Programmer's Reference Guide

BCM848XX

MACSec Theory of Operation

848XX-PG100-R

5300 California Avenue • Irvine, CA 92617 • Phone: 949-926-5000 • Fax: 949-926-5203

February 12, 2014

Revision History

Revision	*Date*	*Change Description*
848XX-PG100-R	02/12/14	Initial release

Broadcom Corporation
5300 California Avenue
Irvine, CA 92617

Table of Contents

List of Figures

List of Tables

About This Document

Purpose and Audience

This document describes the architecture and features of the Broadcom® BCM848XX PHY device. This document is intended to provide an overview of the device architecture and how it could be used at the system level. In addition, high level packet or frame flows are described for each supported feature. This document is intended for system architects, programmers, and those interested in learning how the BCM848XX device operates.

This document does not provide programming examples. For programming examples, see the *BCM848XX MACsec User Guide* listed in “References” on page 8.

This document does not detail electrical specifications or register information. Although relevant registers and tables are mentioned as appropriate when a particular feature is described in the document, refer to the Data Sheet for the BCM848XX device and the *BCM848XX Programmer’s Register Reference Guide* listed in “References” on page 8 for detailed electrical and register information.

Acronyms and Abbreviations

In most cases, acronyms and abbreviations are defined on first use.

Acronyms and abbreviations in this document are also defined in Appendix A: “MACsec Terminology,” on page 55.

For a comprehensive list of acronyms and other terms used in Broadcom documents, go to: http://www.broadcom.com/press/glossary.php.

Document Conventions

The following conventions may be used in this document:

Convention	***Description***
Bold	User input and actions: for example, type **exit**, click **OK,** press **Alt+C**
`Monospace`	Code: `#include <iostream>` HTML: `<td rowspan = 3>` Command line commands and parameters: `wl [-l] <command>`
< >	Placeholders for *required* elements: enter your <username> or `wl <command>`
[ ]	Indicates *optional* command-line parameters: `wl [-l]` Indicates bit and byte ranges (inclusive): [0:3] or [7:0]

References

The references in this section may be used in conjunction with this document.

Note: Broadcom provides customer access to technical documentation and software through its Customer Support Portal (CSP) and Downloads & Support site (see Technical Support).

Document (or Item) Name	*Number*	*Source*
Broadcom Items		
[1] BCM848XX Programmer’s Register Reference Guide	848XX-PR100-R	Broadcom CSP
[2] 10GBASE-T Transceiver with Integrated MACsec Rev B1 (BCM848XX Data Sheet)	848XX-DS00-R	Broadcom CSP
[3] BCM848XX MACsec User Guide	1CS848XX-UM200-R	Broadcom CSP

Technical Support

Broadcom provides customer access to a wide range of information, including technical documentation, schematic diagrams, product bill of materials, PCB layout information, and software updates through its customer support portal (https://support.broadcom.com). For a CSP account, contact your Sales or Engineering support representative.

In addition, Broadcom provides other product support through its Downloads & Support site (http://www.broadcom.com/support/).

Section 1: Introduction

The BCM848XX is a fully integrated 40 Gigabit-Ethernet (GbE) transceiver with standards-compliant MACsec functionality on all the ports. The device can be configured at the line (physical media) side as follows:

- 1 × 40G port
- 4 × 10G/1G SFP+
- 4 × triple speed SGMII

At the system (switch) side, the device can be configured as follows:

- 1 × XLAUI
- 4 × XFI

The device is a multirate PHY targeted for MMF and copper Twinax application interfacing to both limiting based and linear-based SFP+ and SFP modules. On-chip clock synthesis is performed by the high-frequency, low-jitter phase-locked loops for the PMD and XFI output retimers. Individual PMD and XFI clock recovery is performed in the device by synchronizing directly with the respective incoming data streams. An external 156.25 MHz oscillator is required for the reference clock input.

MACsec Support in the BCM848XX

The IEEE MACsec standards define port-based link layer security services and protocols that are used to protect Ethernet-based networks. MACsec is a link-layer security mechanism defined by IEEE 802.1AE that defines MAC layer traffic protection via MAC Security Entity (SecY). Although IEEE 802.1AE is a completed standard, it has no provisions for key exchange and relies on another specification, IEEE 802.1X-2010, for key management.

The IEEE 802.1AE standard specifies the format of the MACsec Protocol Data Unit (MPDU) and the MACsec service provided by SecY in the MAC sublayer. It also specifies the interface requirement for security association (SA) management. The IEEE 802.1X-2010 standard defines the key agreement protocol as another entity (KaY) to support the establishment of secure channel (SC) and SA for SecY. Both standards define the MIB management interface for network management of security services.

MACsec secures LANs from wiretapping, impersonation, and replay attacks. It also limits the extent of denial-of-service attacks. MACsec does not protect against attacks of trusted components themselves, provide end-to-end security, or replace IEEE 802.11i.

MACsec provides confidentiality and per packet authentication on Layer 2 links. When used with policy-based Layer 2 security mechanisms, it helps to secure Layer 2 of the network infrastructure. MACsec provides security over a single hop between two IEEE 802.1AE-conformant points of attachment. The hop can be either physical or logical. For a logical hop, it can be between customer equipment across a provider bridging network.

The BCM848XX supports the following specific MACsec features:

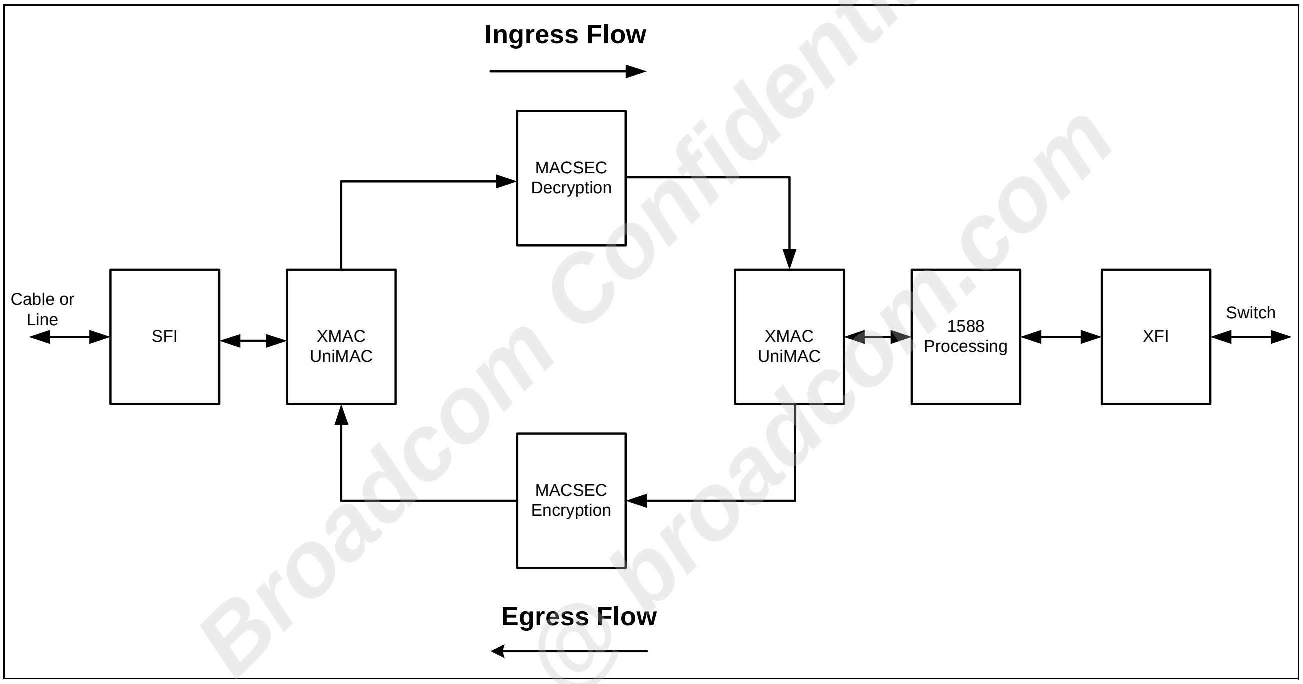
- Security:
  - IEEE 802.1AE Compliance
  - AES-GCM with 128-bit key
  - Confidentiality Offset from 0–63 bytes in byte granularity
  - Replay Window enabled on per SA basis
  - Replay Window size programmable per SA basis up to 2^{32-1}
  - Soft and hard key expiry interrupt per SA
  - Support running test vectors in AES-GCM mode for FIPS algorithm certification (NITS)
  - Start and stop timers per SA
  - Egress auto-an roll over (rolls over to the next enabled AN)
  - Security status interrupt, generate interrupt for each type of SecY failure and consistency check failure
- Performance:
  - Maximum supported frame size is 16K
  - Optional padding small frames to 64B post MACsec decapsulation. Small frames are padded to legal Ethernet frames on the system side
  - Configurable option to accept frames between 18B to 63B (not including CRC) from switch link ports (egress direction)
  - Per User-Priority MTU (16K MAX) checks on packets received from the line-side ports and packets transmitted to line-side ports
  - Support 1G/10G line rates for minimum 64B packets
  - Minimize latency in both ingress and egress direction
  - Support 25 MHz MDC for MDIO interface Clause 45
  - Support Variable IPG on transmit and receive
- Data Path:
  - Support 16 SCs/32 SAs per direction per port
  - Support 1, 2, and 4 AN modes
  - 32KB Buffer per port per direction
  - Fixed latency mode
  - Store-and-forward operation where entire packet is stored before forwarding
  - Bypass all MACsec functionality
  - PFC and PAUSE frames pass through mode
  - Pause termination and generation in store-and-forward mode
  - Post-decryption frame processing on ingress (line side to switch)
  - Pre-Encryption frame processing on egress (switch to line side)
  - Support loopback mode within the MACsec core

Section 2: Chip and System Views

Chip Logical View Block Diagram

Figure 1 shows the device's top-level block diagram. It is a simplified functional diagram, and the purpose is to show different packet flows based on the modes of operation and the features being used (such as MACsec or 1588). The black lines show the ingress (from line to switch) and egress (from switch to line) packet flows when the device is configured in MACsec mode, with or without using the 1588 function.

Figure 1: Simplified Top Level Block Diagram



Modes of Operation

The BCM848XX contains four identical channels, and the modes of operation described in this section are independent per channel, except for the 40G repeater mode. After power on reset, all channels come up in the 10GBASE-R mode. The MACsec and 1588 timestamping features are disabled by default. You must enable these features separately, as supported by the device, because these features might be mutually exclusive. See the following subsections for details.

When the device is powered up, the line-side link up process depends on the presence or absence of a module. If there is already a module present, there are two ways to detect the module type and configure the device in one of the operating modes:

- The module’s presence initiates the PHY to generate an interrupt that wakes up the station manager. As part of the interrupt service routine, the station manager reads the module type and determines the speed mode requirement for that type. The device driver should then configure the PHY and the additional feature set that is needed for that link through the MDIO interface.
- A built-in auto module detect firmware feature inside the PHY can also perform the same steps, except the feature configuration. When the PHY detects the module presence, apart from generating the interrupt, a built-in logic will use the Broadcom serial control (BSC), Philips® I^{2}C-compatible, interface to read the Module ID and then configure the PHY port accordingly. Then the device manager needs to configure only the additional feature set.

10GBASE-R Mode

The 10 Gbps link is established without any auto negotiation between the link partners. The system-side link comes up independent of the line-side link status. Any changes in link status at the line side or system side causes an interrupt to be generated. The software could enable the egress traffic as soon as the line and system-side link up status are detected.

During normal operation, if either the line-side or system-side link goes down, a PHY interrupt is issued. When the line-side link is down, the PCS layer puts out the local Fault message through the XGMII interface, which is transmitted out through the system-side interface. The software can either read the link status or observe at the MAC layer to perform the Link Fault Handling as specified in IEEE 802.3ae, Section 46.3.4.

When the device is configured in 10GBASE-R mode, MACsec and 1588 can be enabled independently. If both these features are ON, then the MACsec will be in fixed delay mode.

1000BASE-X Mode

The 1000BASE-X speed mode can be configured via the MDIO interface. The line side is configured in 1000BASE-X with Clause 37 AN mode and the system side in 1 Gbps link mode. The 1000BASE-X line-side link is established through Clause 37 auto negotiation between the link partners. Any changes in link status at the line side or system side causes an interrupt to be generated. The software could read the line-side link status and enable the egress traffic.

During normal operation, if the line-side link goes down, a PHY interrupt is issued.

When the device is configured in the 1000BASE-X mode, the MACsec and 1588 features can be enabled independently. If both of these features are ON, then the MACsec will be in fixed delay mode.

SGMII1000/100/10 Mode

The SGMII speed mode can be configured via the MDIO interface. The line side is configured in triple speed SGMII MAC mode, and the system side is configured in SGMII PHY mode. In the SGMII mode of operation, after the line-side SGMII interface links up and establishes the negotiated capabilities, an interrupt is generated. At this point the software verifies the link status, configures the system-side SGMII base page with the appropriate link up capabilities (tx_config_Reg[15, 12:10]), and restarts SGMII auto negotiation. This enables the device and the switch chip to establish the system-side link. After the system side is link up, another interrupt is generated, so practically speaking, the line side receives the capability from its LP, software configures the system-side interface with this information, and the system side pushes those capabilities to the switch chip. Link up on line and system side enables egress and ingress traffic flow. The system-side link can be configured to come up in three ways:

- When the ports are configured, the system-side port will be configured in 1G SGMII and tx_Config_Reg[15] will be set to 1. This will allow the system-side link to come up independent of the line-side link status.
- When the ports are configured, the system-side port will be configured in 1G SGMII, but tx_Config_Reg[15] will not be set to 1. The software will set this bit to 1 only when the line-side links up. This will behave exactly like how the links come up in triple speed SGMII mode.
- During normal operation if the line-side link goes down, a PHY interrupt is issued.

The system-side link can be brought down by several ways. One recommend way when operating in SGMII mode is by setting tx_config_Reg[15] to 0 and restarting auto negotiation. But if the system-side link is kept up, then when the line-side links up again, the software driver must compare both the links. If there is a difference in link configurations, then the system-side link must be reestablished, with new capability. The *BCM848XX Programmer's Register Reference Guide* provides the necessary device ID, register addresses and bit information.

When the PHY is configured in 1000BASE-X mode, MACsec and 1588 can be enabled independently. If both of these features are ON, then the MACsec will be in fixed delay mode.

40G Repeater Mode

The device can be configured in 40G Repeater mode. In this mode all four channels behave like one single port. The MACsec and 1588 TimeSync features are not available in this mode.

10G Repeater Mode

The device can be configured in 10G Repeater mode. The MACsec and 1588 TimeSync features are not available in this mode.

1G Repeater Mode

The PHY can be configured in 1G Repeater mode. The MACsec and 1588 TimeSync features are not available in this mode.

SGMII Repeater Mode

The PHY can be configured in SGMII Repeater mode. The MACsec and 1588 TimeSync features are not available in this mode.

Interrupts

At the top or package level, the BCM848XX supports four hardware interrupt pins. Each interrupt pin corresponds to a channel (or port) and is mulitplexed with different functions as listed below:

- EDC (similar to other PHY device)
- Link Alarm Status Interrupt (LASI) (similar to other PHY device)
- 1588
- MACsec

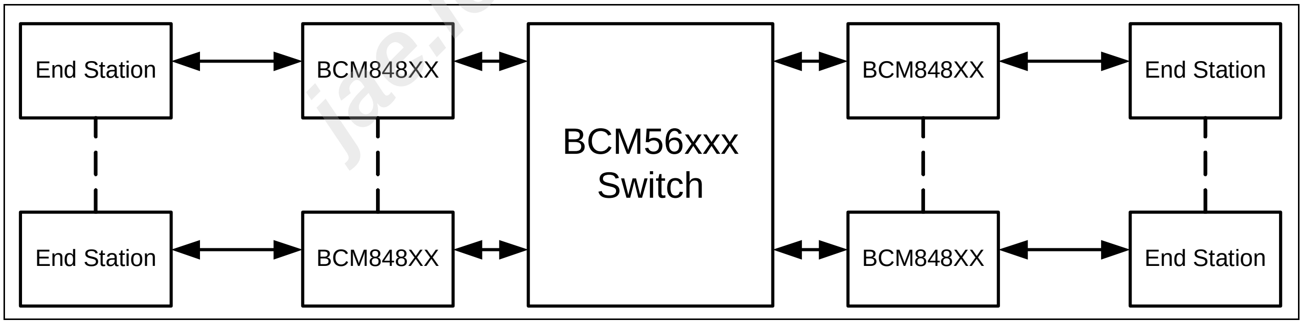
Refer to the *BCM848XX Programmer’s Register Reference Guide* for the different possible interrupt reasons within each function or feature.

MACsec System View

This section provides an example of how the device is being used in MACsec applications at the system level with the StrataXGS® family of switches.

