RTL839x/RTL835x
LAYER 2 PLUS MANAGED 52*10/100/1000M-PORT SWITCH CONTROLLER

SWRED
Application Note

Rev. 1.0
17 December 2012

Realtek Semiconductor Corp.
No. 2, Innovation Road II, Hsinchu Science Park, Hsinchu 300, Taiwan
Tel.: +886-3-578-0211 Fax: +886-3-577-6047
www.realtek.com

USING THIS DOCUMENT
This document is intended for use by the system engineer when integrating with Realtek switch products. Though every effort has been made to assure that this document is current and accurate, more information may have become available subsequent to the production of this guide. In that event, please contact your Realtek representative for additional information that may help in the development process.

Revision	Release Date	Summary
1.0	2012/12/17	First Release

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Simplified Weighted Random Early Detection

針對所有的 ingress packet 系統均會給予一個 Drop Precedence (DP) value。DP 可以從 DSCP or DEI remap 而來，也可透過 Meter (srTCM/trTCM)來改變 DP，而不同的 DP 可以設定不同的 drop 比率，以實現系統在真正發生擁塞之前就可以不同的比率提早 drop packet，避免系統一旦真正發生擁塞時連續 drop packet，進而減低對 TCP flow 的影響。DP 0,1,2分別映射到Green、Yellow和Red三種color。

1.1 Drop Precedence Remarked By trTCM/srTCM

在 trTCM 的應用上，traffic rate 低於 LB0_RATE 的時候，系統是處於 Green 安全狀態。Traffic 超過 LB0_RATE 時，系統是處於 Yellow 警戒狀態。一旦 traffic 超過 LB1_RATE 時，系統則是處於 Red 擁塞狀態。例如：設定 LB0_RATE=50Mbps, LB1_RATE=100Mbps，當送進 40Mbps traffic 時，因爲小於 LB0_RATE，DP 會維持不變(packet default DP is 0)。當 traffic 增加到 70Mbps 時，因爲 LB0_RATE <70Mbps<LB1_RATE，DP 值會被 remarked 爲 1(Yellow)。如果送進 traffic 大於 100Mbps，由於已經超過 LB1_RATE，DP 值會被 remarked 爲 2(Red)。

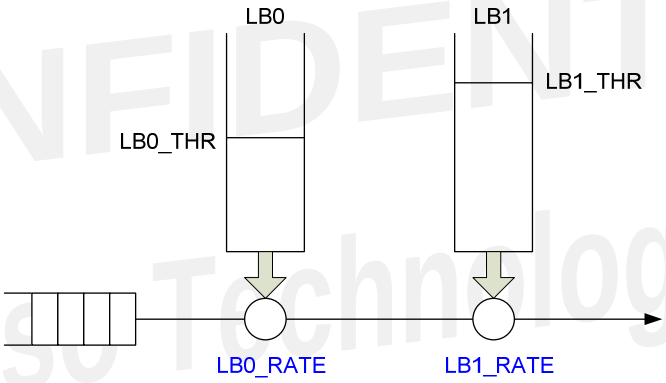

Figure 1: Two Rate Three Color Maker (trTCM)

如果是在 srTCM 的應用上，系統能明顯區別開來的是 Green 和 Red 兩種狀態，原因是 srTCM 在標示系統 Green、Yellow 和 Red 三種狀態，是以 LB0_THR 和 LB1_THR 兩個 threshold 做區分(如**錯誤! 找不到參照來源。**)，traffic 累積超過 LB0_THR 到小於 LB1_THR 的時間非常的短暫，不容易區分出 Yellow 的狀態。若以應用上來看，當 rate 設定爲 50Mbps，送進的 traffic<50Mbps 時，系統維持 DP 不變，如果 traffic>50Mbps，DP 值才會被設定成 2(Red)。如果 traffic rate 是 80Mbps，因爲超過 50Mbps 的部份僅有 30Mbps，所以系統在 srTCM 修正後的 DP 0 和 DP 2 的 traffic 比例爲 5:3(50Mbps: 30Mbps)。而如果 traffic rate 爲 100Mbps，DP 0 和 DP 2 的比例則會是 1:1。

