ETHERNET OAM

A Technical Overview

Copyright © 2010, Juniper Networks, Inc.

Table of Contents

Table of Figures

Copyright © 2010, Juniper Networks, Inc.

Executive Summary

As Ethernet emerged from being a LAN technology into a carrier-class technology, Ethernet OAM has been developed to ease the operations, administration, and maintenance of complex Ethernet service provider networks and lower their operational expenditures. This paper gives a technical overview of Ethernet OAM. It discusses link fault management, connectivity fault management, and performance monitoring. In the use case section, it discusses cases in which Ethernet OAM is deployed and how Ethernet OAM adds value to the deployment of related technologies. The technologies discussed are MPLS fast reroute, bidirectional detection forwarding, Layer 2 VPNs, virtual private LAN service, and Ethernet ring protection switching.

Market Motivation

Ethernet first gained dominance as a LAN medium. The high volume has driven down the cost of Ethernet devices. To take advantage of the economy of scale created by Ethernet's dominance and operators' familiarity with Ethernet, operators deploy Ethernet service in WAN and metropolitan area networks (MANs) to meet the demand driven by tremendous growth in data traffic and bandwidth-intensive IP-based applications. By deploying Ethernet service, operators lower capital expenditures compared to capital expenditures incurred from deployment of traditional services such as SONET, SDH, and Frame Relay. Ethernet devices are placed at mission-critical points in operators' networks. In order to offer carrier-class Ethernet service, operators use Ethernet OAM mechanisms to deliver quality of experience that their customers expect and meet strict service-level agreements (SLAs) for five 9's resiliency and throughput performance contracted to customers. Ethernet OAM is driven by the following requirements:

- Ethernet OAM should be similar to, or better than, the ones operators have been accustomed to using with traditional carrier services to manage their networks. ATM, for example, provides extensive OAM capabilities.
- In addition to the capital expenditures realized by deploying Ethernet devices, operators use Ethernet OAM to reduce operational expenditures and downtime in the network.
- Ethernet OAM enables operators to manage services that span multiple interconnecting provider bridged networks. Different and independent administrative domains can administer these networks.

Ethernet OAM performs the functions identified in Table 1.

Table 1: OAM Functions

CATEGORY	FUNCTIONS
Operations	• Automatically and proactively monitors environment. • Quickly detects, isolates faults—and recovers from them. • Alerts administrators.
Administration	• Monitors performance. • Facilitates capacity planning.
Maintenance	• Performs upgrades. • Deploys new features. • Monitors network health.

Copyright © 2010, Juniper Networks, Inc.

Ethernet OAM Overview

To address the requirements set forth by Ethernet operators, various standardization bodies developed standards for Ethernet OAM to operate at different OAM layers. Table 2 shows how the OAM layers, their functions, and corresponding standards work.

Table 2: OAM Layers

OAM LAYERS	FUNCTIONS	STANDARDS
Transport Layer	• Ensures that two directly connected peers maintain bidirectional communication. • Must detect "link down" failure and notify higher layer for protocol to reroute around the failure. • Monitors link quality to ensure that performance meets an acceptable level.	• IEEE 802.3 The latest IEEE 802.3 incorporates IEEE 802.3ah, which originally defined OAM link fault management mechanisms for Ethernet in the First Mile.
Connectivity Layer	• Monitors path between two non-adjacent devices.	• IEEE 802.1ag • ITU-T Y.1731 • MEF Specification
Service Layer	• Measures and represents the status of the services as seen by the customer. • Produces metrics such as throughput, round-trip delay, and jitter that need to be monitored to meet the SLAs contracted between the provider and the customer.	• ITU-T Y.1731 • MEF Specification

Figure 1 shows an example of how the different layers of OAM can be used in a network. In the example, transport layer OAM mechanisms are supported between devices A and B and between devices B and C. Connectivity layer OAM mechanisms are supported between devices A and C. Service layer OAM mechanisms are supported between customer premise X and customer premise Y.

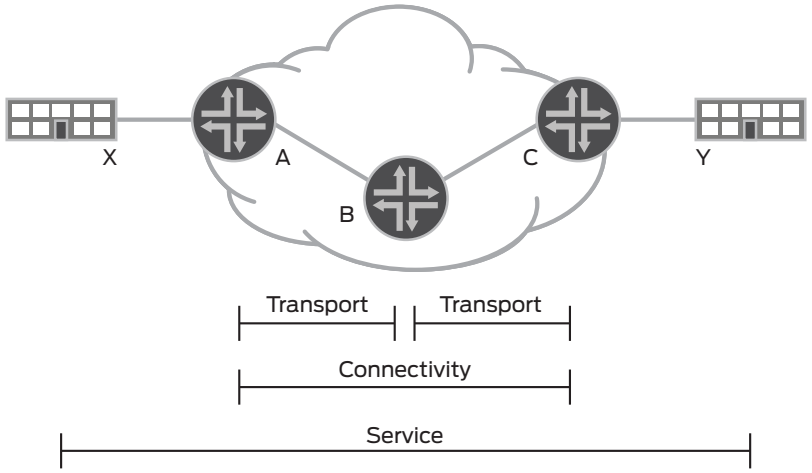

Figure 1. OAM at transport layer, connectivity layer, and service layer

Table 3 lists the protocols that operate at the different OAM layers. In the following sections, each protocol is discussed in detail.

Table 3: OAM Protocols

OAM LAYERS	PROTOCOLS
Transport Layer	Link fault management
Connectivity Layer	Connectivity fault management including linktrace, continuity check, and loopback protocols
Service Layer	Performance monitoring including frame loss, frame delay, and throughput measurements

Copyright © 2010, Juniper Networks, Inc.

