间系统——双重 IS-ISv6

IS-ISv6 是链路状态路由选择协议，它从邻接节点收集路由选择信息，生成链路状态数据库，在网络内部使用 SPF 算法（Dijkstra）计算出到目的网络的最佳路径。

交换机可选实现双重 IS-ISv6。

双重 IS-IS 在 IETF RFC1142 和 IETF RFC1195，draft-ietf-isis-ipv6-06 中规定。

实现双重 IS-ISv6 的交换机必须实现双重 IS-ISv6 MIB（见 IETF RFC4444）。



### 9.3 外部网关协议

#### 9.3.1 概述

外部网关协议用于自治系统之间，实现特定自治系统内一组网络与相邻自治系统之间的网络可达性信息交换。

#### 9.3.2 边界网关协议——BGP4+

##### 9.3.2.1 定义

边界网关协议（BGP4+）是自治系统域间路由协议，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。此外，该信息应足够丰富以用作构建 AS 互连图，在 AS 互连图中，必须裁减路由环路，必须被实施相应的策略控制。

交换机根据在组网中的实际需要选择实现 BGP4+路由协议。交换机实现 BGP4+必须符合 IETF RFC1771 和 IETF RFC2858, 以及 YD/T 1342-2005《IPv6 路由协议——支持 IPv6 的边界网关协议(BGP4)》。

实现 BGP4+的交换机必须实现 BGP4 MIB，遵从 IETF RFC1657 的规定。

建议实现 BGP4+的交换机遵从 IETF RFC1772 第 6 章中的规定。

##### 9.3.2.2 协议介绍

BGP4+提供对非常复杂的路由策略的支持，但不要求所有对 BGP4+的实现都支持这样的策略。BGP4+至少要实现：

[1] 应允许 AS 控制 BGP4+学到的路由是否广播到相邻 AS。BGP4+的实现应至少在单个网络粒度上实现上述规定。BGP4+的实现同样应在自治系统粒度上实现上述规定，上述自治系统可能是产生路由的自治系统，或者可能是将路由广播到本地系统的自治系统（相邻自治系统）。

[2] 应允许 AS 当存在多条路径时倾向于使用某条特定路径。该功能应通过允许管理员更改自治系统度量值来实现，使路由选择进程选择一条最低度量的路由（路由的度量定义为有关该路由 AS\_PATH 路径属性所有 AS 的度量之和）。

[3] 应允许 AS 忽略 AS\_PATH 路径属性中包含某特定 AS 的路由。这样的功能可以使用 2) 中提到的技术实现：将某 AS 度量赋为无限大。路由选择进程必须不理睬度量为无限大的路由。

[4] 支持 BGP4+路由反射，并符合 IETF RFC1966 的规定。

[5] 支持 BGP4+团体属性，并符合 IETF RFC1997 的规定。

[6] 支持自治系统联盟。

[7] 支持 IETF RFC2439 规定的路由振荡抑制。

[8] 支持 IETF RFC2283 规定的 BGP 4+的多协议扩展。

[9] 支持 IPv6 的 BGP4+路由协议利用 BGP4 多协议扩展定义的 MP\_REACH\_NLRI 和 MP\_UNREACH\_NLRI BGP 属性来传送 IPv6 的路由信息。具体的规定见 YD/T1342-2005。

##### 9.3.3 没有外部协议的自治系统域间路由

在两个独立的标准内部路由协议之间不使用标准外部路由协议交换两个自治域中的路由信息是可能的。这样做通常的方法是在一个边界路由器上独立运行两个内部路由协议进程，在两个进程间交换重分布路由信息。

如同 EGP 与 BGP 交换信息，如果没有适当的控制，在一个路由器两个 IGP 间交换路由信息很容易产生路由环路。

## 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交换机。交换机应提供一种途径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度量。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：

- 前缀长度；
- 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

网络中每个交换机都是基于路由转发表来作转发决定。在一个简单网络中，路由转发表可以通过静态配置方式来维护，而当网络变得复杂时，则必须通过动态更新的方式来维护。

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用于交换机创建路由转发表的信息的传播。这种控制策略可以通过认证通告路由信息源，路由信息的有效性来实现。交换机路由转发表就是将可用的路由消息经过过滤后计算得到的。

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路由转发表稳定性。

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，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交换机用作转发包的路径进行控制。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；
- 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，交换机必须指定从哪一个相邻交换机可以接收路由信息。