The MACsec system diagram shown in Figure 2 consists of several BCM848XX devices, a BCM56xxx switch, and numerous user end stations to implement a switched LAN application. In a switched LAN, the switch could have security policies to forward traffic only among certain stations or VLANs. However, without per-packet authentication or MACsec, there is no mechanism to verify that a packet is sent by a particular station or VLAN. When the device is configured with the MACsec feature, a CA can be set up between a station and a switch port. All packets sent between the end station and switch port could be authenticated and encrypted. The device performs the following MACsec functions in the ingress and egress packet flows: (1) MACsec protected ingress packet from End stations are decapsulated before forwarding to the switch, and (2) egress packets from the switch are added with MACsec protection before forwarding to the end stations.

Figure 2: MACsec System Diagram



Section 3: MACsec

This section provides an overview of MACsec standard and technology. The majority of the content is extracted directly from the two IEEE standards defined in [IEEE 802.1AE] and [IEEE 802.1X-2010]. This section does not contain any information specific to the device's implementation.

The IEEE MACsec standards define port-based link layer security services and protocols to be used to protect Ethernet-based networks. The IEEE 802.1AE (8021AE) standard specifies the format of the MACsec Protocol Data Unit (MPDU) and the MACsec service provided by MAC Security Entity (SecY) in the MAC Sublayer. It also specifies the interface requirement for security association (SA) management. The IEEE 802.1X-2010 standard defines the key agreement protocol as another entity (KaY) to support the establishment of secure channel (SC) and SA for SecY. Both standards define the MIB management interface for network management of security services.

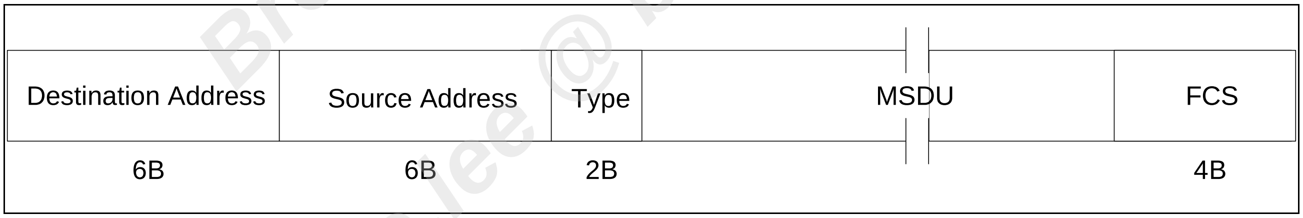
MACsec provides the following:

- Connectionless data integrity
- Data origin authenticity
- Confidentiality
- Replay protection
- Bounded receive delay

MAC Security Protocol and Frame Format

MACsec provides security service to the MAC Service Data Unit (MSDU). Figure 3 shows the user data (MSDU) encapsulated with standard Ethernet-II packet fields.

Figure 3: Ethernet-II Packet Format



Destination Address	Source Address	Type	MSDU	FCS
6B	6B	2B		4B

MACsec encapsulation adds a SecTAG in front of the user data, and an Integrity Check Value (ICV) between the user data and FCS, as shown in Figure 4 on page 16. The Type field forms the first 2 bytes of the SecTAG and has the value of the newly defined MACsec EtherType.

Confidentiality can be optionally provided on the packet data between the SecTAG and the ICV. The ICV is calculated over the MAC Addresses, additional inserted fields before the SecTAG, the SecTAG and packet data to offer integrity protection on the entire packet.

Additional tags and headers after the SecTAG are supported on byte boundary from 0 to 63 bytes. Typically, these additional tags and headers such as VLAN Tags and LLC headers will be part of the secured data after the SecTAG. Figure 4 shows a packet with a VLAN Tag after the SecTAG and Figure 5 shows MACsec encapsulation of an IEEE 802.3 SNAP packet.

Figure 4: Secured VLAN Tag in MACsec Packet

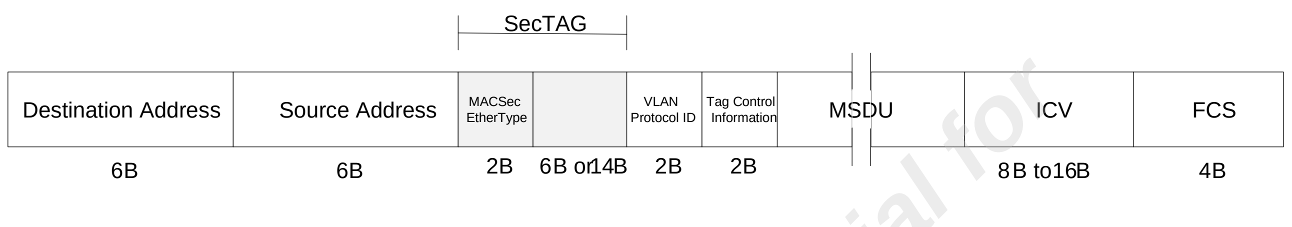
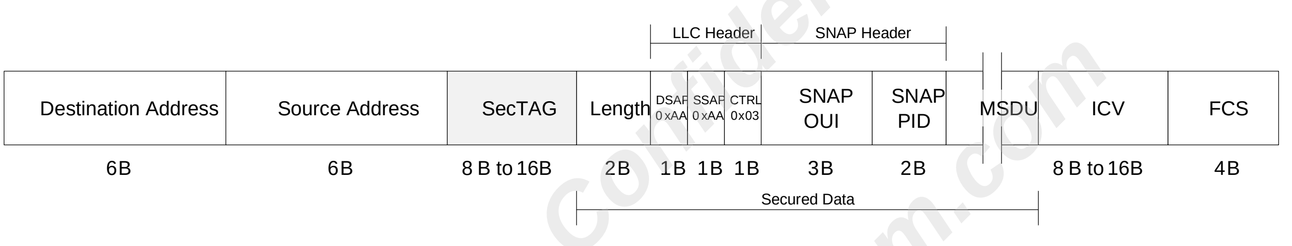


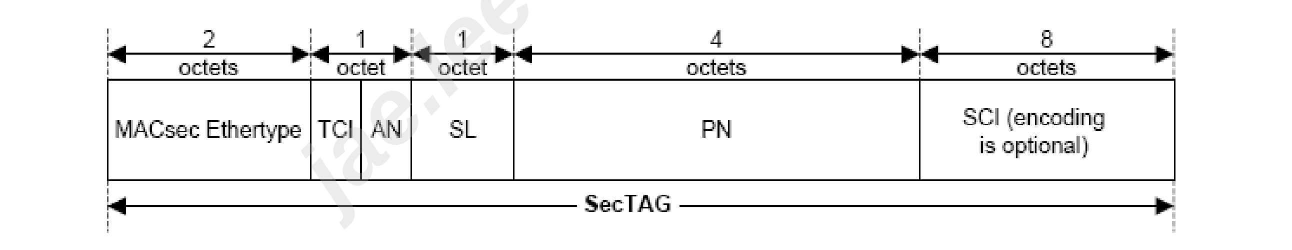
Figure 5: MACsec Encapsulation of IEEE 802.3 SNAP Packet



The size of the SecTAG is either 16 bytes or 8 bytes depending on the inclusion of the SCI field. The format of the SecTAG is shown in Figure 6.

The SecTAG starts with a new MACsec EtherType with value 88-E5. The Tag Control Information (TCI) field contains six flags detailed in Figure 6. The AN field is a two-bit security association number. The SL (short length) field, if it is non-zero, specifies the unpadded User Data (MSDU) length for packets less than 48 bytes. The PN field is a 32-bit packet number for replay protection. The Secure Channel Identifier (SCI) is the concatenation of the 48-bit MAC Source Address (MAC_SA) and the 16-bit port number.

Figure 6: SecTAG Format



The TCI field defines the following flags:

- V=0, version number default to zero
- ES, end station
- SC, SCI included
- SCB, EPON Single Copy Broadcast
- E, encryption protected
- C, changed text

The combinations of E-bit and C-bit are described in Table 1.

Table 1: *SecTAG E-bit and C-bit Encoding*

Combination	***Description***
E=0, C=0	No encryption, UserData is not modified, ICV is 16 bytes
E=0, C=1	No encryption, UserData is modified or ICV is not 16 bytes
E=1, C=0	None-SecY encoding reserved for KaY
E=1, C=1	Packet is encrypted and integrity protected

The combination of ES, SC, and SCB are described in Table 2.

Table 2: *SecTAG ES, SC, and SCB Encoding*

Combination	***Description***
ES=1, SCB=0=> SC=0	End station transmit, SCI is not included, MAC Source Address is used to determine SCI, SCI={SA, 0x0001}
ES=X, SCB=1=> SC=0	EPON Single Copy Broadcast, SCI is not included, MAC Source Address is used to determine SCI, SCI={SA, 0x0000}
SC=1=> ES=0, SCB=0	SCI is included explicitly in the 16-byte SecTAG

Ethernet places a minimum size restriction of 64B on the transmitted packets. Undersized packets are padded with 0s before transmission. The FCS is calculated over the original packet plus the padding.

With MACsec, the MSDU must o be encapsulated with MACsec fields, specifically the SecTAG and the ICV. The resulting packet, if undersized, is then padded with 0s. The FCS is calculated over the MACsec encapsulated packet and padding. The SL field of the SecTAG, specifies the length of the secured data between the SecTAG and ICV, if the length is less than or equal to 48 bytes. The SL field enables the ICV to be located in the packet, when padding is inserted between the ICV and the FCS on transmission.

Packets received that are less than 64 bytes are dropped and the appropriate statistics counters incremented.

The addition of the SecTAG and ICV can increase the size of the resulting packet so that it is no longer within the maximum frame size limitation of the link. Such frames will be discarded by the receiver. The maximum size of the MAC Service Data Unit (MSDU) should be lowered so that the resulting size of the MPDU, after MACsec encapsulation, is within the maximum frame size limitation of the transmission link.

Cipher Suite

MACsec defines a mandatory default cipher suite, AES-GCM-128. The cipher suite can be used to provide integrity-only protection or optionally both integrity and confidentiality protection. The cipher key is a 128-bit AES key. The 96-bit initial vector (IV) used by the AES-GCM mode of operation is defined as follows:

IV = {SCI || PN}

The integrity protection is over the DA, the SA, the SecTAG (with SCI) and the UserData. The confidentiality protection is over UserData (with the consideration of ConfidentialityOffset). The ICV is 16 bytes long.

The maximum length of the UserData is 2^{16-1} bytes. The maximum number of packets per SA session is 2^{32-1}.

Principles of MACsec Operation

Logically the SecY portion of the MACsec layer has three logical ports: the Common Port, the Controlled Port and the Uncontrolled Port.

MACsec uses an insecure MAC Service Access Point at the Common port to provide a Secure MAC Service Access Point to the client of its Controlled Port, and an insecure MAC Service Access Point to the client of its uncontrolled port.

All packets received on the Common Port are available to the client of the Uncontrolled Port. Only packets received on the Common Port which passes the SecY security processing will be forwarded to the client of the Controlled Port. Typically the client of the Uncontrolled Port will drop all traffic except packets for network configuration and MACsec key agreement, thus enabling keys to be setup and network configured before the main traffic flow is enabled on the Controlled Port.

The MACsec protocol is designed to operate in both point-to-point LAN and shared-media LAN environments. The SecY entity is associated with a port which is in term associated with a single connectivity association (CA) at a given time. The SecY establishes multiple secure channels (SC) with its peers in the same CA. The SC is unidirectional. It is determined by the transmit SecY. Each SC can have up to 4 Security Associations (SA). (Note that the device supports 3 modes: 1:1, 2:1 and 4:1 for SC to AN mapping) The transmit SecY assumes certain time delay (one second) between the establishment of its SA and the same SA at the receiver.

MACsec provides security service to requests made at the controlled port using the active transmit SA pointed to by the SA identifier (constructed from {SCI || AN}). In doing so, MACsec inserts the SecTAG into the frame, performing integrity computation and inserting the ICV into the frame. MACsec optionally encrypts the UserData. The FCS is updated after the packet has been protected. MACsec tracks the packet number (PN) and stops transmitting packets of an SA after the NextPN number reaches the maximum allowable value. MACsec also updates management information.

When a frame is received on the common port, MACsec performs SA look-up based on the SA identifier if the frame is protected by MACsec. The frame is decrypted and authenticated. Non-authenticated frames are discarded. Authenticated frames are optionally checked for replay attacks. Replayed frames are discarded. Afterwards, the replay window is updated. The management information for the receive channel is also updated.

In switched network, point-to-point connection is typically used. In which case, a minimum of two SCs are required: one transmitting SC and one receiving SC.

MACsec Key Management

The MACsec SAK management is defined by IEEE 802.1X-2010 standard, which is still in development. As far as the SecY module is concerned, a Layer Management Interface (LMI) has to be provided to load and unload the SA and check the operational status of the port. The SecY specification further defines the following constraints:

- A minimum conformance requirement of two active receive SAs for a minimum single receive SC must be supported.
- The receive channel must support SA swapping without interrupting traffic.
- The transmit channel may refresh SA by temporarily turning off ControlledPortEnabled configuration.
- The time bound within which the receiver can accept interleaved SA is 0.5 second.

MACsec SecY Architecture

This section describes the architecture of the MACsec SecY module inside the device. SecY is the service module name for the link layer traffic protection as specified in the IEEE 802.1AE standard.

The SecY module relies on the line-side MAC layer to realize the following functionality:

- CRC checking and removal for ingress frames
- Undersized frame filtering on ingress
- Oversized frame truncation on ingress
- Processing of received pause frames and management of the pause timer counter
- CRC generation on egress frames
- Enforcement of minimum Inter-Frame-Packet Gap (IPG) on transmit
- Determination that the Common Port is Operational
- Statistics collection for counters common to both the Common Port and the Uncontrolled Port

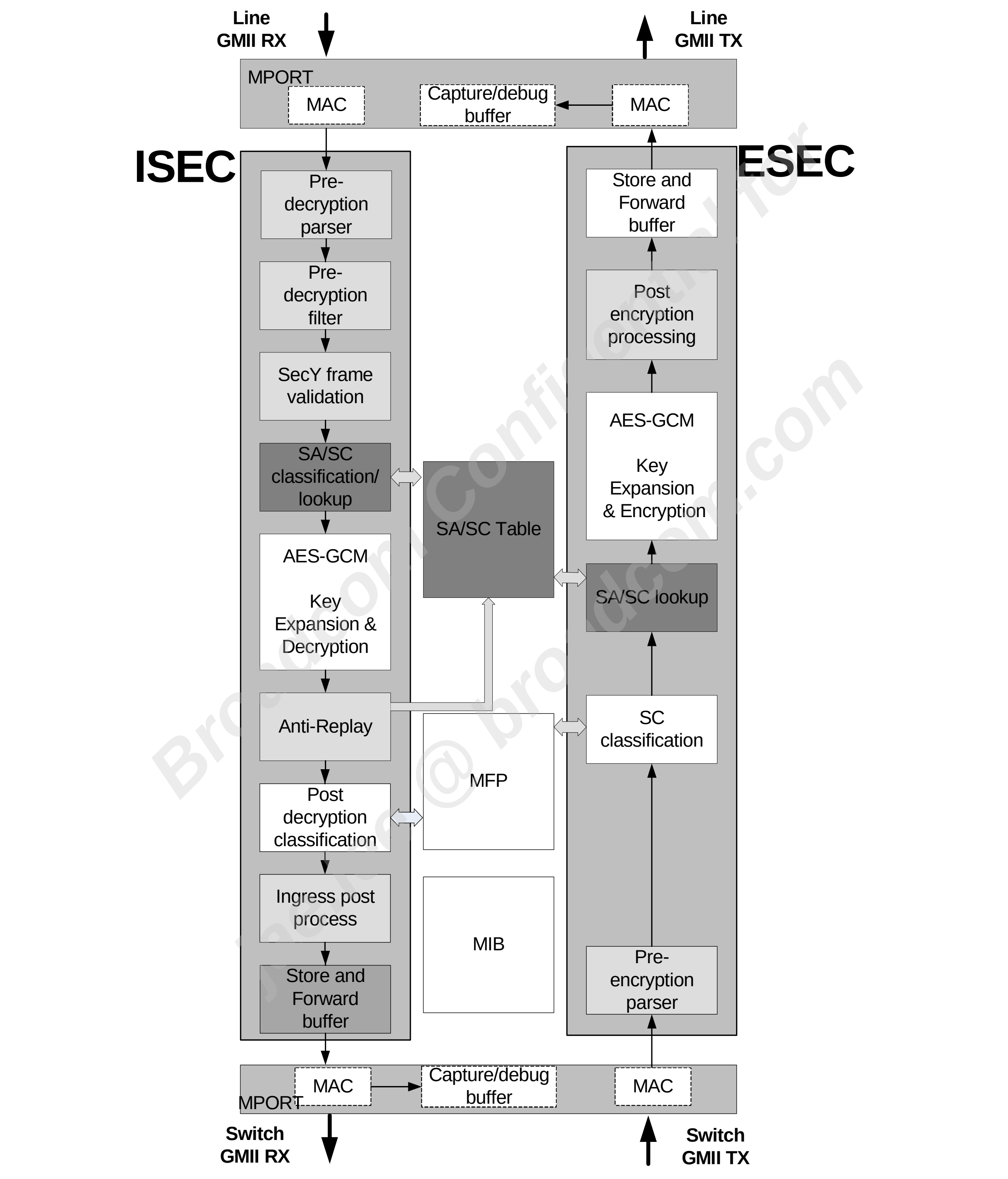
The SecY module relies on the system-side MAC layer to perform the following tasks:

- Transmit and receive Ethernet frames over XAUI/XFI
- Regenerate CRC after MACsec frame processing
- Perform full-duplex flow control
- Pad short packets after the MACsec header is stripped off

The line-side MACs are addressed as ports 0 through 3 and the system-side MACs are addressed as ports 4 through 7.