Link Fault Management

At the transport layer, link fault management (LFM) OAM mechanisms are used for monitoring link operation for physical point-to-point or emulated point-to-point IEEE 802.3 links connecting peer OAM entities. IEEE 802.3 defines OAM mechanisms including keep lives and loopbacks for LFM. Per IEEE 802.3, LFM supports:

- Remote failure indication during fault conditions.
- Remote loopback mode used for fault isolation and link performance testing. An Ethernet device with OAM enabled sends loopback requests to its peer. The peer in loopback mode loops all protocol data units (PDUs) back except OAM PDUs.
- Link monitoring, which supports event notification and polling of IEEE 802.3 Clause 30 MIB variables contain information about the link.
- Optional implementation and activation of OAM. OAM can be enabled on a subset of ports or all ports on a given system.
- OAM discovery mechanism to determine if the remote peer has OAM enabled and if its configured parameters and supported functions are compatible with those on its peer. During the discovery process, an Ethernet device can enable OAM functionality based on the information stored in received OAM PDUs and based on local and remote states and configuration settings.
- Extension mechanism available for higher-level management applications. The extension mechanism uses organization-specific OAM PDU; organization-specific information type, length, and value (TLV); and organization- specific event TLV.

OAM PDUs are IEEE 802.3 slow protocol frames containing OAM control and status information used to monitor, test, and troubleshoot links. IEEE 802.3 slow protocol frames are used in order to restrict bandwidth consumption and performance demand. The restrictions are as follows:

- No more than 10 frames must be transmitted in any one-second period per slow protocol subtype.
- The maximum number of slow protocol subtypes is 10.
- The MAC client data generated by any of these protocols must be no larger than maximum basic data size of IEEE 802.3.

OAM uses slow protocol multicast address as the destination address for OAM PDUs. The subtype value for OAM PDUs is 88-09. Since LFM applies to single links only, OAM PDUs are not forwarded by MAC clients such as bridges and switches. OAM PDUs can carry notification of critical link events. Table 4 lists the critical link events and their description.

Table 4: OAM Critical Link Events and Description

CRITICAL LINK EVENTS	DESCRIPTION
Link fault	Ethernet device has determined a fault has occurred in the receive direction.
Dying gasp	An unrecoverable local failure condition has occurred.
Critical event	An unspecified critical event has occurred.

OAM PDUs can also carry notification of non-critical link events such as errored frame event, errored frame period event, errored frame seconds summary event, and organization-specific event.

An Ethernet device can support OAM in active mode and/or passive mode although the device can run in only one mode at a time. Table 5 lists the features supported in each mode.

Table 5: Features Supported in Active Mode and Passive Mode

FEATURE DESCRIPTION	ACTIVE MODE	PASSIVE MODE
Initiates OAM discovery process	Yes	No
Reacts to OAM discovery process initiation	Yes	Yes
Required to send information OAM PDUs	Yes	Yes
Permitted to send event notification OAM PDUs	Yes	Yes
Permitted to send variable request OAM PDUs	Yes	No
Permitted to send variable response OAM PDUs	Yes	Yes
Permitted to send loopback control OAM PDUs	Yes	No
Reacts to loopback control OAM PDUs	Yes	Yes
Permitted to send organization-specific OAM PDUs	Yes	Yes

Copyright © 2010, Juniper Networks, Inc.

Connectivity Fault Management

Interconnected provider bridged networks pose a set of challenges to operators. Operators offer services that span multiple interconnected provider bridged networks administered by diverse administrative domains. A provider often contracts with another provider for use and management of equipment it needs in a physical location where it does not own any equipment. Providers might give restricted management access to their networks to each other. Unlike local area bridged networks, where equipment might be easily accessible, provider bridged network equipment can be difficult and expensive to access. Connectivity fault management (CFM) addresses these issues. Diverse administrative domains perform CFM functions to detect, isolate, and correct connectivity faults with minimum access to each other's equipment. Table 6 describes the five CFM functions.

Table 6: CFM Funcions

FUNCTION	DESCRIPTION
Path discovery	Administrator uses linktrace protocol to determine the path taken to a target MAC address inside a maintenance association.
Fault detection	Administrator uses continuity check protocol to detect connectivity faults and unintended connectivity between service instances.
Fault verification and isolation	Administrator uses loopback protocol and linktrace protocol to verify and isolate a connectivity fault after the fault is detected.
Fault notification	A MEP triggers unsolicited fault notification when it detects connectivity fault in the maintenance association. If the CFM MIB module is implemented, the MEP sends an SNMP notification.
Fault recovery	Connectivity faults can be recovered by administrator's actions.

CFM operates at the connectivity layer of OAM monitoring paths between non-adjacent devices. CFM partitions a network into hierarchical administrative domains called maintenance domains. A maintenance domain, controlled by a single operator, is the part of a network in which CFM is enabled. Each maintenance domain is assigned a unique level number ranging from 0 to 7. The level numbers are assigned upon mutual agreement by the different administrative domains if the maintenance domain levels are shared among the domains. Maintenance domains can be nested but not overlapped. A maintenance domain is visible only to the immediately higher maintenance domain, also known as the enclosing maintenance domain, and not to any other maintenance domain. This requirement ensures that administrators at every maintenance domain level are shielded from having to know the internals of lower maintenance domain levels to operate their maintenance domain.

Maintenance points (MPs) are CFM protocol shims. A MP can be a maintenance association endpoint (MEP) or a maintenance association intermediate point (MIP). Each MEP has a direction, down or up. Down MEP receives CFM PDUs from the LAN and sends CFM PDUs towards the LAN. Up MEP receives CFM PDUs from a bridge relay entity and sends CFM PDUs towards the bridge relay entity on a bridge. End stations support down MEPs only, as they have no bridge relay entities.

A MEP resides on a station or bridge at the edge of a maintenance domain. Representing a service provided to a customer, a service instance can span more than one interconnected provider bridged network. A unique virtual LAN identifier (VLAN ID) identifies each service instance. A maintenance association is used to monitor connectivity for a single service instance within the maintenance domain. It consists of a set of MEPs configured with the same maintenance association identifier and the same maintenance domain level. Multiple maintenance associations can be configured on a single bridge port. The bridge multiplexes CFM PDUs to the different maintenance associations using VLAN tag and maintenance domain level. A MEP proactively transmits continuity check messages (CCMs). A DOWN MEP multicasts CCM on the logical interface configured for the MEP. An UP MEP multicasts CCM to all logical interfaces in a routing instance. A MEP maintains MEP CCM database processes received CFM frames, dropping them if the maintenance domain level in the CFM frames is at a lower level than the configured maintenance domain level in the MEP. Triggered by operator command, a MEP transmits loopback messages (LBMs) and linktrace messages (LTMs).