## 10 MPLS 协议

交换机体系结构可选支持 MPLS 标记交换，可选实现下述功能：

- 支持 MPLS LER 和 LSR 功能；
- 支持 MPLS 显式路由 LSP；
- 支持 LDP 标记分发协议；
- 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 协议具体要求见 YD/T1162.1-2005《多协议标记交换（MPLS）技术要求》。

交换机应可选实现二层 VPN 和三层 VPN 的功能，实现 CE 和 PE 的功能。二层 VPN 可选实现点到点或者是点到多点的拓扑；三层 VPN 可选实现 BGP/MPLS VPN。

## 11 QoS 控制机制

### 11.1 流量控制

应能够根据TCP/IP分组报头中的一个或多个字段，标识属于特定通信类的分组，然后通过重新设置IP区分服务代码点（DSCP）字段值来标记已被分类的通信。

支持对用户数据进行带宽控制，依据与用户签订的SLA协定，为用户分配带宽资源，对于超出SLA协议的流量可以采取降级或者是丢弃等操作。

交换机可选实现流量整形功能。

### 11.2 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）。
- 3) 必须支持优先级队列（PQ）。

### 11.3 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，建议提供有效的方法消除头部拥塞。

## 12 组播协议

交换机建议实现组播协议。可选支持组播监听协议 MLD（见 IETF RFC3590 和 IETF RFC3810），可选支持和协议无关组播协议——稀疏模式（PIM-SM），可选实现距离矢量组播路由协议 DVMRP（IETF RFC1075）和组播源发现协议 MSDP。域间组播实现可选 MBGP。

## 13 安全功能要求

交换机必须实现一定的安全控制策略，为用户提供数据完整性、可用性、保密性和访问控制等安全服务。此外，交换机还应具有一定的抗攻击性，要能够抵御网络攻击者的恶意攻击，例如要能够抵御分布式的DoS攻击。

### 13.1 链路层安全功能要求

交换机应支持MAC链路层的安全功能，支持基于目标MAC地址的访问控制，使得交换机拒收不是发给自己的数据包。

### 13.2 IPv6 协议的安全功能要求

#### 13.2.1 IPsec 功能

IPsec作为IPv6协议栈的一个基本组成部件，在交换机中必须实现。通过建立IPSec隧道，为用户数据提供完整性、数据源身份认证、保密性以及防重放攻击的保护。应支持AH和ESP协议，支持传输模式和隧道模式，支持安全联盟手工建立和IKE协议自动建立两种方式。

AH协议应支持HMAC-SHA1-96验证算法，可支持HMAC-MD5-96验证算法。

ESP协议应支持HMAC-SHA1-96验证算法，可支持HMAC-MD5-96验证算法。

加密算法应支持空加密算法、3DES-CBC加密算法，可支持DES-CBC、AES-CBC和国家规定的标准分组加密算法。

如果实现密钥的动态分管理协议IKE，则应支持IKE的下列特性：

- 支持安全联盟的手工管理和IKE自动管理。在手工管理安全联盟时，可支持以十六进制配置算法所需密钥，应支持任意长度字符串形式配置密钥；
- 支持预共享密钥验证，可支持数字证书验证和RSA加密Nonce验证；
- 应支持3DES加密算法，应支持SHA1完整性验证算法，可支持MD5完整性验证算法，同时还应支持DES、AES加密算法和国内的分组加密算法；
- 在IKE的DH交换中，应支持MODP-Group1、MODP-Group2；
- 阶段2交换中应支持完美前向保护特性；
- 阶段1应支持主模式和野蛮模式，阶段2应支持快速模式，还应支持信息交换；
- 可支持NAT穿越；
- 在IKE阶段1中应该能指定发起模式；
- 在IKE的阶段2协商中，应支持ID\_IPV4\_ADDRESS和ID\_IPV4\_SUBNET身份载荷。

### 13.2.2 NDP协议安全要求

在NDP邻居发现协议的实现中，应支持关闭重定向和ARP代理功能。

### 13.2.3 访问控制

交换机在进行数据转发时，应支持ACL（访问控制列表）功能，支持基于源/目的地址、源/目的端口、协议五元组过滤；可选支持uRPF源路由过滤功能。

## 13.3 路由协议的安全功能要求

交换机必须通过路由协议交换路由信息来构建路由转发表，因此保证路由信息的完整性和有效性是非常重要的。交换机应支持对路由信息通告源的认证，确保路由信息来自可靠的对等实体；应提供必要安全机制，确保路由信息在网络传递中不会被攻击者篡改、插入或监听，从而导致路由转发表构建错误，或者是泄漏路由信息。