Top Level Architecture and Data Flow

Figure 7: MACsec SecY Block Diagram



The MACsec SecY module top level block diagram is shown in Figure 7 on page 20. The SecY module is logically partitioned in to an Ingress Security Module (ISEC) and an Egress Security Module (ESEC). The ingress and egress logic operate independently as two data pipeline with the exception of the flow control handshakes, SC/SA table, MFP table and MIB counters are shared between ingress and egress paths.

The ingress logic and egress logic are in the middle of two MAC layers, the system-side MACs and the line-side MACs. The SecY module is designed to operate in either 10 Gbps mode or 1 Gbps mode via configuration through the MDIO interface. Two separate MACs are used to support 10 Gbps and 1 Gbps data rate independently. The SecY module operates in either cut-through mode or store-and-forward mode. A 32 KB packet buffer is provided at the ingress path for flow control purposes and is able to support lossless application with packet size up to 9.6 KB. The device is able to support packet size up to 16 KB with lossy flow control. Similarly, a separate 32 KB packet buffer is provide at the egress path. These packet buffers are accessible via the MDIO.

Functionally the SecY module contains the following key components:

- Ingress line-side MACs
- ISEC path which consists of:
  - Ingress traffic pre-filter (Pre-decryption parser and Pre-decryption filter)
  - Ingress PDU validation
  - Ingress CA/SA look-up
  - AES-GCM engine for ingress frame processing
  - Ingress packet verification, anti-replay
  - Ingress post decryption packet classification
  - Ingress post processing
  - Ingress store-and-forward packet buffer
- Ingress system-side MACs
- Egress system-side MACs
- ESEC path which consists of:
  - Egress SC packet classification to separate controlled and uncontrolled MACsec flows
  - Egress CA/SA look-up
  - AES-GCM engine for egress frame processing
  - Egress post encryption processing
  - Egress store-and-forward packet buffer
- Egress line-side MACs
- Registers and tables access via the MDIO, LMI, and SBUS protocols. (Red lines in the diagram)
- Supports packet store-and-forward mode or fixed latency mode.

Ingress and Egress Packet Processing Flows

In the context of this document, the ingress traffic refers to the traffic received from the line-side MAC, processed by the SecY module and forwarded to the switch. The egress traffic refers to the traffic received from the system-side MAC, processed by the SecY module and transmitted over the wire.

Ingress Packet Flow

This section describes the ingress packet and data processing flows. As shown in Figure 8, Figure 9 on page 23, and Figure 10 on page 24, The ingress packet flow starts at the line-side MAC and ends at the system-side MAC. For readability, the flow is split into three separate diagrams. IFP-1, IFP-2, IFP-3, and SC/SA are connectors between the three diagrams.

Figure 8: MACsec Ingress Packet Flow—Part 1

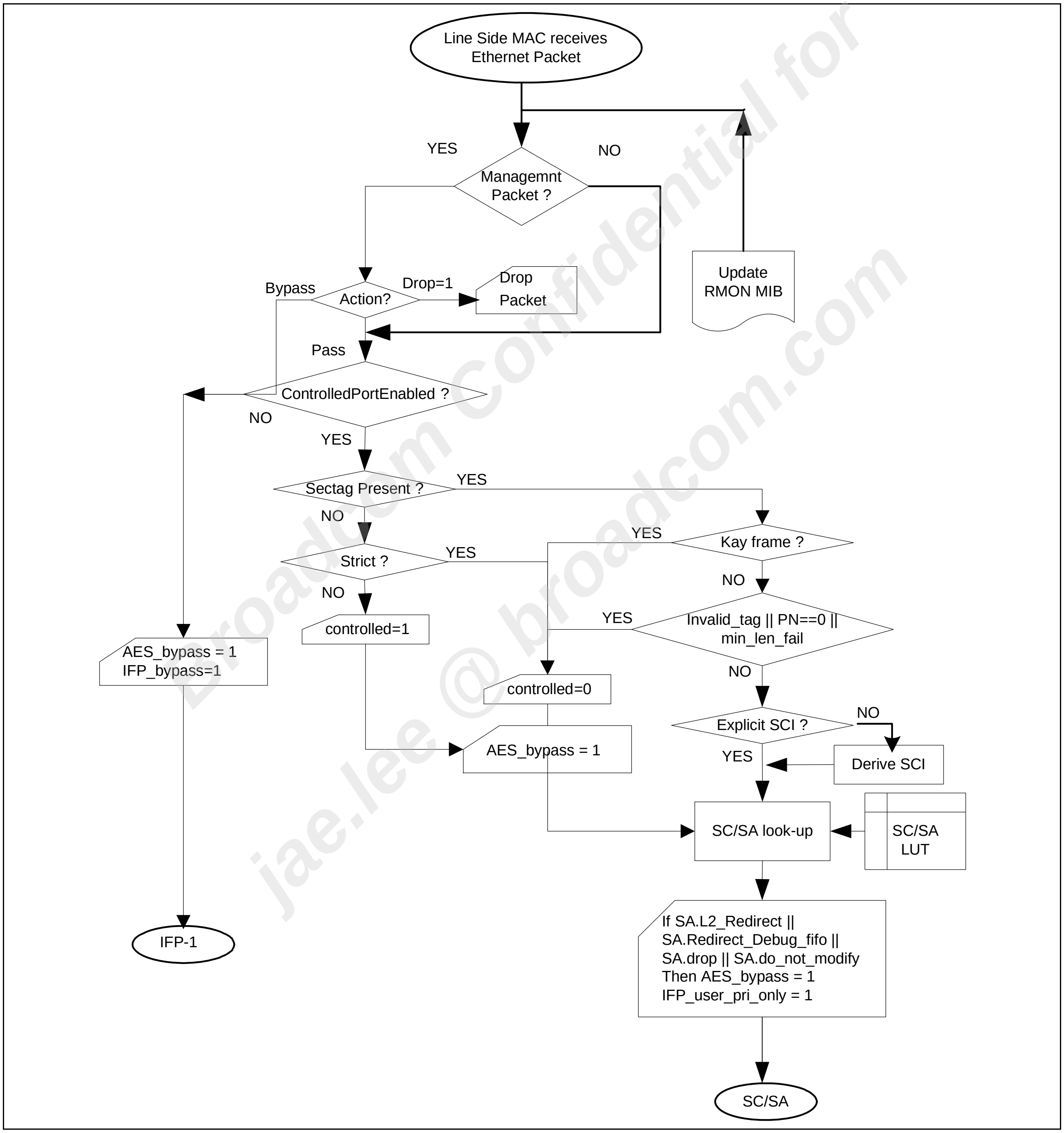


Figure 9: MACsec Ingress Packet Flow—Part 2

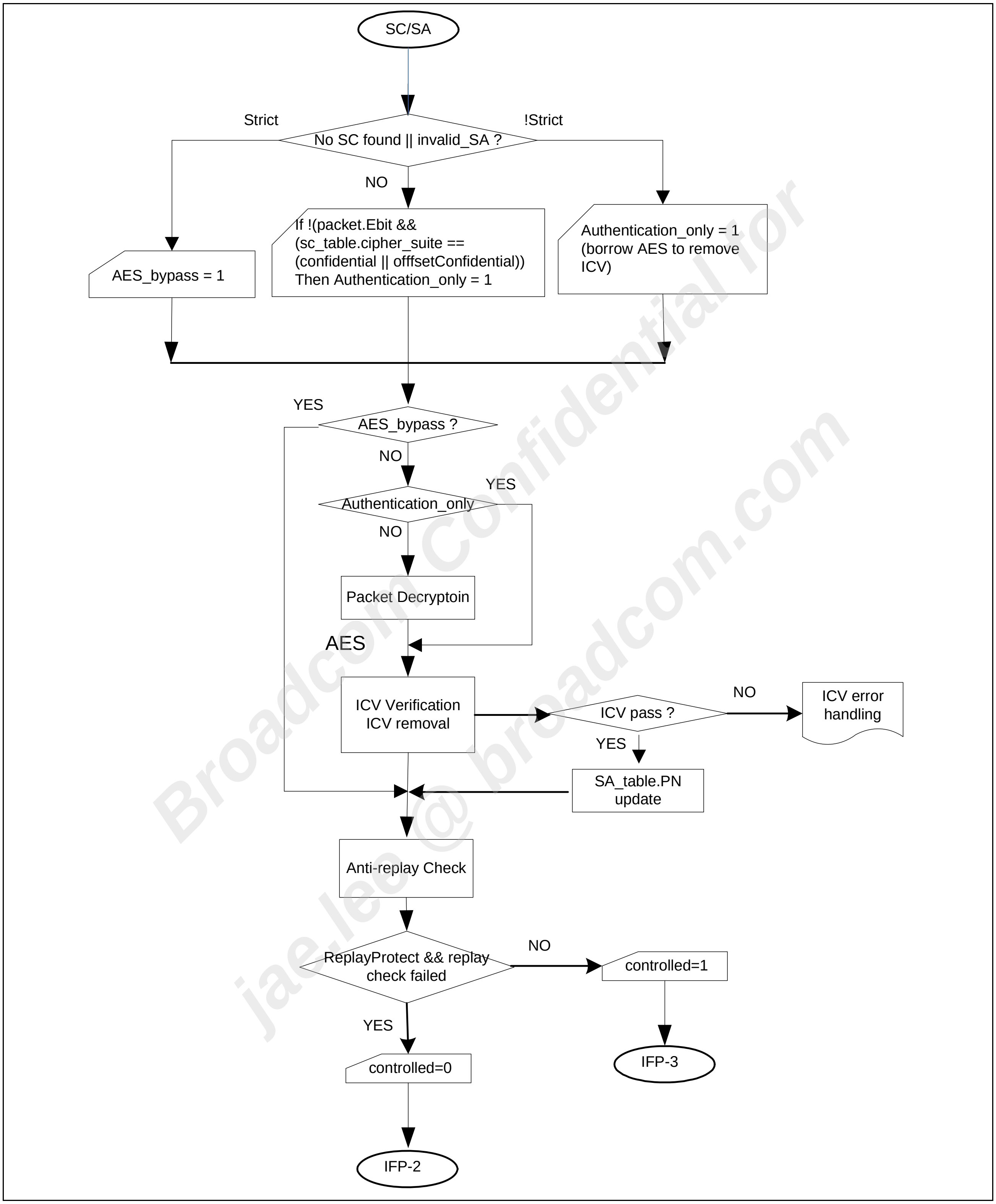
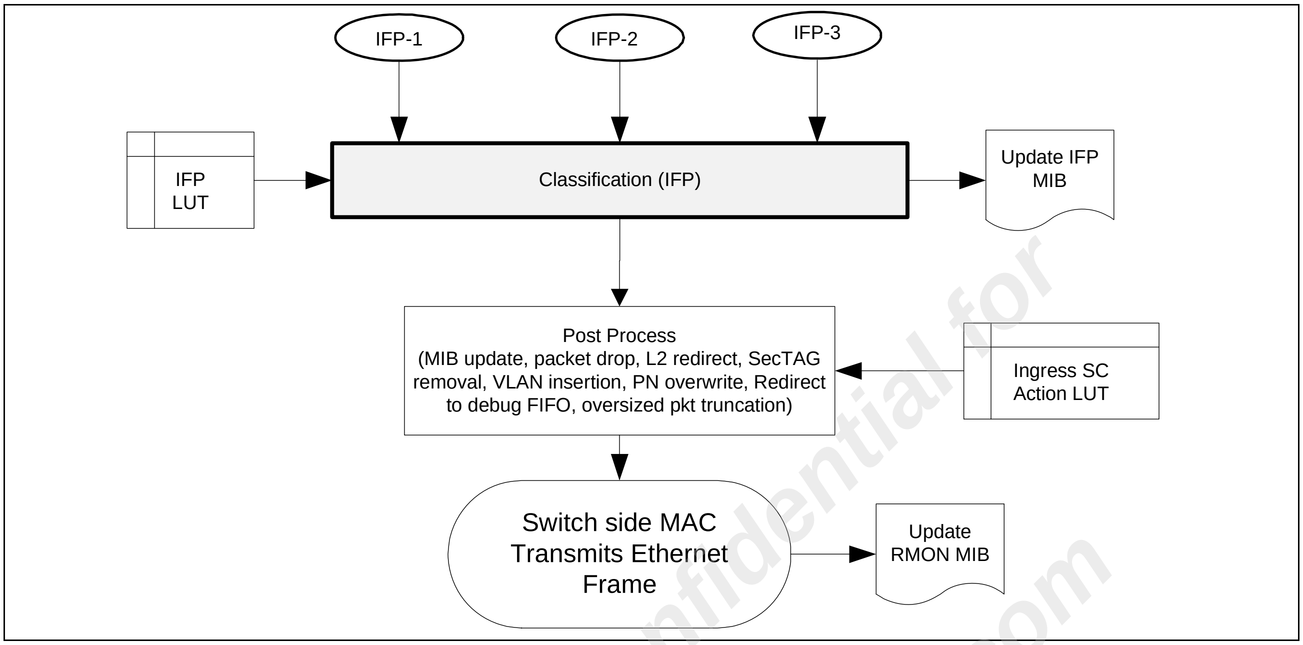


Figure 10: MACsec Ingress Packet Flow—Part 3


IFP-1
IFP-2
IFP-3
IFP LUT
Classification (IFP)
Update IFP MIB
Post Process
(MIB update, packet drop, L2 redirect, SecTAG removal, VLAN insertion, PN overwrite, Redirect to debug FIFO, oversized pkt truncation)
Ingress SC Action LUT
Switch side MAC Transmits Ethernet Frame
Update RMON MIB

Strict = (ValidateFrames==Strict || packet.Cbit)

KaY frame = SecTAG && Ebit==1 && Cbit==0

IFP: Ingress filter classification engine

The ingress packet flow is divided into the following stages and they are described in detail in the next few subsections:

1. Ingress Pre-Filter: consists of 21 rules used to separate MACsec and special management traffic.
2. Ingress PDU Validation and SC/SA Lookup: MACsec SC and SA processing
3. Ingress MACsec Packet Processing: MACsec SecTAG processing, anti-replay check, and error handling.
4. Ingress Post Decryption Packet Classification:ACL processing based on the first 64 bytes of the packet header
5. Ingress Traffic Categorization: Final packet modification based on results from previous stages.

Ingress Pre-Filter

The ingress pre-filer divides management traffic and MACsec-related traffic. The ISEC block consists of a set of pre-filter registers to explicitly allow certain type of management packets that matched any of the rules setup in the pre registers. These registers are maintained per port. Table 3 shows a summary of the pre-filter matching registers with the highest precedence or priority in the first row. If multiple hits occur in this table, the highest precedence entry wins.

Table 3: *Ingress Pre-Filter Table*

Register Name	*# of Rules*	*Description*
N/A	1	Predefined MAC DA to be matched in the packet. Predefined MAC DA 0: 0x01_80_C2_00_00_00
N/A	1	Predefined MAC DA to be matched in the packet. Predefined MAC DA 1: 0x01_00_0C_CC_CC_CC
Ingress Pre MAC DA[7:0]	8	MAC DA to be matched in the packet.
Ingress Pre EtherType	8	EtherType to be matched in the packet. Each register contains two EtherType
Ingress Pre MAC DA[9:8} (Range)	1	These two registers specify the range of MAC DA to be matched.
Ingress Pre MAC DA EtherType	2	MAC DA and EtherType combination to be matched in the packet.

The actions corresponded to the ingress pre-filter registers are specified in the Ingress_Pres_Psr_Ctrl register. Each rule has two bits to control with the following encoding: 0 = disable, 1 = Pass control packet to further processing, 2 = Bypass control packet from ISEC processing, 3 = Drop management packet If hits multiple rules. The priority of these rules is as follows (1 is highest priority):

1. MACDA_HARDCODED_VAL0_EN,
2. MACDA_HARDCODED_VAL1_EN,
3. MACDA_EN all 8 MAC DAs have same priority, note that software should not program the same MAC DAs with different actions,
4. ETYPE_EN, all 8 ETYPEs have same priority, sw should not program same ETYPEs with different actions)
5. MACDA_RANGE_EN
6. MACDA_ETYPE_EN (all 2 MACDA_ETYPEs have same priority, sw should not program same MACDA_ETYPEs with different actions)

Associated with each entry of the table is a 32-bit counter. An example usage of the counters would be to count packets per EtherType. If there are multiple hits for an incoming packet, only the counter in the highest precedence entry is incremented for that packet.

