- Serial Over LAN (SoL) — 300 Kb/s

10.2.1 DMTF NC-SI Mode

The 82599 supports all the mandatory features of the DMTF NC-SI spec rev1.0.0a.

10.2.1.1 Supported Features

Table 10-1 lists the commands supported by the 82599.

Table 10-1 Supported NC-SI Commands

Command	Supported
Clear Initial State	Yes
Get Version ID	Yes
Get Parameters	Yes
Get Controller Packet Statistics	No
Get Link Status	Yes
Enable Channel	Yes
Disable Channel	Yes
Reset Channel	Yes
Enable VLAN	Yes (filtering only by the VLAN ID. No filtering by the User priority)
Disable VLAN	Yes
Enable BCast	Yes
Disable BCast	Yes
Set MAC Address	Yes
Get NC-SI Statistics	Yes, partially
Set NC-SI Flow Control	No
Set Link Command	Yes (support for 10 GbE is not fully defined in the specification)
Enable Global MCast Filter	Yes
Disable Global MCast Filter	Yes
Get Capabilities	Yes

Table 10-1 Supported NC-SI Commands (Continued)

Command	Supported
Set VLAN Filters	Yes
AEN Enable	Yes
Get Pass-Through Statistics	Yes, partially
Select Package	Yes
Deselect Package	Yes
Enable Channel Network Tx	Yes
Disable Channel Network Tx	Yes
OEM Command	Yes

Table 10-2 lists the NC-SI features supported by the 82599:

Table 10-2 Optional NC-SI Features Support

Feature	Supported	Details
AENs	Yes.	*Note:* The driver state AEN might be emitted up to 15 seconds after actual driver change.
Get NC-SI statistics command	Yes, partially	Supports the following counters: 1-4, 7
Get NC-SI pass-through statistics command	Yes, partially	Supports the following counters: 2 Supports the following counters only when the operating system is down: 1, 6, 7
VLAN modes	Yes, partially	Supports only modes 1,3
Buffering capabilities	Yes	8 K
Ethernet MAC address filters	Yes	Supports 2 Ethernet MAC addresses as mixed per port
Channel count	Yes	Supports 2 channels
VLAN filters	Yes	Supports 8 VLAN filters per port
Broadcast filters	Yes	Supports the following filters: • ARP • DHCP • NetBIOS
Multicast filters	Yes	Supports the following filters (supported only when all three are enabled): • IPv6 neighbor advertisement • IPv6 router advertisement • DHCPv6 relay and server multicast

Table 10-2 Optional NC-SI Features Support (Continued)

Feature	Supported	Details
NC-SI flow control command	No	
Hardware arbitration	No	

10.2.2 SMBus Pass Through (PT) Functionality

When operating in SMBus mode, the 82599 provides the following manageability services to the BMC on top of the pass through traffic functionality:

- ARP handling — The 82599 can be programmed to auto-ARP replying for ARP request packets and sending gratuitous ARP to reduce the traffic over the SMBus.
- Teaming and fail-over — The 82599 can be configured to either teaming or non-teaming modes. When operated in teaming mode the 82599 can also provide auto fail-over configurations as detailed in the following sub-sections.
- Default configuration of filters by EEPROM — When working in SMBus mode, the default values of the manageability receive filters can be set according to the PT LAN (Section 6.4.3) and flex TCO EEPROM structure (Section 6.4.5).

10.2.2.1 Pass Through (PT) Modes

PT configuration depends on how the LAN ports are configured. If the LAN ports are configured as two different channels (non-teaming mode) then the 82599 is presented on the manageability link as two different devices (via two different SMBus addresses) on which each device is connected to a different LAN port. In this mode (the same as in the LAN channels), there is no logical connection between the two devices. In this mode, the fail-over between the two LAN ports are done by the external BMC (by sending/receiving packets through different devices). The status reports to the BMC, ARP handling, DHCP and other pass through functionality are unique for each port.

When the 82599 operates in teaming mode, it presents itself on the SMBus as a single device. In this mode, the external BMC is not aware that there are two LAN ports. The 82599 determines how to route the packets that it receives on the manageability channel according to the fail-over algorithm. The status reports to the BMC and other pass through configurations are common to both ports.

In pass through mode most of the manageability traffic is handled by the BMC. However, portion of the network traffic can be offloaded and by the 82599 as described in the following sub-sections. This configuration can be done by issuing configuration commands over the SMBus channel or the 82599 can load it from its EEPROM at power up (or both).

10.2.2.2 LAN Fail-Over in LAN Teaming Mode

Manageability fail-over is the ability to detect that the LAN connection on one port is lost, and enable the other port for manageability traffic. When the 82599 operates in teaming mode, the operating system and the external BMC consider it as one logical network device. The decision on which of the 82599 ports are used is done internally by the 82599 (or by the ANS driver in case of the regular receive/transmit traffic). This section deals with fail-over in teaming mode only. In non-teaming mode, the external BMC should consider the 82599's network ports as two different network devices, and the BMC is solely responsible for the fail-over mechanism.

In teaming mode, the 82599 maps both network ports into a single SMBus slave device. The 82599 automatically handles the configurations of both network ports. Thus, for configurations, receiving and transmitting the BMC should consider both ports as a single entity.

When the currently active transmission port becomes unavailable (such as the link is down), the 82599 automatically switches transmission to the other port. Thus, as long as one of the ports is valid, the BMC will have a valid link indication for the SMBus slave.

Note: As both ports might be active (such as with a valid link) packets might be received on the currently non-active port. To avoid packet duplication, failover should not be enabled when connected to a hub.

Note: Fail over and teaming are not supported in NC-SI mode.

10.2.2.2.1 Port Switching (Fail-Over)

While in teaming mode, transmit traffic is always transmitted by the 82599 through only one of the ports at any given time. The 82599 might switch the traffic transmission between ports under any of the following conditions:

1. The current transmitting port link is not available
2. The preferred primary port is enabled and becomes available for transmission.

10.2.2.2.2 Driver Interactions

When the LAN driver is present, the decision to switch between the two ports is done by the driver. When the driver is absent, this decision is done internally by the 82599.

Note: When the driver releases teaming mode (such as, when the system state changes), the 82599 reconfigures the LAN ports to teaming mode. The 82599 accomplishes this by re-setting the Ethernet MAC address of the two ports to be the teaming address in order to re-start teaming. This is followed by transmission of gratuitous ARP packets to notify the network of teaming mode re-setting.