—— RIPng，依赖IP认证头和IP安全载荷封装来保证路由交互信息的完整性和保密性。

—— OSPFv3，在OSPFv3的协议报头中已经去除了认证和认证类别域，依赖IP认证头和IP安全载荷封装来保证路由交互信息的完整性和保密性。

—— IS-ISv6，应实现基于链路、Level1和level2域的认证，符合IETF RFC1195的规定。

—— BGP4+，可以通过两种方式来提供协议认证，一种是基于TCP的MD5认证，另一种是通过OPEN消息中的认证属性来提供MD5认证，通过消息头的Marker字段值检测对等体之间的失步。

此外应支持必要的路由过滤功能，确保交换机不会被倾泄路由，导致内存溢出，服务中断。

## 13.4 管理的安全功能要求

为实现管理的安全，交换机宜实现下面功能，但不局限在这些方面：

- 身份认证，登陆用户必须提供帐号/口令才能对交换机进行操作，用户地址和操作应记入日志。
- 应提供用户的分级管理功能，实现不同级别的用户具有不同的访问控制权限。
- 应能够关闭远程Telnet登陆服务，或者是限定能够远程Telnet登陆访问的用户IP地址段。
- 应提供SSH远程安全登陆服务，支持SSHv1和SSHv2两种版本；SSHv2应支持用于会话的加

密密钥和认证密钥的动态管理，支持 Diffie-Hellman 组 14 的密钥交换，在密钥交换过程中协商密钥交换算法、对称加密算法和认证算法等；应支持 HMAC-SHA1 认证算法，建议支持 HMAC-SHA1-96 认证算法，可实现 HMAC-MD5、HMAC-MD5-96 等认证算法；应支持 3DES-CBC 对称加密算法，可实现 Blowfish-CBC、IDEA-CBC、CAST128-CBC、AES256-CBC、AES128-CBC 等对称加密算法；对于非对称加密算法，应支持 SSH-DSS，建议实现 SSH-RSA；可限定用户通过哪些 IP 地址使用 SSH 服务对设备进行访问；应支持必要时关闭 SSH 服务。

- 使用 SNMPv3 或安全性更强的管理网络，与网管系统建立可信任连接。
- 提供一个管理端口，从物理上实现管理和服务的分离。
- 具备日志和告警功能。

关于审计和审计记录的具体要求参见 15.4.1。

## 14 性能指标要求

### 14.1 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×2（全双工）。

本标准对三层交换的吞吐量本标准不作规定。

### 14.2 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。

### 14.3 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。

### 14.4 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。

### 14.5 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。

### 14.6 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定。

### 14.7 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。

本规范不强制实现避免队头阻塞，但建议生产厂家实现。

#### 14.8 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。

本标准建议千兆以上端口平均 MAC 地址缓存能力不低于 1024 个。

#### 14.9 地址学习能力

交换机学习新 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大于 1000 个/秒。

#### 14.10 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。

本标准建议吞吐量下的二层帧转发和三层分组转发的时延不超过 1ms。

#### 14.11 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定。

#### 14.12 丢包率

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

本标准建议交换机整机满负荷情况下的三层交换丢包率为 0。

本标准建议交换机整机满负荷情况下的二层交换丢包率为 0。

#### 14.13 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。

#### 14.14 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。

#### 14.15 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定。只作为重要的性能指标供比较。

#### 14.16 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交换机作规定。建议系统的无故障工作时间 MTBF 大于 17520h。建议系统故障恢复时间小于 1h。建议主要部件热备份冗余。

#### 14.17 VLAN 数量

VLAN 数量指系统支持的 VLAN 个数。

本标准建议交换机所支持的 VLAN 数量大于等于交换机的端口数量。

## 15 运行与维护

### 15.1 定义

下面行为必须包含在交换机的 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之中协调上述行为。

### 15.2 交换机初始化

#### 15.2.1 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。

#### 15.2.2 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，并存储在非一不稳定存储器中，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 128）；
- 对应于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。

### 15.3 运行和维护具体规定

#### 15.3.1 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 Telnet 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。

### 15.3.2 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。

### 15.3.3 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主机，管理员能看到可能存在的问题。

## 15.4 安全性考虑

### 15.4.1 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。

### 15.4.2 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。

## 16 网络管理协议

### 16.1 简单网络管理协议-SNMP

#### 16.1.1 SNMP 协议元素

交换机必须支持 IETF RFC1902、IETF RFC1903、IETF RFC1904、IETF RFC1905、IETF 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，IETF RFC1418，IETF 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（IETF 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 IETF RFC1224 中描述的异步告警管理算法来实现上述机制。

