AN-138

QoS Priority Support In the KSZ8842 Family

Introduction

Latency critical applications such as Voice over IP (VoIP) and video typically need to guarantee a high quality of service (QoS) throughout the network. QoS can be supported by various priority schemes offered by the switches and routers.

This application note describes the different priority schemes supported by the KSZ8842 family of switches and how they are configured.

There are three different priority schemes supported by the devices: Port-based priority, 802.1p Tag-based Priority and DiffServ-based priority. The mechanisms for each of these priority schemes differ only with respect to the ingress port. The egress port buffer scheme is common to all three priority schemes.

Common Settings for Port Based, 802.1p Tag Based and DiffServ Based Priorities

The KSZ8842 family of devices offers four priority transmits queues per port. Queue 3 is the highest priority queue as priority 3 and Queue 0 is the lowest priority queue as priority 0. The bits 0 of the registers P1CR1, P2CR1 and P3CR1 are used to enable splitting transmit queues for egress port 1, port 2 and host port respectively.

Priority Scheme

Priority transmit queuing allows a switch to define two priority schemes. One is “always transmit higher priority packets first” mode. The other is the weighted fair queuing (WFQ) mode. When the transmit queue is set to WFQ mode, the transmit queue will follow a scale for the four queues and the bandwidth allocation is Q3:Q2:Q1:Q0=8:4:2:1. If any queue is empty, the highest non-empty queue will get one more weighting. For example, if Q2 is empty, Q3:Q2:Q1:Q0 will become (8+1):0:2:1.

This mechanism assures that during congestion, the higher-priority data does not get delayed by lower-priority traffic. Some examples of priority queuing are:

- Important Voice and video packets are assigned a highest-priority queue level 3.
- General Voice and video packets are assigned a higher-priority queue level 2.
- Web traffic is assigned a lower-priority queue level 1.
- Back-up data traffic is assigned the lowest-priority queue level 0.

The Weighted Fair Queuing (WFQ) mode tries to ensure that the lower-priority packets will not be starved during congestion. WFQ is implemented in the KSZ8842 two port switches; with a host bus as the host port, by controlling the priority scheme select bit 11 in the SGCR2 register. (It is also recommended to enable the Priority Buffer Reserve bit at the same time). These related registers as shown in table 1.

All datasheets and support documentation can be found on Micrel’s web site at: www.micrel.com.

Register	Bit	Name	Description	Default
Global Ctrl Reg 2 SGCR2	0	Priority Buffer Reserve	1 = Each port is pre-allocated 48 buffers exclusively for high priority packets (Q2,Q2,Q1). Effective only when the multiple queues feature is turned on. 0 = Each port is pre-allocated 48 buffers. Used for all priority packets (Q3,Q2,Q1,Q0).	1
Global Ctrl Reg 2 SGCR2	11	Priority Scheme Select	0 = Always TX higher priority packets first. 1 = Weighted Fair Queuing (WFQ) is enabled, Q3,Q2,Q1,Q0 = 8,4,2,1.	0
Port n Ctrl Reg 1 P1CR1,P2CR1,P3CR1	0	TX Multiple Queues Select Enable	1 = The port’s output queue is split into four priority queues. 0 = Single output queue on the port.	0

Table 1. Registers are used for Common and Egress Port Setting

Notes:

1. If the *TX Multiple Queues Select Enable* bit is not enabled in P1CR1, P2CR1 and P3CR1, then only a single output queue will be present at the egress port. Hence, the priority scheme selection will have no effect, irrespective of the other settings for the ingress and egress ports.
2. The settings highlighted in “Note 1” above will be used for all Port based priorities, 802.1p based priorities and DiffServ based priorities.

Port Based Priority

Port based priority is the simplest form of QoS. Each ingress port can be individually classified as one of the priorities 0-3. All packets arriving at the ingress port will be passed to any of the four priority queues at the egress port, depending upon the configuration of the ingress port.

Each ingress port can be configured as one of the priorities 0-3 by using the Port Based Priority Classification Enable bit shown in Table 2.

For example, if register P2CR1 bit 4-3 is set to 10, all of packets from ingress port 2 will be treated as priority 2 level packets and go to priority 2 transmit queue on the egress port which has set into four priority queues.

Register	Bit	Name	Description	Default
Port n Ctrl Reg 1 P1CR1,P2CR1,P3CR1	4-3	Port based priority classification	00 = ingress packets on port n will be classified as priority 0 queue. 01 = ingress packets on port n will be classified as priority 1 queue. 10 = ingress packets on port n will be classified as priority 2 queue. 11 = ingress packets on port n will be classified as priority 3 queue.	00

Table 2. Registers are used for Port Based Priority

Note: "Diffserv", "802.1p" and port priority can be enabled at the same time. The OR'ed result of 802.1p and DSCP overwrites the port based priority.