10.2.2.2.3 Fail-Over Configuration

Fail-over operation is configured through the fail-over configuration structure (see Section 10.2.2.2.4).

The BMC should configure this register after a the 82599 initialization indication (following a firmware reset). The different configurations available to the BMC are detailed in this section.

Note: In teaming mode both ports should be configured with the same receive manageability filters parameters (EEPROM sections for port 0 and port 1 should be identical).

Preferred Primary Port — The BMC might choose one of the network ports (LAN0 or LAN1) as a preferred primary port for packet transmission. The 82599 always switches to the preferred primary port when it is available.

Gratuitous ARPs — In order to notify the link partner that a port switching has occurred, the 82599 can be configured to automatically send gratuitous ARPs. These gratuitous ARPs cause the link partner to update its ARP tables to reflect the change. The BMC might enable/disable gratuitous ARPs, configure the number of gratuitous ARPs or the interval between them by modifying the Fail Over configuration register.

Link Down Timeout — The BMC can control the timeout for a link to be considered invalid. The 82599 waits this timeout before attempting to switch from an inactive port.

10.2.2.2.4 Fail-Over Structure

The fail-over structure (listed in the following table) is loaded on power up from the EEPROM (see Section 6.4.4.5), or through the Set Fail-Over Configuration host command by the LAN driver (see Section 10.5.3.8). The bits in this register can also be modified by the 82599 hardware reflecting its current state.

Field	Bit(s)	RW	Init Val	Description
RMP0EN	0	RO	0x1	RCV MNG port 0 Enable. When this bit is set, it reports that MNG traffic is received from port 0.
RMP1EN	1	RO	0x1	RCV MNG port 1 Enable. When this bit is set, it reports that MNG traffic is received from port 1.
MXP	2	RO	0x0	MNG XMT Port. 0b = MNG traffic should be transmitted through port 0. 1b = MNG traffic should be transmitted through port 1.
PRPP	3	RW	0x0	Preferred Primary Port. 0b = Port 0 is the preferred primary port. 1b = Port 1 is the preferred primary port.
PRPPE	4	RW	0x0	Preferred Primary Port enables.
Reserved	5	RO	0x0	Reserved.
RGAEN	6	RW	0x0	Repeated Gratuitous ARP Enable. If this bit is set, the 82599 sends a configurable number of gratuitous ARP packets (GAC bits of this register) using configurable interval (GATI bits of this register) after the following events: System move to Dx, or fail-over event initiated the 82599.

Field	Bit(s)	RW	Init Val	Description
Reserved	7	RO	0x0	Reserved.
Reserved	8	RO	0x0	Reserved.
TFOENODX	9	RW	0x0	Teaming Fail Over Enable on Dx. Enable fail-over mechanism. Bits 3-8 are valid only if this bit is set.
Reserved	10-11	RO	0x0	Reserved.
GAC	12-15	RW	0x0	Gratuitous ARP counter. Counts the number of gratuitous ARP that should be done after a fail-over event and after a move to Dx. When it is set to zero, there is no limit on the gratuitous ARP packets.
LDFOT	16-23	RW	0x0	Link Down Fail-Over Time. Defines the time in seconds the link should be down before doing a fail over to the other port. This is also the time that the primary link should be up (after it was down) before the 82599 switches back to the primary port.
GATI	24-31	RW	0x0	Gratuitous ARP Transmission Interval. Defines the GAP in seconds before retransmission of gratuitous ARP packets.

10.2.2.3 ARP Handling

Independent of the management interface, the 82599 can be programmed by the BMC to provide ARP services. The 82599 supports auto-ARP replying for ARP request packets and sending Gratuitous ARP. Auto-ARP is done in both ports in either modes: dual-channel and one-channel. In dual-channel mode, each channel uses its own IP and Ethernet MAC address (either the operating system Ethernet MAC address or independent addresses). In one-channel mode, both ports use the same IP and Ethernet MAC address and the ARP is responded to through the port it was received.

The following ARP parameters are loaded from the EEPROM on power up or configured through the management interface:

- ARP auto-reply enabled
- ARP IP address (to filter ARP packets)
- ARP Ethernet MAC Addresses (for ARP response)

When an ARP request packet is received on the wire and ARP auto-reply is enabled, the 82599 checks the targeted IP address (after the packet has passed L2 checks and ARP checks). If the targeted IP matches the 82599 IP configuration, then it replies with an ARP response. The 82599 responds to the ARP request targeted to the ARP IP address with its ARP Ethernet MAC address. In a case where there is no match, the 82599 silently discards the packets. If the 82599 is not configured to do auto-ARP response, it forwards the ARP packets to the BMC.

When the external BMC uses the same IP and MAC of the operating system, the ARP operation should be coordinated with the operating system operation. In this mode, the external BMC has the responsibility and ARP auto-reply should be disabled.

Note: When configured in NC-SI mode, the 82599 does not provide ARP services. All ARP handling is done by the BMC.

10.3 Manageability Receive Filtering

10.3.1 Overview and General Structure

For completeness, this section summarizes the MAC and VLAN filters described in Section 7.1.1.1 and Section 7.1.1.2. In addition, this section describes the manageability receive packet filtering flow. The description applies to any of the 82599 LAN ports. Receive packet filtering can have one of the following routing results:

- Discard packets (packets that do not pass the host nor manageability filtering)
- Send packets to host memory (default hardware setting)
- Send packets to the external BMC (two modes):
  - Receive All — All received packets are routed to the BMC in this mode. It is enabled by setting the RCV_TCO_EN bit (which enables packets to be routed to the BMC) and RCV_ALL bit (which routes all packets to the BMC) in the MANC register.
  - Receive Filtering — In this mode only some of the packet types are directed to the manageability block. The BMC should set the RCV_TCO_EN bit together with the required packet types bits in the manageability filtering registers. Note that the RCV_ALL bit must be cleared).
- Send packets to both the external BMC and host memory:
  - The BMC can enable this mode by setting the EN_MNG2HOST bit in the MANC register and enable specific packet types in the MANC2H register.

The BMC controls its packet filtering by programming the receive manageability filters listed in the following table. These registers are not write-accessible by the host (protecting the BMC from erroneous/malicious host software).

