Quality of Service (QoS) in VoIP networks

Mangle, Queue Tree

Aris Vink
Presales consultant, MikroTik trainer at SDC B.V.


25 JAAR
SDC
SDC
Connecting your business
SDC

Who is SDC B.V. ?

- Distributor in ICT components
- Founded in 1992, located in Doetinchem(NL)
- Family company
- Supportive with solutions and service
- ICT partnerchannel in the Netherlands

Marketing&sales

Logistics

Academy

Professional services

Quality of Service in VoIP networks

- Quality of Service(QoS) introduction
- Circuit switched vs Packet switched networks
- QoS criteria for Voice over IP
- Problems in VoIP environments
- QoS traffic control
- Hierarchical Token Bucket (HTB)
- Qualify VoIP traffic with Mangle
- HTB implementation Queue Tree

Quality of Service(QoS)

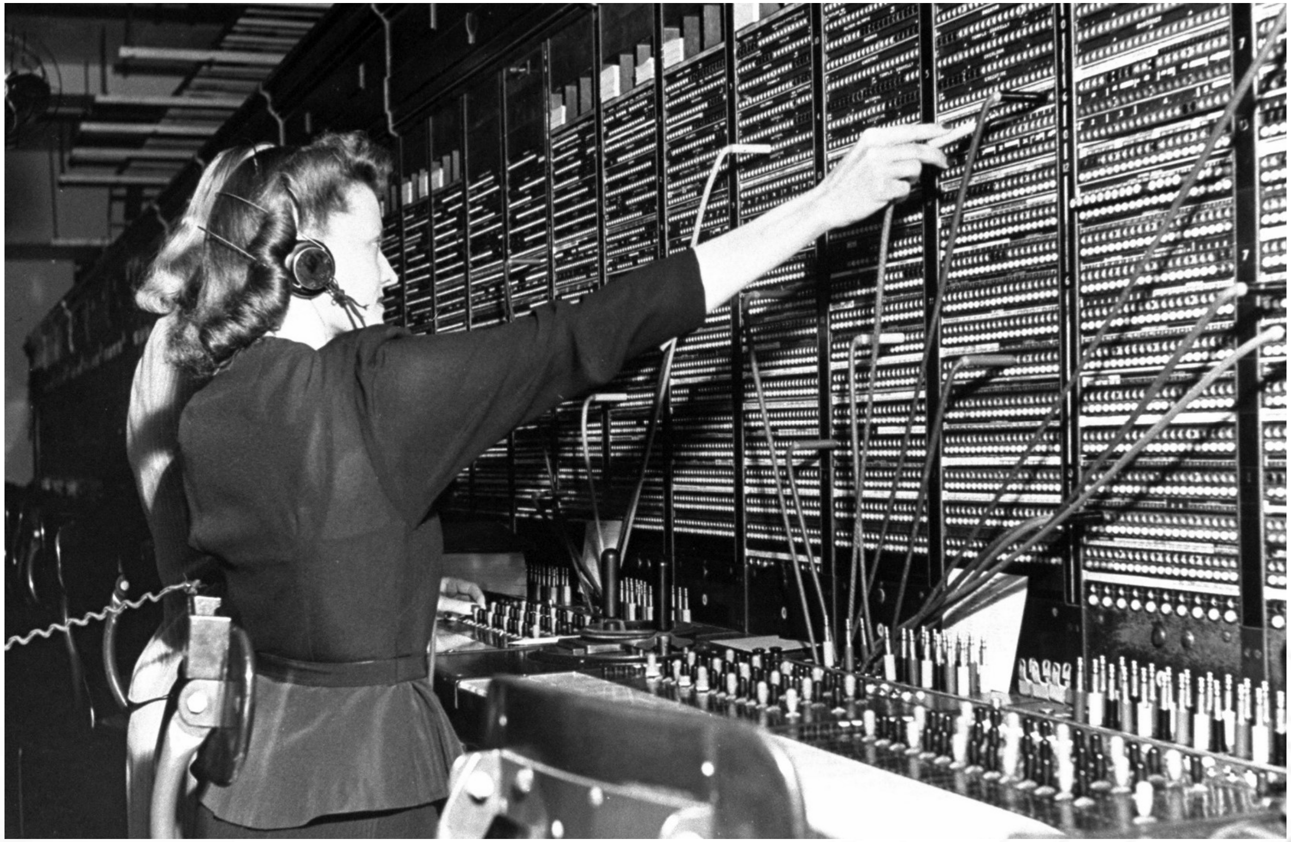
Quality of Service(QoS) is a general term that describes the measurement of the overall performance of a service such as telephony, video, file transfer or other network services.

In the field of **computer networking**, quality of service refers to **traffic prioritization** and **resource reservation** control mechanisms. QoS is needed in order to **guarantee** a certain level of **performance** to a data flow

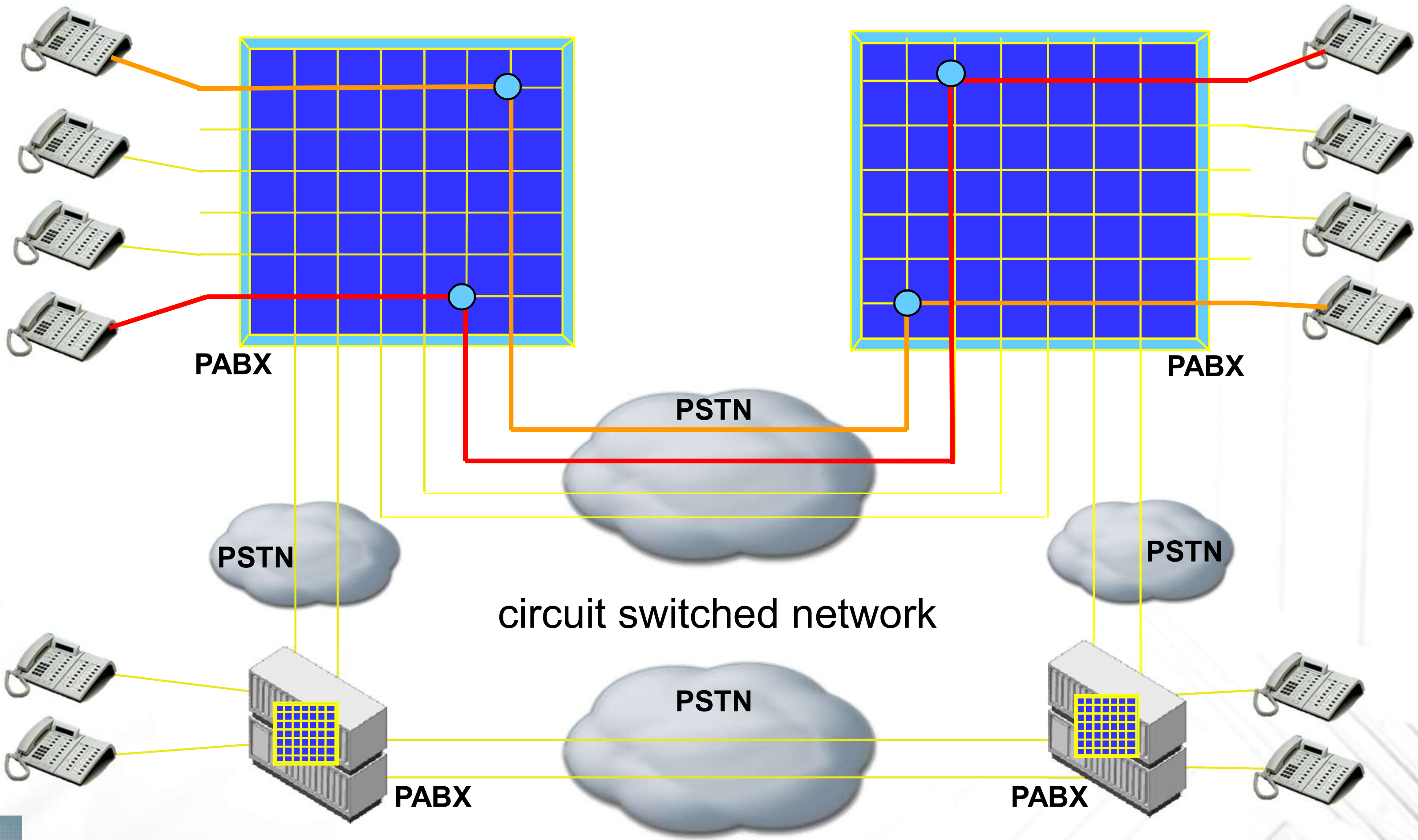
Why the need for Quality of Service(QoS)

For instance Quality of Service is necessary to guarantee the bandwidth of an IP-phone. A phone call does not need much bandwidth but must always be instantly available for phone calls!