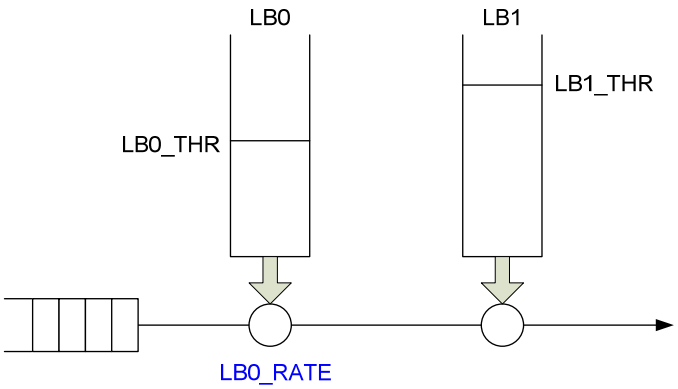

Figure 2: Single Rate Tree Color Maker (srTCM)

1.2 SWRED Cooperate With trTCM

系統提供兩種 egress packet drop 機制：Tail Drop(TD)、Simplified Weighted Random Early Detection(SWRED)，TD 就是一般 packet 超過 drop threshold 時，就 drop packets。而 SWRED 則是利用 DP 值取得不同的 drop threshold 以及 drop probability，進一步計算後，再決是否 drop 或 forward。

Table 1: Port-based Configuration of SWRED

Field Name	Bits	Description
DROP_RATE_DP	8	Drop rate (D_{rate}) of all ports (excluding CPU port) for drop precedence *d* (*d*=0~2). Maximum is 255.
THMAX_DP	12	Maximum drop threshold of all ports (excluding CPU port) for drop precedence *d* (*d*=0~2). Unit: page
THMIN_DP	12	Minimum drop threshold of all ports (excluding CPU port) for drop precedence *d* (*d*=0~2). Unit: page

Table 2: Queue-based Configuration of SWRED

Field Name	Bits	Description
Q_DROP_RATE_DP	8	Drop rate (D_{rate}) of egress queue n (n=0~7) for drop precedence *d* (*d*=0~2). Maximum is 255.
Q_THMAX_DP	12	Maximum drop threshold of egress queue n (n=0~7) for drop precedence *d* (*d*=0~2). Unit: page
Q_THMIN_DP	12	Minimum drop threshold of egress queue n (n=0~7) for drop precedence *d* (*d*=0~2). Unit: page

系統在SWRED module上面針對egress port和egress queue分別提供port-based(Table 1)和queue-based(Table 2)兩種設定。假設port-based和queue-based的設定如下面表格設定，port-based的DP 0和DP 2的drop rate都爲0，只有DP 1的drop rate=255。另外，因爲SWRED會同時計算port-based和queue-based兩種drop probability，爲了讓情況簡化，將queue-based的drop rate都設成0， minimum threshold和max threshold都設成最大值，也就是讓queue-based失效。
下表的設定會讓DP=1的packet，無論被送到哪個port，都有24.93%(255/1023)的機率會被drop。會讓DP=2的packet，無論被送到哪個port，都一定會被drop。

若以前面trTCM的例子來看，當50Mbps(LB0_RATE) < traffic rate< 100Mbps(LB1_RATE)發生時，packet的DP會被修改爲1，因此traffic會有24.93%的機率被drop(註1)。當traffic rate > 100Mbps(LB1_RATE)發生時，packet的DP會被修改爲2，因此traffic is always dropped(註1)。

Port-based Setting	DP=0	DP=1	DP=2
DROP_RATE_DP	0	255	0
THMAX_DP	4095	4095	0
THMIN_DP	4095	0	0