802.1p Tag Based Priority

802.1p priority can be enabled by the 802.1p Priority Classification Enable bit in the ingress port Control Registers P1CR1, P2CR1 and P3CR1.

Ethernet packets can have an optional 4-byte 802.1q VLAN tag inserted between the source address (SA) and the length/type fields. As shown in Figure 1, there is a 3-bit priority field embedded in the 4-byte tag, the 3-bit priority field is used for 802.1p priority classification.

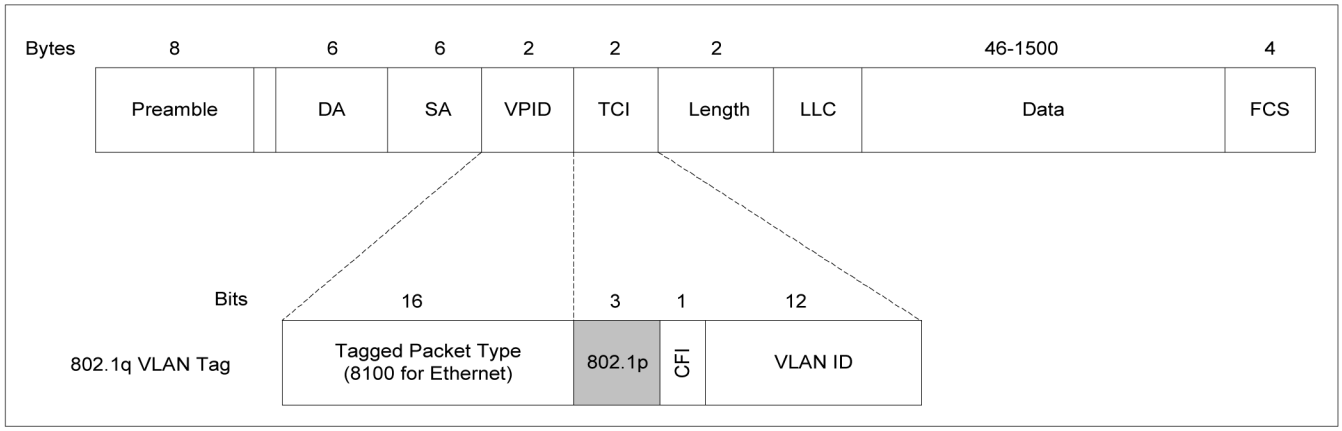


Figure 1. Ethernet Packet with 802.1q VLAN Tag

The 3-bit priority field in the VLAN tag is used to set the priority level (0-3) for each packet. The number value from 0 to 7 is configured in the switch and compared to the binary equivalent of the incoming packet's priority field in the VLAN tag. The 3-bit priority field value (0-7) can be decoded as priority level (0-3) by the SGCR 6 register which can be programmed by the user (See Table 3 for details). The priority field value of the incoming tagged packets will be re-classified to four priority levels based on the SGCR6 setting. The priority level classification is set using the SGCR6 register. The related registers are as shown in Table 3 for 802.1p based priority.

Register	Bit	Name	Description	Default
Port n Ctrl Reg 1 P1CR1,P2CR1,P3 CR1	4-3	Priority Classification Enable	1 = Enable 802.1p priority classification for ingress packets on port. 0 = Disable 802.1p priority.	0
Global Ctrl Reg 6 SGCR6	15-14	Tag_0x7	IEEE 802.1p mapping. The value is used as the frame's priority when its IEEE Tag has a value of 0x7.	0X3 for priority 3
	13-12	Tag_0x6	IEEE 802.1p mapping. The value is used as the frame's priority when its IEEE Tag has a value of 0x6.	0X3 for priority 3
	11-10	Tag_0x5	IEEE 802.1p mapping. The value is used as the frame's priority when its IEEE Tag has a value of 0x5.	0X3 for priority 2
	9-8	Tag_0x4	IEEE 802.1p mapping. The value is used as the frame's priority when its IEEE Tag has a value of 0x4.	0X3 for priority 2
	7-6	Tag_0x3	IEEE 802.1p mapping. The value is used as the frame's priority when its IEEE Tag has a value of 0x3.	0X3 for priority 1
	5-4	Tag_0x2	IEEE 802.1p mapping. The value is used as the frame's priority when its IEEE Tag has a value of 0x2.	0X3 for priority 1
	3-2	Tag_0x1	IEEE 802.1p mapping. The value is used as the frame's priority when its IEEE Tag has a value of 0x1.	0X3 for priority 0
	1-0	Tag_0x0	IEEE 802.1p mapping. The value is used as the frame's priority when its IEEE Tag has a value of 0x0.	0X3 for priority 0
Port n Ctrl Reg 2 P1CR2, P2CR2, P3CR2	3	User Priority Field (Ceiling)	1 = If the packet's "priority field" is greater than the "user priority bits" in port n's VID Control register bits [15:13], replace the packet's " priority field" with the "user priority bits" in port n's VID Control register bits [15:13]. 0 = Do not compare and replace the packet's "user priority field."	0
Port n VID Ctrl Reg P1VIDCR, P2VIDCR, P3VIDCR	15-13	User Priority bits	Port n tag [15-13] for priority field of ingress tagged packet to be compared or replaced.	000

Table 3. Registers are used for Tag Based Priority

Note: The OR'ed result of 802.1p and DSCP (see below) priority classification overrides any port priority.

