Scaling a Broadcast Domain of Ethernet: Extensible Transparent Filter using SDN

Kotaro Kataoka, Nitin Agarwal, Aditya V Kamath

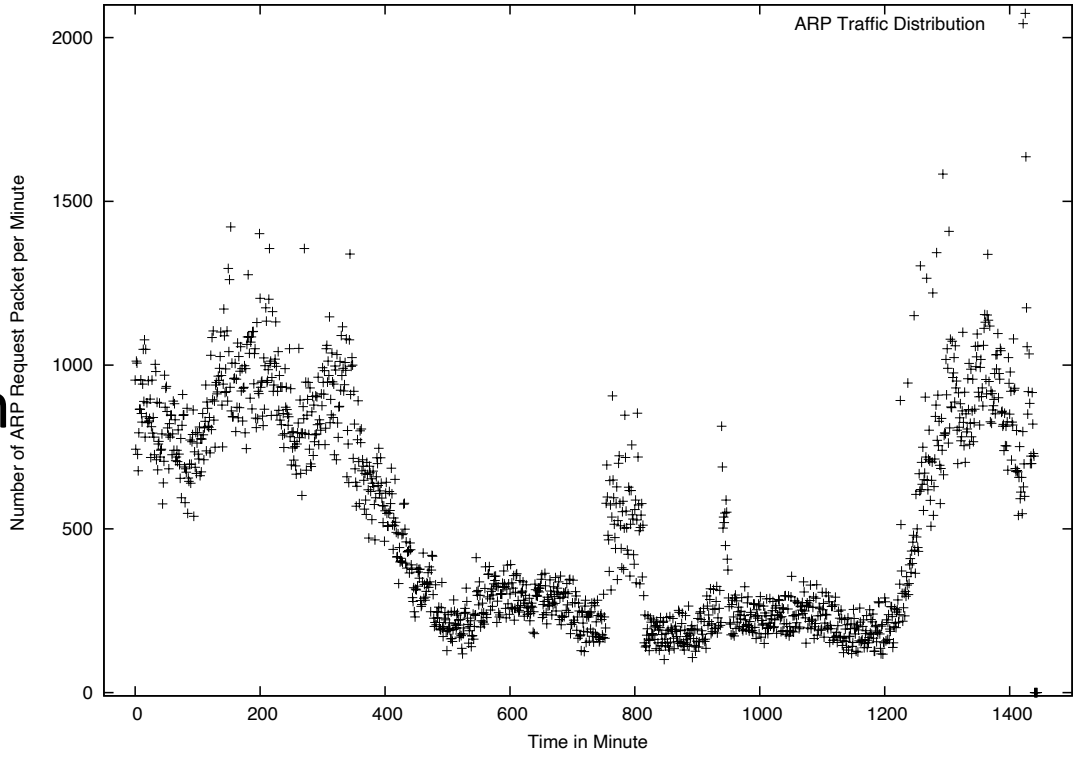
Indian Institute of Technology Hyderabad

Introduction

- Less L2 segments and L3 subnets mean simplicity
- IPv6 (big prefix) is coming into the picture
- ARP, DHCP and other discovery services use broadcast packets, but it’s noisy
- Broadcast in Wi-Fi does not use RTS/CTS and makes network even noisier
- Small chunks of network will reduce the impact of broadcast the available number of routed subnet will be subject to the performance of switches or routers

How many broadcast messages?

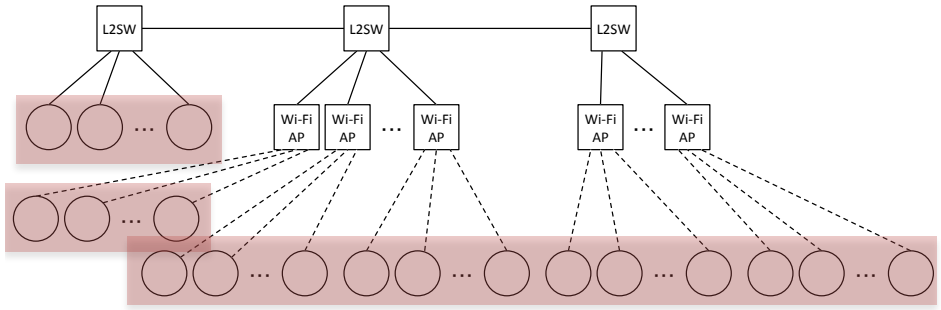
- Number of ARP Requests in /21 subnets in IITH
  - Avg. 515 ppm
  - Max. 2074 ppm
  - 595 unique IP addresses
- 3 most popular targets in the same subnet
  - WEB Proxy
  - Default Gateway
  - DNS Server