Telephony history



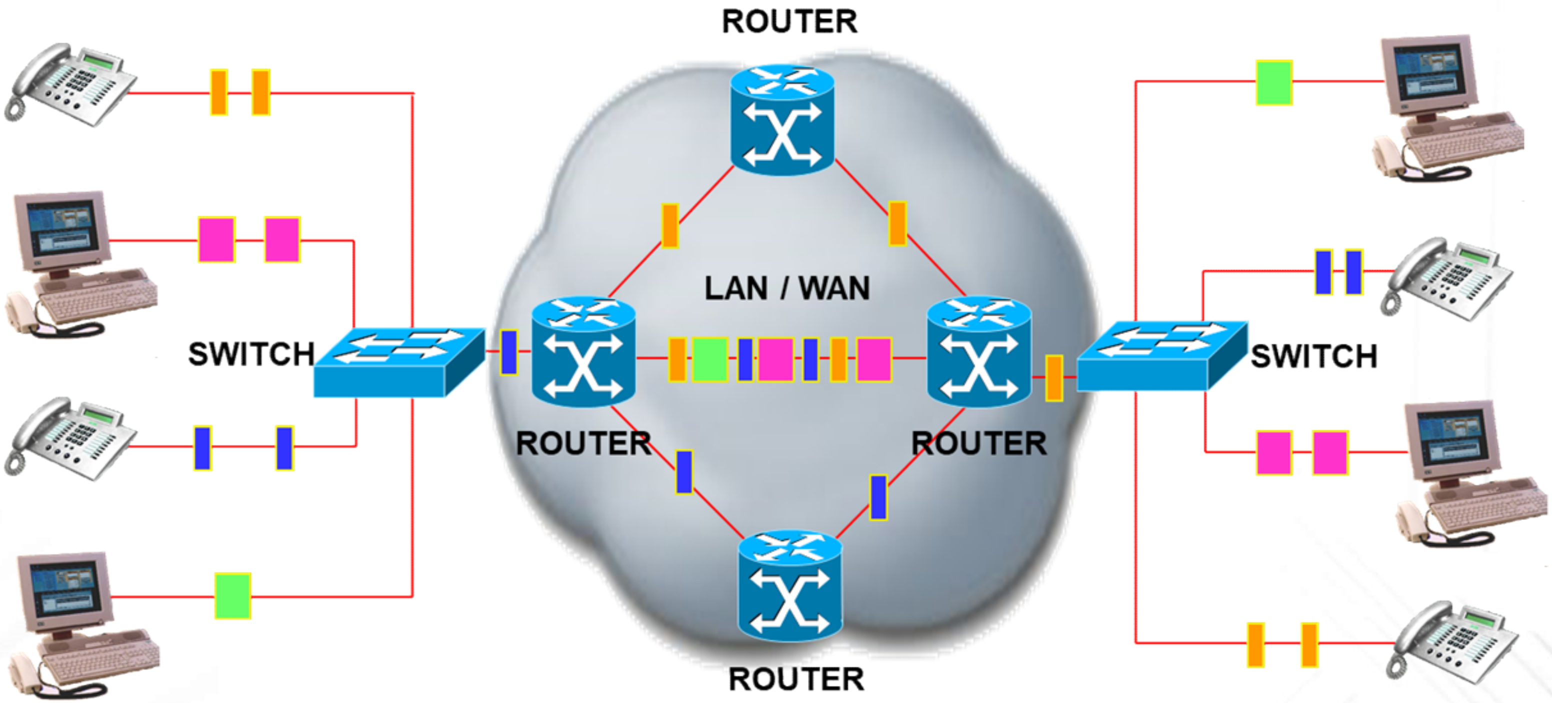
Circuit switched network (64kbit PCM telephony)



Circuit switched network

- A **direct(fysical) connection** is made between phones. For setting up the circuit(call setup) a **seperate signalling network** is nessesary.
- A circuit switched network has a fixed(short) latency and for every session a fixed bandwidth. Circuit switched networks by standard have **QoS**.
- Unused bandwith within a circuit is unavailable for other services. Circuit switched networks are **unefficient** in this regard.

Packet switched network



Packet switched network

- IP phones exchange voice data over a **shared network**
- Since there is **no real connection**(circuit) the voice packets are **coded(IP)** so they can be routed through the network to the destination
- To **guarantee QoS** for telephony traffic identification, shaping en prioritizing has to be performed end to end.
- Services like speech, video and data share the same fysical network enabling **more efficient** use of network **bandwith**.

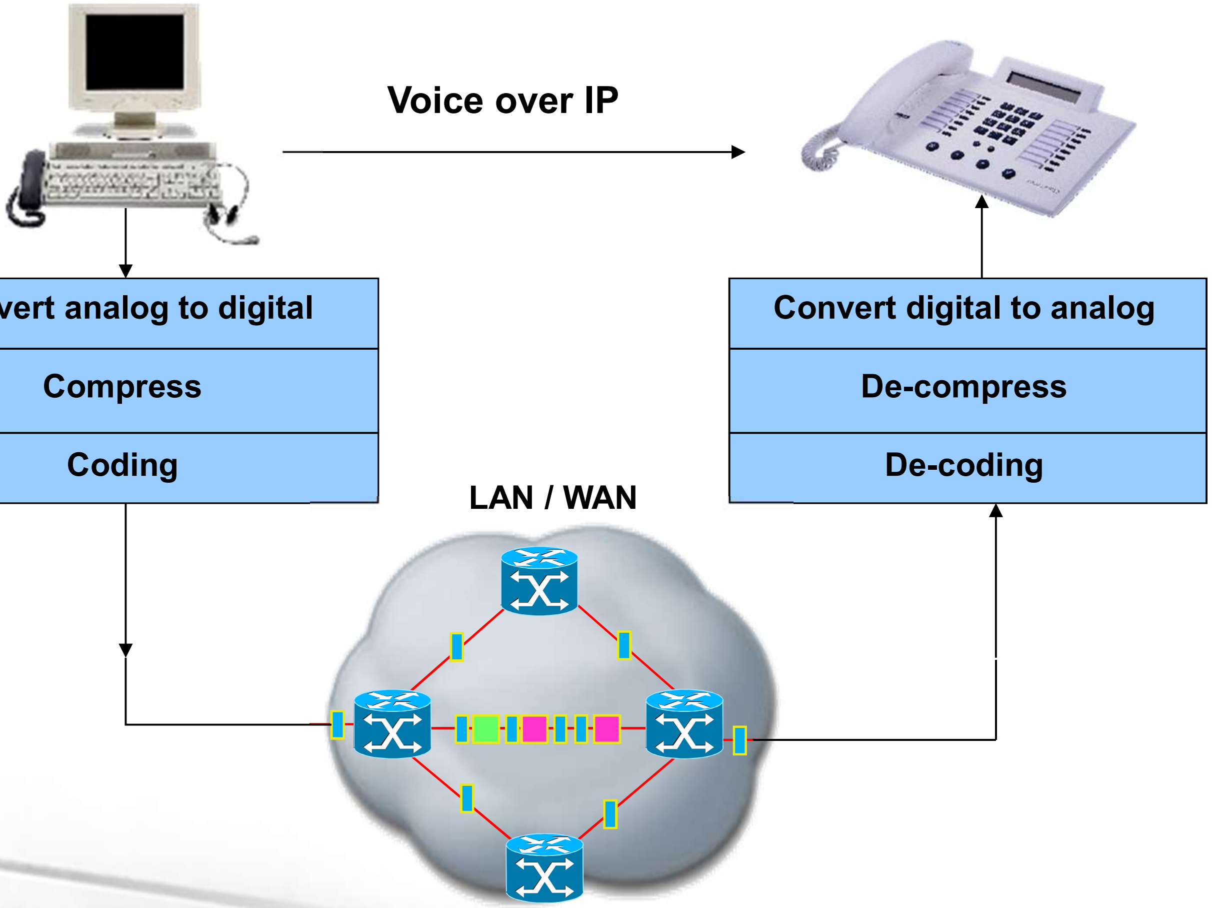
QoS criteria for Voice over IP

- Sufficient bandwidth
- Availability of bandwidth(at all times)
- Low packet loss
- Low round trip delay(low latency)
- Minimal Jitter

Problems in VoIP environments

- The quality of a VoIP call is heavily dependent on the network environment.
- **Latency, Jitter, & Packet Loss** can cause quality of experience issues for **Voice over IP (VoIP)** phone calls
- **Packet loss** is very common in IP networks, but certain networks such as WiFi can be particularly prone to high levels of packet loss.

Bandwith usage



Bandwith usage

code	Bandwidth for speech	+	Bandwidth for "IP"-Overhead	=	Bandwidth usages in SIP for: one cannel	Bandwidth usages in SIP for: bi-directional connection
G.711	64 kBit/s		ca. 19 kBit/s		ca. 83 kBit/s	ca. 166 kBit/s
G.723.1	6,3 kBit/s		ca. 19 kBit/s		ca. 25 kBit/s	ca. 50 kBit/s

RTP(Realtime Transport Protocol)

SIP uses RTP for the end to end transport of audio and video data(**payload**). RTP adds extra information to the UDP header and in some cases can replace the UDP header.

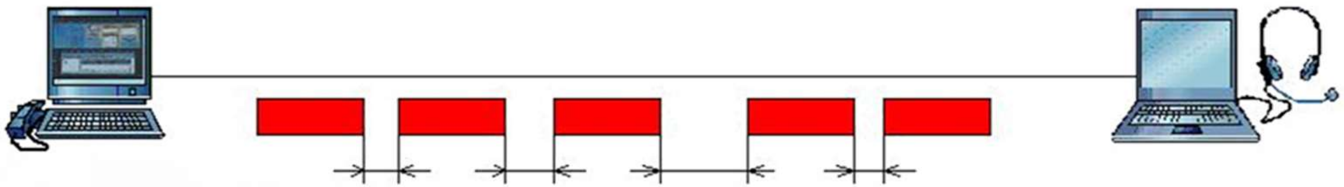
- Sequence number
- Timestamp
- Payload type

RTCP(Realtime Transport Control Protocol)

RTP functionality is expanded by Real Time Control Protocol (RTCP). The goal of RTCP is to provide feedback to the VoIP endpoints about the quality of the data transmission. RTCP regularly sends reports with receipt statistics.

These statistics contain packet loss and jitter.

VoIP endpoints use this information to adapt the jitter buffer and/or the used audiocodec.



Latency

- **Latency** is the time it takes the RTP (media) packets to traverse the network. Too much latency causes callers to speak over the top of each other.
- Callers start to notice the effect of **latency** around 250ms, above ~600ms the experience is unusable. There will always be some latency, the objective is to minimize it and keep total roundtrip time below 250ms.
- Ideally latency should be below 100ms because, while it is noticable at 250ms, other services and issues beyond your control might add delay causing the cumulative total to be over 250ms.

