Fast IP Address Lookup and Packet Classification

Jonathan Turner & Subhash Suri
(and Zubin Dittia, Will Eatherton, George Varghese, Cheenu Venkatachary)
Washington University
Computer Science Department
{jst,suri}@cs.wustl.edu
IP Address Lookup

- Routing tables at router input ports contain (prefix, next hop) pairs.
- Unicast address in packet is compared to stored prefixes, starting with left-most bit.
- Prefix that matches largest number of address bits is desired match.
- Packet is forwarded to the specified next hop.
- Next hop fields change as a result of topology changes and traffic changes.
- Set of prefixes changes infrequently.

<table>
<thead>
<tr>
<th>prefix</th>
<th>next hop</th>
</tr>
</thead>
<tbody>
<tr>
<td>10*</td>
<td>7</td>
</tr>
<tr>
<td>01*</td>
<td>5</td>
</tr>
<tr>
<td>110*</td>
<td>3</td>
</tr>
<tr>
<td>1011*</td>
<td>5</td>
</tr>
<tr>
<td>0001*</td>
<td>0</td>
</tr>
<tr>
<td>01011*</td>
<td>7</td>
</tr>
<tr>
<td>00010*</td>
<td>1</td>
</tr>
<tr>
<td>001100*</td>
<td>2</td>
</tr>
<tr>
<td>1011001*</td>
<td>3</td>
</tr>
<tr>
<td>1011010*</td>
<td>5</td>
</tr>
<tr>
<td>0100110*</td>
<td>6</td>
</tr>
<tr>
<td>01001100*</td>
<td>4</td>
</tr>
<tr>
<td>10110011*</td>
<td>8</td>
</tr>
<tr>
<td>10011000*</td>
<td>10</td>
</tr>
<tr>
<td>01011001*</td>
<td>9</td>
</tr>
</tbody>
</table>
Address Lookup Using Tries

- Prefixes stored in binary trie.
- Green nodes denote terminal nodes for prefixes.
- Search for address, by using address bits to traverse path from route.
- Remember most recent green node visited.
- If search ends at leaf, then exact match.
- If search ends because no matching branch, go back to last green node.
- Number of memory accesses proportional to address length (32 in IPv4, 128 in IPv6).
Multibit Tries

- Multibit tries match several bits at once.
- Best prefix length or next hop info stored in trie.
- Greatly speeds search at cost of more memory.
  - 8 bit stride gives 4 memory accesses steps for IPv4; first two are typically in cache (for software)
  - requires much more memory than trie
  - optimize space usage for given depth using variable length stride and dynamic programming
- Like standard tries, supports in-place modification.

address: 101 100 011 000
Coding Subtrees with Bitmaps

- Compact form of multibit trie node
- Siblings stored contiguously.
  - memory mgmt. issues
- Separate array points to next hop info.
- 4 bit stride works best.
  - typically use <8 bytes per stored prefix
  - also, <8 bytes per node, gives fast access
  - 9 accesses for v4; reduces to 6 with 9 bit starter array
- Good for hardware implementation.
Implementing Fast IP Lookup

- **PP Extender** implements IP lookup (and packet classification)
- **Input FPGA** finds packets, extracts address, updates header.
  - initial version will have single VCI for IP packets, may extend later
  - non IP cells pass through without change
- **Address lookup yields new VCI for cells in IP packets.**
  - all cells in packet mapped to new VCI (preconfigured by control software)
  - ATM layer lookup in switch routes to proper output

Jonathan Turner 7/23/99