Priority Re-Mapping

The KSZ8842 family of devices has the ability to re-map the ingress packets 802.1p priority field by setting bit 3 of User Priority Field in registers P1CR2, P2CR2 and P3CR2 and bits [15-13] of User Priority Bits in registers P1VIDCR, P2VIDCR and P3VIDCR. An example of the importance of priority re-mapping is shown as follows.

In the case that port 1 is connected to a PC and port 2 is connected to a VoIP router, a problem may occur if you have a 'corrupt' PC transmitting data packets containing high priority 802.1p tags. This causes the VoIP and PC (data) packets to both be tagged as high priority and hence, there is no differentiation between them. Acceptable QoS for the voice traffic can no longer be guaranteed.

Priority re-mapping is available on the KSZ8842 family of ports. Each ingress port can be set as specified in the User Priority Bits in the registers P1VIDCR, P2VIDCR or P3VIDCR. If the incoming packet's 802.1p priority field is greater than the user defined value in the User Priority Bits in the register P1VIDCR, P2VIDCR or P3VIDCR, then the packet's priority field is replaced with the user defined value in the User Priority Bits. Priority re-mapping is enabled using the *User Priority Field (Ceiling)* bit in P1CR2, P2CR2 and P3CR2, as shown in Table 3.

DSCP (DiffServ-based) Priority

The KSZ8842 devices support DiffServ-based priority in IPv4 and IPv6 IP packets. In this note, IPv4 packets are used as an example.

The differentiated service code point (DSCP) priority operates in the Layer 3, IP protocol. The IP datagram header is embedded within the Ethernet data field (see Figure 2).

The DSCP priority bits are located inside the type of service (TOS) field, within the standard IPv4 header.

The IPv4 header is shown below in more detail. The TOS byte is the second byte located after the header length field (HLEN).

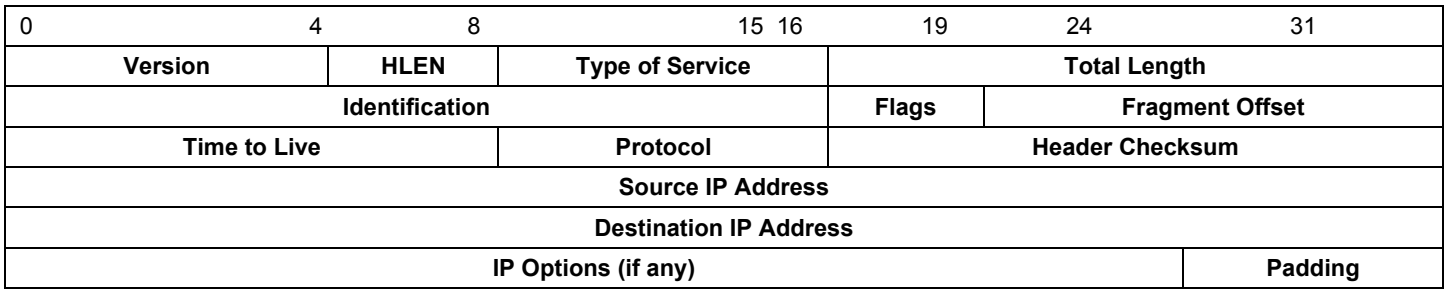
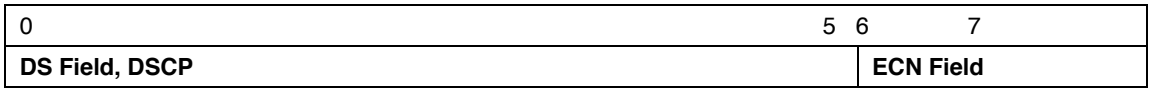


Figure 2. Format of IPv4 Datagram Header

Bits 0 to 5 of the ToS field are then taken and fully decoded in 64 separate QoS service codes as shown in Figure 3.