Filters	Functionality	When Reset?
Filters enable	General configuration of the manageability filters.	Internal Power On Reset
Manageability to host	Enables routing of manageability packets to host.	Internal Power On Reset
Manageability decision filters [7:0]	Configuration of manageability decision filters.	Internal Power On Reset
MAC address [3:0]	Four unicast MAC manageability addresses.	Internal Power On Reset
VLAN filters [7:0]	Eight VLAN tag values.	Internal Power On Reset
UDP/TCP port filters [15:0]	16 destination port values.	Internal Power On Reset
Flexible 128-byte TCO filters	Length values for four flex TCO filters.	Internal Power On Reset
IPv4 and IPv6 address filters[3:0]	IP address for manageability filtering.	Internal Power On Reset
L2 EtherType filters [3:0]	Four L2 EtherType values.	Internal Power On Reset

Manageability filtering follows these steps and are detailed in the following sections:

1. L2 Ethernet MAC address and VLAN filtering
2. L3/L4 manageability filters — Port, IP, flex filters (packets must also match the above L2 filtering).

Filtering exceptions:

- Fragmented packets can be routed to manageability but not parsed beyond the IP header.
- Packets with L2 errors (CRC, alignment, etc.) are never forwarded to manageability.

Note: Jumbo packets above 2 KB are not expected to be received by the manageability data path. If the manageability unit uses a dedicated Ethernet MAC address/VLAN tag, it should not use further L3/L4 filters on top of it. Otherwise, packets that match the L2 filters but fail the L3/L4 filters are routed to the host.

The complete filtering flow is described in the following flow diagram:

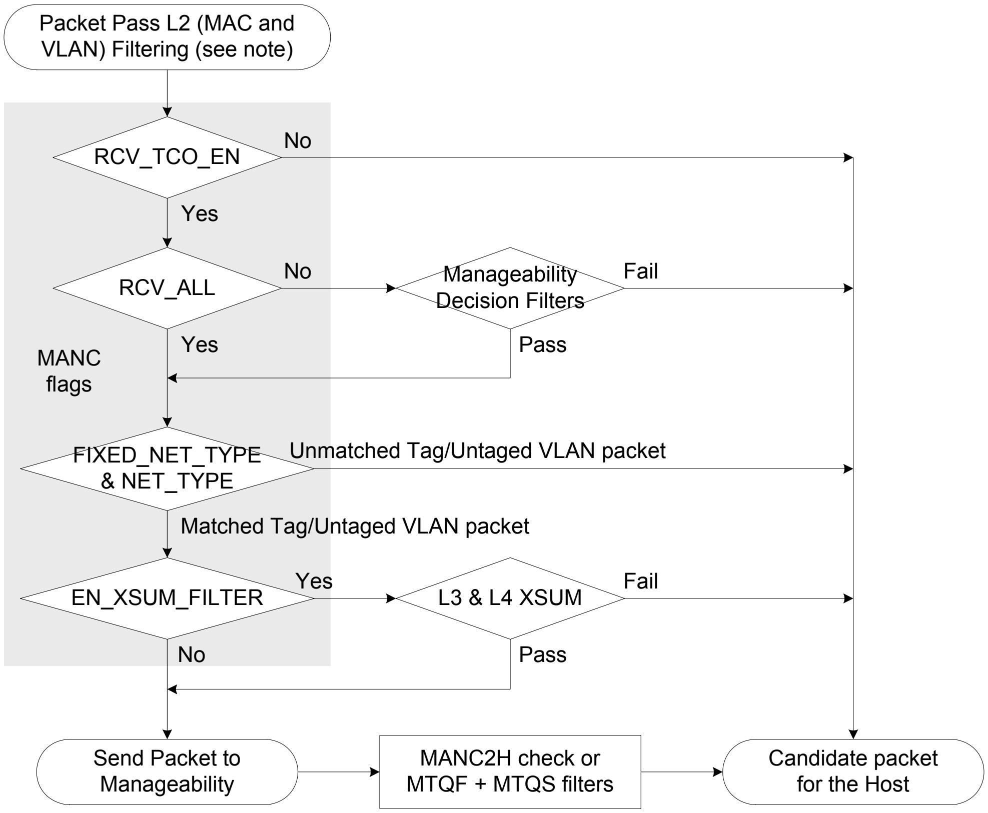
Packet Pass L2 (MAC and VLAN) Filtering (see note)
RCV_TCO_EN
No
Yes
RCV_ALL
No
Manageability Decision Filters
Fail
Yes
Pass
MANC flags
FIXED_NET_TYPE & NET_TYPE
Unmatched Tag/Untaged VLAN packet
Matched Tag/Untaged VLAN packet
EN_XSUM_FILTER
Yes
L3 & L4 XSUM
Fail
No
Pass
Send Packet to Manageability
MANC2H check or MTQF + MTQS filters
Candidate packet for the Host

Figure 10-2 Flow Diagram

Note: L2 MAC address and VLAN filtering are described in Section 7.1.1.1 and Section 7.1.1.2.

10.3.2 L2 EtherType Filters

Packets are compared against the EtherType filters programmed in the METF.EType (up to 4 filters) and the result is incorporated to the decision filters.

Each of the manageability EtherType filters can be configured as pass (positive) or reject (negative) polarity. When negative polarity filters are used, all negative filters should be included in all enabled decision filters.

Examples for usages of the L2 EtherType filters are:

- Block routing of packets with the NC-SI EtherType from being routed to the BMC. The NC-SI EtherType is used for communications between the BMC on the NC-SI link and the 82599. Packets coming from the network are not expected to carry this EtherType and such packets are blocked to prevent attacks on the BMC.
- Determine the destination of 802.1X control packets. The 802.1X protocol is executed at different times in either the BMC or by the host. The L2 EtherType filters are used to route these packets to the proper agent.

10.3.3 VLAN Filters - Single and Double VLAN Cases

The 82599 supports eight VLAN filters per port defined by the MAVTV[n] and controlled by the MANC register as described in the text that follows.