By default, packets that do not match the filter table and are without a SecTAG are dropped. These packets are logged in the InPktsNoTag and InPktsUntagged MIB registers for the controlled port depending on whether the ValidateFrames is set to Strict or not. A single counter is updated per dropped packet.

Optionally, for servicing SecY uncontrolled clients, these packets can be presented to the Ingress Post Decryption Packet Classification Engine for further policy filtering if the IngressUncontrolledFiltering flag in the Master Control Register is set.

Ingress PDU Validation and SC/SA Lookup

After the pre-filter table matching, the next step in the ingress packet flow is PDU validation and SC/SA lookup.

PDU Validation

PDU validation checks the consistency of the information embedded in the SecTAG. It also checks the length of the packet for undersized packets. While the length is being checked, the ShortLength field is examined to determine the location of the ICV for ICV verification. When the PDU verification fails, the packet may be dropped and the event logged into the MIB database.

If the incoming packet exceeds the value programmed in the Maximum Transmission Unit Register, the logic corrupts the CRC of the packet and records the event in the MIB counter. The MTU check is per packet priority level.

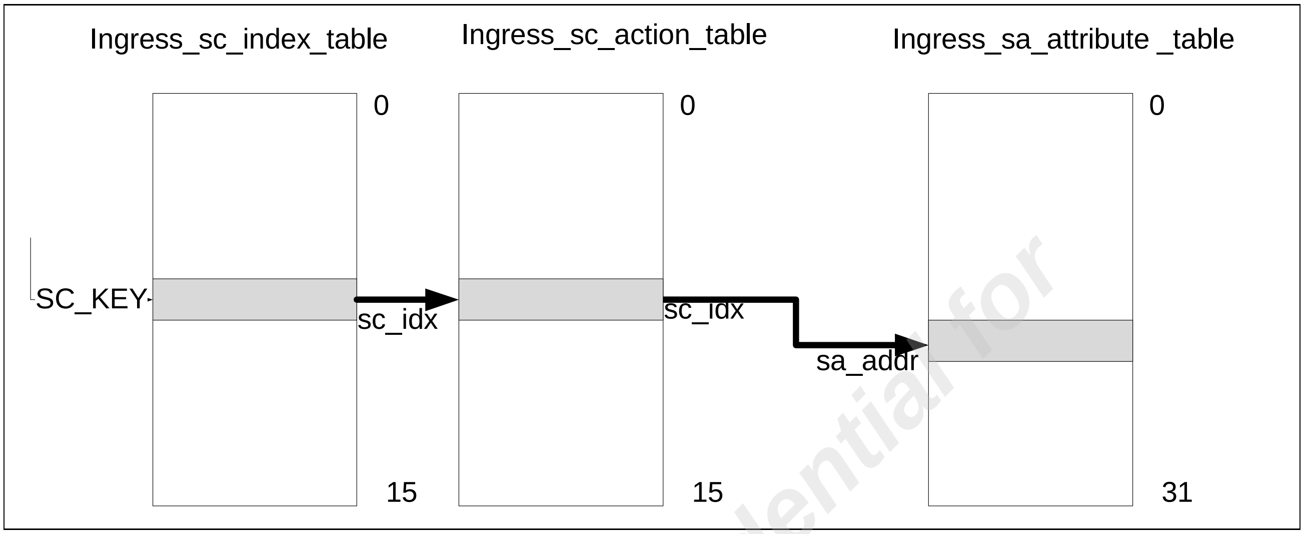
SC/SA Lookup

During SC/SA lookup, A SC_KEY is formed and used to lookup the SC_TABLE for a matching entry. The ingress SC_KEY is based on the fields defined in bits [203:0] in Table 4 on page 27. Some fields are byte maskable using the KEY_BYTE MASK field and some fields are bit maskable using the KEY_BIT_MASK field in Table 4.

Figure 11 on page 27 shows the SC/SA lookup process. The address into the SA Attribute Table is generated based on the per port SC_MODE and is defined as follows:

```
if (SC_MODE == 1:1)
    sa_addr = sc_idx[3:0]
else if (SC_MODE == 2:1)
    sa_addr = {sc_idx[3:0],sectag.AN[0]}
else
    sa_addr = {sc_idx[2:0],sectag.AN[1:0]}
```

Figure 11: Ingress SC/SA Lookup



[Table 4](#) shows the Ingress SC INDEX Table format and Table 5 on page 28 shows the Ingress SA Attribute Table.

Table 4: *Ingress SC Index Table*

Field Name	***Bit Range***	***Description***
VALID	[250]	1'b0: invalid entry, 1'b1: valid entry
AN1	[249:248]	One of the 2 valid AN, only valid in 2 SA per SC mode
AN0	[247:246]	One of the 2 valid AN, only valid in 2 SA per SC mode
CIPHER_SUITE_PROTECTION	[245:244]	2'b00: Integrity, 2'b01: Confidentiality, 2'b10: OffsetConfidentiality, 2'b11: reserved
KEY_BIT_MASK	[243:228]	Masks the fields - TCI_AN, PKT_TYPE, L2_FRAME_TYPE, OUTER_TAGGED, OUTER_TAG_TYPE, SECTAG_PRESENT, EGR_IGR_LPBK on a per-bit basis
KEY_BYTE_MASK	[227:204]	Masks the rest of the fields on a per byte basis, VLAN is treated as 2 bytes even though it is a 12-bit field
EGR_INGR_LB	[203]	Indicates if the packet is looped backed from Egress
OUTER_TAGGED	[202]	1 indicates packet contains a outer VLAN tag and VLAN field is valid
OUTER_TAG_TYPE	[201]	1 - indicates the outer VLAN tag is a QTAG, 0 - indicates the outer VLAN tag is a STAG
SECTAG_PRESENT	[200]	1 indicates packet contains a SecTAG
PKT_TYPE	[199:198]	00 - non-MACsec packet, 01 - MACsec packet, 1x - Reserved
L2_FRAME_TYPE	[197:196]	00 - EII, 01 - SNAP, 10 - LLC, 11 - Reserved
TCI_AN	[195:188]	6-bit TCI, 2-bit AN
VLAN	[187:176]	VID of the Outer VLAN Tag of a packet
MAC SA	[175:128]	MAC SA extracted from packet
MAC DA	[127:80]	MAC DA extracted from packet
EtherType	[79:64]	EtherType

Table 4: *Ingress SC Index Table (Cont.)*

Field Name	*Bit Range*	*Description*
SCI	[63:0]	64-bit SCI number

Table 5: *Ingress SA Attribute Table*

Field Name	*Bit Range*	*Description*
VALID	[157:156]	2'b00: invalid entry, 2'b01: reserved, 2'b11: valid and fresh, 2'b10: valid and in use
SEGTAG_MODE	[155:153]	0: Preserve SecTAG 1: Reserved 2: Preserve SecTAG and replace SecTAG.PN with {16'h0,MACSECID[15:0]}.MACSECID will be derived from INGRESS_SC_ACTION_TABLE and register INGRESS_SPI_CONFIG. 3: Remove SecTAG 4: Replace SecTAG with VLAN tag. The VLAN_ID will come from INGRESS_SC_ACTION_TABLE 7-5: Reserved
DO_NOT_MODIFY	[152:150]	1-indicates no modification of the packet. AES engine should be bypassed 1 bit for each packet type: Bit0 is reserved Bit1 for non-MACsec packets Bit2 for MACsec packets
L2_REDIRECT	[149:147]	1- Indicates to redirect packet. 1 bit for each packet type (see above). For the redirect packets (including the redirect packets from this table and IFP action table), L2_ENCAP mode will be used to encapsulate the packet with a new L2 header. A per-port configuration register SP_Ingress_L2_Sectag_Override_Cntrl.SECTAG_PN_OVRRIDE_EN will determine whether or not to update the SecTAG.PN with Error Status Code. PN MSB will be replaced by Status Code and LSB will be replaced by 16-bit zeros.
REDIR_DEBUG_FIFO	[146:144]	1- indicates to redirect to capture FIFO 1 bit for each packet type: -- Bit0 for management (control) packets -- Bit1 for non-MACsec packets -- Bit2 for MACsec packets
DROP	[143:141]	1- indicates to drop packet 1 bit for each packet type: -- Bit0 for management (control) packets -- Bit1 for non-MACsec packets -- Bit2 for MACsec packets

Table 5: *Ingress SA Attribute Table (Cont.)*

Field Name	***Bit Range***	***Description***
DROP_IF_NOT_LB	[140:138]	1- indicates to drop packet if it is not looped back from egress 1 bit for each packet type: -- Bit0 for management (control) packets -- Bit1 for non-MACsec packets -- Bit2 for MACsec packets
ReplayProtectWindow	[137:106]	Replay Protect Window
Protection_Offset	[105:98]	The offset of to apply confidentiality protection from the start of the MSDU
Replay Control	[97]	0- Replay Protect disabled. 1- Replay Protect enabled.
Unused	[96]	Reserved bit, not used
Next_PN	[95:64]	Next packet sequence number (NextPN)
START_TIMER	[63:32]	The system time when this receiving SA last started receiving MACsec frames.
STOP_TIMER	[31:0]	The system time when this receiving SA last stopped receiving MACsec frames.

In the IEEE 802.1AE standard based lookup mode, the SCI is derived from the following sources:

- From the SCI if it is present in the packet.
- If the SCI is not present (SecTAG.SC = 0), an SCI can be derived by checking the flags in the SecTAG:
  - SecTAG.ES
  - SecTAG.SCB
- If neither is set and OperPointToPointMAC is set, a default SCI at SC Table Entry 0 is used

In the L2 Header lookup mode, the SCI is extracted directly from the packet if TCI.SC =1, else it is set to 0.

Once the SC is determined, the AN field of the packet, plus the SC index, comprises the SA index used to locate the SAK table (described in "Egress SC/SA Lookup" on page 43). If the L2_Redirect bit is set in the SAK Entry Table, the packet is encapsulated with an L2 header. If the L2_Redirect bit is not set and the SAK is valid, the SAK is expanded into round keys and stored in the key buffer for further packet processing. Otherwise, the corresponding event is logged.

Ingress SA-Based L2-Redirect

If the L2_Redirect bit in the SA entry is set, the packet is tagged as L2–redirected and does not go through SecY Processing. Further, a L2–Redirected packet bypasses all subsequent processing including SecY Ingress Classification Engine. The entire original frame is L2 encapsulated with the MAC DA, MAC SA, and EtherType from the registers and forwarded to the switch.

A controlled packet that has a SecTAG is always forwarded to the ingress classification engine regardless of whether an error condition is detected or the packet is tagged as a packet to be dropped. The actual drop decision will be made in this stage. The ingress post–processing can override the SecY processing drop decision and redirect the packet to the switch.

Ingress MACsec Packet Processing

Ingress packet processing includes PDU validation, packet decryption and ICV verification, short packet handling and anti-replay protection and error packet handling, and so on.

Packet Decryption and Authentication ICV Verification

Packet is decrypted if the PDU verification passes. The authentication ICV is recomputed by the AES-GCM engine based on the authentication payload prior to decryption. The computed ICV is compared against the ICV transmitted in the packet. If authentication fails, the packet is logged. The ICV is removed from the packet after the ICV verification.

Ingress Anti-Replay Verification

The Ingress anti-replay protection is done post decryption when the ReplayProtect is enabled. A 32-bit programmable replay window size can be specified on a per-SA basis via the ReplayProtectWindow.

The post-decryption replay verification is done after the packet is processed through the AES engine. It checks if the PN number is less than SA.NextPN – ReplayProtectWindow to determine if the packet is a replayed packet. The same action is taken when the packet is deemed as a replayed packet. Further, this step updates the NextPN and the field in the corresponding SA using the following approach.

An SA expires when its NextPN reaches 0xFFFFFFFF. At most, 2^{32-1} packets can be transmitted with a given SA. Normally, when a valid packet is received with a PN equal to or greater than the NextPN field stored in the SA descriptor, the NextPN field is updated with Max(NextPN+1, Packet PN). When the PN in the packet is 0xFFFFFFFF, and the NextPN is also 0xFFFFFFFF, the field is not updated. Instead, the NextPN_Saturation bit is set. The KaY entity can invalidate the SA at anytime by installing a new SA.

Ingress Error Packet Handling

The ingress MACsec packet processing identifies error cases and tags the packets. Packets tagged as dropped represent the error conditions that should ultimately cause the packets to be discarded. However, the actual action does not take place until the Ingress Packet Classification stage. When SecY is operating in cut-through mode, packets subject to certain error types cannot be dropped by SecY. The packets must be discarded by the switch. As shown in Figure 6 on page 16, there are many conditional checks required by the standard when the packet traverses the MACsec processing steps. Regardless of the SecY's ValidateFrames setting, the following error packets are always dropped as controlled port packet:

- Invalid SecTAG or packet format detected in early PDU validation
- Early replay check failed or final replay check failed

If SecY's ValidateFrames is configured to STRICT, then packets with the following errors are also dropped as controlled port packet:

- No matching SC/SA or the matched SC/SA is not valid
- ICV check failed

A controlled packet with SecTAG is always forwarded to the classification engine regardless of whether an error condition is detected or the packet is tagged as a packet to be dropped.

Ingress Post Decryption Packet Classification

All packets after SecY processing could be matched against an entry in the Ingress Post Decryption Packet Classification Engine. This includes both traffic identified as Uncontrolled and Controlled.

The Ingress Post Decryption Packet Classification Engine is organized similarly to the Egress Packet Classification Engine. The TCAM and the Action Table RAM are shared between the ingress Packet Classification Engine and the egress Packet Classification Engine although the content of the memories are organized differently. The Direction bit in the Filter Processor (FP) entry defines whether the entry is applied for either ingress or egress packet flow. See Figure 7 on page 20.

Ingress Filter Processor Parser

The ingress FP parser parses the incoming packet to form the lookup key to the TCAM. The lookup key consists of fixed fields from the packet and user-defined fields (UDF) extracted at configurable offsets from the first 64 bytes of the packet. The fields within the key are as shown in Table 6.

Table 6: *Ingress FP Key Definition*

Selector Name	*Bit Range*	*Description*
Key_Valid	[468]	Indicates whether this is entry is valid 1= Valid; 0 = Invalid
Key_Mask	[467:234]	Per-bit key comparison enable for the Key fields defined in [233:0] 1: Compare; 0: Ignore
Direction	[233]	0: Ingress; 1: Egress This bit should set to 0 for IFP
Port number.	[232:229]	Port Number
SecTAG_Status	[228:227]	'b00: Frame does not contain a SecTAG 'b01: Frame contains a SecTAG 'b10: Reserved 'b11: Reserved
Frame_format	[226:225]	Type of Ethernet frame. 0: Ethernet II packet (LENTYPE>=CP_ETYPE_MAX_LEN) 1: SNAP packet (aa-aa-03-00-00-00) 2: LLC packet (LENTYPE < CP_ETYPE_MAX_LEN and!SNAP) 3: Reserved

Table 6: ***Ingress FP Key Definition (Cont.)***

Selector Name	*Bit Range*	*Description*
Vlan_tag_status	[224:221]	Type of VLAN Tags: 4'h0:untagged packet 4'h1:Single VLAN Tag. It is Inner Tag 4'h4:Single VLAN Tag. It is ST-VLAN Tag 4'h5:Single VLAN Tag. It is Outer TPID1 (QTAG) 4'h6:Single VLAN Tag. It is Outer TPID2 (STAG) 4'h7:Single VLAN Tag. It is Outer TPID3 4'h8:Double VLAN Tag. Outer Tag is ST-VLAN Tag 4'h9:Double VLAN Tag. Outer Tag is Outer TPID1 (QTAG) 4'ha:Double VLAN Tag. Outer Tag is Outer TPID2 (STAG) 4'hb:Double VLAN Tag. Outer Tag is Outer TPID3 others: reserved
Pkt_type	[220:219]	Type of the packet. 00 non-MACsec packet 01 MACsec packet 1x Reserved
Reserved	[218:217]	Reserved
Egr_ingr_lb_bit	[216]	Indicates the packet is loopbacked from egress direction
security_status	[215:208]	0x1: controlled_port_packet. Indicate the controlled port packet 0x2: replay_fail. Replay failed error 0x4: SA miss. SA miss error 0x8: SC miss. SC miss error 0x10: VersionChkFail. SecTAG version check failure 0x20: TciChkFail. Illegal TCI combination failure 0x40: IllegalSL. Illegal Short Length which is by SL byte value in SecTAG is equal to or larger than 48. 0x80: IllegalST. Illegal SecTAG (OR of IllegalSL, TciChkFail, VersionChkFail, PN=0)
Reserved	[207:205]	Reserved
SA_index[4:0]	[204:200]	Ingress SA Table index
Reserved	[199:184]	Reserved
INNER_TAG	[183:168]	Inner VLAN Tag {PRI[2:0], CFI, VID[11:0]}
UDF	[167:136]	4-byte user-defined field. It's defined by theIngress_UDF register
Reserved	[135:132]	Reserved
SecTAG.C_E	[131:130]	SecTAG TCI C and E bits
Reserved	[129:128]	Reserved
ETHERTYPE	[127:112]	EtherType for Ethernet II/SNAP packets. This field represents {DSAP, SSAP} for LLC packets.
OUTER_TAG	[111:96]	Outer VLAN Tag {PRI[2:0], CFI, VID[11:0]}
SA	[95:48]	Source MAC address
DA	[48:0]	Destination MAC address

Ingress FP Actions

The Ingress FP Action Table supports the following actions on a per–flow basis as shown in Table 7.