A MIP is a passive functional entity located at intermediate points in a maintenance domain. Frames flow through MIPs in transit from MEP to MEP inside a maintenance domain. A MIP does not initiate OAM frames. Rather, it responds to LBM and LTM. It can maintain an optional MIP CCM database.

Linktrace Output Multiplexer is the third CFM protocol shim type. A linktrace responder on a bridge port sends LTMs through the Linktrace Output Multiplexer, thus preventing the LTMs from being sent back to the bridge forwarding process and getting forwarded.

Copyright © 2010, Juniper Networks, Inc.

When link aggregation is configured, a MP can be created to monitor the aggregated link. Additional MPs can be created to monitor connectivity of the individual links.

CFM supports point-to-point and point-to-multipoint services, which makes it suitable for true bridge environments. It supports three protocols with three message types.

Table 7: CFM Protocols and Message Types

CFM PROTOCOL	CFM MESSAGE TYPE	CFM MESSAGE TYPE ACRONYM
Linktrace Protocol	Linktrace Message	LTM
Continuity Check Protocol	Continuity Check Message	CCM
Loopback Protocol	Loopback Message	LBM

Linktrace Protocol

Operators trigger linktrace protocol to perform path discovery and fault isolation in their networks. As part of the linktrace protocol, a MEP multicasts a LTM using linktrace message group destination MAC address, 01-80-C2-00-00-3y. The maintenance domain level of the LTM plus eight is used for the "y" address bits. For example, a LTM at level 0 has destination address 01-80-C2-00-00-38, and a LTM at level 7 has destination address 01-80-C2-00-00-3F. The destination address of the replying MP is embedded in the payload of the LTM.

A MEP transmits a LTM over a maintenance association to neighboring MIPs, from MIP to MIP, to the terminating MP at the end of the path. Only one egress port on a bridge sends LTMs. The LTM traverses through the bridged network until it reaches a MEP at an equal or higher maintenance domain level or a MIP at an equal maintenance domain level. A MEP at a higher maintenance domain level discards the LTM. The LTM at an equal maintenance domain level is sent to the linktrace responder. The linktrace responder is responsible for processing LTMs, forwarding LTMs, and sending back linktrace replies (LTRs). The MIPs along the path and the terminating MP send unicast LTRs, after a random delay of less than one second, back to the originating MEP. With the random delay in place, the originating MEP is less likely to be inundated by incoming LTRs. If an MIP is not the terminating MP, it forwards LTMs as well as replies to them.

Each LTM has a LTM transaction identifier. LTM transaction identifiers that are transmitted inside LTMs are unique for a MEP for at least five seconds so that LTRs from slow MPs can be matched with the corresponding LTMs. Using the LTRs collected, the originating MEP builds the sequence of MPs traversed by the initial LTM. The administrator can then determine the path taken from the MEP to the destination MAC address by examining the sequence of MPs. The difference between the path taken by the LTM and the expected sequence helps pinpoint the location of a fault. It is possible for multiple LTRs to return along multiple paths. In such cases, the reply time to live field in the LTRs is used to order returning LTRs along a single path.

Because target MAC address of an LTM can be deleted from the filtering databases immediately after a topology change, or aged out and be deleted shortly after being learned, CFM has three possible ways to alleviate this problem.

1. Send LTMs immediately after detecting a fault to prevent the target MAC address from being aged out.

2. Use optional MIP database to store information about the target MP at the MIPs along the path.

3. Send periodic LTMs to maintain path information.

Continuity Check Protocol

Continuity Check protocol is used to detect connectivity failures and unintended connectivity. Periodically, each MEP transmits a multicast continuity check message (CCM) embedded with identity of the MEP and maintenance association. MEPs use CCM group destination address, 01-80-C2-00-00-3y, for the destination MAC address in CCM frames. The maintenance domain level of the CCM is used for the "y" address bits. For example, if the maintenance domain level of the CCM is 0, then the CCM destination address is 01-80-C2-00-00-30. If the maintenance domain level of the CCM is 7, then the destination address is 01-80-C2-00-00-37. The frames are sent with sequence numbers. Multicast frames reduce bandwidth requirements in a full mesh. In addition, they allow detection of accidentally cross-connected MEPs belonging to different service instances. The transmission rate for CCMs is configurable. As a MEP receives CCMs from other MEPs, it determines any discrepancies between the information received and waited for. The MEP processes CCMs at maintenance domain level equal or lower than its maintenance domain level and uses the information to update its CCM database. Continuity check function on MIPs and the MIP CCM database are optional. A MIP processes CCM at it maintenance domain level and updates the MIP CCM database.

Copyright © 2010, Juniper Networks, Inc.

The following are types of connectivity fault that can occur in the network.

- MEP failure or a network failure is detected when a MEP fails to receive three consecutive CCMs from any one of the MEPs in its maintenance association. The MEP uses its own configured timer to detect CCM loss. If the timer interval is set to 3.3 ms or 10 ms, the MEP can detect failures within 50 ms. It sends CCMs with Remote Detection (RDI) bit set after detecting loss of CCM from a MEP. Upon receiving the first CCM with RDI bit set, the peer MEP detects the RDI condition. When the defect condition clears, the MEP sends CCMs with the RDI bit cleared. If a MEP does not receive the RDI, the maintenance association has no defect. This allows the administrator to inspect a MEP and determine if there is any defect in the maintenance association.
- A configuration error is detected when a MEP receives a CCM with an incorrect transmission interval.
- A configuration error or cross-connect error is detected when a MEP receives a CCM with an incorrect MEP identifier or maintenance association identifier.
- A configuration error or cross-connect error is detected when a MEP receives a CCM with a maintenance domain level lower than that of the MEP.
- A MEP receives indication of a failed bridge port or aggregate port from a CCM containing a port status TLV or interface status TLV.