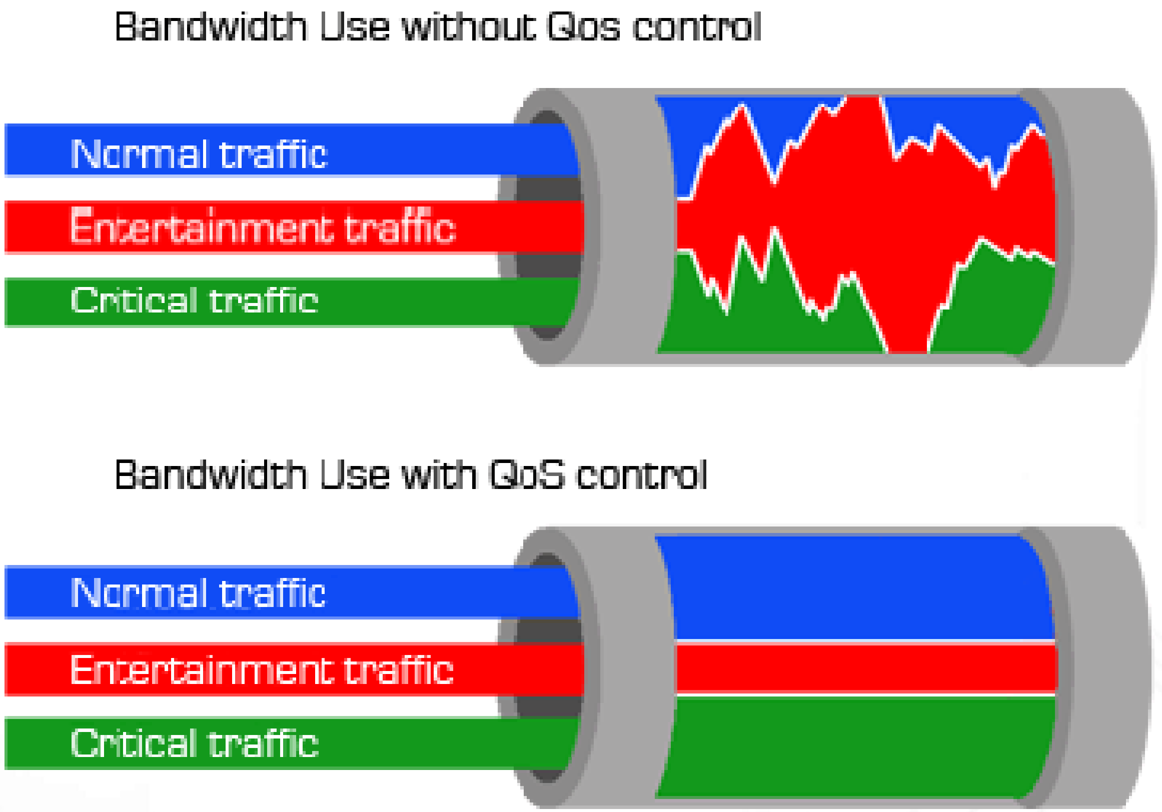
Jitter

- **Jitter** is when the latency through the network is not fixed but varies. Jitter causes that packets don’t arrive in the same order they were sent. For small amounts of jitter, this can be resolved in the **jitter**buffer. The length of the jitter buffer introduced causes increased latency.
- Too much jitter cannot be resolved by a reasonable length jitter buffer without introducing too much delay, so instead results in **jitter induced packet loss** causing choppy audio.

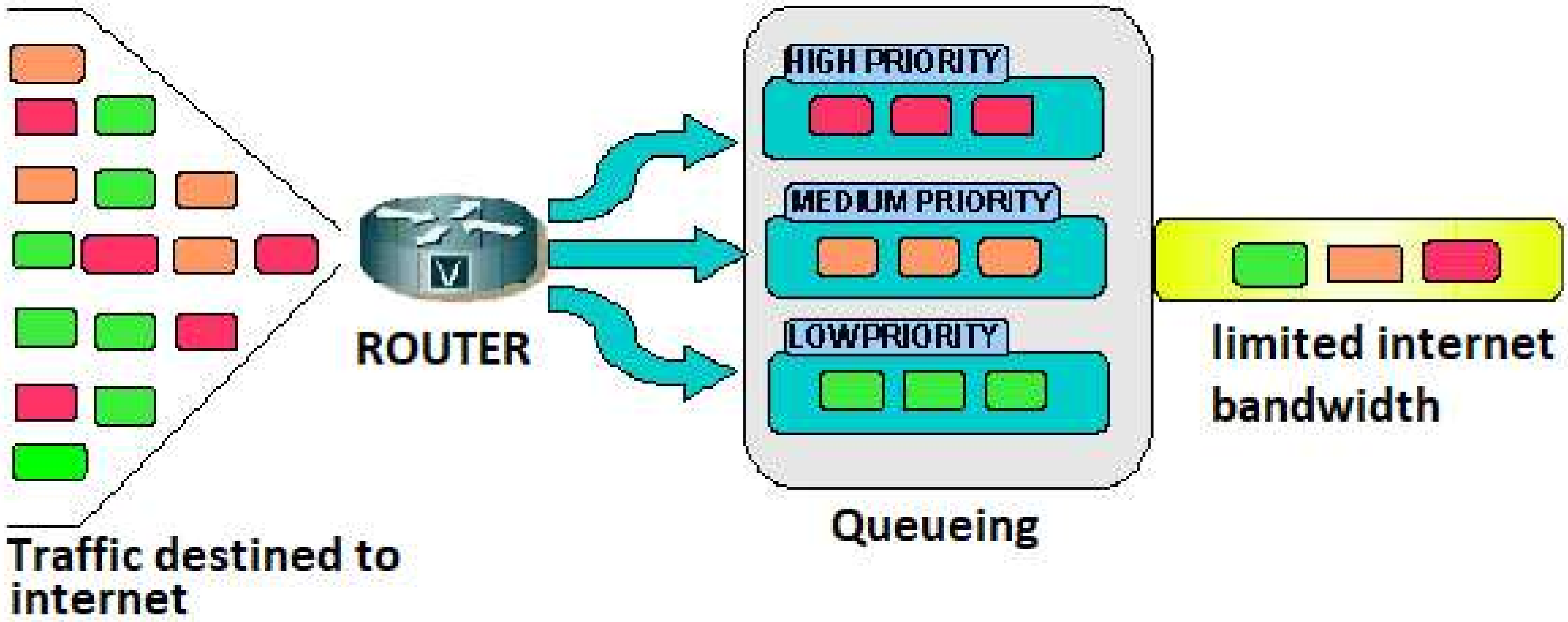
QoS traffic control

- Traffic shaping
- Prioritizing

QoS traffic shaping.



QoS traffic prioritizing



QoS traffic control

Traffic control is done on the **outbound interface** (we have no direct control over traffic that is being sent to us)

- **Rate limiting** is done by **dropping** some low priority packets so we have capacity for higher priority packets
- We need to know how much bandwidth is available
- We are not reordering the packets! The packets will leave the router in the **exact sequence** as they are received (provided that we are forwarding them)

Hierarchical Token Bucket (HTB)

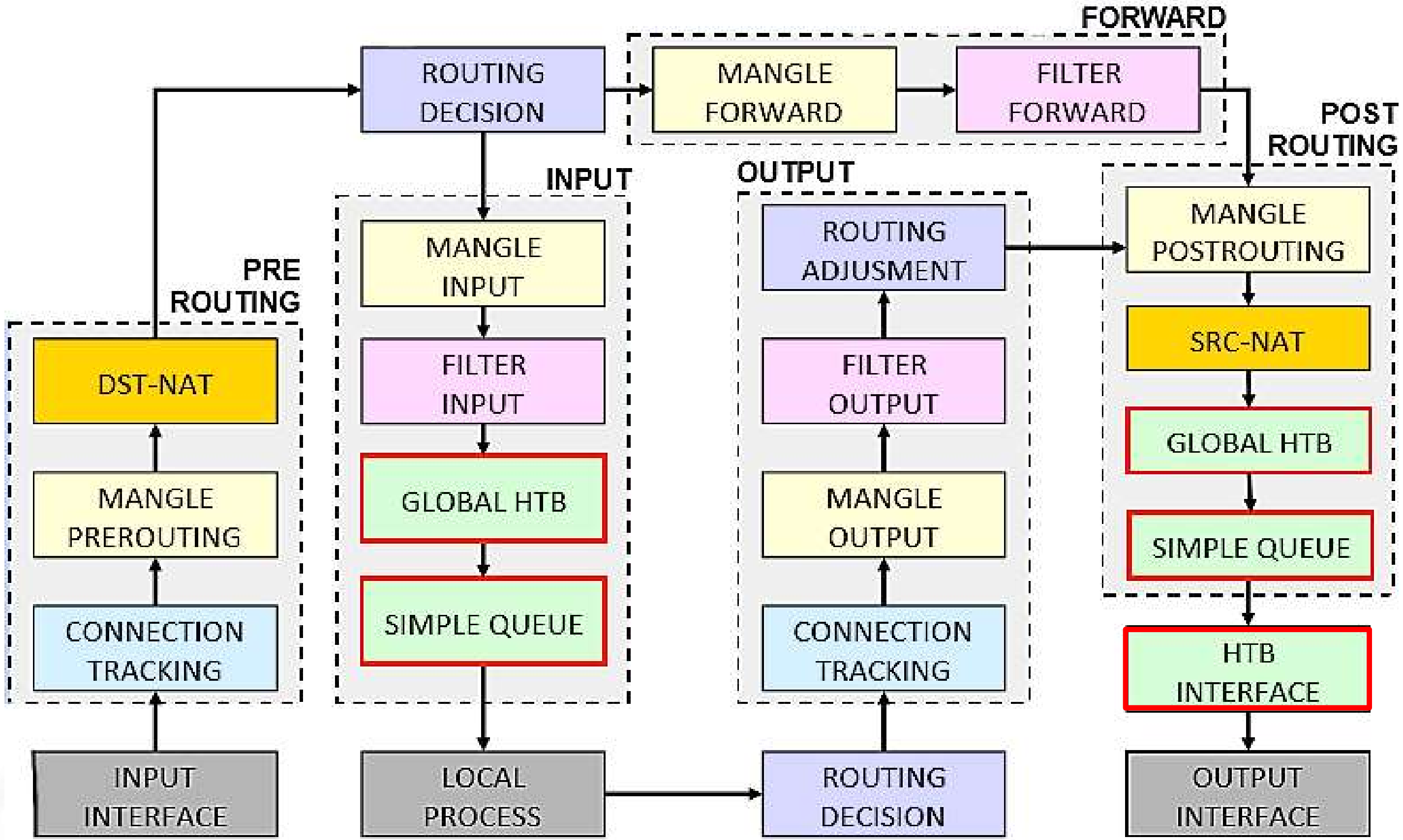
Hierarchical Token Bucket(HTB) is a classful queuing method for handling different kinds of traffic.

- All QoS implementation in **RouterOS** is based on Hierarchical Token Bucket.
- HTB provides a way to create a hierarchical queue structure and determine relations between parent and child queues and relations between child queues.