Table 7: ***Ingress IFP Action Table***

Sub Field Name	Range	Description
POLICY	[31:29]	3'b000:block/drop packet 3'b001:drop if packet failed ICV check 3'b010:do not drop if packet failed 3'b011:add a 4-byte special VLAN Tag with VLAN ID=0,and {CFI,PRI}={2'b11, E, C} (This action can only apply on unprotected packets) 3'b100:NO-OP 3'b101:Redirect. L2_ENCAP mode will be used to encapsulate the packet with a new L2 header. A per-port configuration register SP_Ingress_L2_Sectag_Override_Cntrl.SECTAG_PN_OVRRIDE_EN will determine whether or not to update the SecTAG.PN with Error Status Code. PN MSB will be replaced by Status Code and LSB will be replaced by 16-bit zeros. 3'b110:Redirect if packet failed ICV checks. L2_ENCAP mode will be used to encapsulate the packet with a new L2 header. A per-port configuration register SP_Ingress_L2_Sectag_Override_Cntrl.SECTAG_PN_OVRRIDE_EN will determine whether or not to update the SecTAG.PN with Error Status Code. PN MSB will be replaced by Status Code and LSB will be replaced by 16-bit zeros.
DROP_IF_NOT_LB	[28]	Drop packet if LB (egress-to-ingress loopback) bit is not set
REDIR_DBUG_FIFO	[27]	Redirect the packet to the ingress debug capture FIFO
COPY_DBUG_FIFO	[26]	If set, copy the packet to the ingress debug capture FIFO. Optionally flush packet at EOP depending on packet errors
MATCH_INDEX	[25:19]	To be carried in the reason code to the ingress debug capture FIFO
BYPASS	[18]	If set, do not modify the packet
Reserved	[17:0]	Reserved

The drop if packet failed and do not drop if packet failed actions can be used to override the DropFailedPkt global configuration on a per flow basis. The add 4 bytes of special VLAN TAG with VLANID = 0 action can be used to indicate to the switch that the packet came in with the special TPID value after the MAC SA.

When packet is tagged as 'Controlled' and dropped by the controlled packet processing logic per standard, the packet may still required by SecY clients of the uncontrolled port. SecY design allows this category of packets to be specially tagged and forwarded to the switch if the DropFailedPackets flag in the Master Control Register is not set. The logic that removes the error packets from the packet stream takes into consideration of the DropFailedPackets flag.

The ingress post processing logic can be configured such that all packets with SecTAG would have their SecTAG preserved (Ingress_SecTAG == 2'b00) when they are forwarded to the switch. When the SecTAG is preserved, the SecTAG.PN number is set to zero to indicate the packet is an error packet when the error is detected in early stage or when the SecY is operating in store-and-forward mode. If Ingress_SecTAG == 2'b01, the SecTAG EtherType (0x88E5) is replaced with a special EtherType specified in the INGRESS_SPECIAL_ETYPE register. The IFP REDIRECT action could also be used to replace the SecTAG EtherType with a special EtherType.

Alternatively, the ingress post processing logic can replace the SecTAG in a packet with a ST-VLAN tag (Ingress_SecTAG == 2'b10). The ST-VLAN tag for ingress is capable of indicating packet categories, error status as well as preserving certain information that is carried in the SCI of the original packet. The format of the ingress ST-VLAN tag is defined in Table 8.

Table 8: *Ingress ST-VLAN Tag Format*

Field Name	***Bits***	***Description***
TAG Protocol ID	31:16	TPID to indicate this is an ST-VLAN tag for SecY.
Status	15:12	4'b0000: Good packet. [E,C] = 00 4'b0001: Good packet [E, C] = 01 4'b0010: Reserved 4'b0011: Good packet [E, C] = 11 4'b0100: Bad packet—invalid SecTAG 4'b0101: Bad packet—no SCI or SA found 4'b0110: Bad packet—replay check failed 4'b0111: Bad packet—ICV check failed 4'b1xxx: Indicates the packet came in with a VLAN tag that matches the TPID specified for the ST-VLAN tag
Flow ID	11:0	12-bit value mapped from the SCI

As a third option, the ingress post processing logic can be additionally configured to always remove the SecTAG from a packet (Ingress_SecTAG == 2'b00) without using the ST–VLAN tag. In this case, it is expected that all error packets will be either dropped or L2–redirected.

Ingress Traffic Categorization

In the final ingress post processing step, an ingress packet is categorized by traffic type and is processed as described in Table 9 on page 35. The Master_Ctrl register, Pre-filter table, ingress FP action table and SA table jointly determine the final destination of a packet. The following policy terms are used in Table 9 on page 35:

- Drop
  - In store-and-forward mode, drop of a packet means the packet is dropped and never forwarded to the Switch.
  - In cut-through mode, drop of a packet means the packet is CRC corrupted (inverted per new requirement) and forwarded to the Switch.
- Redirect from IFP (SA.REDIRECT=1 or IFP.POLICY=REDIRECT)
  - L2_REDIRECT – Encap frame with new L2 Ethernet header from SP_INGRESS_L2_REDIRECT registers when SECTAG_ETYPE_OVRRIDE_EN=0.

  - L2_REDIRECT and CHANGE PN - Encap frame with new L2 Ethernet header from SP_INGRESS_L2_REDIRECT registers when SECTAG_ETYPE_OVRRIDE_EN=0. When SECTAG_PN_OVRRIDE_EN=1, SECTAG.PN will be replaced with the error code.
- STRICT: Master_Ctrl.ValidateFrames == Strict or packet.Cbit==1
- SEC_TAG MODE Operation (from SA and Master_CTRL.Ingress_SECTAG_MODE) following operations:
  - No Replacement of SecTAG
  - Overwrite PN with SPI and change EtherType to XXXX
  - Overwrite PN with SPI leave EtherType.
  - Remove SecTAG
  - Replace SecTAG with a special VLAN tag. if SA.SEC_TAGMODE=Replace with specialVLAN
- IEEE CTRL definition:
  - Untagged_packet STRICT mode: CTRL=0;   non-STRICT mode:CTRL=1
  - SC/SA miss STRICT mode:CTRL=0; non-STRICT mode:CTRL=1
  - Invalid SecTAG CTRL=0
  - KAY frame CTRL=0

Priority of multiple actions from IFP/SA tables are treated in this order: DO_NO_MODIFY, drop, redirect. The SA table actions has priority over IFP. If there is an SA action, the IFP lookup is bypassed and IFP actions are not possible.

A packet classified to be the control packet by the Pre-filter Table will skip the SA Table lookup and IFP Key Table lookup.

A packet with the valid SA and the SA actions “DO_NOT_MODIFY/DROP/REDIRECT/REDIR_DBUG_FIFO/DROP_IF_NOT_LB” will skip the IFP Key Table lookup.

Table 9: *Traffic Category*

Traffic Category	*Action*
CRC packet	Drop
Uncontrolled packet (match found in Pre-filter table with BYPASS action)	Forward frame unmodified (highest priority)
Uncontrolled packet (match found in Pre-filter table with DROP action)	Drop frame
Kay packet (no SA/SC)	• Forward frame unmodified if IFP.BYPASS (ICV is not removed) • Drop if IFP.drop • Redirect if IFP.POLICY=L2_redirect • Drop frame if IFP.Drop_IF_NOT_LB is set for traffic type and not looped back • Replace SecTAG with special VLAN (vid=0, PRFI/C=0b11,E,C) if IFP.POLICY=add_VLAN. • Forward without change.

Table 9: *Traffic Category (Cont.)*

Traffic Category	***Action***
Normal (no error) controlled packet including KaY (known SC/SA)	First, decrypt if needed, and ICV verify and remove. Then perform one of the following actions: • Forward frame unmodified if IFP.BYPASS or SA.DO_NOT_MODIFY (ICV is not removed) • Drop if SA/IFP.drop • Redirect if SA.l2_redirect or IFP.POLICY=L2_redirect • Drop frame if SA/IFP.Drop_IF_NOT_LB is set for traffic type and not looped back • Forward based on SA.SECTAG_MODE • Update counter based on SA.CTRL (IFP.POLICY=add_VLAN should not be set by software)
No SecTAG packet	• Forward frame unmodified if SA.DO_NO_MODIFY or IFP.BYPASS • SA/IFP.drop (based on packet type: untagged, untagged managed) - • Redirect if SA.l2_redirect or IFP.POLICY=L2_redirect • Drop frame if SA/IFP.Drop_IF_NOT_LB is set for traffic type and not looped back • Add a special VLAN (vid=0, PRFI/C=0xF) if IFP.POLICY=add_VLAN. • If Strict and drop failed then drop packet • Update counter based on SA.CTRL if sc/sa hit else use uncontrolled counter. Forward packet without change
Invalid_SecTAG packet	• Forward frame unmodified if SA.DO_NO_MODIFY or IFP.BYPASS • Drop if SA/IFP.drop. • Redirect if SA.l2_redirect or IFP.POLICY=L2_redirect • Drop frame if SA/IFP.Drop_IF_NOT_LB is set for traffic type and not looped back. • Drop if Master_Ctrl.DropFailedPkts && !IFP.POLICY=not_drop_packet_failed. • Forward frame based on Master_control. Ingress_SecTAG • Update counters based on IEEE.CTRL
No SC found packet or Invalid SA packet	• Forward frame unmodified if IFP.DO_NOT_MODIFY • Drop if IFP.drop (can use this for per packet type drops) • Redirect if IFP.POLICY=L2_redirect • Drop frame if IFP.Drop_IF_NOT_LB is set for traffic type and not looped back If STRICT Master_Ctrl.DropFailed && !IFP.POLICY=not_drop_packet_failed • Drop frame - do not apply CTS else Forward based on Master_control. Ingress_SecTAG Update counter based on IEEE.CTRL (IFP.POLICY=add_VLAN should not be set by software)

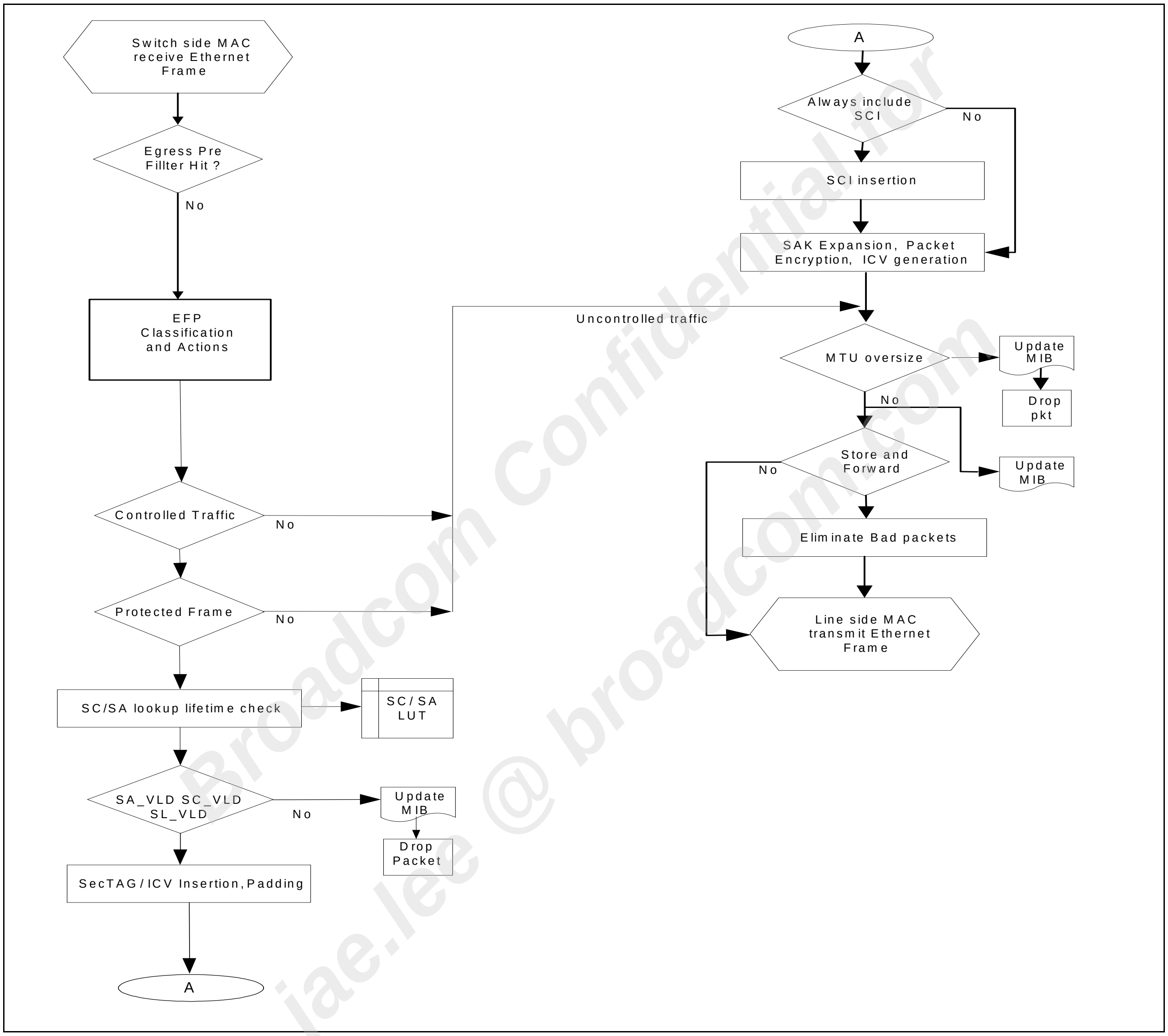
Table 9: ***Traffic Category (Cont.)***

Traffic Category	*Action*
Replay check failed packet	Remove ICV, then one of the below actions: • Forward frame unmodified (except ICV is removed) if SA.DO_NO_MODIFY or IFP.BYPASS • Drop if SA/IFP drop • Redirect if SA/IFP.redirect • Drop frame if SA/IFP.Drop_IF_NOT_LB is set for traffic type and not looped back If SA.ReplayProtect && Drop if Master_Ctrl.DropFailedPkts && !IFP.not_drop_packet_failed • Drop frame else • Forward based on SA.SECTAG_MODE • Update counter based on SA.CTRL (IFP.POLICY=add_VLAN should not be set by software)
ICV check failed packet (SA is found)	• Forward frame unmodified if SA.DO_NO_MODIFY keep ICV • Forward frame unmodified if IFP.DO_NO_MODIFY except for ICV removal based on strict mode • SA/IFP.drop or IFP.drop_ICV_failed (in SAF mode) • Redirect if SA/IFP.redirect or IFP.redirect_icv_failed(only in store-and-forward mode) • Drop frame if SA/IFP.Drop_IF_NOT_LB is set for traffic type and not looped back If STRICT (ICV is always preserved) && Master_Ctrl.DropFailedPkts && ! IFP.not_drop_packet_failed. • Drop frame. else • Forward based on SA. Ingress_SecTAG if in store-and-forward mode (Note: ICV is always removed) • Update counters based on SA.CTRL (IFP.POLICY=add_VLAN should not be set by software)

Egress Packet Flow

Figure 12 provides a functional description of the egress packet and processing flow.

Figure 12: Egress Data Processing Flow



The egress packet flow consists of the following stages, which are described in more detail in the next few subsections:

1. Egress MACsec Packet Classification: ACL based on the first 64 bytes of the packet header.
2. Egress SC/SA Lookup: MACsec SC and SA processing.
3. Egress Packet Processing: Final packet modification based on results from previous stages.

Egress MACsec Packet Classification

The egress Packet Classification Engine is organized similarly to the ingress Packet Classification Engine. The TCAM and the Action Table RAM are shared between the egress Packet Classification Engine and the ingress Packet Classification Engine although the content of the memories are organized differently. The Direction bit in the FP entry defines whether the entry is applied for either ingress or egress packet flow.

The purposes of the egress MACsec packet Classification Engine (Egress FP) are as follows:

- Apply egress security policy control (drop, uncontrolled, controlled)
- Determine the value of the SecTAG (TCI bits, PN and SCI) of the controlled traffic
- Determine the security association of the controlled traffic

All packets are filtered by the Egress FP. The Egress FP Policy Actions direct the selection of the outgoing SC, the formation of the outgoing packet, the construction of SecTAG and other user controlled actions.