#### 16.1.2 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一不稳定的存储器内。

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。

#### 16.1.3 标准 MIBS

所有关于交换机配置的 MIB 都应实现：

- MIB-II（IETF RFC1213）中的系统、接口、IP、ICMP 和 UDP 组必须实现。
- 接口扩展 MIB（IETF RFC1229）必须实现。

- 如果交换机实现 TCP（例如，远程登录），MIB-II（IETF RFC1213）中的 TCP 组必须实现。
- 以太网—链路 MIB（IETF RFC1643）必须实现。
- 网桥 MIB（IETF RFC1493）必须实现。
- 远程网络监视 MIB（IETF 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（IETF RFC2613）可选支持。

#### 16.1.4 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 IETF RFC1155，并且必须以 IETF RFC1212 指定的方式描述。

#### 16.2 简单网络管理协议版本 3（SNMPv3）

交换机应可选支持简单网络管理协议版本 3，其实现应符合 IETF RFC3414 中的规定。

SNMPv3 描述了简单网络管理协议基于用户的安全模型，在实现上应提供如下安全服务：

- 数据完整性，数据没有以未经授权的方式被更改或者是破坏；
- 数据源认证，数据接收者能够认证数据源的身份；
- 数据保密性，数据信息不被未授权的个人、实体或者是进程所使用；
- 通信消息的实时性和防重放保护。

SNMPv3 安全协议应实现三个不同的模块，每个模块完成相应功能，提供相应的安全服务。

- 认证模块：实现数据完整性和数据源认证功能；
- 实时模块：实现消息的延迟或重放保护；
- 私密模块：保护消息的内容不被暴露。

在基于用户的安全模型中，认证协议必须支持 HMAC-MD5-96，可选支持 HMAC-SHA-

96；私密协议必须支持 CBC-DES 对称加密协议。

#### 17 环境要求

具有 IPv6 路由功能的交换机设备的环境要求应符合 GB/T 2423.1-2001《电工电子产品环境试验 第 2 部分：试验方法 试验 A：低温》、GB/T 2423.2-2001《电工电子产品环境试验 第 2 部分：试验方法 试验 B：高温》、GB/T 2423.3-2006《电工电子产品环境试验 第 2 部分：试验方法 试验 Cab：恒定湿热试验》、GB/T 2423.4《电工电子产品环境试验规程 试验 Db：交变湿热试验方法》和 GB/T 2423.9-2001《电工电子产品环境试验 第 2 部分：试验方法 试验 Cb：设备用恒定湿热》中的规定。

## 附录 A

### (规范性附录)

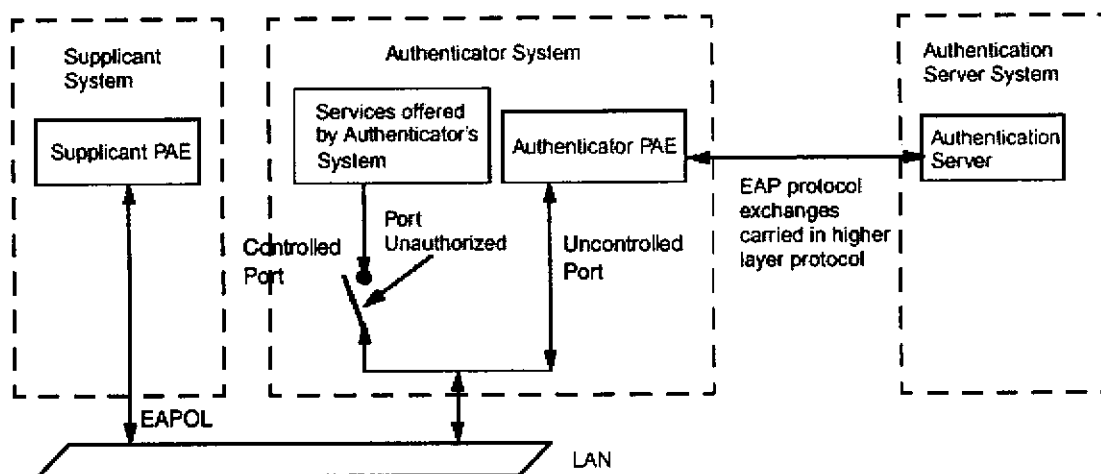
### 802.1X

#### 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 的体系结构中包括三个部分：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，实现方式有下列三种：

- 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 地址和 VLANID，和物理端口组合作为受控端口的标识。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小于15s。