- When MANC.NET_TYPE = 1b and MANC.FIXED_NET_TYPE = 1b (pass only VLAN tagged packets)
  - A packet without any VLAN or a single VLAN header is not routed to manageability
  - A packet with 2 VLANs is a candidate for manageability
- When MANC.NET_TYPE = 0b and MANC.FIXED_NET_TYPE = 1b (pass only un-tagged packets)
  - A packet without any VLAN or a single VLAN header is a candidate for manageability
  - A packet with 2 VLANs is not routed to manageability
- When MANC.FIXED_NET_TYPE = 0b (both tagged and untagged packets are candidates for manageability)
  - A packet with no VLAN header skips successfully to the next filtering level
  - A packet with a single VLAN or 2 VLANs are filtered by its VLAN header as described in Section 7.1.1.2

10.3.4 L3 and L4 Filters

ARP Filtering: The 82599 supports filtering of both ARP request packets (initiated externally) and ARP responses (to requests initiated by the BMC or the 82599).

Neighbor Discovery Filtering: The 82599 supports filtering of neighbor discovery packets. Neighbor discovery filters use the IPV6 destination address filters defined in the MIPAF registers (such as match to any of the enabled IPv6 addresses).

Port 0x298/0x26F Filtering: The 82599 supports filtering by fixed destination ports numbers: 0x26F and 0x298.

Flex Port Filtering: The 82599 implements 16 flex destination port filters. The 82599 directs packets whose L4 destination port matches the value of the respective word in the MFUTP registers. The BMC must ensure that only valid entries are enabled in the decision filters that follow.

Flex TCO Filters: See Section 10.3.4.1.

IP Address Filtering: The 82599 supports filtering by IP address through dedicated IPv4 and IPv6 address filters to manageability. Two modes are possible, depending on the value of the MANC.EN_IPv4_FILTER bit:

- EN_IPv4_FILTER = 0b: The 82599 provides four IPv6 address filters.
- EN_IPv4_FILTER = 1b: The 82599 provides three IPv6 address filters and four IPv4 address filters.
- The MFVAL register indicates which of the IP address filters are valid (contains a valid entry and should be used for comparison).

Checksum Filter: If bit MANC.EN_XSUM_FILTER is set, the 82599 directs packets to the BMC only if they match all other filters previously described as well as pass L3/L4 checksum (if it exists).

10.3.4.1 Flexible 128 Bytes Filter (TCO Filters)

10.3.4.1.1 Overview

The flexible 128 filters are a set of filters designed to enable dynamic filtering of received packets. These filters are part of the manageability receive filters. The filters do not make a decision on the packet's destination. They participate in the decision mechanism for each received packet (Section 10.3.5).

Each filter enables a flexible testing of the first 128 bytes of the packet against a given value. The filter also enables testing of specific bytes by defining a byte-wise mask on the filter.

The 82599 provides four flex TCO filters. Each filter looks for a pattern match within the 1st 128 bytes of the packet. The BMC must ensure that only valid entries are enabled in the decision filters.

Note: The flex filters are temporarily disabled when read or written by the host. Any packet received during a read or write operation is dropped. Filter operation resumes once the read or write access completes.

10.3.4.1.2 Structure

Each filter is composed of the following fields:

1. Flexible Filter Length: This field indicates the number of bytes in the packet header that should be inspected. This field also indicates the minimal length of packets in order to be inspected by the filter. A packet below that length is not inspected by the filter. Valid values for this field are: 8*n, where n=1…8.
2. Data: This is a set of up to 128 bytes comprising the values that the header bytes of each packet are tested against.
3. Mask: This is a set of 128 bits corresponding to the 128 data bytes that indicate for each corresponding byte if is tested against its corresponding byte.

Overall, each filter tests the first 128 bytes (or less) of a packet, where not necessarily all bytes must be tested.

10.3.4.1.3 Programming

Programming each filter is done using the following two commands (NC-SI or SMBus) in a sequential manner:

1. Filter Mask and Length. This command configures the following fields.
   a. Mask: A set of 16 bytes containing the 128 bits of the mask. Bit 0 of the first byte corresponds to the first byte on the wire.
   b. Length: A 1-byte field indicating the length.
2. Filter Data.

   The filter data is divided into groups of bytes. as follows:

Group	Test Bytes
0x0	0-29
0x1	30-59
0x2	60-89
0x3	90-119
0x4	120-127

Each group of bytes needs to be configured using a separate command, where the group number is given as a parameter.

The command has the following parameters:

a. Group number. A 1-byte field indicating the current group addressed.
b. Data bytes. Up to 30 bytes of test-bytes for the current group.

10.3.5 Manageability Decision Filters

The manageability decision filters are a set of eight filters with the same structure (MDEF[7:0] and MDEF_EXT[7:0]). The filtering rule for each decision filter is programmed by the BMC and defines which of the L2, VLAN, and manageability filters participate in the decision (host software can't modify their setting). A packet that passes at least one set of decision filters is directed to manageability and possibly to the host as well. The inputs to each decision filter are:

- Packet passed a valid management L2 unicast address filter.
- Packet is a broadcast packet.
- Packet has a VLAN header and it passed a valid manageability VLAN filter.
- Packet matched one of the valid IPv4 or IPv6 manageability address filters.
- Packet is a multicast packet.
- Packet passed ARP filtering (request or response).
- Packet passed neighbor discovery filtering.
- Packet passed 0x298/0x26F port filter.
- Packet passed a valid flex port filter.
- Packet passed a valid flex TCO filter.
- Packet passed or failed an L2 EtherType filter.

The structure of each of the decision filters is shown in Figure 10-3 . A boxed “x.y” number indicates that the input is conditioned on a mask bit “y” defined in register index “x”, while x=0 denotes MDEF and x=1 denotes MDEF_EXT. The decision filter rules are as follows:

- Any bit set in the MDEF and MDEF_EXT registers enables its corresponding filter. Any filter that is not enabled in the MDEF and MDEF_EXT registers is ignored. If all bits in the MDEF and MDEF_EXT registers of a specific decision filter are cleared, it is disabled and ignored.
- All enabled AND filters must pass for the decision filter to match.
- If at least one OR filter is enabled, then at least one of the enabled OR filters must pass for the decision filter to match.