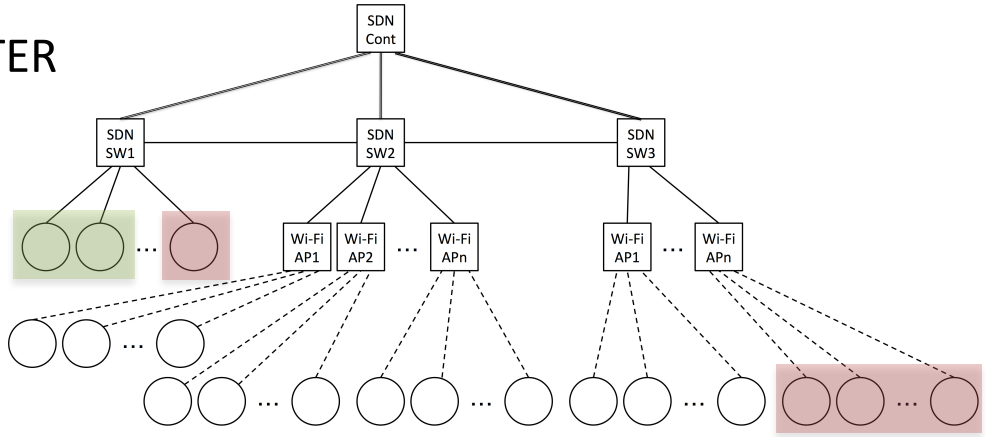
Proposal

- Isolate the impact of broadcast within only one switch port or collision domain

