Scalable Management of Enterprise and Data Center Networks

Jennifer Rexford
jrex@cs.princeton.edu
Princeton University

Joint work with Minlan Yu, Mike Freedman, Jia Wang
Large Enterprise Networks

- Switches (1K - 5K)
- Hosts (10K - 100K)
- Applications (100 - 1K)
- Skype
- YouTube
Large Data Center Networks

Switches
(1K - 10K)

Servers and VMs
(100K)

Applications
(100 - 1K)
Considerations:
- Performance
- Security
- Mobility
- Energy-saving
- Cost reduction
- Debugging
- Maintenance

Flexible Policies

Customized Routing

Measurement Diagnosis

Access Control
Scaling Edge Network Management

- Large Networks (hosts, switches, apps)
- Simple Switches (cost, energy)
- Flexible Policies (routing, security, measurement)
Traditional Network

Management plane:
offline, sometimes manual

Control plane:
Hard to manage

Data plane:
Limited policies

New trend: Flow-based switches & logically centralized control
Data plane: Flow-based Switches

- Rules that perform simple actions on packets
  - Match: Match on bits in the packet header
  - Actions: Drop, forward, count
  - Store rules in high speed memory (TCAM)
Control Plane: Logically Centralized

DIFANE: A scalable way to apply fine-grained policies

RCP [NSDI’05], 4D [CCR’05], Ethane [SIGCOMM’07], NOX [CCR’08], Onix [OSDI’10], Software defined networking
Pre-install Rules in Switches

- Problems:
  - No host mobility support
  - Switches do not have enough memory
Install Rules on Demand (Ethane)

- **First packet misses the rules**
- **Buffer and send packet header to the controller**
- **Forward**
- **Install rules**

**Problems:**
- Delay of going through the controller
- Switch complexity
- Misbehaving hosts
DIFANE architecture (two stages)
Stage 1

The controller *proactively* generates the rules and *distributes* them to authority switches.
Partition and Distribute the Flow Rules

Controller

Ingress Switch

Authority Switch A

Authority Switch B

Authority Switch C

Flow space

Distribute partition information

accept

reject

Egress Switch

Authority Switch A

Authority Switch B

Authority Switch C
Stage 2

The authority switches keep packets always in the data plane and reactively cache rules.
PacketRedirection and Rule Caching

Ingress Switch

Egress Switch

Authority Switch

First packet

Following packets

Feedback: Cache rules

Redirect

Forward

Hit cached rules and forward

A slightly longer path in the data plane is faster than going through the control plane
Locate Authority Switches

• Partition information in ingress switches
  – Using a small set of coarse-grained wildcard rules
  – ... to locate the authority switch for each packet

• A distributed directory service of rules
  – Hashing does not work for wildcards
Packet Redirection and Rule Caching

Ingress Switch

First packet

Following packets

Cache Rules

Partition Rules

Redirect

Authority Switch

Feedback: Cache rules

Forward

Egress Switch

Hit cached rules and forward
### Three Sets of Rules in TCAM

<table>
<thead>
<tr>
<th>Type</th>
<th>Priority</th>
<th>Field 1</th>
<th>Field 2</th>
<th>Action</th>
<th>Timeout</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cache Rules</strong></td>
<td>1</td>
<td>00**</td>
<td>111*</td>
<td>Forward to Switch B</td>
<td>10 sec</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>00**</td>
<td>001*</td>
<td>Forward to Switch B</td>
<td>Infinity</td>
</tr>
<tr>
<td><strong>Authority Rules</strong></td>
<td>109</td>
<td>0***</td>
<td>000*</td>
<td>Redirect to auth. switch</td>
<td></td>
</tr>
<tr>
<td><strong>Partition Rules</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Cache Rules**
  - In ingress switches *reactively* installed by authority switches

- **Authority Rules**
  - In authority switches *proactively* installed by controller

- **Partition Rules**
  - In every switch *proactively* installed by controller
DIFANE Switch Prototype

Built with OpenFlow switch

Cache Rules

DIFANE Switch Prototype

Built with OpenFlow switch

Cache Rules

Recv Cache Updates

Send Cache Updates

Only in Auth. Switches

Control Plane

Data Plane

Authority Rules

ParTTon Rules

Notification

Just software modification for authority switches
Caching Wildcard Rules

- Overlapping wildcard rules
  - Cannot simply cache matching rules

Priority: R1>R2>R3>R4
Caching Wildcard Rules

- **Multiple authority switches**
  - Contain independent sets of rules
  - Avoid cache conflicts in ingress switch
Partition Wildcard Rules

• Partition rules
  – Minimize the TCAM entries in switches
  – Decision-tree based rule partition algorithm

Cut B is better than Cut A
Peak Throughput

- One authority switch; First Packet of each flow

DIFANE is self-scaling: Higher throughput with more authority switches.
Summary: DIFANE in the Sweet Spot

Distributed

Traditional network (Hard to manage)

Logically-centralized

OpenFlow/Ethane (Not scalable)

DIFANE: Scalable management
Controller is still in charge
Switches host a distributed directory of the rules