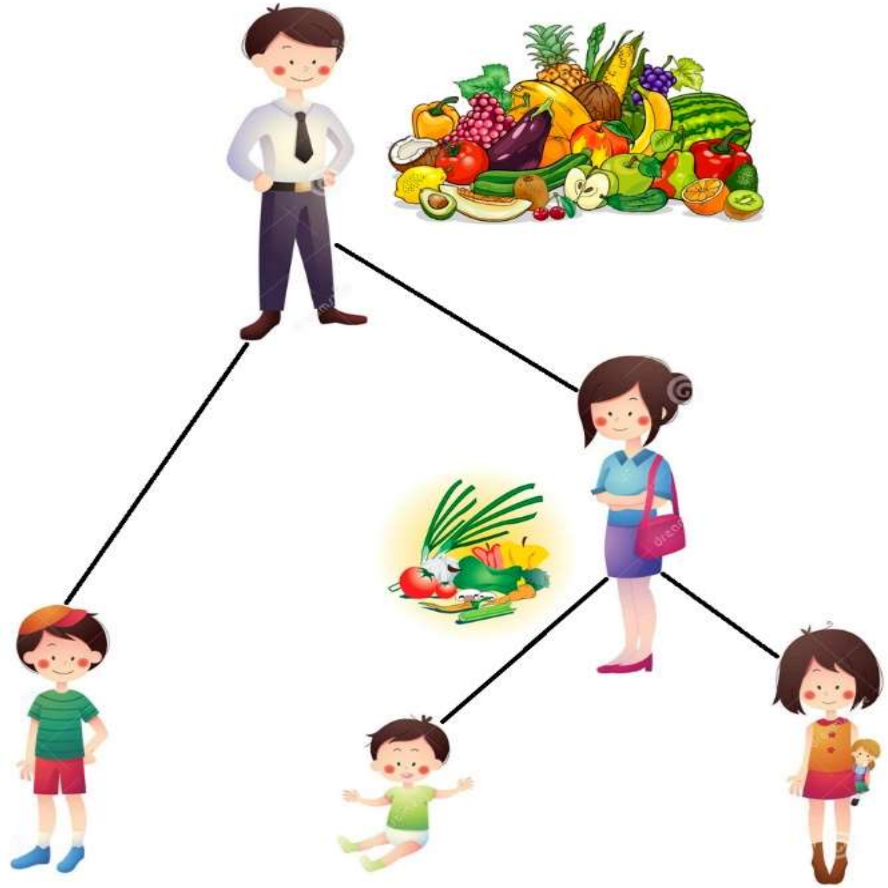
Hierarchical Token Bucket (HTB)

- When a packet travels through the router, it passes **2 HTB trees (global and interface)** and simple queues
- When a packet travels to the router, it passes only **global HTB** and simple queues.
- When a packet travels from the router, it passes **2 HTB trees** and simple queues.

HTB packet flow routerOS V6



HTB queue structure(children and parents)



HTB queue structure (children and parents)

- As soon as queue has at least one child it becomes a parent queue.
- All child queues (it doesn't matter how many levels of parents they have) are on the same bottom level of HTB
- Child queues do the actual traffic consumption, parent queues are only responsible for traffic distribution

HTB rate limiting

HTB has two rate limits:

- **CIR (Committed Information Rate)** – (**limit-at** in RouterOS) worst case scenario,the flow will get this amount of traffic no matter what (assuming there is enough bandwidth)
- **MIR (Maximal Information Rate)** – (**max-limit** in RouterOS) best case scenario, rate that the flow can get up to, if the queue's parent has spare bandwidth available

At first HTB will try to satisfy every child queue's **limit-at** only then it will try to reach max-limit

HTB rate limiting

Maximal rate of the parent should be equal or bigger than sum of committed rates of the children

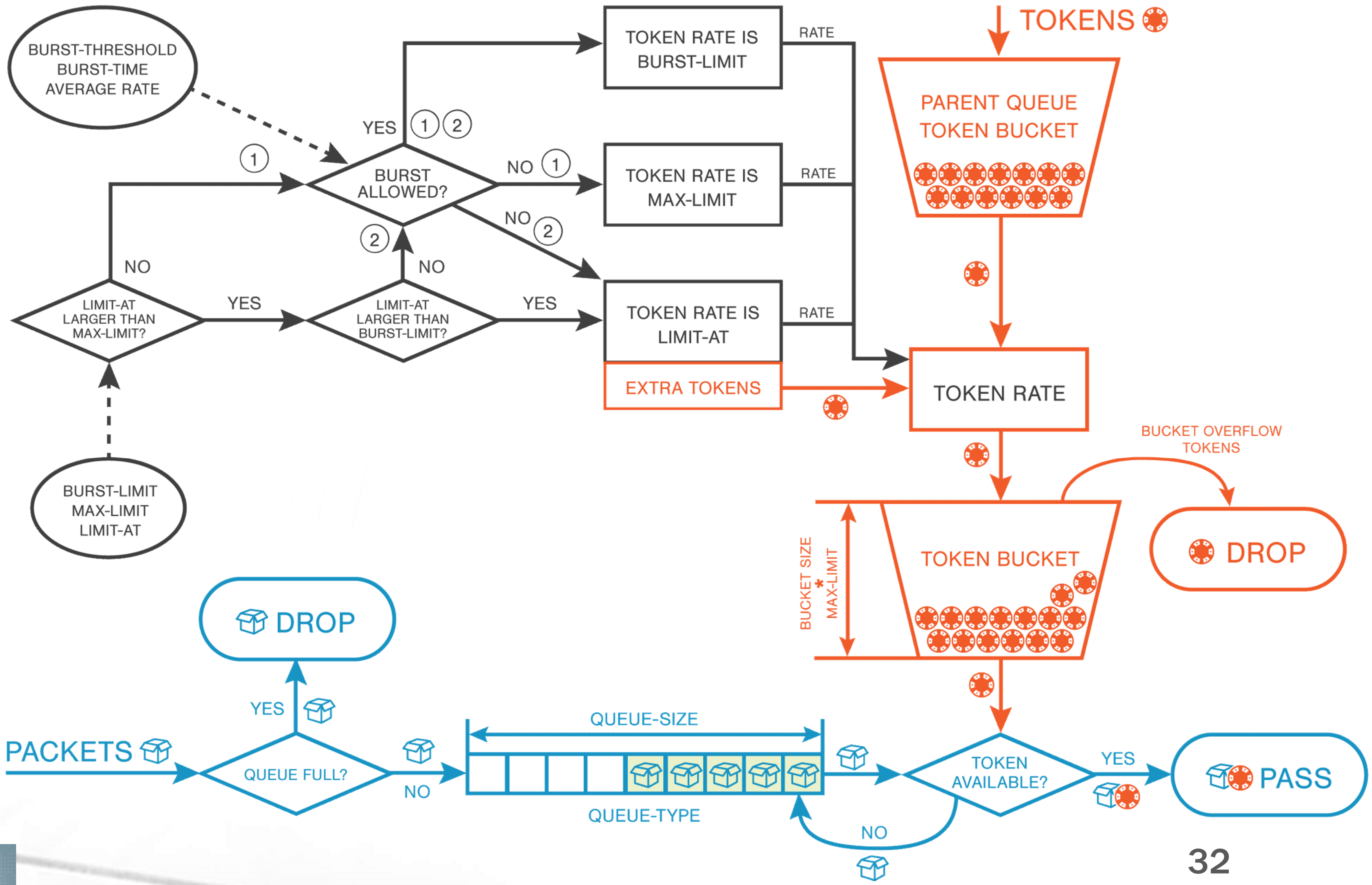
- MIR (parent) ≥ CIR(child1) +...+ CIR(childN)

Maximal rate of any child should be less or equal to maximal rate of the parent

- MIR (parent) ≥ MIR(child1)
- MIR (parent) ≥ MIR(child2)
- MIR (parent) ≥ MIR(childN)

HTB priority

- Works only for child queues to differentiate them.
- 8 is the lowest priority, 1 is the highest priority.
- Queue with higher priority will get the chance to satisfy its max-limit before other queues.
- Actual traffic prioritization will only work if limits are specified. A queue without limits will not prioritize anything.



MUM Mikrotik User Meeting

SDC

QoS example scenario HTB

Gigaset N510 IP Pro
- 6 DECT handsets
- 6 SIP registrations
- 4 simultanious SIP calls

QoS example scenario HTB

Qualify VoIP traffic with Mangle

- Identifying VoIP traffic
- Marking VoIP connections
- Marking VoIP packets within connection

Reduce CPU load

VoIP call consists of 2 kinds of traffic

- Call setup and signalling (UDP 5060)
- RTP Voice payload (UDP ports depended on VoIP system)

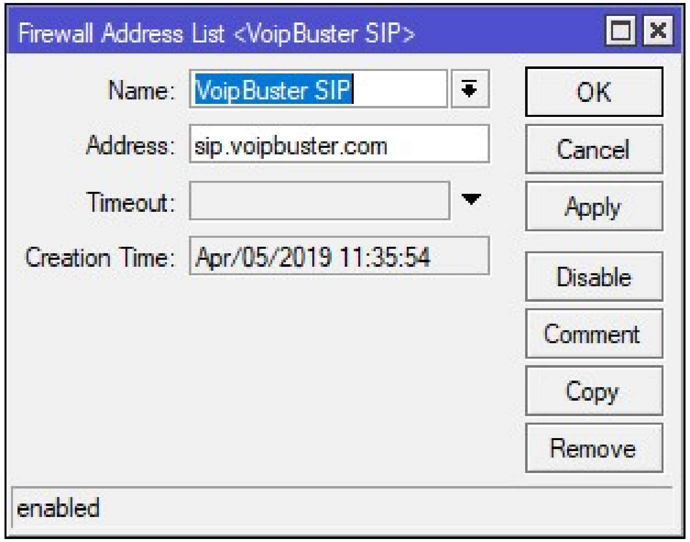
Qualify VoIP traffic with Mangle

VoipBuster SIP settings:

- SIP port : 5060
- Registrar : sip.voipbuster.com
- Proxy server : sip.voipbuster.com

Gigaset N510 IP Pro RTP port numbers:

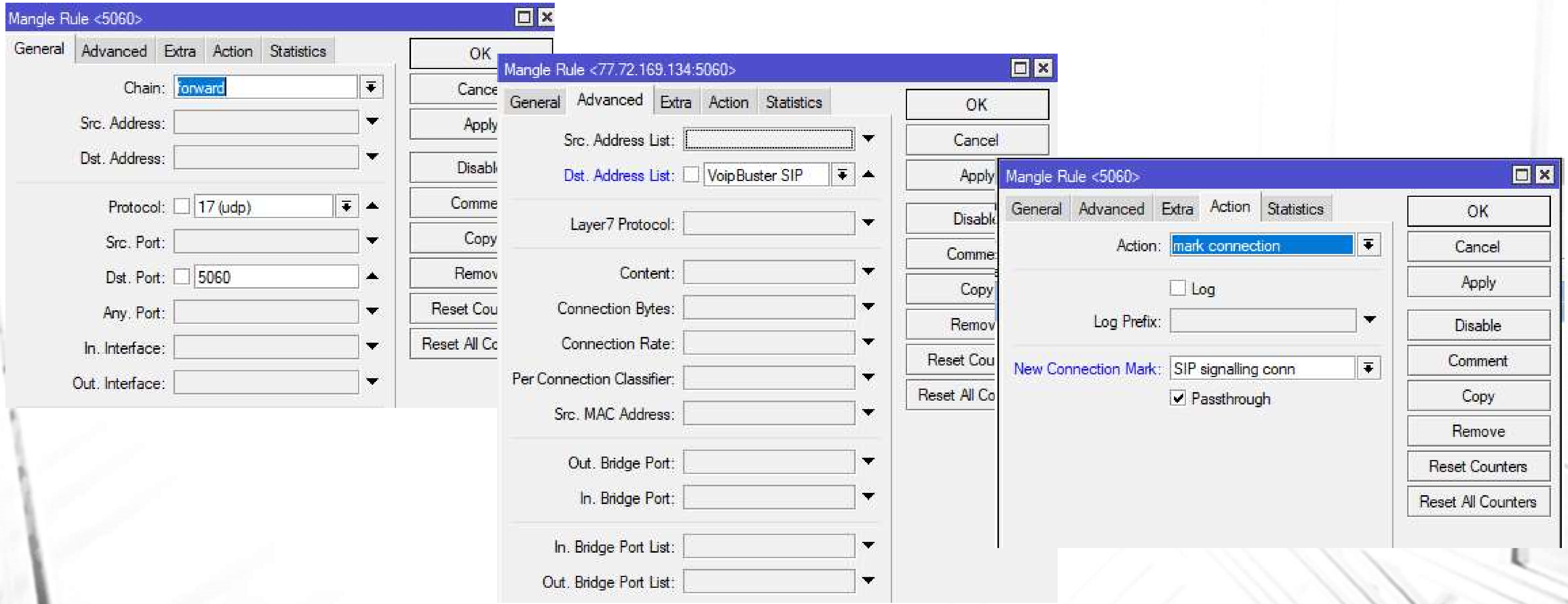
Qualify VoIP traffic with Mangle



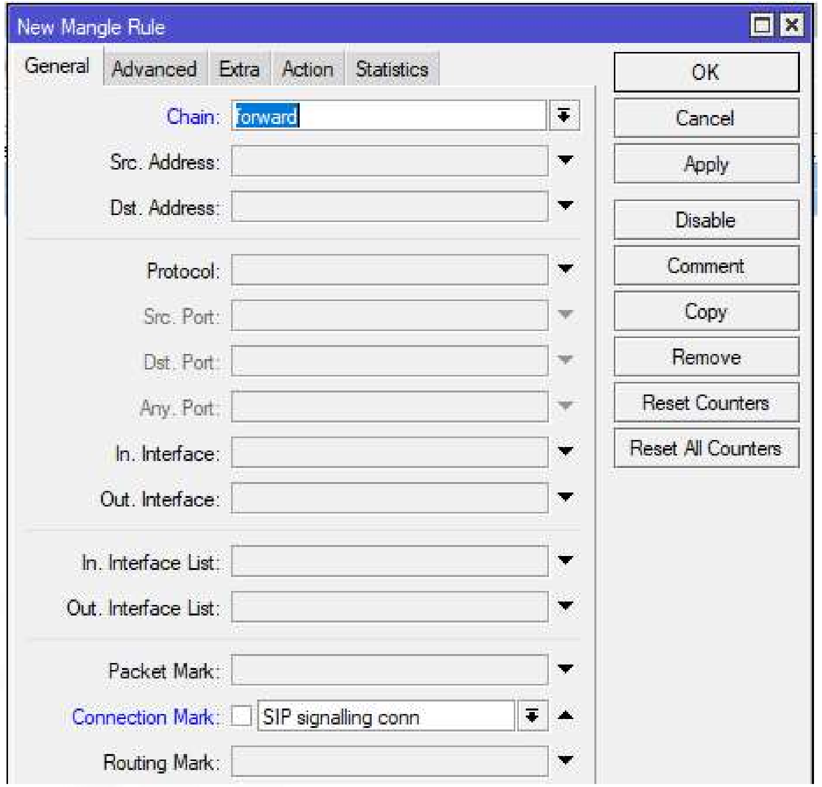
Add address list for discovery SIP server addresses Voipbuster

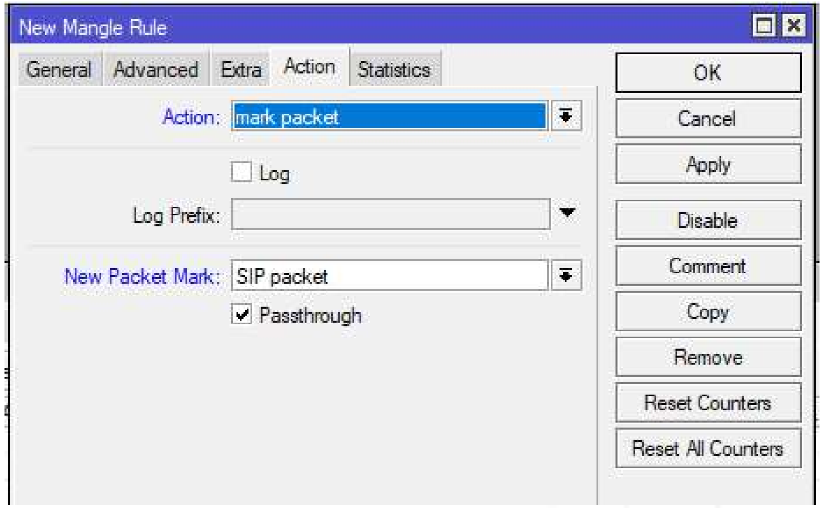
	VoipBuster SIP	sip.voipbuster.com		Apr/05/2019 11:35:54
	;;; sip.voipbuster.com			
D	VoipBuster SIP	77.72.169.134		Apr/05/2019 11:35:54
	;;; sip.voipbuster.com			
D	VoipBuster SIP	77.72.169.129		Apr/05/2019 11:35:54