BEFORE

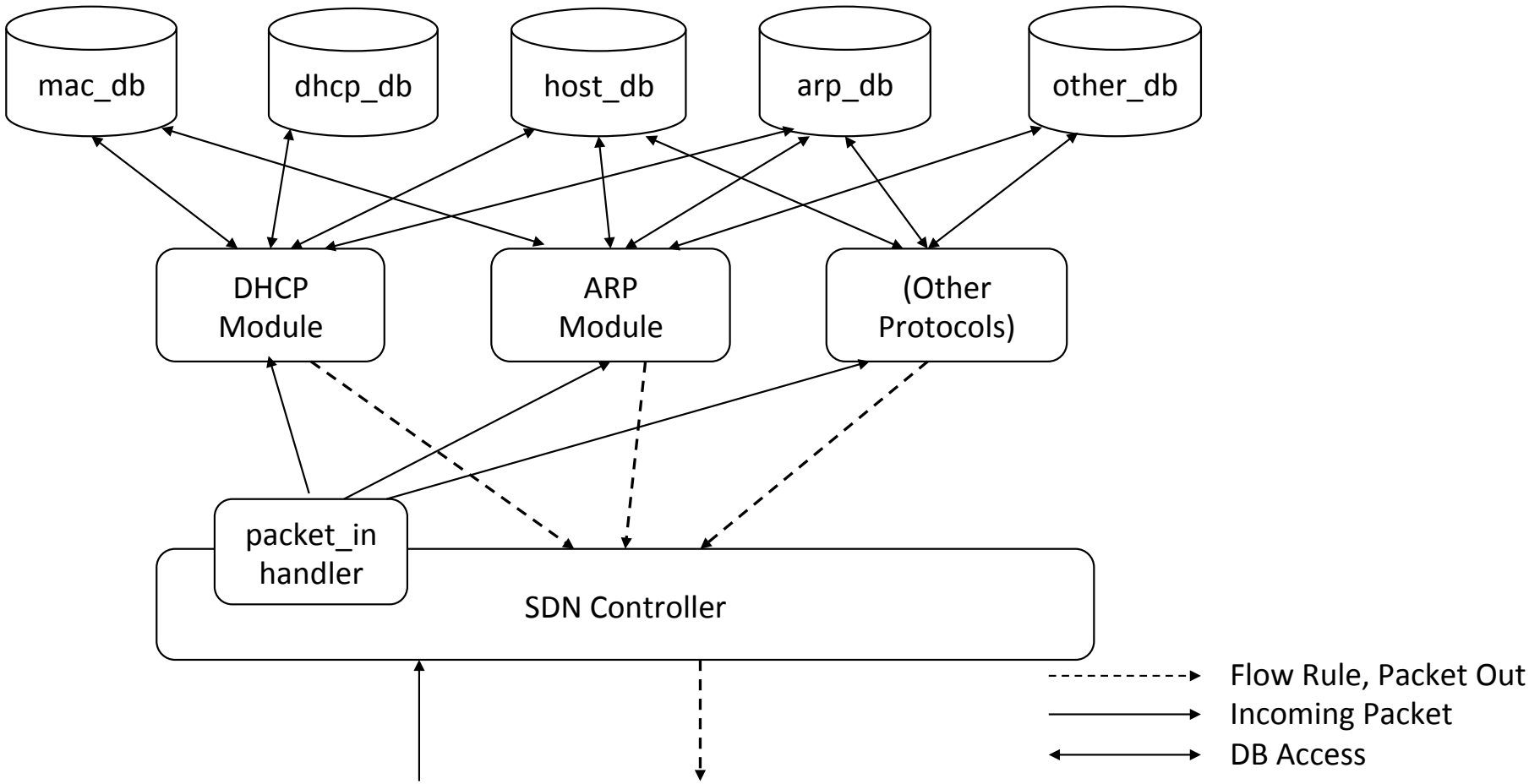


AFTER

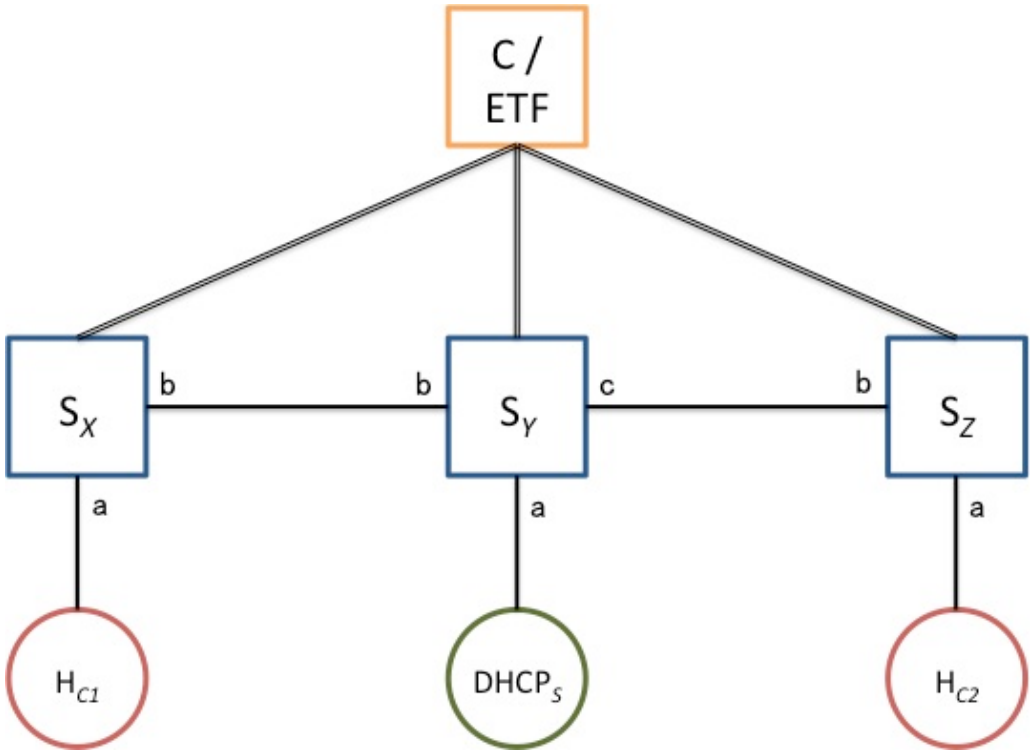


- Let ARP and DHCP behave as they do
  - ARP should be handled by end systems
  - DHCP can do much more than configuring subnet, host address, default gateway and DNS server
- SDN switch/controller learn the host from its very fist packet
  - Data Path ID (Switch), Port No., and MAC address

