HTB Implementation on RouterOS QoS

Prepared by: Valens Riyadi
Citraweb Nusa Infomedia, Indonesia
www.mikrotik.co.id
Introduction

• Valens Riyadi - valens@mikrotik.co.id
• Company: Citraweb Nusa Infomedia
  – Wireless ISP - www.citra.net.id
  – Web Developer - www.citra.web.id
• Head of National Internet Resources of Indonesian ISP Association / IDNIC
• Founder and Volunteer of Airputih Foundation, an IT Emergency Task Force on Disaster Area
Outline

• Basic Concept
• Simple HTB
• Terms in HTB
• Common Mistakes
• More detail information
Basic Concept

• QoS in RouterOS is not only about bandwidth limitation, but how to distribute the bandwidth fairly

• Things have to consider:
  – How to mangle
    • Check packet flow and firewall documentation on manual and wiki pages
    • Will not covered in this presentation
  – HTB (Hierarchical Token Bucket)
Basic Concept

• HTB (Hierarchical Token Bucket) is part of QoS, to make a hierarchical queue structure and determine relations between queues (priority, burst possibility, etc)

• HTB is meant as a more understandable, intuitive and faster replacement for the CBQ qdisc in Linux.

• HTB assigned to any physical interface or virtual interface (global-in, global-out, global-total)
Sample of HTB

- A has 2 children:
  - B
  - C
- C has 2 children:
  - D
  - E
Type of Queues

• Leaf queue
  – Most lower level, has no child
  – Make actual traffic consumption
  – All leaf queues are treated on equal basis
  – All leafs located on the most bottom level of HTB
Type of Queues

• Inner queue
  – Have at least one child
  – Responsible only for traffic distribution
Terms in HTB

• Committed Information Rate (CIR)
  – Limit-at in RouterOS
  – In worst case scenario, flow will get this amount of traffic no matter what (assuming we can actually send so much data)
  – CIR works on both leaf and inner queue
Terms in HTB

• Maximal Information Rate (MIR)
  – Max-limit in RouterOS
  – Rate that flow can get up to, if there queue's parent has spare bandwidth
Terms in HTB

• Priority
  – Work only for leaf (child) queue
  – 1 .. Highest priority, and 8 … lowest priority
  – Priority will work only if limits are specified
  – Priority calculated only after limit-at achieved
Sample Configuration

- We have 5 leaf queues:

<table>
<thead>
<tr>
<th>Queue</th>
<th>Limit-at</th>
<th>Max-limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>10M</td>
<td>20M</td>
</tr>
<tr>
<td>C2</td>
<td>1M</td>
<td>20M</td>
</tr>
<tr>
<td>C3</td>
<td>1M</td>
<td>20M</td>
</tr>
<tr>
<td>C4</td>
<td>1M</td>
<td>20M</td>
</tr>
<tr>
<td>C5</td>
<td>1M</td>
<td>20M</td>
</tr>
</tbody>
</table>
Winbox Configuration

Router fail to give C1 limit-at (10M)
With Parent

<table>
<thead>
<tr>
<th>Name</th>
<th>Parent</th>
<th>Packet Mark</th>
<th>Priority</th>
<th>Limit At (bits/s)</th>
<th>Max Limit (bits/s)</th>
<th>Avg. Rate</th>
<th>Q1</th>
</tr>
</thead>
<tbody>
<tr>
<td>queue-parent</td>
<td>ether3</td>
<td></td>
<td>1</td>
<td>5M</td>
<td>5.0 Mbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>queue-</td>
<td>C1</td>
<td>1</td>
<td>2M</td>
<td>2.0 Mbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>queue-</td>
<td>C2</td>
<td>8</td>
<td>2M</td>
<td>760.7 kbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>queue-</td>
<td>C3</td>
<td>8</td>
<td>2M</td>
<td>751.6 kbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>queue-</td>
<td>C4</td>
<td>8</td>
<td>2M</td>
<td>736.6 kbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5</td>
<td>queue-</td>
<td>C5</td>
<td>8</td>
<td>2M</td>
<td>751.6 kbps</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
More Hierarchy

A

B1

C1

C2

B2

C3

C4

C5
C1 have better priority, get up to max-limit, all other capacity split for other leaf queue
**Limit-at on Inner Queue**