Mark VoIP signalling connections

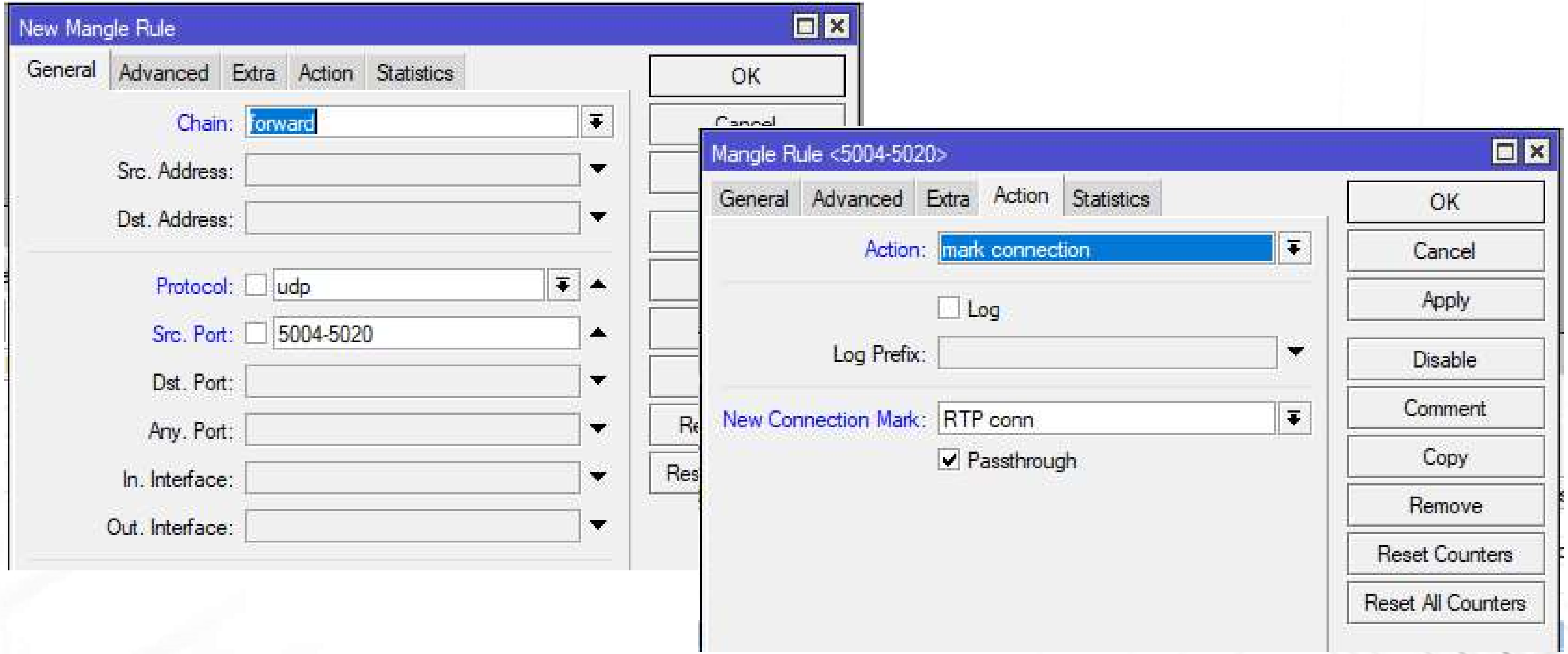


Mark VoIP signalling packets

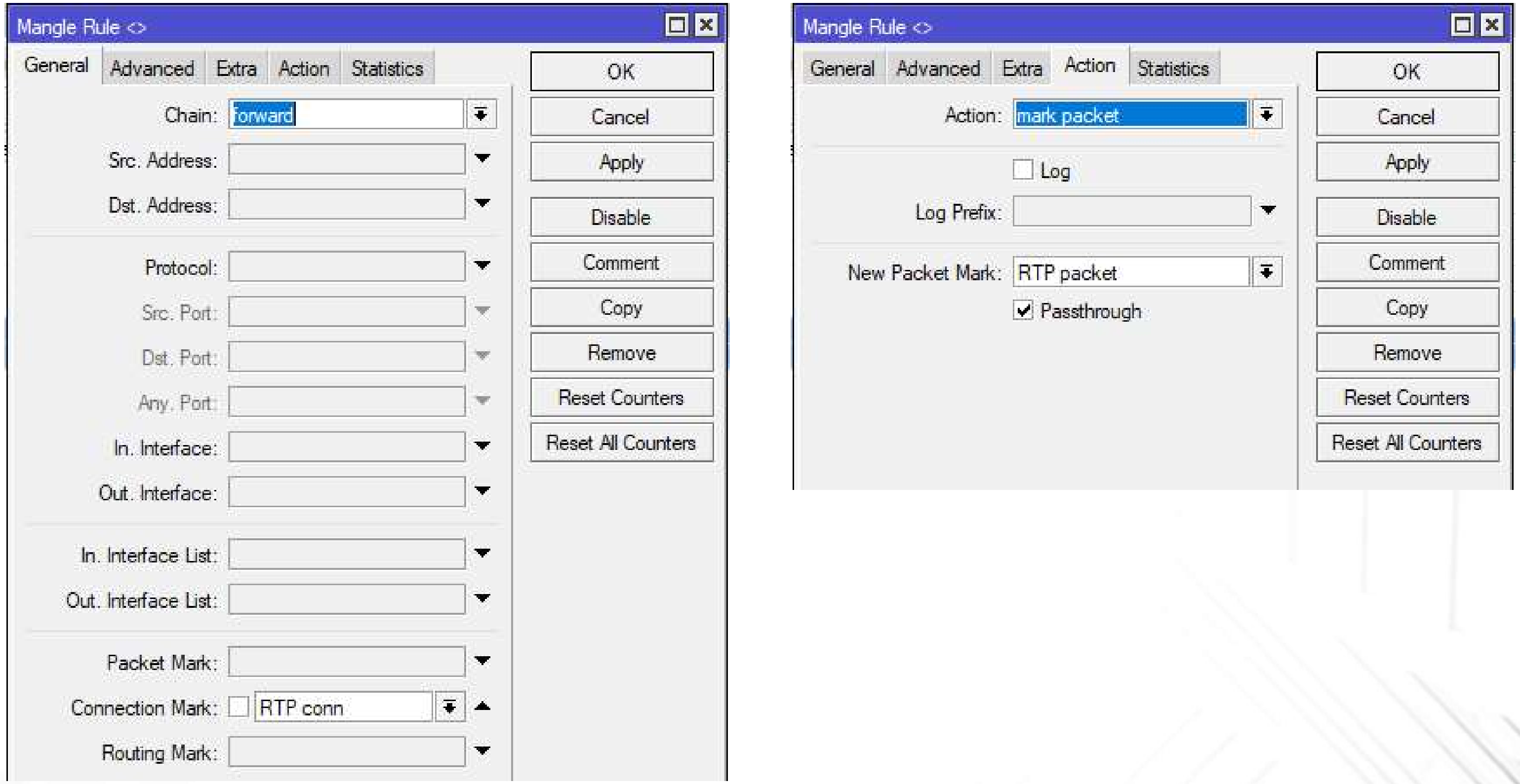




Mark VoIP RTP connections



Mark VoIP RTP packets



Qualify VoIP traffic with Mangle

Firewall

Filter Rules | NAT | Mangle | Raw | Service Ports | Connections | Address Lists | Layer7 Protocols

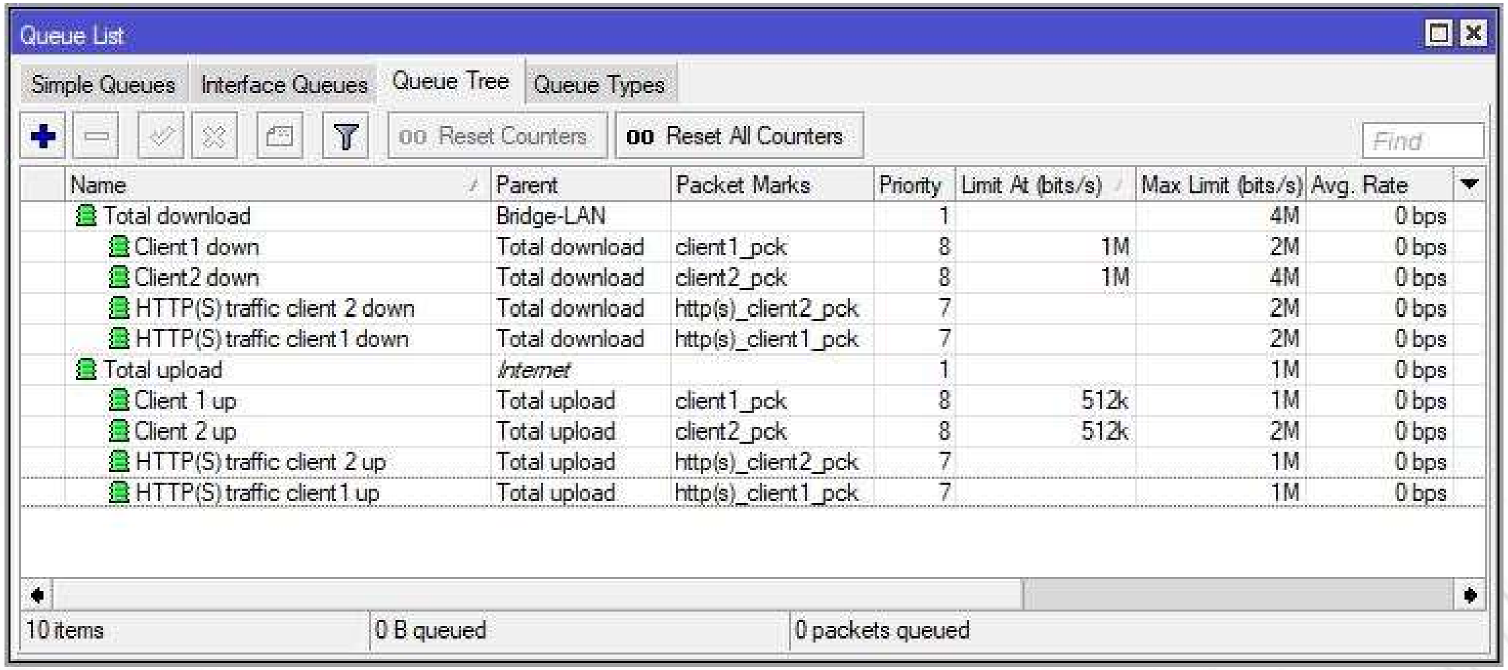
00 Reset Counters | 00 Reset All Counters | Find

#	Action	Chain	Src. Address	Dst. Address	Protocol	Src. Port	Dst. Port	In. Inter...	Out. Int...	In. Inter...	Out. Int...	Src. Ad...	Dst. Address List	New Packet Mark	Bytes	Packets
;;; Connection mark SIP signalling																
0	mark connection	forward			17 (udp)		5060						VoipBuster SIP		6.5 KiB	26
;;; Packet mark SIP signalling																
1	mark packet	forward												SIP signalling pac...	11.5 KiB	35
;;; Connection mark RTP																
2	mark connection	forward			17 (udp)	5004-5020									171.5 KiB	878
;;; Packet mark RTP																
3	mark packet	forward												RTP packet	343.6 KiB	1 759

Provide the mangle rules with clear comments!

HTB implementation Queue Tree

- Queue Tree is a direct implementation of HTB



Queue List

Simple Queues | Interface Queues | Queue Tree | Queue Types

00 Reset Counters | 00 Reset All Counters | Find

Name	Parent	Packet Marks	Priority	Limit At (bits/s)	Max Limit (bits/s)	Avg. Rate
Total download	Bridge-LAN		1		4M	0 bps
Client1 down	Total download	client1_pck	8	1M	2M	0 bps
Client2 down	Total download	client2_pck	8	1M	4M	0 bps
HTTP(S) traffic client 2 down	Total download	http(s)_client2_pck	7		2M	0 bps
HTTP(S) traffic client1 down	Total download	http(s)_client1_pck	7		2M	0 bps
Total upload	*Internet*		1		1M	0 bps
Client 1 up	Total upload	client1_pck	8	512k	1M	0 bps
Client 2 up	Total upload	client2_pck	8	512k	2M	0 bps
HTTP(S) traffic client 2 up	Total upload	http(s)_client2_pck	7		1M	0 bps
HTTP(S) traffic client1 up	Total upload	http(s)_client1_pck	7		1M	0 bps

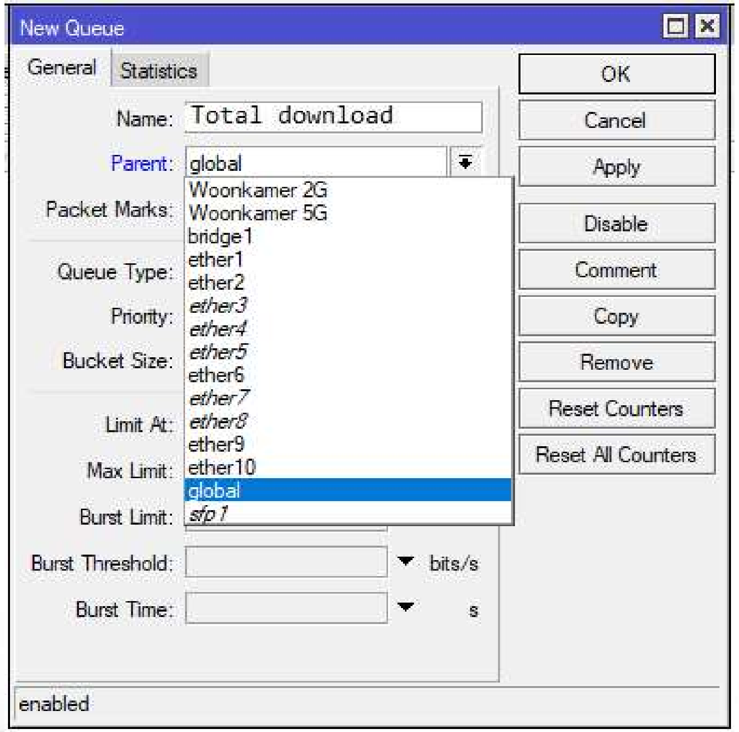
10 items | 0 B queued | 0 packets queued

HTB implementation Queue Tree

- Queue Tree is one directional only and can be placed in any of the available HTBs
- Queue Tree queues don't have any order – all traffic is processed simultaneously
- All child queues must have packet marks from "/ip firewall mangle" facility assigned to them
- If placed in the same HTB, Simple queue will take all the traffic away from the Queue Tree queue