Loopback Protocol

Loopback protocol is used to verify and isolate connectivity faults. An administrator can trigger a MEP to send one or more loopback messages (LBMs) with an arbitrary amount of data. If the MEP does not receive a valid LTR corresponding to the LBM, the administrator knows a connectivity fault exists. The receiving MP turns the LBM at its maintenance domain level only into a loopback reply (LBR) back to the originating MEP. In response to a multicast LBM frame, the receiving MP waits a random delay between zero and one second before sending a LBR. The source address from the LBM is used as the destination address for the LBR. The source address of the LBR is the MAC address of the receiving MP. The receiving MP changes the opcode in the frame from LBM to LBR. The originating MEP keeps count of the LBRs from other MPs by incrementing in-sequence counter and out-of-sequence counter. It correlates received LBRs with transmitted LBMs using loopback transaction identifier in the loopback frames. A LBR is valid if it has an expected transaction identifier and is received by the originating MEP within five seconds after transmitting the initial LBM.

Nested Maintenance Domains Example

Figure 2 illustrates an example of entities in nested maintenance domains. The provider offers a service to the customer that uses equipment from two independent operators. The different maintenance domains negotiate(?) on their assigned maintenance domain level numbers because the maintenance domain level numbers cannot be shared by the maintenance domains. However, if the OAM flows are distinguishable by VLAN tag or priority tag in the Ethernet MAC encapsulation of the frames, all eight maintenance domain levels are available to be assigned independently. For example, two maintenance domains can be assigned the same maintenance domain level number if their OAM flows have different VLAN tags.

The example shows four maintenance domain levels—customer maintenance domain at maintenance domain level 5, provider maintenance domain at maintenance domain level 3, operator maintenance domain at maintenance domain level 2, and physical maintenance domain at maintenance domain level 0. Note the configuration of UP MEPs, DOWN MEPs, MIPs, maintenance associations, and maintenance domains. In the example, at the customer maintenance domain level 5, two pieces of customer equipment—devices 1 and 6, each configured with a DOWN MEP—are connected to a provider's network that comprises equipment from two operators, Operator A and Operator B. The two MEPs receive CFM PDUs from the LAN and send them to the LAN. Two MIPs are configured at devices 2 and 5 since these are intermediate points in the customer maintenance domain. At the provider maintenance domain level 3, the UP MEPs residing on bridges 2 and 5 serve as endpoints in the maintenance domain. These UP MEPs receive traffic from the bridge relay entity and send traffic to the bridge relay entity. The MIPs at that level reside on bridges 3 and 4. Two maintenance domains are created, one for each operator at the operator maintenance domain level. Bridges 2 and 3 are provider bridges in Operator A's network. Bridges 4 and 5 are provider bridges in Operator B's network. Operator A maintenance domain and Operator B maintenance domain are enclosed inside the provider maintenance domain. The two operators' maintenance domains do not overlap. The two operators' networks are connected by LAN or emulated LAN between bridges 3 and 4. CFM PDUs flow in the path shown in the figure.

Copyright © 2010, Juniper Networks, Inc.

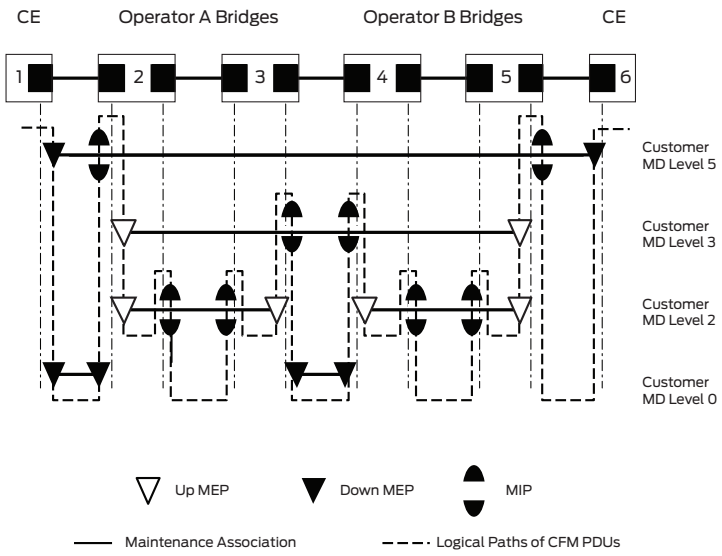

Figure 2. Entities in nested maintenance domains

Y.1731

ITU-T developed Y.1731 as a recommendation for OAM functions and mechanisms for Ethernet-based networks. Both Y.1731 and IEEE 802.1ag define specifications for connection fault management. They are compatible and aligned with each other. Y.1731 defines additional OAM functions such as ETH-AIS, ETH-LCK, ETH-TEST, ETH-APS, ETH-MCC, ETH-EXP, and ETH-VSP, and performance monitoring.

ETH-AIS

Ethernet alarm indication signal function (ETH-AIS) allows a customer who deploys an Ethernet service to tell if a connectivity fault exists at the current level or at a level below. If the fault occurs at the current level, the customer can address the fault. If the fault occurs at a level below, the customer can contact the provider to address the fault. AIS frames are sent to a level above the level where the fault occurs.

Per Y.1731, a server MEP represents the compound function of the server layer termination function and server/ETH adaptation function, which is used to notify the ETH layer MEPs upon failure detection by the server layer termination function or server/ETH adaptation function. The server layer termination function is expected to run OAM mechanisms specific to the server layer. ETH-AIS suppresses alarms following detection of defect conditions at the server (sub) layer. ETH-AIS are not applicable in Spanning Tree Protocol (STP) environments. Transmission of frames with ETH-AIS information can be enabled or disabled on a MEP (or on a server MEP). Frames with ETH-AIS information are issued at a client maintenance level by a MEP, including a server MEP, upon detecting defect conditions. For example, the defect conditions might include:

- Signal fail conditions in the case that ETH-CC is enabled
- AIS condition or LCK condition in the case that ETH-CC is disabled

When a defect condition is detected, a MEP sends AIS frames periodically in a direction opposite to its peer MEP(s). It continues sending the AIS frames until the defect condition no longer exists. In cases where the client layer supports multiple maintenance associations, AIS frames must be sent to those client layer maintenance associations within one second of detection of signal fail condition at the server layer. When a MEP receives the AIS frames, it enters into AIS condition and suppresses loss of continuity alarms associated with all of its peers. For multipoint Ethernet connectivity, a MEP suppresses alarms to every peer MEP while for point-to-point Ethernet connectivity, a MEP suppresses alarms to its only peer. The MEP clears AIS condition after not receiving AIS frames for 3.5 times the transmission period. When AIS condition does not exist and loss of continuity condition is detected, the MEP generates loss of continuity alarms again.