DSCP: Differentiated Services Code Point

ECN: Explicit Congestion Notification (Unused)

Figure 3. Differential Services (DS) Code Point within the ToS Field of an IP Datagram

The Differentiated Service Code is then compared against the corresponding bit in the *Priority Control Registers 1 to 8 (TOSR1-TOSR8)* of the KSZ8842 devices (16 bits x 8 registers = 128 bits totally). The corresponding 2-bits in TOSR1-TOSR8 registers stands for one code point of DSCP. 2-Bits have 4 priority levels, where 00 is priority 0, 01 is priority 1, 10 is priority 2 and 11 is priority 3. Using the 128-bits of TOSRn registers, it is possible to make 64 DSCP that have 4 priority levels for each DSCP.

DSCP priority is enabled using the *DiffServ Priority Classification Enable* bit in each port Control Register P1CR1, P2CR1 and P3CR1. They are shown in Table 4.

Register	Bit	Name	Description	Default
Port n Ctrl 0 P1CR1, P2CR1, P3CR1	6	Diffserv Priority Classification Enable	1 = Enable diffserv priority classification for ingress packets on port. 0 = Disable diffserv priority.	0

Table 4. Registers are used for DiffServ Priority

Note: The OR'ed result of 802.1p and DSCP priority classification overrides any port priority.

Register	Bit	DSCPs of the Corresponding Registers									Description	Default
TOSR1, TOSR2, TOSR3, TOST4, TOSR5, TOSR6, TOSR7, TOSR8	15-14	DSCP n=	7	15	23	31	39	47	55	63	00 = priority 0 01 = priority 1 10 = priority 2 11 = priority 3	0
		TOSR n=	1	2	3	4	5	6	7	8		
	13-12	DSCP n=	6	14	22	30	38	46	54	62	00 = priority 0 01 = priority 1 10 = priority 2 11 = priority 3	0
		TOSR n=	1	2	3	4	5	6	7	8		
	11-10	DSCP n=	5	13	21	29	37	45	53	61	00 = priority 0 01 = priority 1 10 = priority 2 11 = priority 3	0
		TOSR n=	1	2	3	4	5	6	7	8		
	9-8	DSCP n=	4	12	20	28	36	44	52	60	00 = priority 0 01 = priority 1 10 = priority 2 11 = priority 3	0
		TOSR n=	1	2	3	4	5	6	7	8		
	7-6	DSCP n=	3	11	19	27	35	43	51	59	00 = priority 0 01 = priority 1 10 = priority 2 11 = priority 3	0
		TOSR n=	1	2	3	4	5	6	7	8		
	5-4	DSCP n=	2	10	18	26	34	42	50	58	00 = priority 0 01 = priority 1 10 = priority 2 11 = priority 3	0
		TOSR n=	1	2	3	4	5	6	7	8		
	3-2	DSCP n=	1	9	17	25	33	41	49	57	00 = priority 0 01 = priority 1 10 = priority 2 11 = priority 3	0
		TOSR n=	1	2	3	4	5	6	7	8		
	1-0	DSCP n=	0	8	16	24	32	40	48	56	00 = priority 0 01 = priority 1 10 = priority 2 11 = priority 3	0
		TOSR n=	1	2	3	4	5	6	7	8		

Table 5. Registers are used for 64 DSCP Priority Level Settings

All Priority Control Registers (TOSR1-TOSR8) are as shown in Table 5 for detail priority levels.

For example,

If DSCP=001000 (Bin) = 8 (Dec), this implies that TOS Priority Control Register TOSR2, bits 1-0 will be examined, and the priority of those bits will set as the priority level of the packet.

Conclusion

The KSZ8842 family of switches is ideal for handling Quality of Service requirements. With emerging applications such VoIP and Video Broadcasting, the network must be ready to handle different type of services in a cost effective manner. The network should be designed with an end-to-end QoS capability for the future. As shown in this paper, Micrel's Ethernet product family of switches provides a rich set of QoS functionality to meet the needs for emerging triple play applications.

MICREL, INC. 2180 FORTUNE DRIVE SAN JOSE, CA 95131 USA
TEL +1 (408) 944-0800 FAX +1 (408) 474-1000 WEB http://www.micrel.com

The information furnished by Micrel in this data sheet is believed to be accurate and reliable. However, no responsibility is assumed by Micrel for its use. Micrel reserves the right to change circuitry and specifications at any time without notification to the customer.

Micrel Products are not designed or authorized for use as components in life support appliances, devices or systems where malfunction of a product can reasonably be expected to result in personal injury. Life support devices or systems are devices or systems that (a) are intended for surgical implant into the body or (b) support or sustain life, and whose failure to perform can be reasonably expected to result in a significant injury to the user. A Purchaser's use or sale of Micrel Products for use in life support appliances, devices or systems is a Purchaser's own risk and Purchaser agrees to fully indemnify Micrel for any damages resulting from such use or sale.

© 2006 Micrel, Incorporated.