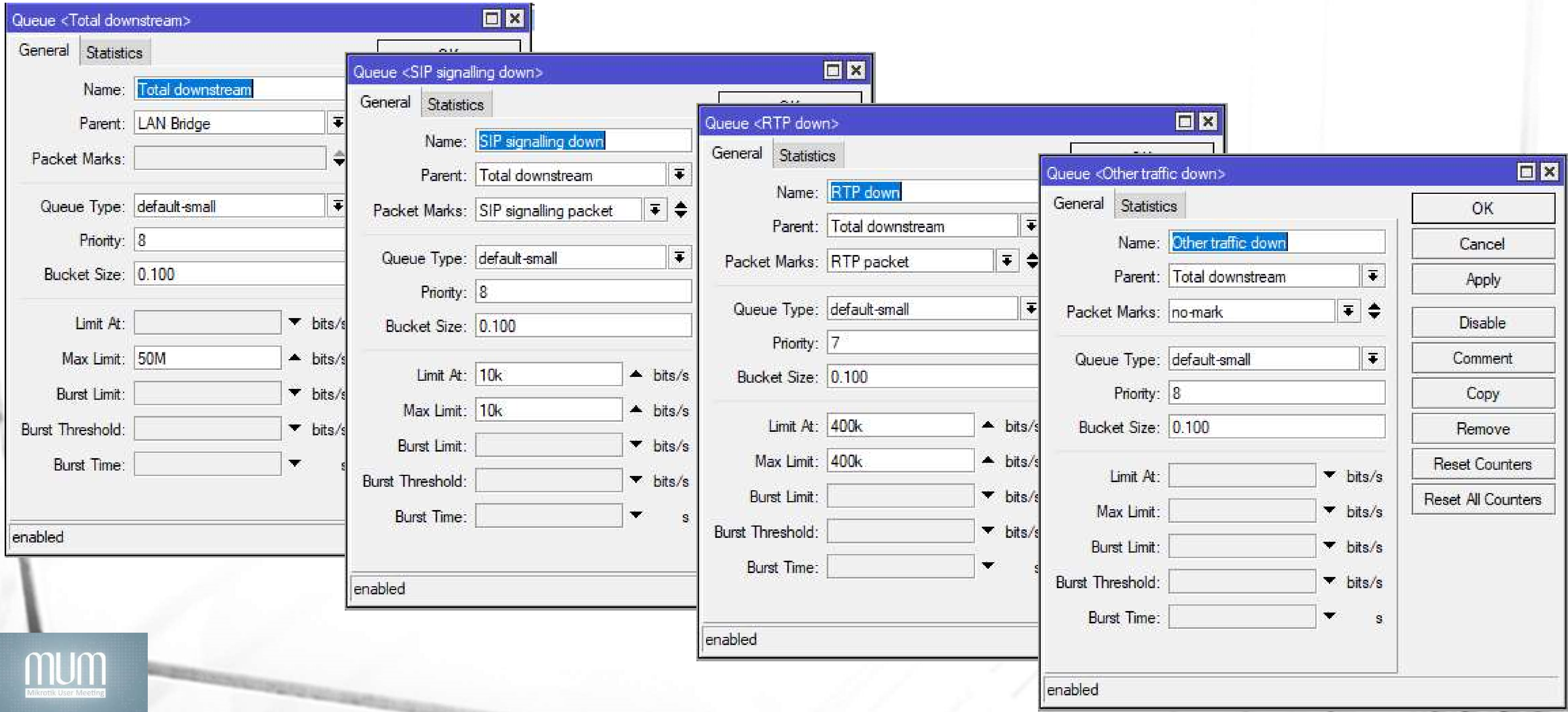
HTB implementation Queue Tree

The Queue Tree can be placed in ‘**global**’ or in ‘**interface queue**’

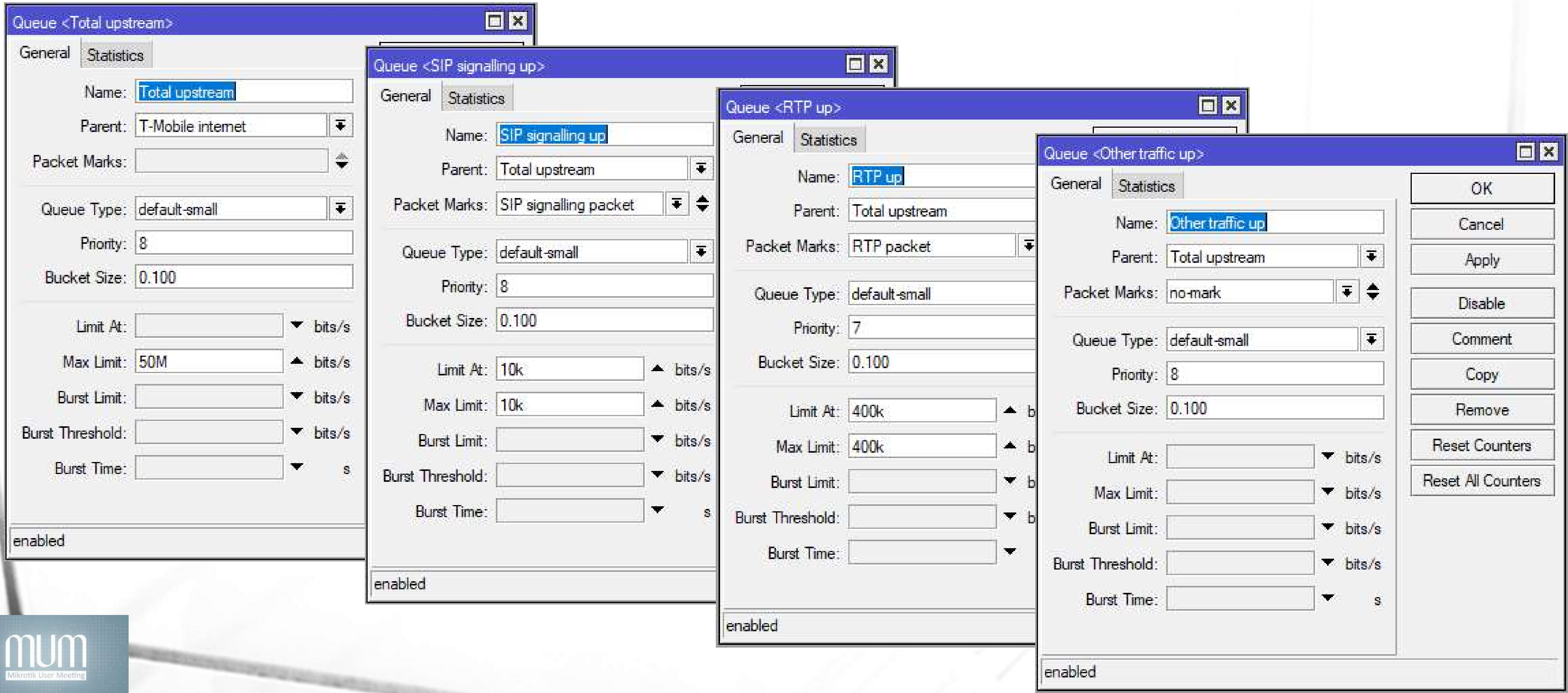


Each Interface HTB only receives traffic that will be leaving through a particular interface – there is no need for to separate upload and download in mangle

Queue Tree downstream



Queue Tree upstream



HTB implementation Queue Tree

Queue Tree structure

Queue List

Simple Queues | Interface Queues | Queue Tree | Queue Types

00 Reset Counters | 00 Reset All Counters | Find

Name	Parent	Packet Marks	Limit At (bits/s)	Max Limit (bits/s)	Avg. Rate	Queued Bytes	Bytes	Packets
Total downstream	LAN Bridge			50M	20.9 kbps	0 B	447.4 MiB	427 671
Other traffic down	Total downstream	no-mark			20.9 kbps	0 B	447.3 MiB	427 066
RTP down	Total downstream	RTP packet	400k	400k	0 bps	0 B	118.7 KiB	568
SIP signalling down	Total downstream	SIP signalling packet	10k	10k	0 bps	0 B	8.3 KiB	15
Total upstream	T-Mobile internet			50M	512 bps	0 B	411.3 MiB	443 030
Other traffic up	Total upstream	no-mark			512 bps	0 B	411.2 MiB	442 402
RTP up	Total upstream	RTP packet	400k	400k	0 bps	0 B	118.1 KiB	565
SIP signalling up	Total upstream	SIP signalling packet	10k	10k	0 bps	0 B	13.2 KiB	63

8 items (1 selected) | 0 B queued | 0 packets queued

Queue colors in Winbox:

- **0% - 50% available traffic used - green**
- **51% - 75% available traffic used - yellow**
- **76% - 100% available traffic used - red**

Check VoIP traffic

- **Checking for dropped packets in Queue Tree**

  Show Colums ‘dropped’

Queue List

Simple Queues | Interface Queues | Queue Tree | Queue Types

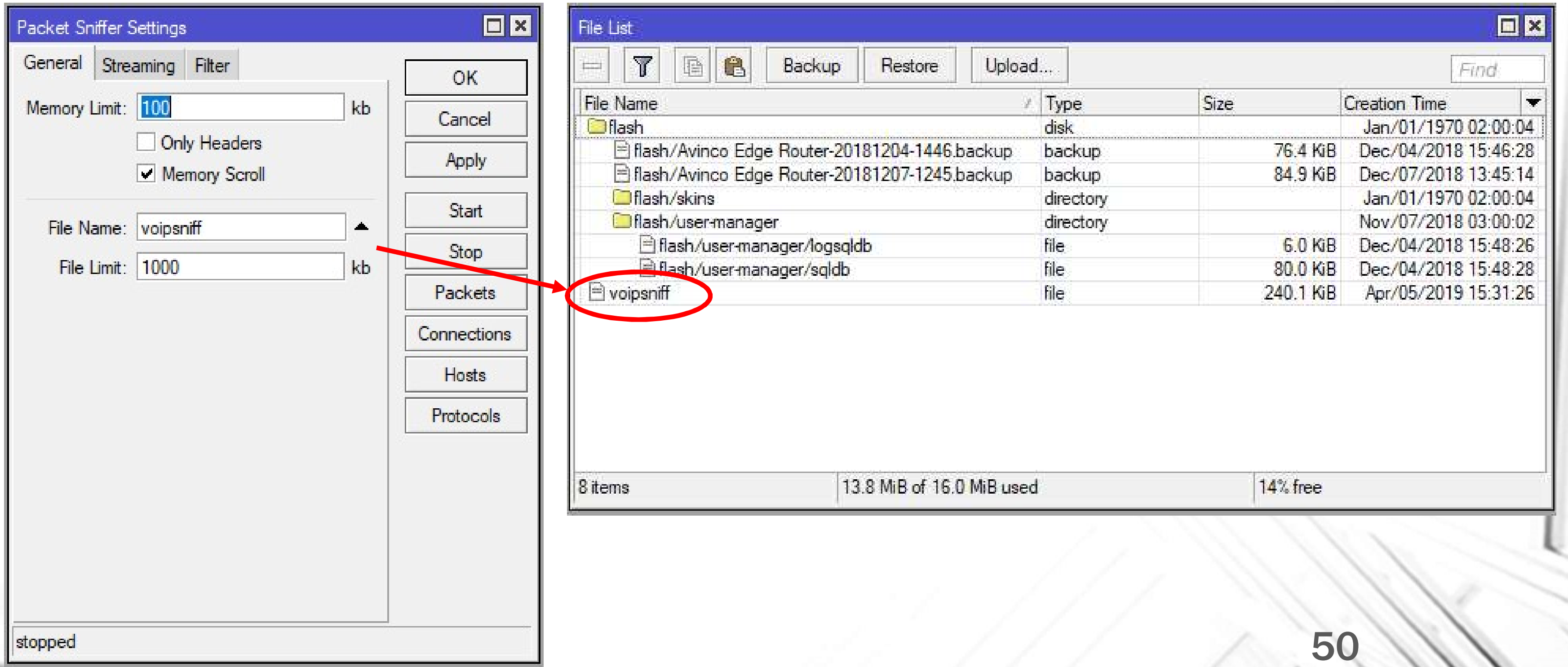
oo Reset Counters | oo Reset All Counters | Find