Copyright © 2010, Juniper Networks, Inc.

ETH-RDI

Both 802.1ag and Y.1731 specify Ethernet Remote Defect Indication function (ETH-RDI). ETH-RDI has been discussed earlier in this paper.

ETH-LCK

Ethernet lock signal function is used to signal administrative locking of a server (sub) layer MEP and interruption of data traffic forwarding toward the MEP waiting for the traffic. The transmission and reception of LCK frames is similar to that of AIS frames except that the condition communicated is an administrative locking condition and not a defect condition.

ETH-Test

Administrators trigger in-service and out-of-service diagnostics tests by using the Ethernet test signal function (ETH-test). A MEP sends test frames with specified throughput, test patterns, and frame size. When an out-of-service test is performed at a MEP, client data traffic is stopped. The MEP sends LCK frames at the immediate client maintenance level in the same direction as the direction of the test frame transmission. When an in-service test is performed, client data traffic continues to flow. Care must be taken to ensure that the bandwidth used by the test frames is limited. Each test frame has a sequence number. A MEP cannot repeat sequence numbers within one minute. The receiving MEP accepts the test frames, and detects and reports errors associated with the test frames such as bit errors.

ETH-APS

The Ethernet automatic protection switching function (ETH-APS) controls protection switching operations to enhance reliability. ETH-APS can be either linear APS or Ring APS. Ring APS is discussed further in the use case section. Linear APS is out of the scope for this document.

ETH-MCC

The Ethernet maintenance communication channel function (ETH-MCC) provides a maintenance communication channel between a pair of MEPs. ETH-MCC enables MEPs to perform remote management. Y.1731 does not specify the application of ETH-MCC.

ETH-EXP

Ethernet experimental OAM (ETH-EXP) is used on a temporary basis within an administration domain. Y.1731 does not specify the application of ETH-EXP. Interoperability of the experimental OAM functionality is currently not applicable across different administrative domains.

ETH-VSP

Ethernet vendor-specific OAM (ETH-VSP) is used for vendor-specific OAM functionality across a vendor's own equipment. Vendor-specific OAM functionalities are not currently interoperable across different vendors' equipment.

Performance Monitoring

Y.1731 specifies OAM functions for performance monitoring of Ethernet networks enabling operators to meet strict SLAs. Running at the service layer of OAM, performance monitoring functions ensure that customers receive the level of Ethernet service they pay for. Per Y.1731, OAM functions for performance monitoring allow measurement of three parameters—frame loss ratio, frame delay, and frame delay variation. These performance parameters apply to service frames, which conform to an agreed-upon level of bandwidth profile conformance. IETF RFC 2544 specifies throughput measurement, which is an important component of performance monitoring.

ETH-LM

Frame Loss Measurement function (ETH-LM) maintains counters of received and transmitted service frames between a pair of MEPs. These counters are used to calculate frame loss ratio, which is a ratio of the number of service frames not delivered, divided by the total number of service frames during a time interval. The number of service frames not delivered is the difference between the number of service frames arriving at the ingress Ethernet flow point and the number of service frames delivered at the egress Ethernet flow point in a point-to-point Ethernet connection.

Copyright © 2010, Juniper Networks, Inc.

Dual-ended LM and single-ended LM are two ways ETH-LM can be performed. To perform dual-ended LM, each MEP proactively sends periodic CCM frames to its peer MEP. Each peer MEP terminates the CCM frames and performs near-end and far-end loss measurements using local counters and counter values in the received CCM frames.

To perform single-ended LM, a MEP sends LM request (LMM) frames to its peer MEP upon an on-demand administrative trigger. The peer MEP responds with LM reply (LMR) frames. Using counter values in LMR frames and its local counter value, a MEP performs near-end and far-end loss measurements. The following are the dual-ended and single-ended frame loss formulas.

RxFCl is the value of the local counter for in-profile data frames received from the peer MEP.

TxFCl is the value of the local counter for in-profile data frames transmitted toward the peer MEP.

tc is the reception time of the current frame.

tp is the reception time of the previous frame.

Dual-Ended LM Calculation

Frame lossfar-end = |TxFCb[tc] – TxFCb[tp]| – |RxFCb[tc] – RxFCb[tp]|

Frame lossnear-end = |TxFCf[tc] – TxFCf[tp]| – |RxFCl[tc] – RxFCl[tp]|

TxFCf is the value of the local counter TxFCl at the time of transmission of the CCM frame.

RxFCb is the value of the local counter RxFCl at the time of reception of the last CCM frame from the peer MEP.

TxFCb is the value of TxFCf in the last received CCM frame from the peer MEP.

Single-Ended LM Calculation

Frame lossfar-end = |TxFCf[tc] – TxFCf[tp]| – |RxFCf[tc] – RxFCf[tp]|

Frame lossnear-end = |TxFCb[tc] – TxFCb[tp]| – |RxFCl[tc] – RxFCl[tp]|

TxFCf is the value of the local counter TxFCl at the time of LMM frame transmission.

RxFCf is the value of local counter RxFCl at the time of LMM frame reception.

TxFCb is the value of local counter TxFCl at the time of LMR frame transmission.

ETH-DM

When a MEP is enabled to perform the frame delay and frame delay variation measurement function (ETH-DM), it periodically sends frames with ETH-DM information to its peer MEP. It receives frames with ETH-DM information from its peer MEP. MEPs can use one of two methods to perform ETH-DM, one-way ETH-DM or two-way ETH-DM.

For one-way ETH-DM to work properly, clocks on the peer MEPs must be synchronized. The sending MEP sends 1DM frames including timestamp at transmission time. The receiving MEP calculates the frame delay using the timestamp at the reception of the 1DM frame and the timestamp in the 1DM frame. For one-way frame delay variation measurement, clock synchronization on the peer MEPs is not required. The out-of-phase period can be removed by the difference of subsequent frame delay variation measurements.

