A Traffic-aware Top-N Firewall Ruleset Approximation Algorithm

Ivan H.Y. Lam, Donghan Wang, H. Jonathan Chao

Department of Electrical and Computer Engineering, Polytechnic Institute of New York University

1. Centralized management and distributed rule sets

- Firewall rules centrally managed with distributed enforcement
- If a ruleset is too big for a data-path device either distributed caching of rules or pre-populate a subset of rules.

2. Top-N Rules

- Cache eviction by simple LRU
- Or according to hit-rates
- Trivial if rules has no dependency

3. Rule dependencies

- Two (partially) overlapping rules with different actions – conflicting rules.
- Prioritize rules and use the action from the highest priority rule
- Dependency arise when choosing subset of rules

4. Goal

- An algorithm that dynamically selecting a subset of N rules (Top-N) with the highest hit-rates that:
  1. Accepts rulesets with dependencies
  2. Suitable for online computation
  3. Imposes only light burden on traffic monitoring
  4. Dynamically adapts to traffic changes

5. Deployment Scenario

- Sub-rulesets stored in data-plane devices (fast path)
- A miss (default rule) looks up at global ruleset (slow path)
- Keeps hit-rates, periodically updates sub-rulesets

6. Top-N Target List

<table>
<thead>
<tr>
<th>Rule</th>
<th>Hits 1</th>
<th>Hits 2</th>
<th>Hits 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>250</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>R2</td>
<td>90</td>
<td>30</td>
<td>15</td>
</tr>
<tr>
<td>R3</td>
<td>150</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>120</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td>R5</td>
<td>300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R6</td>
<td>250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7</td>
<td>30</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>R8</td>
<td>150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R9</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R10</td>
<td>30</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

7. Dependency Graph

- Rules can be overlapped in the matching space – Multiple matches for a packet

8. Top-N Approximation

- An approximation table is pre-computed for all the possible combinations of scenarios that two overlapping rules both depended on by a target rule

9. Approximation Tables

- Three types of overlap between two rules in 1-dimension

10. Evaluation

- Uniformly distributed hit-rates
- Exponentially distributed hit-rates

11. Conclusion

- Proposed a greedy heuristic algorithm for top-N selection
- Optimize for higher cumulative hit-rates in sub-ruleset
- Either resolves or includes dependencies for correctness
- Hit-rate statistics are simple and readily available
- Preliminary simulations show reasonably close to optimal
- Runtime is much faster than brute-force approach