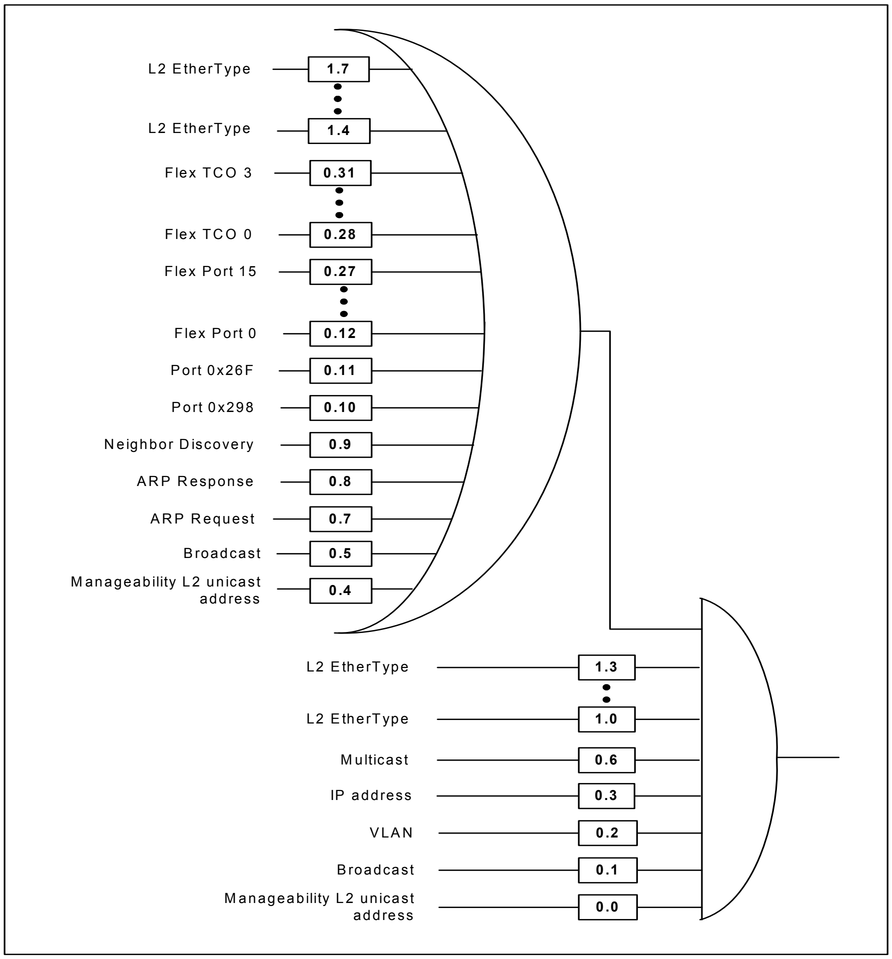

Figure 10-3 Manageability Decision Filters

10.3.6 Possible Configurations

This section describes possible ways of using the management filters. Actual usage might vary.

Dedicated MAC packet filtering

- Select one of the eight rules for broadcast filtering
- Set bit 0 of the decision rule to enforce Ethernet MAC address filtering
- Set other bits to qualify which packets are allowed to pass through. For example:
  - Set bit 2 to qualify with manageability VLAN
  - Set bit 3 to qualify with a match to an IP address
  - Set any L3/L4 bits (30:7) to qualify with any of a set of L3/L4 filters

Broadcast packet filtering

- Select one of the eight rules for broadcast filtering
- Set bit 1 of the decision rule to enforce broadcast filtering
- Set other bits to qualify which broadcast packets are allowed to pass through. For example:
  - Set bit 2 to qualify with manageability VLAN
  - Set bit 3 to qualify with a match to an IP address
  - Set any L3/L4 bits (30:7) to qualify with any of a set of L3/L4 filters

VLAN packet filtering

- Select one of the eight rules for VLAN filtering
- Set bit 2 of the decision rule to enforce VLAN filtering
- Set other bits to qualify which VLAN packets are allowed to pass through. For example:
  - Set any L3/L4 bits (30:7) to qualify with any of a set of L3/L4 filters

IPv6 filtering is done via the following IPv6-specific filters:

- IP unicast filtering — requires filtering for link local address and a global address. Filtering setup might depend on whether an Ethernet MAC address is shared with the host or dedicated to manageability:
  - Dedicated Ethernet MAC address (such as dynamic address allocation with DHCP does not support multiple IP addresses for one Ethernet MAC address). In this case, filtering can be done at L2 using two dedicated unicast MAC filters.
  - Shared Ethernet MAC address (such as static address allocation sharing addresses with the host). In this case, filtering needs to be done at L3, requiring two IPv6 address filters, one per address.

- A Neighbor discovery filter — The 82599 supports IPv6 neighbor discovery protocol. Since the protocol relies on multicast packets, the 82599 supports filtering of these packets. IPv6 multicast addresses are translated into corresponding Ethernet multicast addresses in the form of 33-33-xx-xx-xx-xx, where the last 32 bits of the address are taken from the last 32 bits of the IPv6 multicast address. Therefore, two direct MAC filters can be used to filter IPv6 solicited-node multicast packets as well as IPv6 all node multicast packets.

Receive filtering with shared IP — CPMP

When the BMC shares the MAC and IP address with the host, receive filtering is based mainly on identifying specific flows through port allocation. The following setting can be used:

Select one of the eight rules:

- Set a manageability dedicated MAC filter to the host Ethernet MAC address and set bit 0 in the MNG_ FILTER_RULE register.
- If VLAN is used for management, load one of more management VLAN filters and set bit 2 in the MNG_ FILTER_RULE register
- ARP filter / neighbor discovery filter is enabled when the BMC is responsible to handle the ARP protocol. Set bit 7 or bit 8 in the MNG_ FILTER_RULE register for this functionality.
- Program flex port filters with the port values for management flows such as DHCP, HTTP, HTTPS, SMWG, SoL/IDER/KVM, WS-MAN, Telnet, USB redirection, SSH, DNS, and more. Set the respective bits 26:11 in the MNG_ FILTER_RULE register.
- An IP address filter can be loaded as well by setting bit 3 in the MNG_ FILTER_RULE register.
- Management flex filters are programmed to correspond to remaining flows such as DNS update response packets. Set appropriate bits 30:27 in the MNG_ FILTER_RULE register.

10.4 LinkSec and Manageability

For details on LinkSec and the role of manageability in it, see Section 7.8.