If clocks on peer MEPs are not synchronized, a MEP can measure frame delay using two-way ETH-DM. When two-way DM is enabled, a MEP sends ETH-DM request (DMM) frames including timestamp at transmission time. The receiving MEP copies the timestamp into ETH-DM Reply (DMR) and sends that DMR back to the sending MEP. The sending MEP receives the DMR and calculates the two-way frame delay using the timestamp in the DMR and the timestamp at reception of the DMR. Frame delay variation measurement is done by calculating the difference between two subsequent two-way frame delay measurements.

One-Way DM Calculation

Frame Delay = RxTimef – TxTimeStampf

RxTimef is the time at reception of the 1DM frame.

TxTimeStampf is the timestamp at the transmission time of the 1DM frame.

Copyright © 2010, Juniper Networks, Inc.

Two-Way DM Calculation

Frame Delay = (RxTimeb – TxTimeStampf) – (TxTimeStampb – RxTimeStampf)

RxTimeb is the time at reception of the DMR frame.

TxTimeStampf is the timestamp at the transmission time of the DMM frame.

TxTimeStampb is the timestamp at the transmission of the DMR frame.

RxTimeStampf is the timestamp at the reception of the DMM frame.

Throughput Measurement

To perform throughput measurement, a MEP sends unicast loopback or test frames at increasing rate until frames start getting dropped. The rate at which the frames start getting dropped is reported. Frame size is configurable. The throughput measurement can be one-way or two-way.

Use Cases

This section describes how Ethernet OAM is deployed in Ethernet networks working in conjunction with the following technologies.

- MPLS Fast Reroute
- Bidirectional Forwarding Detection
- Layer 2 VPN
- Virtual Private LAN Service
- Ethernet Ring Protection

A brief overview of each of the aforementioned technologies is discussed. For further information on the technologies, refer to references.

MPLS Fast Reroute

In MPLS, fast rerouting is accomplished by pre-computing and pre-establishing a number of detours along the label-switched path (LSP). When a network failure occurs on the current LSP path, the traffic is quickly routed to one of the detours. While MPLS fast reroute does not depend on Ethernet OAM, MPLS fast reroute—working together with Ethernet OAM—can detect faults and recover from faults quickly. Juniper Networks® Junos® operating system provides a command-line interface configuration construct called action profile to associate an action to an event trigger. An administrator can configure an action profile, which associates Ethernet OAM fault events to link-down action. Upon fault detection on the link associated with the interface, the interface is marked down. When the interface comes down, MPLS fast reroute occurs.

Ethernet OAM can be used to detect fault on aggregated Ethernet child links. The state of each child link of an aggregated Ethernet bundle can affect the state of the aggregated bundle. For example, if a fault is detected on a child link—causing bandwidth on the aggregated Ethernet bundle to go below the aggregated Ethernet bandwidth threshold—the interface associated with the entire bundle is marked down, thus triggering MPLS fast reroute. The ability to detect faults on individual child links is particularly useful in cases where other protocols such as Bidirectional Forwarding Detection (BFD) cannot detect faults on child links of aggregated Ethernet bundles.

Multilayer OAM—Ethernet OAM and Bidirectional Forwarding Detection (BFD)

Both BFD and CFM can be used to monitor connectivity between a pair of Ethernet devices. This section describes BFD and explains how BFD and CFM can be deployed.

BFD is a hello protocol that provides a low-overhead and short-duration mechanism to detect fault in a path between a pair of forwarding engines. It is transport aware and service agnostic. It does not need to run on Ethernet-only networks. Timers are specified in microseconds so that sub-second fault detection can be performed. As a single mechanism to detect liveliness over any media at any protocol layer, BFD delivers a consistent service to control protocols acting as its clients. BFD indicates session state transition to and from up state to its clients. When the control protocol is a routing protocol, a fault detected in a path causes the routing protocol to reroute traffic around the path. BFD works in two modes, asynchronous mode and demand mode. In asynchronous mode, unicast BFD packets are sent at regular intervals. In demand mode, unicast BFD packets are sent explicitly upon an administrative trigger.

Copyright © 2010, Juniper Networks, Inc.

A BFD session is bootstrapped by its client. A BFD session has no auto-discovery mechanism to detect neighbors. When a BFD session is established over a point-to-point MPLS LSP, the BFD session is bootstrapped by LSP ping. Although BFD is used for fault detection on the data plane of the LSP, LSP ping can still be used for verifying the data plane against the control plane on the LSP.

LFM, CFM, and BFD can be deployed in any combination depending on the requirements of the network. The deployment of one of these protocols does not preclude the deployment of the other in the same network. For monitoring at Ethernet link level, LFM should be used. Before one decides whether CFM or BFD or both should be deployed in a network, the following points should be considered.

- CFM is service aware and must run on Ethernet networks. It can work with VPLS and pseudowires. BFD can work for networks other than Ethernet. It is suitable for Layer 3 and MPLS.
- BFD uses ping and traceroute for loopback and trace functions. CFM uses loopback and linktrace.
- Entities in CFM such as MEPs, MIPs, maintenance associations, and maintenance domains give flexibility to the administrator but also add complexity to management tasks. BFD configuration is simpler than that for CFM.
- As mentioned earlier, BFD runs on aggregated Ethernet bundles and not on the child links. CFM works on child links as well as aggregated Ethernet bundles.
- For BFD, session down causes IGP and BGP to reroute. For CFM, interface down causes protocols to reroute.
- For MPLS OAM, data plane failure detection for LSPs can trigger fast rerouting. For Ethernet OAM, interface down can trigger fast rerouting.
- Both CFM and BFD detect faults in microseconds, meeting time constraints of applications such as VoIP for which end-to-end failures must recover in 300 ms.

Figure 3 shows LFM, CFM, and BFD running together in an Ethernet network. In the example, LFM runs on each Ethernet link. CFM continuity check and BFD detect connectivity faults between PE1 and PE2.