Queue-based Setting	DP=0	DP=1	DP=2
Q_DROP_RATE_DP	0	0	0
Q_THMAX_DP	4095	4095	4095
Q_THMIN_DP	4095	4095	4095

註1：在此簡化的測試情形下，系統並不會發生congestion，但SWRED drop packet的另一條件爲系統需處於

congestion狀態，此時需將Flow Control的ignore RX port congest的設定打開，讓SWRED module忽略system congestion的條件。

DIAGSHELL COMMAND REFERENCE

```
/* Set an ACL meter entry for tcTCM which cir(LB0_RATE) = 50Mbps and pir(LB1_RATE) = 100Mbps */
acl set entry phase 0 entry 0 state valid
acl set entry phase 0 entry 0 action meter state enable
acl set entry phase 0 entry 0 action meter 0
acl set meter mode block 0 byte
acl set meter burst-size byte trtcm lb0-burst-size 300 lb1-burst-size 1000
acl set meter entry 0 trtcm color-unaware cir 3125 pir 6250

/* Enable ignoring RX port congest state */
flowctrl set egress port all igr-congest-check state enable

/* Set SWRED to be egress drop algorithm */
flowctrl set congest-avoidance algorithm swred

/* SWRED system port-based configuration */
flowctrl set congest-avoidance system-threshold drop-precedence 0 drop-probability 0
flowctrl set congest-avoidance system-threshold drop-precedence 0 max-threshold 4095 min-threshold
4095
flowctrl set congest-avoidance system-threshold drop-precedence 1 drop-probability 255
flowctrl set congest-avoidance system-threshold drop-precedence 1 max-threshold 4095 min-threshold 0
flowctrl set congest-avoidance system-threshold drop-precedence 2 drop-probability 0
flowctrl set congest-avoidance system-threshold drop-precedence 2 max-threshold 0 min-threshold 0

/* SWRED system queue-based configuration */
flowctrl set congest-avoidance queue-threshold queue-id all drop-precedence 0 drop-probability 0
flowctrl set congest-avoidance queue-threshold queue-id all drop-precedence 0 max-threshold 4095
min-threshold 4095
flowctrl set congest-avoidance queue-threshold queue-id all drop-precedence 1 drop-probability 0
flowctrl set congest-avoidance queue-threshold queue-id all drop-precedence 1 max-threshold 4095
min-threshold 4095
flowctrl set congest-avoidance queue-threshold queue-id all drop-precedence 2 drop-probability 0
flowctrl set congest-avoidance queue-threshold queue-id all drop-precedence 2 max-threshold 4095
min-threshold 4095
```

1.3 SWRED Cooperate With srTCM

假設port-based和queue-based的設定修改爲下面表格的設定，使port-based的drop失效，drop改以queue-based做爲依據。Queue-based的設定上，設定queue *N,M如下表所示*，其他剩餘的queues則將drop rate設成0、drop thresholds設爲最大，使其失效。另外搭配srTCM機制，設定srTCM的LB0_RATE=50Mbps，當送進100Mbps traffic至Queue *N* 時，有50% traffic會被remark成DP 2，而這50%的traffic則會有24.93%的機率會被drop。也就是100Mbps，大約只有75Mbps的traffic會被送出。而如果送進100Mbps traffic至Queue *M*，則有50Mbps DP 2的traffic會被drop，也就是只有50Mbps的traffic會被送出。

Port-based Setting	**DP=0**	**DP=1**	**DP=2**
DROP_RATE_DP	0	0	0
THMAX_DP	4095	4095	4095
THMIN_DP	4095	4095	4095

Queue-based Setting for Queue *N*	DP=0	DP=1	DP=2
Q_DROP_RATE_DP	0	0	255
Q_THMAX_DP	4095	4095	4095
Q_THMIN_DP	4095	4095	0

Queue-based Setting for Queue *M*	DP=0	DP=1	DP=2
Q_DROP_RATE_DP	0	0	0
Q_THMAX_DP	4095	4095	0
Q_THMIN_DP	4095	4095	0