The following types of packet formats are expected to be generated by the system side and fed into the MACsec PHY device for egress MACsec operation.

- **Native Packet:** A packet that carries no explicit tag field for MACsec egress operation.
- **Special VLAN-Tagged Packet:** A packet that carries a special 16-bit tag (indicated by a configurable proprietary EtherType) as its outermost tag to indicate explicit information for MACsec egress operation. The Special VLAN Tag also contains C and E bits to control whether the packet should be encrypted or authenticated only.
- **SecTAG-Tagged Packet:** A packet that carries an 8-byte or 16-byte SecTAG (indicated by the MACsec specific EtherType). The SecTAG.TCI field contains C and E bits to control whether the packet should be encrypted or authenticated only.

Table 10 shows the format of the Special VLAN Tag.

Table 10: ***Egress ST-VLAN Tag Format***

Field Name	*Bits*	*Description*
TAG Protocol ID	31:16	TPID to indicate this is a ST-VLAN tag for MACsec PHY
Reserved	15:14	Reserved
C	13	C bit in SecTAG TCI
E	12	E bit in SecTAG TCI
Flow Identifier	11:0	12-bit value to identify an SCI

Egress Filter Processor Parser

The Egress FP parser parses the incoming packet to form the lookup key to the TCAM. The lookup key consists of fixed fields from the packet and user-defined fields (UDF) extracted at configurable offsets from the first 64 bytes of the packet. The fields within the key are as shown in Table 11.

Table 11: ***EFP Key Definition***

Selector Name	*Bit Range*	*Description*
Key_Valid	[468]	Indicates whether this is entry is valid. 1= Valid; 0 = Invalid

Table 11: *EFP Key Definition (Cont.)*

Selector Name	***Bit Range***	***Description***
Key_Mask	[467:234]	Per-bit key comparison enable for the Key fields defined in [233:0] 1 = compare; 0 = ignore
Direction	[233]	0: Ingress 1:egress. This bit should set to 1 for EFP.
Port number	[232:229]	Port Number
SecTAG_Status	[228:227]	'b00: Frame does not contain a SecTAG 'b01: Frame contains a SecTAG 'b10: Reserved 'b11: Reserved
Frame_format	[226:225]	Type of Ethernet frame. 0: Ethernet II packet (LENTYPE>=CP_ETYPE_MAX_LEN) 1: SNAP packet (aa-aa-03-00-00-00) 2: LLC packet (LENTYPE < CP_ETYPE_MAX_LEN and!SNAP) 3: Reserved
Vlan_tag_status	[224:221]	Type of VLAN Tags found on frame. 4'h0: Untagged packet 4'h1: Single VLAN Tag. It is Inner Tag 4'h4: Single VLAN Tag. It is ST-VLAN Tag 4'h5: Single VLAN Tag. It is Outer TPID1 (QTAG) 4'h6: Single VLAN Tag. It is Outer TPID2 (STAG) 4'h7: Single VLAN Tag. It is Outer TPID3 4'h8: Double VLAN Tag. Outer Tag is ST-VLAN Tag 4'h9: Double VLAN Tag. Outer Tag is Outer TPID1 (QTAG) 4'ha: Double VLAN Tag. Outer Tag is Outer TPID2 (STAG) 4'hb: Double VLAN Tag. Outer Tag is Outer TPID3 Others: Reserved
Pkt_type	[220:219]	Type of the packet. 00 non-MACsec packet 01 MACsec packet 1x Reserved
Reserved	[218:217]	Reserved
EFP_KEY_SLICE2	[216:200]	Depends on the setting of Egress_slice2_sel setting, bits [216:200] of the EFP lookup key is defined as: If Egress_slice2_sel = 0, slice 2 is defined as: 216 - reserved 215:200 - PN[15:0]. The lower 16 bits of SecTAG PN field. If Egress_slice2_sel = 1, slice 2 is defined as: 216 - reserved 215:200 - INNER_TAG[15:0]. The inner VLAN Tag {PRI[2:0], CFI, VID[11:0]}.

Table 11: ***EFP Key Definition (Cont.)***

Selector Name	***Bit Range***	***Description***
EFP_KEY_SLICE1	[199:136]	Depends on the setting of Egress_slice1_sel setting, bits [199:136] of the EFP lookup key is defined as: If Egress_slice1_sel = 0, EFP_KEY_SLICE1 is defined as: 199:136 - SCI[63:0]. SecTAG SCI field. If Egress_slice1_sel = 1, slice 1 is defined as: 199:168 - UDF1. The 4 bytes user-defined field. It's defined by register Egress_UDF1. The UDF is parsed from the first 64 byte of the packet starting at the beginning of the packet. 167:136 - UDF0. The 4 bytes user-defined field. It's defined by register Egress_UDF0. The UDF is parsed from the first 64 byte of the packet starting at the beginning of the packet.
SecTAG.TCI_AN	[135:128]	SecTAG TCI and AN fields
ETHERTYPE	[127:112]	EtherType for Ethernet II/SNAP packets. This field represents {DSAP, SSAP} for LLC packets.
OUTER_TAG	[111:96]	Outer VLAN Tag {PRI[2:0], CFI, VID[11:0]}. For the ST-VLAN Tag, the format is {ST_VLAN_reserved[15:14], C, E, Flow_Identifier[11:0]}.
SA	[95:48]	Source MAC address
DA	[48:0]	Destination MAC address

Egress FP Actions

If there is a hit in the EFP, the decision to forward, modify or drop the packet is based on the actions defined in the EFP action table. The FP supports actions shown in Table 12.

There is a 48-bit counter associated with each flow table entry (128 counters total per design). The Classification Engine generates a flow hit ID and a port ID for each packet classified. They are sent to the MIB management module.

Table 12: *Egress FP Actions*

Sub Field Name	*Range*	*Description*
POLICY	[31:27]	[31]: If set, packet is controlled. The only exception is the KaY frame. If the non-MACsec packet is received from the system-side device and the packet is programmed to be a KaY frame (C=0,E=1) either by C and E bits in this table or from the C and E bits in the special VLAN Tag, this bit needs to be set for the egress engine to insert a dumb SecTAG with C=0, E=1, SC=1, PN=0 and SCI=0. The packet is still sent out as the uncontrolled port traffic. [30]: If bit[31] is set, selects between C and E bit from the special VLAN Tag/SecTAG or from the EFP Action Table 1'b1: Use C and E bits from the packet 1'b0: Use C and E bits from EFP Action Table [29]: If bit[31] is set, indicates whether to use the SC and SCI fields from the incoming packet's SecTAG. The SCI_Index always comes from the EFP Action Table 1'b1: Use SC bit and SecTAG.SCI field from the egress SC Table as the outgoing packet's SecTAG.SCI and SecTAG.TCI field 1'b0: Use TCI.SC bit and SecTAG.SCI field from the packet if present in the outgoing packet [28]: If bit [31] is set, indicates whether to use the ES and SCB fields from the incoming packet's SecTAG. 1'b0: Use TCI.ES bit and TCI.SCB bit from packet if present in the outgoing packet 1'b1: Use TCI.ES bit and TCI.SCB bit from the egress SC Table as the outgoing packet's SecTAG.TCI.ES and SecTAG.TCI.SCB field [27]: indicates block/drop the packet 1'b1: drop the packet 1'b0: uncontrolled bypass (NO-OP)
TCI_C	[26]	C bit in SecTAG TCI
TCI_E	[25]	E bit in SecTAG TCI
TCO_V	[24]	V bit in SecTAG TCI. If the packet is a SecTAG packet, the outgoing packets SecTAG.TCI.V will come from the packet. Otherwise, use V bit in EFP Action Table
RDIR_DBUG_FIFIO	[23]	1'b0: Normal mode 1'b1: Redirect the packet to egress debug capture FIFO
Reserved	[22:17]	Reserved

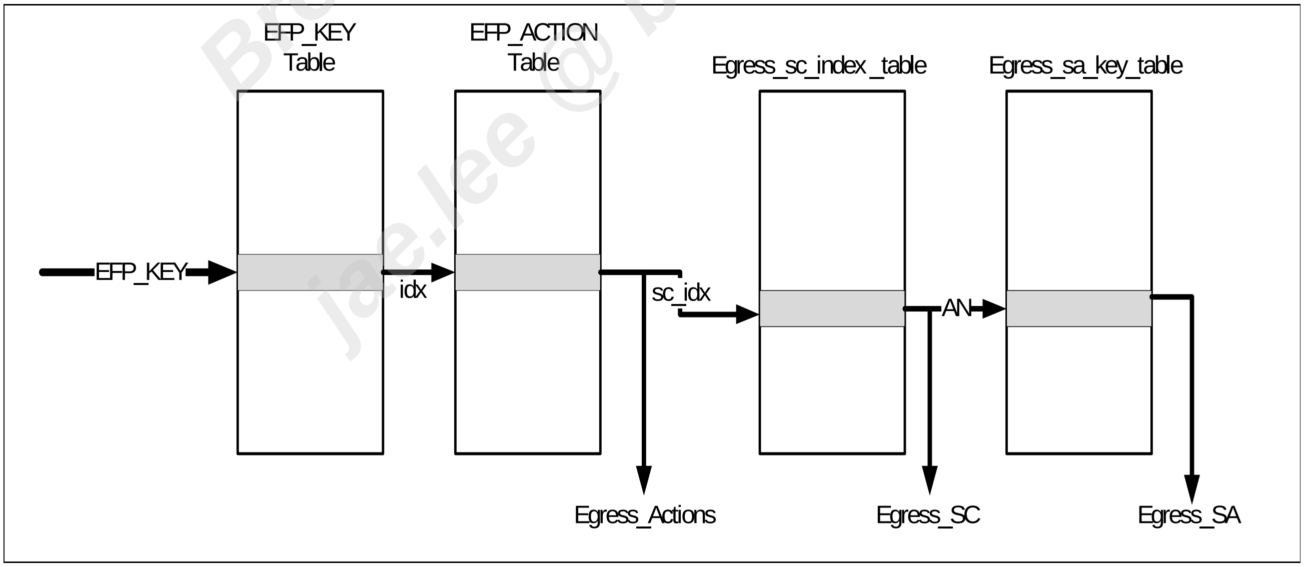
Table 12: *Egress FP Actions (Cont.)*

Sub Field Name	***Range***	***Description***
SCI_Index	[16:10]	SCI_Index – Secure Channel Index. If redir_dbug_fifo=1, this field is a match index to be carried in the reason code. Otherwise lower 4-bits is the index to the Egress SC Table
ACTION	[9:6]	Action type [9]: When set, remove the outer special VLAN tag. This action can be set only when the packets contain the Special VLAN Tag (such as match vlan_tag_status). In addition, this action is only applied to the controlled port packets. [8]: Reserved [7]: When set, override lower 12-bit of SCI with the VID value of the outer VLAN tag [6]: When set, override lower 12-bit of SCI with the VID value of the inner VLAN tag
Reserved	[5:0]	Reserved

Egress SC/SA Lookup

Before SecY processing can be applied to controlled traffic, the SecY module has to determine which secure channel the packet belongs to. If the packet classification resulted in a match providing an SC_Index from the Action Table, and if the POLICY dictates the traffic is controlled traffic, the SC_Index is used to index into the SC Table. When ProtectFrames is set to FALSE, the SC/CA Lookup step is bypassed. Figure 13 shows the egress SC/SA lookup process. The SC_Index is obtained from the EFP Action table. The Egress SC index Table contains AN and is used to determine the index into the Egress SA KeyTable. Note that the Egress SA Key table contains only one field which is the SA Key.

Figure 13: Egress SC/SA Lookup



The SC Index Table is defined in Table 13. Each entry contains the associated SCI and an AN number corresponding to the active SAK.

Table 13: ***SC Index Table***

Field Name	***Range***	***Description***
SCI	71	1'b0: Invalid 1'b1: Valid
	[70:66]	[70:69]: CipherSuiteProtection 2'b00: Integrity 2'b01: Confidentiality 2'b10: OffsetConfidentiality 2'b11: Reserved [68]: UseES [67]: UseSCB [66]: AlwaysIncludeSCI
	[65:64]	Encoding AN for transmit, current AN for receive
	[63:0]	64-bit SCI Number

A per port configuration bit determines if two or four SAKs are stored in the SA memory associated with the corresponding SC. The AN number in the SC determines which SAK is to be applied to the packet. Each SC entry contains two timestamps to be updated. If the SC entry is fresh, both the SC Started Time and Stopped Time are updated. If the SC entry is inUse, only the SC Stopped Time is updated each time the SC is being used. If no valid SC is found, the packet is dropped.

Following the SC look-up step, the SAK is fetched from the EGRESS_SA_KEY_TABLE which contains the SA Key. The SAK is checked to ensure that the packet number has not reached the saturation value (unless Rollover Mode is used). If so, the SAK is forwarded to the AES-GCM engine for data processing. Otherwise, the packet is dropped and an interrupt can be triggered to inform the host. When the SAK is valid, the NextPN number (in the EGRESS_SA_ATTRIBUTE_TABLE) is forwarded to the SecTAG generation logic. The NextPN field is incremented and stored back to that entry. Table 14 shows the EGRESS_SA_ATTRIBUTE_TABLE.

Table 14: ***Egress SA Attribute Table***

Field Name	***Bit Range***	***Description***
Status	[127:126]	2'b00: Invalid 2'b01: Valid and fresh 2'b10: Valid and in use 2'b11: Reserved
–	[125:123]	Reserved
Disable_ReplayControl_ Rollover	[122:121]	2'b00: Normal Mode—Invalidate SA when NextPN expires 2'b01: Rollover Mode—When NextPN reaches all 32'hFFFF_FFFF, rollover to 32'h1. 2'b10: Reserved. 2'b11: Reserved.

Table 14: ***Egress SA Attribute Table (Cont.)***

Field Name	*Bit Range*	*Description*
Loopback_L2Redirect	[120]	1'b0: Forward packet as normal 1'b1: Loop packet back to ingress after SecY processing.
Protection_Offset	[119:112]	The offset of to apply confidentiality protection from the start of the MSDU. Fixed latency mode must be enabled if this offset is larger than 20 when the link speed is 1 Gbps in cut-through mode.
–	[111:97]	Reserved
Next_PN_Saturation	[96]	When 1, this bit indicates the Next_PN has reached all F's and rolled over because Disble_Rollover is set to 0
NextPN	[95:64]	Next packet sequence number
SA Started Time	[63:32]	Transmit: This is the system time when this transmitting SA last started transmitting frames. Receive: This is the system when this receiving SA last started receiving MACsec frames.
SA Stopped Time	[31:0]	Transmit: This is the system time when this transmitting SA last stopped transmitting frames. Receive: This is the system when this receiving SA last started receiving MACsec frames.

Egress Packet Processing

Egress MACsec processing includes SecTAG generation and insertion, short packet handling, packet encryption, and authentication with ICV insertion. A packet loopback path is also included to route the egress packet to the ingress path and back to the system side.

For a packet that does not match in the FP, it is either sent out unmodified or dropped, depending on a per port configurable option. If it matches an FP rule, but the POLICY indicates drop or uncontrolled bypass, it is processed accordingly. For all other packets, packet modifications are done as described in the following sections.

Native Packet Processing

For a Native Packet, if bit[31] from the policy table is set, the packet needs to be encrypted and a SecTAG and ICV inserted into the packet. The FP Policy table provides the SCI_Index which selects the outgoing SC. The packet's C and E bits come from the FP Policy table. The protection mode indicated by the C and E bits are checked against the security policy of the SC, stored in the SC Table to determine if the packet is encrypted or authenticated only.

The SecTAG is formed according to the ACTION field from the FP Policy Table and the fields in the SC Table. The SC Table provides the 64-bit SCI and indication whether to include the SCI in the SecTAG. The ACTION field indicates whether to override the lower 12 bits of the SCI with a 12-bit VID from the Outer or Inner VLAN TAG.

The resulting 8-byte or 16-byte SecTAG will be inserted into the packet at the OFFSET given by the FP Policy Table. The SL is calculated accounting for the OFFSET value.

Special VLAN-Tagged Packet Processing

C and E bits are either taken from the Special VLAN Tag or the FP Policy Table, depending on bit[30] of the policy_data. The mode indicated by the resulting C and E bits are checked against the CipherSuiteProtection capability of the SC, stored in the SC TABLE.

The SecTAG is formed according to the ACTION field from the FP Policy Table and the fields in the SC Table. The SC Table provides the 64-bit SCI and indication whether to include the SCI in the SecTAG. The ACTION field indicates whether to override the lower 12 bits of the SCI with a 12-bit VID from the inner or outer VLAN Tag.

The resulting 8-byte or 16-byte SecTAG will be inserted into the packet at the OFFSET given by the FP Policy Table. The SL is calculated accounting for the OFFSET value.

The Special VLAN Tag can be optionally removed based on the FP ACTION.

SecTAG-Tagged Packet Processing

The packet processing is similar to the Special VLAN Tag case, except the incoming packet's SecTAG may contain additional SL information.