Name	Parent	Packet Marks	Limit At (bits/s)	Max Limit (bits/s)	Avg. Rate	Queued Bytes	Bytes	Packets	Dropped
Total downstream	LAN Bridge			200M	12.8 kbps	0 B	452.5 MiB	439 005	0
Other traffic down	Total downstream	no-mark			12.8 kbps	0 B	452.3 MiB	437 958	0
RTP down	Total downstream	RTP packet	400k	400k	0 bps	0 B	205.4 KiB	983	0
SIP signalling down	Total downstream	SIP signalling packet	10k	10k	0 bps	0 B	22.9 KiB	42	0
Total upstream	T-Mobile internet			200M	496 bps	0 B	412.0 MiB	447 086	0
Other traffic up	Total upstream	no-mark			496 bps	0 B	411.8 MiB	445 940	0
RTP up	Total upstream	RTP packet	400k	400k	0 bps	0 B	204.2 KiB	977	0
SIP signalling up	Total upstream	SIP signalling packet	10k	10k	0 bps	0 B	33.1 KiB	169	0

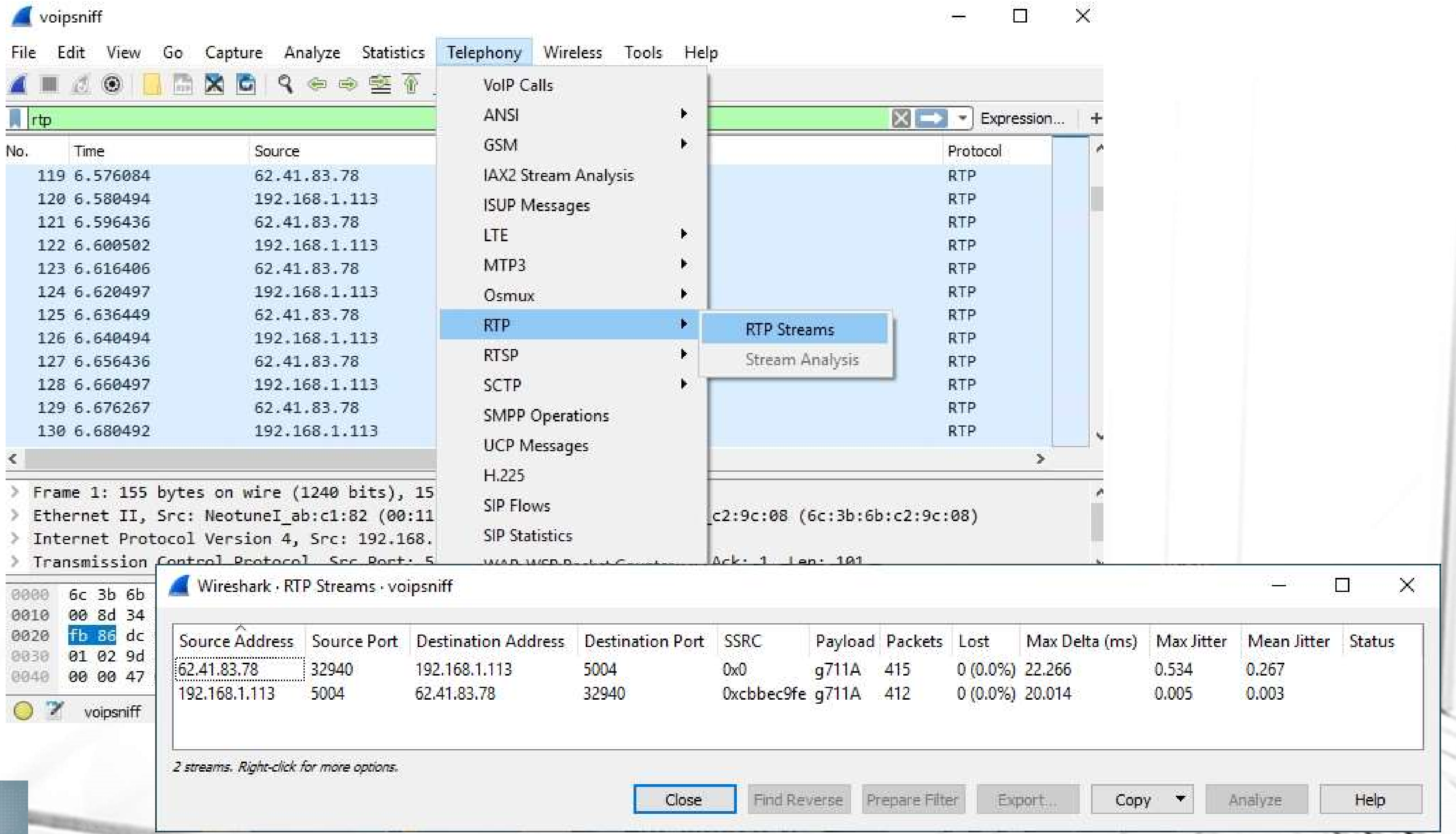
8 items (1 selected) | 0 B queued | 0 packets queued

Check VoIP traffic by sniffing

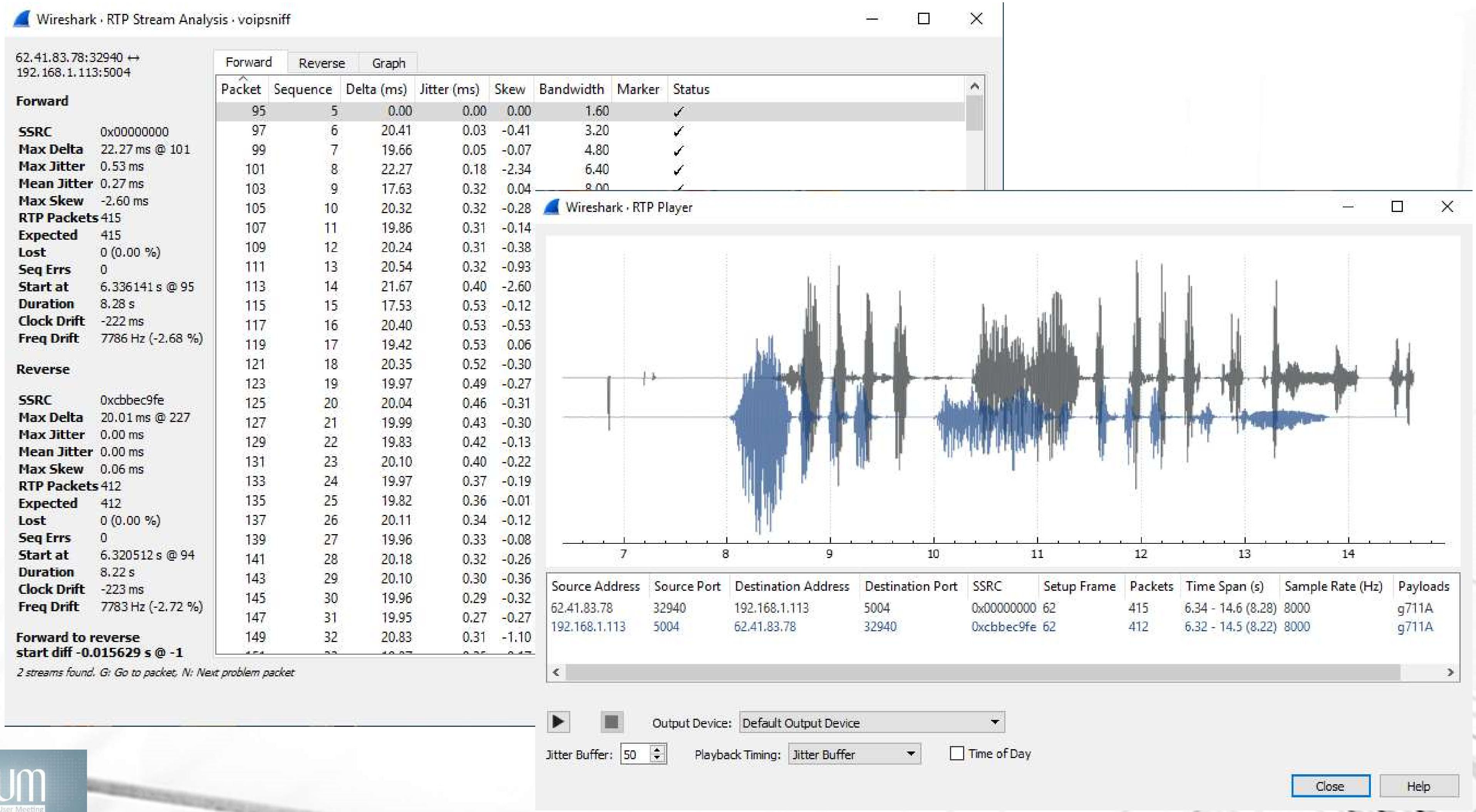
- Activate Packet Sniffer for VoIP traffic



Check VoIP traffic by sniffing



Check VoIP traffic by sniffing



Thank you for your attention!