Queue-based Setting for Other Queues	DP=0	DP=1	DP=2
Q_DROP_RATE_DP	0	0	0
Q_THMAX_DP	4095	4095	4095
Q_THMIN_DP	4095	4095	4095

DIAGSHELL COMMAND REFERENCE

```
/* Set an ACL meter entry for scTCM which cir(LB0_RATE) = 50Mbps */
acl set entry phase 0 entry 1 state valid
acl set entry phase 0 entry 1 action meter state enable
acl set entry phase 0 entry 1 action meter 0
acl set meter mode block 0 byte
acl set meter burst-size byte srtcm lb0-burst-size 300 lb1-burst-size 1000
acl set meter entry 0 srtcm color-unaware cir 3125

/* Enable ignoring RX port congest state */
flowctrl set egress port all igr-congest-check state enable

/* Set SWRED to be egress drop algorithm */
flowctrl set congest-avoidance algorithm swred

/* SWRED system port-based configuration */
flowctrl set congest-avoidance system-threshold drop-precedence 0 drop-probability 0
flowctrl set congest-avoidance system-threshold drop-precedence 0 max-threshold 4095 min-threshold
4095
flowctrl set congest-avoidance system-threshold drop-precedence 1 drop-probability 0
flowctrl set congest-avoidance system-threshold drop-precedence 1 max-threshold 4095 min-threshold
4095
flowctrl set congest-avoidance system-threshold drop-precedence 2 drop-probability 0
flowctrl set congest-avoidance system-threshold drop-precedence 2 max-threshold 4095 min-threshold
4095

/* SWRED system queue-based configuration for TX queue N(0) */
flowctrl set congest-avoidance queue-threshold queue-id 0 drop-precedence 0 drop-probability 0
flowctrl set congest-avoidance queue-threshold queue-id 0 drop-precedence 0 max-threshold 4095
min-threshold 4095
flowctrl set congest-avoidance queue-threshold queue-id 0 drop-precedence 1 drop-probability 0
flowctrl set congest-avoidance queue-threshold queue-id 0 drop-precedence 1 max-threshold 4095
min-threshold 4095
flowctrl set congest-avoidance queue-threshold queue-id 0 drop-precedence 2 drop-probability 255
flowctrl set congest-avoidance queue-threshold queue-id 0 drop-precedence 2 max-threshold 4095
min-threshold 0

/* SWRED system queue-based configuration for TX queue M(1) */
```

```
flowctrl set congest-avoidance queue-threshold queue-id 1 drop-precedence 0 drop-probability 0
flowctrl set congest-avoidance queue-threshold queue-id 1 drop-precedence 0 max-threshold 4095
min-threshold 4095
flowctrl set congest-avoidance queue-threshold queue-id 1 drop-precedence 1 drop-probability 0
flowctrl set congest-avoidance queue-threshold queue-id 1 drop-precedence 1 max-threshold 4095
min-threshold 4095
flowctrl set congest-avoidance queue-threshold queue-id 1 drop-precedence 2 drop-probability 0
flowctrl set congest-avoidance queue-threshold queue-id 1 drop-precedence 2 max-threshold 0
min-threshold 0

/* SWRED system queue-based configuration for others */
flowctrl set congest-avoidance queue-threshold queue-id 2-7 drop-precedence 0 drop-probability 0
flowctrl set congest-avoidance queue-threshold queue-id 2-7 drop-precedence 0 max-threshold 4095
min-threshold 4095
flowctrl set congest-avoidance queue-threshold queue-id 2-7 drop-precedence 1 drop-probability 0
flowctrl set congest-avoidance queue-threshold queue-id 2-7 drop-precedence 1 max-threshold 4095
min-threshold 4095
flowctrl set congest-avoidance queue-threshold queue-id 2-7 drop-precedence 2 drop-probability 0
flowctrl set congest-avoidance queue-threshold queue-id 2-7 drop-precedence 2 max-threshold 4095
min-threshold 4095
```