Incoming_sectag.SL indicates where the 16-byte ICV field should be in the outgoing packet. (The ICV field is not carried in the packet.)

- When the SL=0, the ICV calculated by the egress cipher suite is appended at the end of packet data (excluding original CRC).
- When the SL!=0, the ICV calculated by the egress cipher suite is appended at the offset (indicated by SL) from end of SecTAG before being padded by the MAC if necessary. (The original padding bytes and CRC should be removed before appending the ICV.)

Maintaining PN Number for Egress Packets

Each egress packet is assigned a 32-bit PN number in the SecTAG. The PN number is maintained in the SA and incremented prior to applying MACsec to the packet. The PN number represents the lifetime of the SA. Only a valid SA should be applied to protect the traffic. Typically, when the PN number maintained by the SA reaches 0xFFFFFFFF, the SA is marked as expired and the packet is dropped. An early warning threshold can be set via configuration register so that an interrupt can be triggered to allow host to re-key the SA before it expires. Alternatively, the PN number in the SA is allowed to roll over, in which case, the early warning threshold does not take effect. The PN number is incremented to 0x1 after it reaches 0xFFFFFFFF. Please refer to the SA Management Section for more details of SA lifetime handling.

Egress Short Packet Padding

When the MSDU to be protected is smaller than 48 bytes, the ShortLength field of the SecTAG contains the length of the MSDU. Otherwise, the ShortLength field is set to zero. When the packet is transmitted, it is padded by the MAC to satisfy the minimum packet size required. The receiver uses the ShortLength field to correctly identify the location of the ICV.

Egress Packet Encryption and Authentication ICV Generation

After SecTAG Insertion, the packet is forwarded to the AES-GCM engine for security processing. The MACsec logic provides control information to specify the starting location of the security payload to apply authentication and encryption. The AES-GCM engine generates the ICV after the packet is encrypted. The ICV is 16 bytes long. It is appended to the end of the packet.

Egress Post Processing

There is very little post processing to be done for egress traffic. The SCI is removed from the packet stream if transmission of the SCI is not required. In addition, the MIB counters for the controlled port must be updated when the payload length information is collected at the end. Afterwards, the egress packet is queued into the egress packet buffer for transmit.

Egress Controlled Port MTU Check

If the resulting packet after SecTAG and ICV insertion exceeds the value programmed in the Maximum Transmission Unit Register, the logic corrupts the CRC of the packet and records the event in the MIB counter. The MTU check is per packet priority level.

Egress Packet Transmission

The uncontrolled port traffic and controlled port traffic are combined by the MACsec transmission control logic into a single stream of packets and forwarded to the line-side MAC for transmission. The line-side MAC is responsible for re-generating the FCS and inserting it to the end of packet.

Egress Packet Loopback

Egress packets could be loopback from ESEC to ISEC. The design includes a 512-byte FIFO to store the loopback packet. This is a diagnostic feature to test the SecY module's ingress and egress paths. The port can remain active on the link side. In this case, the maximum size of the loopback packets is 512 bytes. The size of a loopback packet is programmable with a configurable option to drop or truncate any packet exceeding the programmed value. The loopback FIFO is accessible via the MDIO interface in addition to the ingress and egress packet flow paths.

CRC Removal and Generation

The CRC is always removed when a packet is received by a MAC and regenerated when the packet is transmitted by the corresponding MAC at the opposite side with the exception of the intentional CRC corruption for error reporting purpose when the received packet is being forwarded back to the switch.

SecY Management

This section discusses the management mechanisms deployed for SecY.

Interrupt Handling

Interrupt capability is provided by the SecY module with a single external pin. The module is designed such that each traffic direction can generate its own interrupt. Internally, an Interrupt Control/Status Register is provided for software to accurately determine the cause of the interrupt and to be able to mask individual interrupt events.

The following interrupt events are defined as follows:

- SA soft lifetime expiration
- Packet classification flow counter soft threshold
- MIB counter soft threshold

SA Management

SA Lifetime

Each security association key (SAK) that is used to protect traffic in a security association has a lifetime due to security concerns with regard to counter mode encryption. The lifetime of the key is determined by a 32-bit packet number (PN). The PN is embedded in the SecTAG and it is incremented monotonically for every packet transmitted with the same SAK until it reaches its saturation value of '0xFFFFFFFF'. Afterwards, the SAK must be refreshed.

The SecY module uses interrupt to inform the host that the SAK has expired. A configurable threshold is used to allow enough time for the host to perform a new round of key refreshment. When the PN number reaches the threshold level, the corresponding SAK is considered 'soft' expired.

When the PN number reaches the saturation value, the corresponding SAK is considered 'hard' expired. A soft-expired SAK will continue to be applied to MACsec packet processing. A configurable interrupt can be generated when the soft-expiration condition is detected. A hard-expired SAK will be invalidated immediately. A separate configurable interrupt is used for reporting hard expiration. For any SAK expiration, the corresponding SC index is latched into the interrupt status register of the corresponding port so that the host can determine which key to refresh promptly.

For 10-Gigabit links, the lifetime of the SA is approximately 4.8 minutes based on the maximum packet rate of 14.88 Mpps.

The SA management of MACsec is also designed to support the notion of non-interrupting service for both transmitting and receiving packets. This is achieved by support more than 1 SAKs for each SC. One key is used for actively protecting the traffic. The other SAKs are used for backup. The software can setup more than 1 backup SAKs at a time and configure the hardware in Auto_AN_Switch mode. On transmit, when the current AN's NEXT_PN reaches 0xFFFFFFFF, the hardware will automatically switch to the backup SAK. This reduces the frequency that the software would have to be interrupted to setup a new SAK.

The IEEE 802.1AE standard does not require non-interrupting service when transmitting the packets. The host is allowed to bring down the MAC_Operational flag for a short period of time in order to switch keys. However, this is hard to manage without the traffic being stopped inside the switch. The delay is also sensitive to the latency of the MDIO interface. For this reason, non-interrupting service is provided for packet transmission as well as packet receiving. The SecY module does allow two receiving SAKs to be used overlapping each other.

The transmit SAKs are identified by the active AN number programmed into the CA memory for each SC. The AN number is two bits wide. The host changes the AN number to switch to a new SAK. This provides minimum switch-over time required by the standard. The receiving SAKs are identified by the AN number embedded in the SecTAG of the packet.

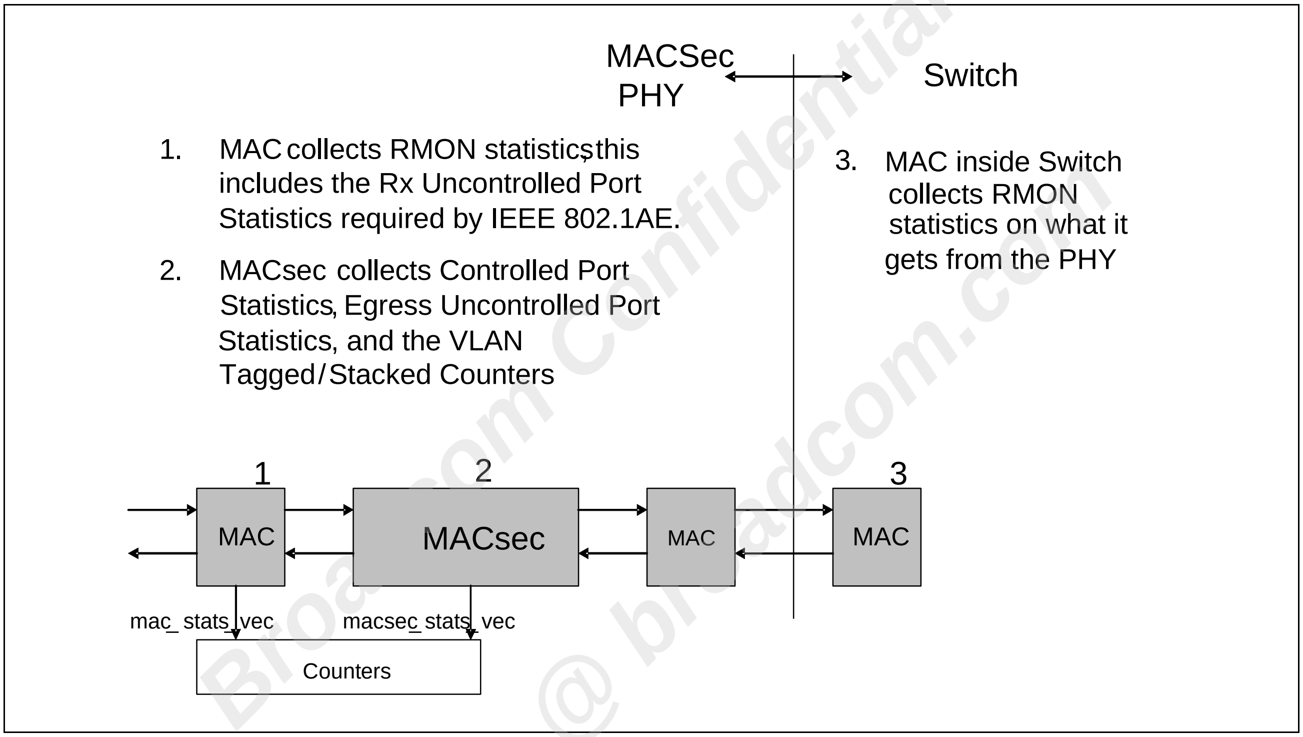
Global Timer

A global timer based on the **Global Time Tick Register** and the **Global Pre-Scale Register** is provided to the SecY module. The initial value of the global time tick and pre-scaling factor can be configured through these two global registers. The global timer is used as a reference time tick to track the start and Stop time of the SC when it is applied to protect the traffic. The default value of the Global Pre-Scale Register sets the time tick to one second interval based on a line clock of either 156 MHz or 125 MHz. The global timer is mainly used to time stamp the SA for management purpose.

MIB Database Management

MIB counters defined by IEEE 802.1AE standard for the common ports are shared with the line-side MACs. The MIB counters for the uncontrolled port for the received direction are either identical to the counters provided by the common port or offer no real significance due to the fact that no error checking is performed by the SecY module for the uncontrolled port, so these MIB counters are not implemented as a separate set of counters. The MIB counters for the uncontrolled port for the transmit direction are counted by the SecY module. The MIB counters for the controlled port are associated with the SecY implementation; they are updated based on the information provided by the SecY module. Figure 14 shows the point of statistics collection in a system with MACsec-enabled PHY. The line counters (RMON) are collected from the line-side MAC.

Figure 14: Statistics Collection Points


MACSec
PHY
Switch
1. MAC collects RMON statistics, this includes the Rx Uncontrolled Port Statistics required by IEEE 802.1AE.
2. MACsec collects Controlled Port Statistics, Egress Uncontrolled Port Statistics, and the VLAN Tagged/Stacked Counters
3. MAC inside Switch collects RMON statistics on what it gets from the PHY
1
2
3
MAC
MACsec
MAC
MAC
mac_ stats_vec
macsec_stats_vec
Counters

The SecY contains a MIB management module. The MIB management module contains a custom memory to store all the MIB counters. The MIB management module receives MIB status update vectors at different stages of the data processing pipeline for both egress and ingress processing.

All standard defined 64-bit counters defined by IEEE 802.1AE are implemented as 64-bit hardware counters to support on-demand access by the software. Refer to the device's programmer's register reference guide for detail MIB register information.

Flow Control

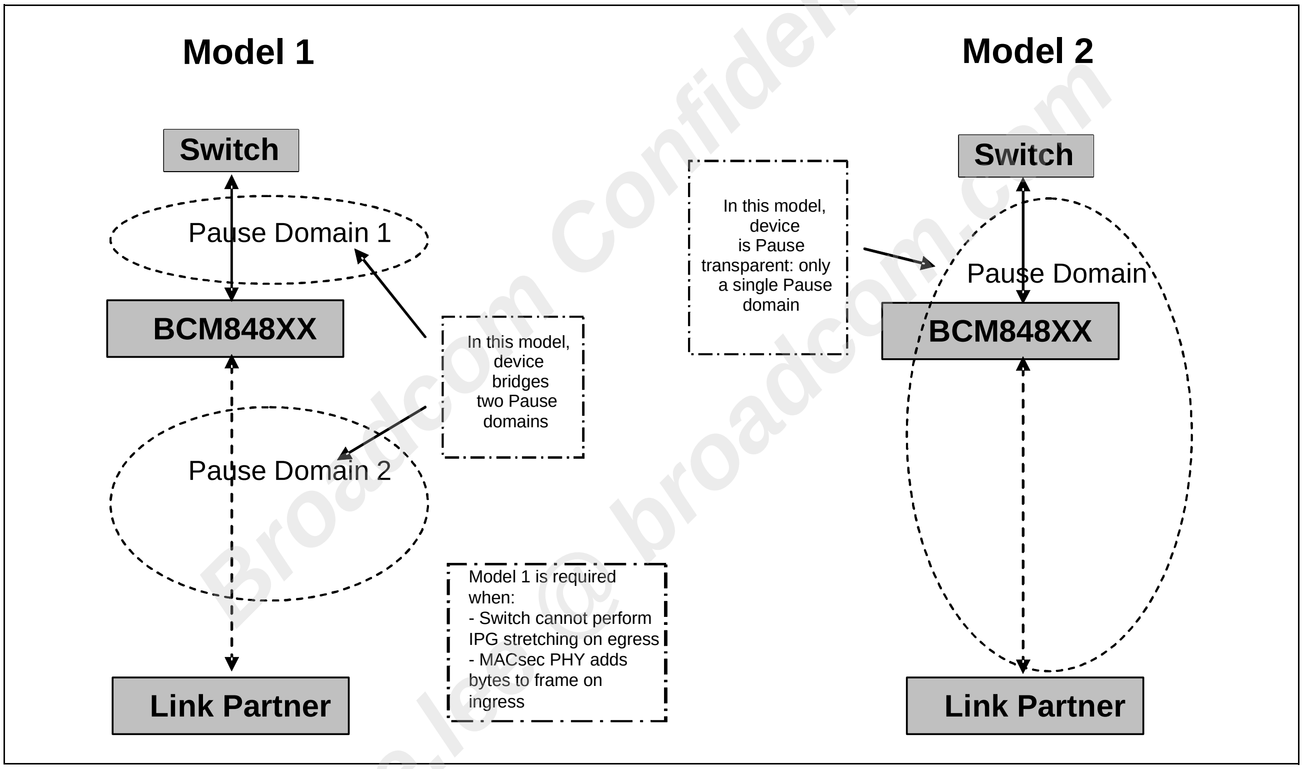
This section covers the flow control schemes supported by the 10G SecY module.

Flow Control Models

The integration of MACsec makes the PHY device less transparent than normal. The latency of traffic flowing through the PHY device increases significantly. It is estimated the latency going through two MACs and the MACsec engine with classification enabled would add approximately 20 to 30 cycles (125 MHz) of latency to the data flow. Further, MACsec inserts SecTAG and ICV for egress traffic. All of these contribute to the need for a robust flow control support in the MACsec PHY.

MACsec supports two flow control models as shown in Figure 15.

Figure 15: Flow Control Model Support



In this model, device bridges two Pause domains

In this model, device is Pause transparent: only a single Pause domain

Model 1 is required when:
- Switch cannot perform IPG stretching on egress
- MACsec PHY adds bytes to frame on ingress

In Model 1, the MACsec PHY decouples the flow control domains to the switch from the flow control domain to the link partner. MACsec PHY terminates all pause frames, and will generate a pause frame when its internal buffers reach a programmable threshold. In Model 2, the MACsec PHY does not generate pause frames and does not respond to any pause frames from the Switch or Link Partner. Instead, it forwards received pause frame to the switch or link partner as normal packets. In this model, the Switch must perform IPG stretching to account for the insertion of SecTAG and ICV.

Packet Buffer Locations

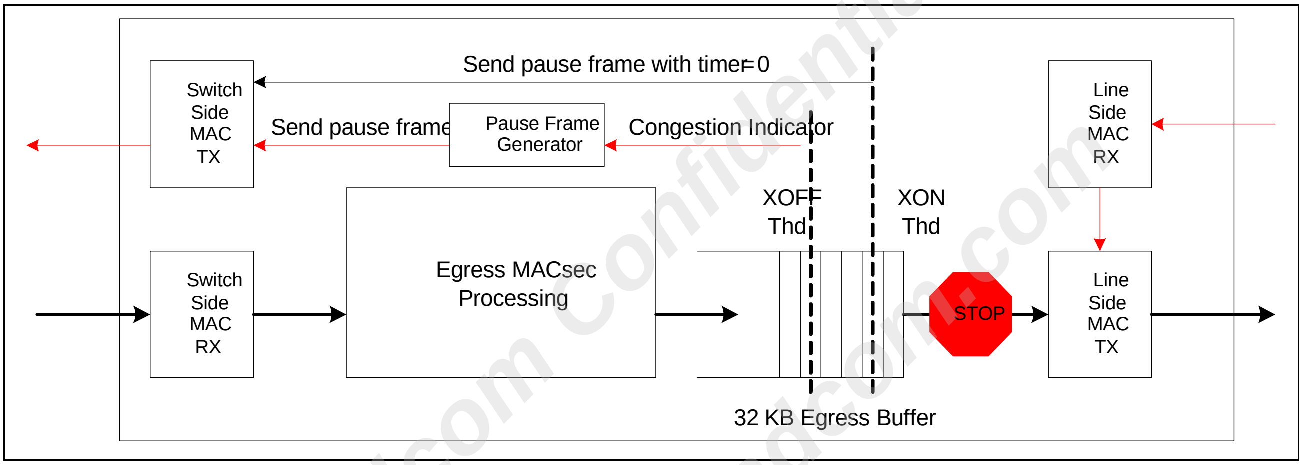
On egress, the packet buffer is placed close to the line-side transmit MACs. On ingress, the packet buffer is placed after the Ingress Classification Engine and close to the system-side MACs.