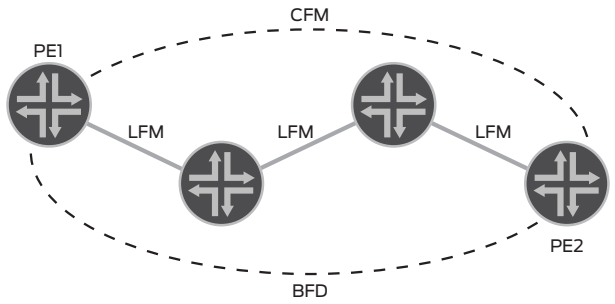

Figure 3. LFM, CFM, BFD over Ethernet

Layer 2 VPN

Figure 4 shows how CFM is used to monitor Layer 2 VPN (L2VPN) carrying Ethernet service between CE1 and CE2 over a pseudowire between PE1 and PE2. The pseudowire is a LSP connecting the two PE routers. A customer maintenance domain is configured with down MEPs at CE1 and CE2 and MIPs at PE1 and PE2. A provider maintenance domain is configured with up MEPs at PE1 and PE2. Two maintenance associations are configured at the physical maintenance domain level.

L2VPN is a type of VPN service used to transport a customer's private Layer 2 traffic (for example, Ethernet, ATM, or Frame Relay) over the service provider's shared IP/MPLS infrastructure. The service PE router must have an interface with circuit cross-connect (CCC) encapsulation to switch the customer edge (CE) traffic to the public network.

Copyright © 2010, Juniper Networks, Inc.

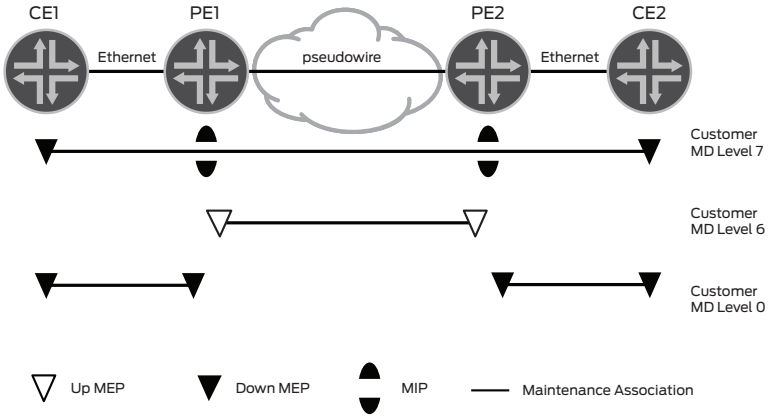

Figure 4. CFM over L2VPN

Virtual Private LAN Service

Virtual Private LAN Service (VPLS) provides Ethernet multipoint-to-multipoint service to multiple dispersed sites connected over an IP or MPLS network. The sites communicate with each other as if they were attached to the same Ethernet LAN. VPLS is a VPN technology that emulates LAN by establishing full-mesh connectivity among participating provider edge (PE) routers. Each participating PE router, acting as an Ethernet bridge, runs a VPLS instance that belongs to a VPLS domain. Every VPLS instance in a VPLS domain builds a pseudowire to every other VPLS instance in that VPLS domain. In MPLS, a pseudowire is a LSP connecting a pair of PE routers.

Since CFM monitors Ethernet networks at a per-service level, it can be used to monitor a VPLS instance. Figure 5 shows a network with a VPLS core connecting two access networks, Access Network-1 and Access Network-2. A full mesh of LSPs is built among the PE routers. A maintenance association is defined to monitor connectivity between PE1 and A1 within Access Network-1 over SVLAN S-VID1. The dotted line represents the current path from CE1 to CE3. An action profile is specified such that loss of connectivity that occurs between PE1 and A1 over SVLAN S-VID1 results in PE1 withdrawing labels to other PE routers. If PE1 loses connectivity to all other PE routers, it could result in notification to A1 to reroute traffic.

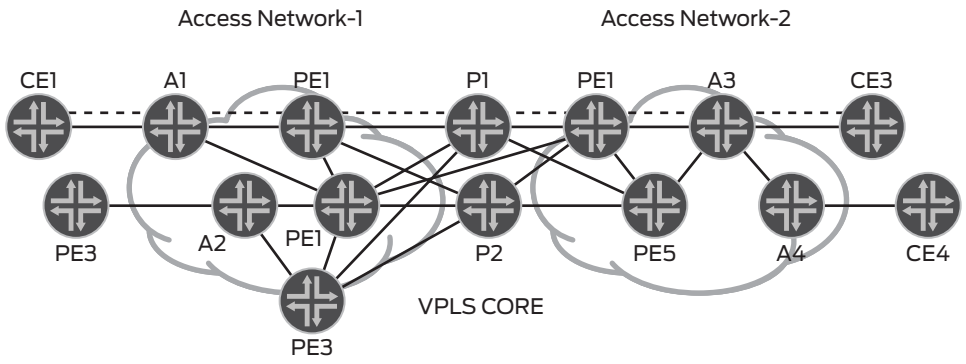

Figure 5. CFM over VPLS

Copyright © 2010, Juniper Networks, Inc.

Ethernet Ring Protection Switching

ITU-T developed G.8032, which specifies Ethernet ring protection switching (ERPS) as a protection mechanism for Ethernet traffic in a ring topology, ensuring no loop is formed in the ring. ERPS uses Ethernet OAM mechanisms for its operation. For typical rings, protection and recovery switching occurs within 50 ms. The minimum number of nodes in a ring is two. Each ring node has ring ports and is connected to two adjacent nodes via two independent ring links. A MEP resides on each ring port, monitoring the attached link. In ERPS, one ring link is designated ring protection link (RPL), which is blocked under normal circumstances so that loop does not form. The ring node that controls the RPL is called the RPL owner.

In Figure 6, nodes A, B, C, and D form an Ethernet ring. Node D is the RPL owner. The RPL is shown blocked, which means traffic cannot flow through it. Two MEPs on two ring ports reside in each ring node. The ring ports are attached to the ring links (RL). These ring links are operating normally in the example.