<table>
<thead>
<tr>
<th>Name</th>
<th>Parent</th>
<th>Packet Mark</th>
<th>Priority</th>
<th>Limit At (...)</th>
<th>Max Limit...</th>
<th>Avg. Rate</th>
<th>Queued Byte</th>
</tr>
</thead>
<tbody>
<tr>
<td>queue-parent</td>
<td>ether3</td>
<td></td>
<td>1</td>
<td>5M</td>
<td>5.0 Mbps</td>
<td></td>
<td>0 B</td>
</tr>
<tr>
<td>-Q1</td>
<td>queue-parent</td>
<td>C1</td>
<td>1</td>
<td>512k</td>
<td>2M</td>
<td>1981.1 kbps</td>
<td>72.7 KiB</td>
</tr>
<tr>
<td>-B1</td>
<td>queue-parent</td>
<td>8</td>
<td>2M</td>
<td>2M</td>
<td>2.0 Mbps</td>
<td></td>
<td>0 B</td>
</tr>
<tr>
<td>-Q2</td>
<td>B1</td>
<td>C2</td>
<td>8</td>
<td>512k</td>
<td>2M</td>
<td>1003.0 kbps</td>
<td>72.7 KiB</td>
</tr>
<tr>
<td>-Q3</td>
<td>B1</td>
<td>C3</td>
<td>8</td>
<td>512k</td>
<td>2M</td>
<td>1003.1 kbps</td>
<td>72.7 KiB</td>
</tr>
<tr>
<td>-B2</td>
<td>queue-parent</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td>1027.2 kbps</td>
<td>0 B</td>
</tr>
<tr>
<td>-Q4</td>
<td>B2</td>
<td>C4</td>
<td>8</td>
<td>512k</td>
<td>2M</td>
<td>509.1 kbps</td>
<td>72.7 KiB</td>
</tr>
<tr>
<td>-Q5</td>
<td>B2</td>
<td>C5</td>
<td>8</td>
<td>512k</td>
<td>2M</td>
<td>512.1 kbps</td>
<td>72.7 KiB</td>
</tr>
</tbody>
</table>
Priority on Leaf Queue

<table>
<thead>
<tr>
<th>Name</th>
<th>Parent</th>
<th>Packet Mark</th>
<th>Priority</th>
<th>Limit At</th>
<th>Max Limit</th>
<th>Avg. Rate</th>
<th>Queued Byte</th>
</tr>
</thead>
<tbody>
<tr>
<td>queue-parent</td>
<td>ether3</td>
<td></td>
<td>1</td>
<td>5M</td>
<td>2M</td>
<td>5.0 Mbps</td>
<td>0 B</td>
</tr>
<tr>
<td>Q1</td>
<td>queue-parent</td>
<td>C1</td>
<td>1</td>
<td>512k</td>
<td>2M</td>
<td>1981.2 kbps</td>
<td>71.3 KiB</td>
</tr>
<tr>
<td>B1</td>
<td>queue-parent</td>
<td>2M</td>
<td>2</td>
<td>512k</td>
<td>2M</td>
<td>2.0 Mbps</td>
<td>0 B</td>
</tr>
<tr>
<td>Q2</td>
<td>B1</td>
<td>C2</td>
<td>8</td>
<td>512k</td>
<td>2M</td>
<td>999.9 kbps</td>
<td>72.7 KiB</td>
</tr>
<tr>
<td>Q3</td>
<td>B1</td>
<td>C3</td>
<td>8</td>
<td>512k</td>
<td>2M</td>
<td>1000.0 kbps</td>
<td>72.8 KiB</td>
</tr>
<tr>
<td>B2</td>
<td>queue-parent</td>
<td>8</td>
<td></td>
<td>512k</td>
<td>2M</td>
<td>1024.4 kbps</td>
<td>0 B</td>
</tr>
<tr>
<td>Q4</td>
<td>B2</td>
<td>C4</td>
<td>2</td>
<td>512k</td>
<td>2M</td>
<td>512.2 kbps</td>
<td>72.7 KiB</td>
</tr>
<tr>
<td>Q5</td>
<td>B2</td>
<td>C5</td>
<td>2</td>
<td>512k</td>
<td>2M</td>
<td>512.1 kbps</td>
<td>72.7 KiB</td>
</tr>
</tbody>
</table>
Priority on Inner Queue

Priority works only on Leaf Queue, will not work on Inner Queue
Common Mistakes

• Leaf queue without parent
• Priority on Inner queue
Test case

- We have 300 hosts on our network
- In worst scenario, 200 hosts will online on the same time
- At least 20 hosts online at minimum.
- All hosts have same priority
- Total bandwidth = 10 Mbps
Tips

• For an office network that all computer have same speed and same priority:
  – Limit-at  = Total bandwidth / max hosts
  – Max-limit = Total bandwidth / min hosts
• Max-limit = 10 mbps / 20 (min host) = 500 kbps
• Limit-at = 10 mbps / 200 (max host) = 50 kbps
More Detail Information

Thank you!

• Q&A........
• Or email to: valens@mikrotik.co.id