Model 1 Flow Control

Egress traffic congestion can be caused by one of the following scenarios:

- The link is back-pressured by a remote link partner (PHY receives pause frames or detects half-duplex collision), or
- MACsec inserts SecTAG and ICV to Ethernet frames (24 bytes to 32 bytes per packet)

Figure 16: Mode 1 Egress Flow Control



When the link is back pressured by the link partner, the line-side MAC RX logic detects the pause frame and informs the line-side MAC TX logic. The MAC TX stops transmitting at the next available IPG. The egress packet buffer is going to continue to build up when the switch continues to send packets out for transmit. The egress packet buffer contains a XOFF threshold. When the threshold is reached, the system-side pause frame generator sends a pause frame through the system-side MAC TX logic to the switch to throttle the switch output. This is shown in Figure 16.

Even if the link is not back-pressured by the link partner, due to SecTAG and ICV insertion, the packet buffer will eventually reach the XOFF threshold which triggers a pause frame to be sent to the switch. In either case, the switch stops sending packets to SecY until the packets held off in the packet buffer is drained to below the XON threshold. At that time, the system-side MAC TX logic is signaled to send a pause frame with the timer value set to zero to resume the switch transmission.

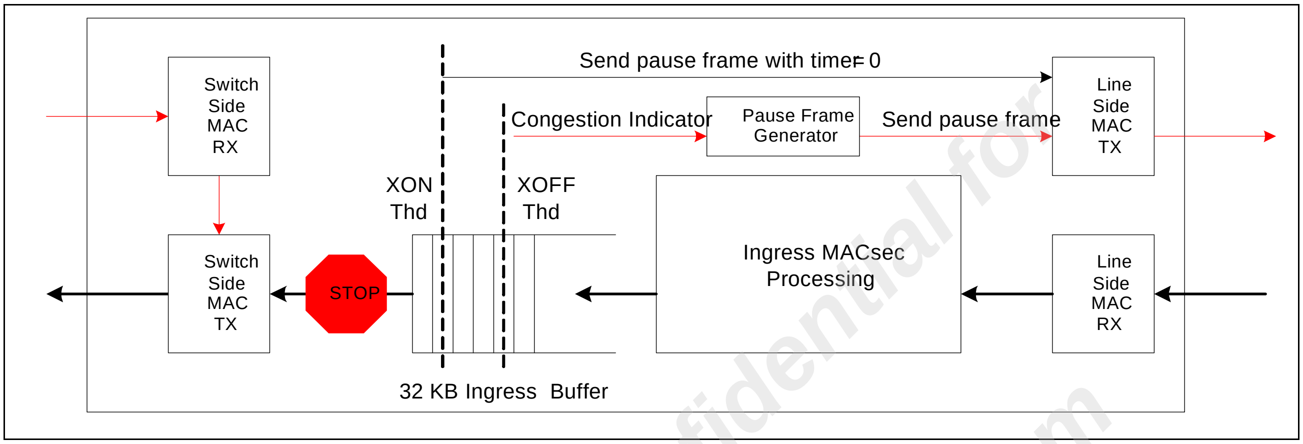
The XOFF threshold must be set up so that there is sufficient room remaining in the buffer to continue to receive packets from the switch until the switch has responded to the pause frame and stopped sending packets. Minimally, the space of two jumbo packets is required.

If the switch is capable of performing IPG stretching, it is always recommended so that the full-duplex flow control doesn't have to be activated. The link operates most effectively when the maximum data rate of encapsulated packets is reached. The link performance degrades when it always has to rely on full-duplex flow control to limit the non-encapsulated traffic.

Similarly, ingress traffic congestion can be caused by one of the following scenarios:

- The switching is back-pressuring the XAUI/XFI interface using full-duplex flow control
- SecY inserts data into Ethernet frames

Figure 17: Mode 1 Ingress Flow Control



Ingress traffic congestion should occur less often than the egress traffic congestion due to the removal of SecTAG and ICV from the Ethernet frame. However, if the SecY inserts additional data into the packet, its receive FIFO may fill up and send out pause frames to its link partner. The ingress flow control typically requires more space to be reserved in the packet buffer with a much lower XOFF threshold level due to larger delay on the media. The ingress flow control is illustrated in Figure 17.

The system-side MACs are capable of performing IPG reduction to minimize the IPG to 8 bytes on average. When this feature is enabled and the reduced IPG is supported by the switch, the ingress congestion caused by SecY inserting data into the packet can be greatly reduced if not completely eliminated.

Model 2 Flow Control

In this mode, the SecY does not generate any pause frame due to internal congestion build-up. To prevent the transmit buffer from overflowing, the switch and the remote link partner must perform IPG stretching. In this mode, the MAC RX logic is configured to transparently pass pause frames to the SecY core logic. The pause frame traverses through the processing pipeline as regular packets until it reaches the packet buffer. The pause frames are then queued differently from regular data frames. They are provided with a high priority dedicated queue. The purpose of this queue is to allow the scheduler to fast forward the pause frame in the next available IPG. This queue contains limited depths. It doesn't generate any flow indication itself. When this queue becomes full, additional pause frames received will be dropped.

When the SecY is operating in this mode and its local switch is sending a pause frame to the link partner, the switch is expected to have sufficient buffering so that it does not stop receiving packets. In this mode, this latency for the link partner to respond to a pause frame is significantly increased. It should be noted that Per-Priority Pause frames are fast-forwarded using the same scheme. This allows Per-Priority Pause frames to be protected by MACsec when needed.

Store-and-Forward Operation

The SecY module can be configured to operate in store-and-forward mode independently in each direction by utilizing the large shared packet buffer. The ingress direction is typically operating in store-and-forward mode for error handling and packet tagging purpose. The egression direction is typically operating in cut-through mode to minimize the latency.

Fixed Latency Mode

The SecY module can be configured to operate in fixed latency mode in both ingress and egress directions. In this mode, packets going through the SecY module have a fixed delay, and it is defined via a programmable register. Fixed latency mode is applicable when the device is configured to support 1588 functions which require packets to have fixed delay through the device.

Appendix A: MACsec Terminology

- **Association Number (AN):** A number that is concatenated with the Secure Channel Identifier to identify a Secure Association.
- **Bounded receive delay:** A guarantee that a frame will not be delivered after a known bounded time, in the case of protocols designed to use the MAC Service this is typically assumed to be less than two seconds.
- **Bridged Local Area Network:** A concatenation of individual IEEE 802 LANs interconnected by MAC Bridges.
- **Cipher Suite:** A set of one or more algorithms, designed to provide any number of the following: data confidentiality, data authenticity, data integrity.
- **Common Port:** An instance of the MAC Internal Sublayer Service used by the SecY to provide transmission and reception of frames for both the controlled and uncontrolled ports.
- **Canonical Format Indicator (CFI):** A 1-bit field. If the value of this field is 1, then the MAC address is in non-canonical format. If the value is 0, then the MAC address is in canonical format. It is always set to zero for Ethernet switches. CFI is used for compatibility between Ethernet and Token Ring networks. If a frame received at an Ethernet port has a CFI set to 1, then that frame should not be bridged to an untagged port.
- **Controlled Port:** The access point used to provide the secure MAC Service to a client of a SecY.
- **Cryptographic key**: A parameter that determines the operation of a cryptographic function such as:
  - The transformation from plain text to cipher text and vice versa.
  - Synchronized generation of keying material.
  - Digital signature computation or validation.
- **Cryptographic mode of operation:** Also referred to as mode. An algorithm for the cryptographic transformation of data that features a symmetric key block cipher algorithm.
- **Data integrity:** A property whereby data has not been altered in an unauthorized manner since it was created, transmitted or stored.
- **Initialization vector (IV):** A vector used in defining the starting point of an encryption process within a cryptographic algorithm.
- **Integrity:** See data integrity.
- **Integrity check value (ICV):** A value that is derived by performing an algorithmic transformation on the data unit for which data integrity services are provided. The ICV is sent with the protected data unit and is recalculated and compared by the receiver to detect data modification.
- **Key:** See cryptographic key.
- **Key management:** The generation, storage, distribution, deletion, archiving, and application of keys in accordance with a security policy.
- **MACsec Management Interface (MMI):** The interface between a protocol entity in a system and the system management, providing for the exchange of parameters with other system entities that are not attached to the service access points used and provided by the protocol entity.
- **MAC Security Entity (SecY):** The entity that operates the MAC Security protocol within a system.
- **MAC Security TAG (SecTAG):** A protocol header, comprising a number of octets and beginning with an EtherType, that is prepended to the service data unit supplied by the client of the protocol, and is used to provide security guarantees.

- **MAC service data unit (MSDU):** A sequence of zero or more octets that compose the data to be communicated with a single MAC Service request or indication.
- **Man-in-the-middle attack:** An attack on the authentication protocol run in which the attacker is positioned between the claimant and verifier so that the attacker can intercept and alter data traveling between the claimant and verifier.
- **Master key:** A secret key that is used to derive one or more cryptographic keys that are used directly to protect data transfer.
- **Message authentication:** If the message arrives authenticated, the cryptographic guarantee is that the message was not modified in transit and that the message originated from an entity with the proper cryptographic credentials.
- **Mode:** See cryptographic mode of operation.
- **Multipoint:** Involving or potentially involving more than one participant in the role of receiver, or in the role of transmitter, in a single data transfer or set of related data transfers.
- **Nonce:** A non-repeating value, such as a counter, used in key management protocols to thwart replay and other types of attack.
- **Packet number (PN):** A monotonically increasing value used to uniquely identify a MACsec frame in the sequence of frames transmitted using an SA.
- **Plaintext key:** An unencrypted cryptographic key.
- **Port Identifier:** A 16-bit number that is unique within the scope of the address of the port.
- **Point-to-Point Protocol (PPP):** The PPP originally emerged as an encapsulation protocol for transporting IP traffic over point-to-point links. PPP also established a standard for the assignment and management of IP addresses, asynchronous (start/stop) and bit-oriented synchronous encapsulation, network protocol multiplexing, link configuration, link quality testing, error detection, and option negotiation for such capabilities as network layer address negotiation and data-compression negotiation. PPP supports these functions by providing an extensible Link Control Protocol (LCP) and a family of Network Control Protocols (NCPs) to negotiate optional configuration parameters and facilities. In addition to IP, PPP supports other protocols, including the Novell® Internetwork Packet Exchange (IPX™) and Cisco® DECnet.
- **Priority Code Point (PCP):** A 3-bit field which refers to the IEEE 802.1p priority. It indicates the frame priority level from 0 (lowest) to 7 (highest), which can be used to prioritize different classes of traffic (such as voice, video, or data).
- **Protocol data unit (PDU):** A unit of data specified in a protocol and consisting of protocol information and, possibly, user data.
- **Secret key:** A cryptographic key used with a secret key cryptographic algorithm that is uniquely associated with one or more entities and should not be made public.
- **Secure Association (SA):** A security relationship that provides security guarantees for frames transmitted from one member of a CA to the other members. Each SA is supported by a single secret key, or a single set of keys where the cryptographic operations used to protect one frame require more than one key.
- **Secure Association Identifier (SAI):** An identifier for an SA, comprising the SCI concatenated with the Association Number (AN).
- **Secure Association Key (SAK):** The secret key used by an SA.
- **Secure Channel (SC):** A security relationship used to provide security guarantees for frames transmitted from one member of a CA to the others. An SC is supported by a sequence of SAs thus allowing the periodic use of fresh keys without terminating the relationship.

- **Secure Channel Identifier (SCI):** A globally unique identifier for a secure channel, comprising a globally unique MAC Address and a Port Identifier, unique within the system allocated that address.
- **Secure Connectivity Association (CA):** A security relationship, established and maintained by key agreement protocols, that comprises a fully connected subset of the service access points in stations attached to a single LAN that are to be supported by MACsec.
- **Short Length (SL):** Is an integer encoded in bits 1 through 6 of octet 4 of the SecTAG and is set to the number of octets in the Secure Data field.
- **Spoofing:** Claiming a fraudulent identity for purposes of mounting an attack.
- **Tag Protocol Identifier (TPID):** A 16-bit field set to a value of 0x8100 to identify the frame as an IEEE 802.1Q-tagged frame.
- **Uncontrolled Port:** The access point used to provide the insecure MAC Service to a client of a SecY.
- **VLAN Identifier (VID):** A 12-bit field specifying the VLAN to which the frame belongs. A value of 0 means that the frame doesn't belong to any VLAN; in this case the 802.1Q tag specifies only a priority and is referred to as a priority tag. A value of hex FFF is reserved for implementation use. All other values may be used as VLAN identifiers, allowing up to 4094 VLANs. On bridges, VLAN 1 is often reserved for management.
- **Wiretapping:** An attack that intercepts and accesses data and other information contained in a flow in a communication system. The term is used to refer to reading information from any sort of medium used for a link or even directly from a node, such as a gateway or subnetwork switch. Active wiretapping attempts to alter the data or otherwise affect the flow; passive wiretapping only attempts to observe the flow and gain knowledge of information it contains.

Abbreviations and Acronyms

- **AES:** Advanced Encryption Standard
- **AN:** Association Number
- **CA:** secure Connectivity Association
- **CFI:** Canonical Format Indicator
- **CRC:** Cyclic Redundancy Check
- **CTR:** Counter mode
- **DA:** Destination Address
- **EAPOL:** EAP over LANs (Extensible Authentication Protocol Over LANs)
- **EPON:** Ethernet Passive Optical Network
- **ES:** End Station
- **FCS:** Frame Check Sequence
- **FIPS:** Federal Information Processing Standard
- **GCM:** Galois Counter Mode
- **Gbps:** Gigabit per second (1 Gbps is equivalent to 1 000 000 000 bits per second)
- **ICV:** integrity check value
- **ISS:** Internal Sublayer Service
- **IV:** Initialization Vector

- **KaY:** MAC Security Key Agreement Entity
- **LACP:** Link Aggregation Control Protocol
- **LAN:** IEEE 802 Local Area Network
- **LLC:** Logical Link Control (IEEE Std 802.2)
- **LLDP:** Link Layer Discovery Protocol
- **LMI:** layer management interface
- **MMI:** MACsec Management Interface
- **MAC:** Media Access Control
- **MACsec:** MAC Security Protocol
- **Mbps:** Megabit per second (1 Mbps is equivalent to 1 000 000 bits per second)
- **MIB:** Management Information Base
- **MPDU:** MACsec Protocol Data Unit
- **MSDU:** MAC Service Data Unit
- **MSTP:** Multiple Spanning Tree Protocol
- **NESSIE:** New European Schemes for Signatures, Integrity, and Encryption
- **NIST:** National Institute of Standards and Technology
- **OLT:** Optical Line Terminator
- **ONU:** Optical Network Unit
- **PAE:** Port Access Entity
- **PDU:** Protocol Data Unit
- **PN:** Packet Number
- **PCP:** Priority Code Point
- **QoS:** quality of service
- **RADIUS:** Remote Authentication Dial-In User Service
- **RSTP:** Rapid Spanning Tree Algorithm and Protocol
- **SA:** Secure Association
- **SAI:** Secure Association Identifier
- **SAK:** Secure Association Key
- **SC:** Secure Channel
- **SCB:** Secure Channel Broadcast
- **SCI:** Secure Channel Identifier
- **SecTAG:** MAC Security TAG (8 or 16 byte MACsec header)
- **SecY:** MAC Security Entity (The entity that operates MAC Security protocol)
- **SL:** Short Length
- **SNMP:** Simple Network Management Protocol
- **TPID:** Tag Protocol Identifier
- **VID:** VLAN Identifier

Broadcom® Corporation reserves the right to make changes without further notice to any products or data herein to improve reliability, function, or design.

Information furnished by Broadcom Corporation is believed to be accurate and reliable. However, Broadcom Corporation does not assume any liability arising out of the application or use of this information, nor the application or use of any product or circuit described herein, neither does it convey any license under its patent rights nor the rights of others.

Connecting everything®

Broadcom Corporation
5300 California Avenue
Irvine, CA 92617

Phone: 949-926-5000
Fax: 949-926-5203
E-mail: info@broadcom.com
Web: www.broadcom.com