Pass-through mode is supported in a LinkSec environment in one of the following modes of operations:

- Management traffic not protected by LinkSec — The management traffic from and to the BMC is carried over a separate Ethernet MAC address and/or a separate VLAN and the network switch is configured to enable such traffic to pass unprotected.
- Management traffic is protected by LinkSec — The 82599 supports a single secure channel for both host and BMC. At a given time, the host and BMC can be active or inactive. When only BMC is active, it acts as the KaY controlling the secured channel. The host can act as the KaY when it is functional and after it acquires control over LinkSec. In this case, the BMC uses the secured channel set by the host. Even when operating in this mode, the BMC can transmit packets on the clear (as required for 802.1x control packets). The BMC must disable MACsec operation before sending such packets and re-enable MACsec operation afterwards. The messages that control MACsec operation are described in Section 10.5.1.15.

The 82599 provides the following functionality that enables management traffic over the same secure channel with the host:

- Handover of LinkSec ownership between the BMC and the host. Several transitions in ownership are possible:
    - Power-on — The 82599 powers up with LinkSec **not** being owned by the BMC. If the BMC is configured for LinkSec, it takes ownership over LinkSec as follows. If the BMC is not configured for LinkSec, the host takes ownership when it boots. If LinkSec is not owned by the BMC, the host is not required for any handshake with the BMC as there are cases where the BMC is not connected to the 82599. If there is a race between the BMC and the host, the BMC wins over LinkSec, and the host is then interrupted so that the LinkSec resources are not accessible.
    - Handover of LinkSec responsibility from BMC to host — The host can initiate a transfer of ownership from the BMC (such as on operating system boot).
    - Handover of LinkSec responsibility from host to BMC — The host can initiate a transfer of ownership to the BMC (such as on entry to low power state). This is done through the host slave command interface.
    - Forced handover of LinkSec responsibility from host to BMC — The BMC can acquire ownership of LinkSec on its own, for example when the host fails to acquire a secure channel. See Section 10.4.1 for the different transition sequences.
- Configuration of LinkSec resources by the BMC — When the BMC owns the secure channel, it configures LinkSec operation through the SMBus or NC-SI vendor-specific commands (see Section 10.5.1.15).
- Alerts — The 82599 initiates an SMBus or NC-SI alert to the BMC on several LinkSec events as follows (see alerts message format in Section 10.5.1.16, Section 10.5.2.2.3, and Section 10.5.2.2.8).
    - Packet arrived with a LinkSec error (no SA match, replay detection, or a bad LinkSec signature).

  - — Key-exchange event — relevant on Tx when the packet number counter reaches the exhaustion threshold as described in Section 7.8.5.1.
  - — Host request for LinkSec ownership.
  - — Host request to relinquish LinkSec ownership.
- Interrupt causes — The 82599 issues a management interrupt to the host on the following LinkSec events:
  - — Acknowledge of handover of LinkSec responsibility from BMC to host.
  - — Forced handover of LinkSec responsibility from host to BMC.

  The host might identify the ownership status by reading the *Operating System Status* field in the LSWFW register.

10.4.1 Handover of LinkSec Responsibility Between BMC and Host

10.4.1.1 KaY Ownership Release by the Host

The following procedure is used by the host in order to release ownership of the LinkSec capability. This procedure is usually done before an ordered shutdown of the host.

- The host should stop accessing the LinkSec registers and set the *Release LinkSec* bit in the LSWFW register.
- Setting the *Release LinkSec* bit causes an interrupt to the firmware that is forwarded to the BMC.
- The BMC then takes ownership as described in Section 10.4.1.2.
- The host can then wait for an interrupt from the firmware indicating that the BMC took the KaY ownership.

10.4.1.2 KaY Ownership Takeover by BMC

As previously mentioned, the BMC can acquire ownership over LinkSec either by ownership relinquish by the host or without any negotiation (such as on power-up and on a forced transition when the host failed to bring up a LinkSec connection). The BMC acquires ownership of LinkSec by taking the following actions:

- Locking access to LinkSec resources to the host by setting the *Lock LinkSec Logic* bit in the LSWFW register.
- Blocking host packets' transmission from the wire by setting the *Block Host Traffic* bit in the LSWFW register.
- Set the *OS Status* field in the LSWFW register to 1b indicating a BMC takeover of the LinkSec logic.
- Issue a manageability event interrupt to the host.

10.4.1.3 KaY Ownership Request by the Host

The following procedure is used by the host in order to request ownership of the LinkSec capability:

- The host should read the LSWFW.OS status field to check if the KaY is currently owned by the BMC.
- if KaY is owned by the BMC, then the host should set the *Request LinkSec* bit in the LSWFW register prior to assuming responsibility over LinkSec connection.
- Setting the *Request LinkSec* bit causes an interrupt to the firmware that is forwarded to the BMC.
- The host should then wait for an interrupt from the firmware indicating that the BMC released the KaY ownership.
- Following the manageability interrupt, the host should check the *OS Statu*s and *Lock LinkSec Logic* fields in the LSWFW register to make sure the BMC released the KaY ownership.

10.4.1.4 KaY Ownership Release by BMC

In order to release ownership of LinkSec, the BMC should take the following actions:

- Disconnect the LinkSec connection with the switch (such as EAP logoff).
- Clear the *Lock LinkSec Logic* bit in the LSWFS register enabling the host setting of the LinkSec registers.
- Clear the *OS Status* bit to 0b in the LSWFS register indicate a LinkSec release.
- Issue a manageability event interrupt to the host.
- Poll the connection state to check if the LinkSec channel was set by the host.

If the BMC decides to deny the release request, it silently ignores the request.

10.4.1.5 Control Registers

The complete set of manageability registers are described in Section 8.2.3.26 and Section 8.2.3.26. The following configuration fields are dedicated for manageability control over LinkSec:

LSWFW Field	LSWFW Field Functionality
Block Host Traffic	Enables or disables host transmit traffic for this PCI function from going to the wire. Default is to enable.
OS Status	Set by firmware to indicate the status of the LinkSec ownership: 0b = LinkSec owned by host (default). 1b = LinkSec owned by BMC.
LinkSec Request	Bit used by host to request KaY ownership.