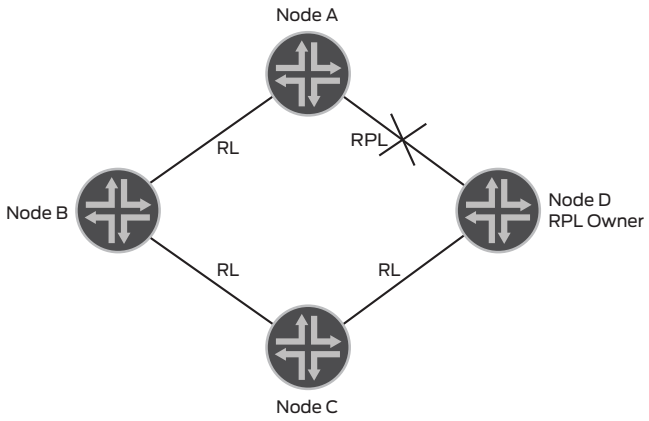

Figure 6. Ethernet ring protection switching

When a MEP detects fault on one of the ring links, signal fail (SF) condition occurs. The nodes adjacent to the failed link initiate the protection mechanism. They send continuous R-APS (SF) messages over both ring ports along both directions while the condition persists. R-APS information is carried in Ethernet OAM APS PDU. The failed link port is blocked. After the RPL owner receives the R-APS message, it unblocks the RPL port, allowing traffic to flow through. Upon the recovery of the failed link, the nodes adjacent to the link send continuous R-APS (no request) messages over both ring ports in both directions. When the RPL owner gets the R-APS (no request) message, it blocks the RPL and sends an R-APS (no request, root blocked) message. This message causes all the other nodes to unblock their blocked ports. The RPL port, however, is kept blocked. The other nodes stop sending R-APS (no request) when they accept R-APS (no request, root blocked).

Conclusion

Ethernet is a MAN and WAN technology as well as a LAN technology. Ethernet OAM provides carrier-class Ethernet to service providers with flexible management tools to manage their complex Ethernet networks. Ethernet OAM is similar to management tools that service providers have been using for traditional technologies such as ATM. This paper describes Ethernet OAM in detail including link fault management, connectivity fault management, and performance monitoring. It includes use cases describing how Ethernet OAM and other related technologies can be deployed in networks.

Copyright © 2010, Juniper Networks, Inc.

Acronyms

APS	Automatic Protection Switching
BFD	Bidirectional Forwarding Detection
BGP	Border Gateway Protocol
CCM	Continuity Check Message
CE	Customer Equipment
CFM	Connectivity Fault Management
ETH	Ethernet MAC layer network
IGP	Interior Gateway Protocol
LFM	Link Fault Management
LMM	Loss Management Message
LMR	Loss Management Reply
LTM	Linktrace Message
LTR	Linktrace Reply
LSP	Label-Switched Path
MAC	Media Access Control
MEP	Maintenance Association Endpoint
MIB	Management Information Base
MIP	Maintenance Association Intermediate Point
MP	Maintenance Association Point
MPLS	Multiprotocol Label Switching
PE	Provider Edge
RL	Ring Link
RPL	Ring Protection Link
SF	Signal Fail
TLV	Type, Length, and Value
VID	VLAN Identifier
VLAN	Virtual Local Area Network
VoIP	Voice Over IP
VPLS	Virtual private LAN service

Copyright © 2010, Juniper Networks, Inc.

References

Draft-ietf-bfd-base-11.txt, Bidirectional Forwarding Detection, D. Katz, D. Ward, 14 January 2010.

IEEE Std 802.1ad-2005, IEEE Standard for Local and Metropolitan Area Networks, Virtual Bridged Local Area Networks, Amendment 4: Provider Bridges, 2005.

IEEE Std 802.1ag-2007, IEEE Standard for Local and Metropolitan Area Networks, Virtual Bridged Local Area Networks, Amendment 5: Connectivity Fault Management, 2007.

IEEE Std 802.3-2008, IEEE Standard for Information Technology-Telecommunications and Information Exchange Between Systems-Local and Metropolitan Area Networks-Specific Requirements-Part 3: Carrier Sense Multiple Access With Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications, 2008.

IETF RFC 4090, Fast Reroute Extensions to RSVP-TE for LSP Tunnels, Pan, et al., May, 2005.

IETF RFC 4761, VPLS using BGP for Autodiscovery and Signaling, Kompella and Rekhter, January 2007.

IETF RFC 4762, VPLS using LDP Signaling, Lasserre and Kompella, January 2007.

ITU-T G.8031 Ethernet Linear Protection Switching, November 2009.

ITU-T G.8032 Ethernet Ring Protection Switching, June 2008.

ITU-T Y.1731 Operation, Administration and Maintenance OAM Functions and Mechanisms for Ethernet-based Networks, June 2008.

MEF 20 User Network Interface (UNI) Type 2 Implementation Agreement, July 2008.

About Juniper Networks

Juniper Networks, Inc. is the leader in high-performance networking. Juniper offers a high-performance network infrastructure that creates a responsive and trusted environment for accelerating the deployment of services and applications over a single network. This fuels high-performance businesses. Additional information can be found at **www.juniper.net**.

Corporate and Sales Headquarters
Juniper Networks, Inc.
1194 North Mathilda Avenue
Sunnyvale, CA 94089 USA
Phone: 888.JUNIPER (888.586.4737)
or 408.745.2000
Fax: 408.745.2100
www.juniper.net

APAC Headquarters
Juniper Networks (Hong Kong)
26/F, Cityplaza One
1111 King's Road
Taikoo Shing, Hong Kong
Phone: 852.2332.3636
Fax: 852.2574.7803

EMEA Headquarters
Juniper Networks Ireland
Airside Business Park
Swords, County Dublin, Ireland
Phone: 35.31.8903.600
EMEA Sales: 00800.4586.4737
Fax: 35.31.8903.601

To purchase Juniper Networks solutions, please contact your Juniper Networks representative at 1-866-298-6428 or authorized reseller.

Copyright 2010 Juniper Networks, Inc. All rights reserved. Juniper Networks, the Juniper Networks logo, Junos, NetScreen, and ScreenOS are registered trademarks of Juniper Networks, Inc. in the United States and other countries. All other trademarks, service marks, registered marks, or registered service marks are the property of their respective owners. Juniper Networks assumes no responsibility for any inaccuracies in this document. Juniper Networks reserves the right to change, modify, transfer, or otherwise revise this publication without notice.

2000364-001-EN Sept 2010 Printed on recycled paper