LSWFW Field	LSWFW Field Functionality
LinkSec Release	Bit used by host to release KaY ownership.
Lock LinkSec Logic	Serves two purposes. It indicates who owns LinkSec (default value is host ownership). Second, it enables or disables host accesses to the LinkSec registers. Default is to enable. The following registers are blocked: LSECTXCAP; LSECRXCAP; LSECTXCTRL; LSECRXCTRL; LSECTXSCL; LSECTXSCH; LSECTXSA; LSECTXPN0; LSECTXPN1; LSECTXKEY0 (4 registers); LSECTXKEY1 (4 registers); LSECRXSCL; LSECRXSCH; LSECRXSA (0 and 1); LSECRXSAPN (0 and 1); LSECRXKEY (4 registers / SA); LSECTXUT; LSECTXPKTE; LSECTXPKTP; LSECTXOCTE; LSECTXOCTP; LSECRXUTnS; LSECRXUTyS; LSECRXOCTE; LSECRXOCTP; LSECRXBAD; LSECRXNOSCInS; LSECRXNOSCIyS; LSECRXNOSCI; LSECRXDELAY; LSECRXLATE; LSECRXOK; LSECRXINVCK; LSECRXINVST; LSECRXNSAST; LSECRXNSA

10.5 Manageability Programming Interfaces

10.5.1 NC-SI Programming

The 82599 supports the mandatory NC-SI commands as listed in Table 10-1. On top of these commands, the 82599 also supports Intel vendor specific commands. The vendor specific commands are based on the NC-SI — OEM Command. These commands are listed in the following sub-sections and are used to enable the BMC to control the 82599 specific features:

- Rx filters:
  - Packet addition decision filters 0x0…0x4
  - Packet reduction decision filters 0x5…0x7
  - MNG2HOST register (controls the forwarding of manageability packets to the host)
  - Flex 128 filters 0x0…0x3
  - Flex TCP/UDP port filters 0..0xA
  - IPv4/IPv6 filters
  - Ether type filters
- Get System Ethernet MAC Address — This command enables the BMC to retrieve the system Ethernet MAC address used by the NC. This Ethernet MAC address can be used for shared Ethernet MAC address mode.
- Keep PHY Link Up (*Veto* bit) Enable/Disable — This feature enables the BMC to block PHY reset, which might cause session loss.
- TCO Reset — Enables the MC to reset the network adapter.
- Checksum Offloading — Offloads IP/UDP/TCP checksum checking from the MC.
- LinkSec logic programming

These commands are designed to be compliant with their corresponding SMBus commands (if existing). All of the commands are based on a single DMTF defined NC-SI command, known as OEM Command described in Section 10.5.1.1.

10.5.1.1 OEM Command (0x50)

The OEM command can be used by the MC to request the sideband interface to provide vendor-specific information. The Vendor Enterprise Number (VEN) is the unique MIB/ SNMP private enterprise number assigned by IANA per organization. Vendors are free to define their own internal data structures in the vendor data fields.

	Bits			
Bytes	**31..24**	**23..16**	**15..08**	**07..00**
00..15	NC-SI Header			
16..19	Manufacturer ID (Intel 0x157)			
20..	Intel Command Number	Optional Data		

10.5.1.2 OEM Response (0xD0)

The sideband interface must return an Unknown Command Type reason code for any un-recognized enterprise number using the following frame format. If the command is valid, the response, if any, is allowed to be vendor-specific. It is recommended to use the 0x8000 range for vendor-specific code.

	Bits			
Bytes	**31..24**	**23..16**	**15..08**	**07..00**
00..15	NC-SI Header			
16..19	Response Code		Reason Code	
20..23	Manufacturer ID (Intel 0x157)			
24..27	Intel Command Number	Optional Return Data		

Table 10-3 OEM Specific Command Response and Reason Codes

Response Code		Reason Code	
Value	**Description**	**Value**	**Description**
0x1	Command Failed	0x5081	Invalid Intel Command Number
		0x5082	Invalid Intel Command Parameter Number

10.5.1.3 Intel Commands

Table 10-4 lists the Intel commands and their associated Intel Command Number values. For detailed description of the commands and their parameters refer to the following sections.

Table 10-4 Intel Command Summary

Intel Command	Parameter	Command Name
0x00	0x00	Set IP Filters Control
0x01	0x00	Get IP Filters Control
0x02	0x0A	Set Manageability to Host
	0x10	Set Flexible 128 Filter 0 Mask and Length
	0x11	Set Flexible 128 Filter 0 Data
	0x20	Set Flexible 128 Filter 1 Mask and Length
	0x21	Set Flexible 128 Filter 1 Data
	0x30	Set Flexible 128 Filter 2 Mask and Length
	0x31	Set Flexible 128 Filter 2 Data
	0x40	Set Flexible 128 Filter 3 Mask and Length
	0x41	Set Flexible 128 Filter 3 Data
	0x61	Set Packet Addition Filters
	0x63	Set Flex TCP/UDP Port Filters
	0x64	Set Flex IPv4 Address Filters
	0x65	Set Flex IPv6 Address Filters
	0x67	Set EtherType Filter
	0x68	Set Packet Addition Extended Decision Filter

Table 10-4 Intel Command Summary (Continued)

Intel Command	Parameter	Command Name
0x3	0x0A	Get Manageability to Host
	0x10	Get Flexible 128 Filter 0 Mask and Length
	0x11	Get Flexible 128 Filter 0 Data
	0x20	Get Flexible 128 Filter 1 Mask and Length
	0x21	Get Flexible 128 Filter 1 Data
	0x30	Get Flexible 128 Filter 2 Mask and Length
	0x31	Get Flexible 128 Filter 2 Data
	0x40	Get Flexible 128 Filter 3 Mask and Length
	0x41	Get Flexible 128 Filter 3 Data
	0x61	Get Packet Addition Filters
	0x63	Get Flex TCP/UDP Port Filters
	0x64	Get Flex IPv4 Address Filters
	0x65	Get Flex IPv6 Address Filters
	0x67	Get EtherType Filter
	0x68	Get Packet Addition Extended Decision Filter
0x04	0x00	Set Unicast Packet Reduction
	0x01	Set Multicast Packet Reduction
	0x02	Set Broadcast Packet Reduction
	0x10	Set Unicast Extended Packet Reduction
	0x11	Set Multicast Extended Packet Reduction
	0x12	Set Broadcast Extended Packet Reduction
0x05	0x00	Get Unicast Packet Reduction
	0x01	Get Multicast Packet Reduction
	0x02	Get Broadcast Packet Reduction
	0x10	Get Unicast Extended Packet Reduction
	0x11	Get Multicast Extended Packet Reduction
	0x12	Get Broadcast Extended Packet Reduction

Table 10-4 Intel Command Summary (Continued)

Intel Command	Parameter	Command Name
0x06	N/A	Get System Ethernet MAC Address
0x20	N/A	Set Intel Management Control
0x21	N/A	Get Intel Management Control
0x22	N/A	Perform TCO Reset
0x23	N/A	Enable IP/UDP/TCP Checksum Offloading
0x24	N/A	Disable IP/UDP/TCP Checksum Offloading
0x30	0x10	Transfer LinkSec Ownership to BMC
	0x11	Transfer LinkSec Ownership to Host
	0x12	Initialize LinkSec Rx
	0x13	Initialize LinkSec Tx
	0x14	Set LinkSec Rx Key
	0x15	Set LinkSec Tx Key
	0x16	Enable Network Tx Encryption
	0x17	Disable Network Tx Encryption
	0x18	Enable Network Rx Decryption
	0x19	Disable Network Rx Decryption
0x31	0x01	Get LinkSec Rx Parameters
	0x02	Get LinkSec Tx Parameters

10.5.1.4 Set Intel Filters Control Command (Intel Command 0x00)

10.5.1.4.1 Set Intel Filters Control – IP Filters Control Command (Intel Command 0x00)

	Bits			
Bytes	**31..24**	**23..16**	**15..08**	**07..00**
00..15	NC-SI Header			
16..19	Manufacturer ID (Intel 0x157)			
20..23	0x00	0x00	IP Filters control (3-2)	
24..27	IP Filters Control (1-0)			

While IP Filters Control has the following format:

Table 10-5 IP Filter Formats

Bit #	Name	Description	Default
0	IPv4/IPv6 Mode	IPv6 (0b): There are 0 IPv4 filters and 4 IPv6 filters IPv4 (1b): There are 4 IPv4 filters and 3 IPv6 filters See Section 8.2.3.25.2 or Section 10.3.4 for details.	1b
1..15	Reserved		
16	IPv4 Filter 0 Valid	Indicates if the IPv4 address configured in IPv4 address 0 is valid. *Note:* The network controller must automatically set this bit to 1b if the Set Intel Filter – IPv4 Filter Command is used for filter 0.	0b
17	IPv4 Filter 1 Valid	Indicates if the IPv4 address configured in IPv4 address 1 is valid. *Note:* The network controller must automatically set this bit to 1b if the Set Intel Filter – IPv4 Filter Command is used for filter 1.	0b
18	IPv4 Filter 2 Valid	Indicates if the IPv4 address configured in IPv4 address 2 is valid. *Note:* The network controller must automatically set this bit to 1b if the Set Intel Filter – IPv4 Filter Command is used for filter 2.	0b
19	IPv4 Filter 3 Valid	Indicates if the IPv4 address configured in IPv4 address 3 is valid. *Note:* The network controller must automatically set this bit to 1b if the Set Intel Filter – IPv4 Filter Command is used for filter 3.	0b
20..23	Reserved		
24	IPv6 Filter 0 Valid	Indicates if the IPv6 address configured in IPv6 address 0 is valid. *Note:* The network controller must automatically set this bit to 1b if the Set Intel Filter – IPv6 Filter Command is used for filter 0.	0b
25	IPv6 Filter 1 Valid	Indicates if the IPv6 address configured in IPv6 address 1 is valid. *Note:* The network controller must automatically set this bit to 1b if the Set Intel Filter – IPv6 Filter Command is used for filter 1.	0b

Table 10-5 IP Filter Formats (Continued)

Bit #	Name	Description	Default
26	IPv6 Filter 2 Valid	Indicates if the IPv6 address configured in IPv6 address 2 is valid. *Note:* The network controller must automatically set this bit to 1b if the Set Intel Filter – IPv6 Filter Command is used for filter 2.	0b
27	IPv6 Filter 3 Valid	Indicates if the IPv6 address configured in IPv6 address 3 is valid. *Note:* The network controller must automatically set this bit to 1b if the Set Intel Filter – IPv6 Filter Command is used for filter 3.	0b
28..31	Reserved	Reserved	

10.5.1.4.2 Set Intel Filters Control – IP Filters Control Response (Intel Command 0x00, Filter Control Index 0x00)

	Bits			
Bytes	**31..24**	**23..16**	**15..08**	**07..00**
00..15	NC-SI Header			
16..19	Response Code		Reason Code	
20..23	Manufacturer ID (Intel 0x157)			
24..27	0x00	0x00		

10.5.1.5 Get Intel Filters Control Command (Intel Command 0x01)

10.5.1.5.1 Get Intel Filters Control – IP Filters Control Command (Intel Command 0x01, Filter Control Index 0x00)

This command controls different aspects of the Intel filters.

	Bits			
Bytes	**31..24**	**23..16**	**15..08**	**07..00**
00..15	NC-SI Header			
16..19	Manufacturer ID (Intel 0x157)			
20..21	0x01	0x00		

10.5.1.5.2 Get Intel Filters Control – IP Filters Control Response (Intel Command 0x01, Filter Control Index 0x00)

	Bits			
Bytes	**31..24**	**23..16**	**15..08**	**07..00**
00..15	NC-SI Header			
16..19	Response Code		Reason Code	
20..23	Manufacturer ID (Intel 0x157)			
24..27	0x01	0x00	IP Filters Control (3-2)	
28..29	IP Filters Control (1-0)			

IP Filter Control: See Table 10-5.

10.5.1.6 Set Intel Filters Formats

10.5.1.6.1 Set Intel Filters Command (Intel Command 0x02)

	Bits			
Bytes	**31..24**	**23..16**	**15..08**	**07..00**
00..15	NC-SI Header			
16..19	Manufacturer ID (Intel 0x157)			
20..21	0x02	Filter Parameter	Filters Data (optional)	

10.5.1.6.2 Set Intel Filters Response (Intel Command 0x02)

	Bits			
Bytes	**31..24**	**23..16**	**15..08**	**07..00**
00..15	NC-SI Header			
16..19	Response Code		Reason Code	
20..23	Manufacturer ID (Intel 0x157)			
24..	0x02	Filter Parameter	Return Data (Optional)	