Extensible Transparent Filter

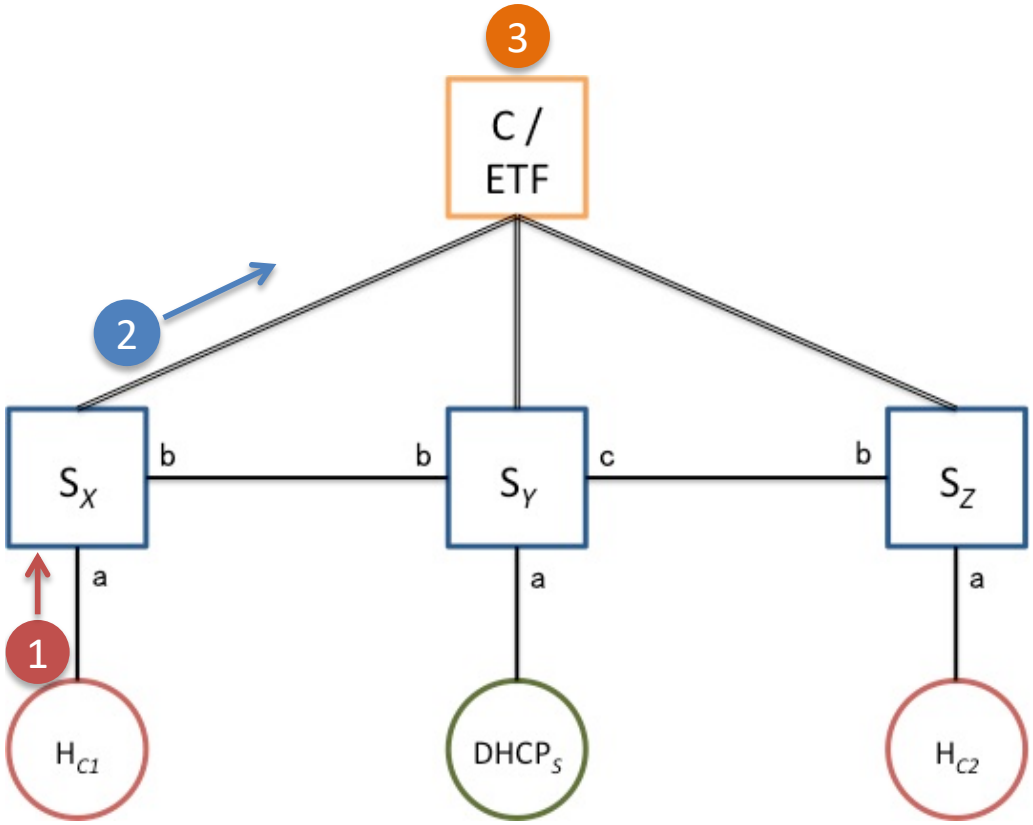


How does DHCP work?



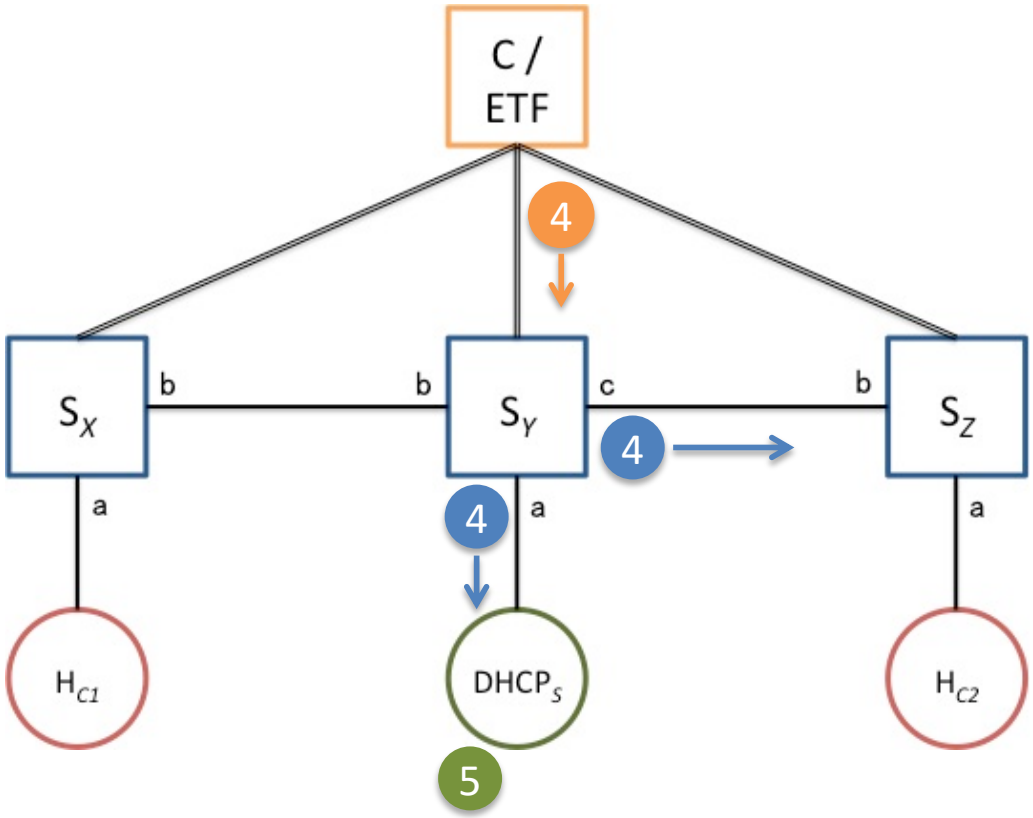
- HC1 wants network configuration
- ETF does not know both of HC1 and DHCP Server
- DHCP Server is up and running
- All the process is kicked by a client

Discovering a DHCP Server (1/3)



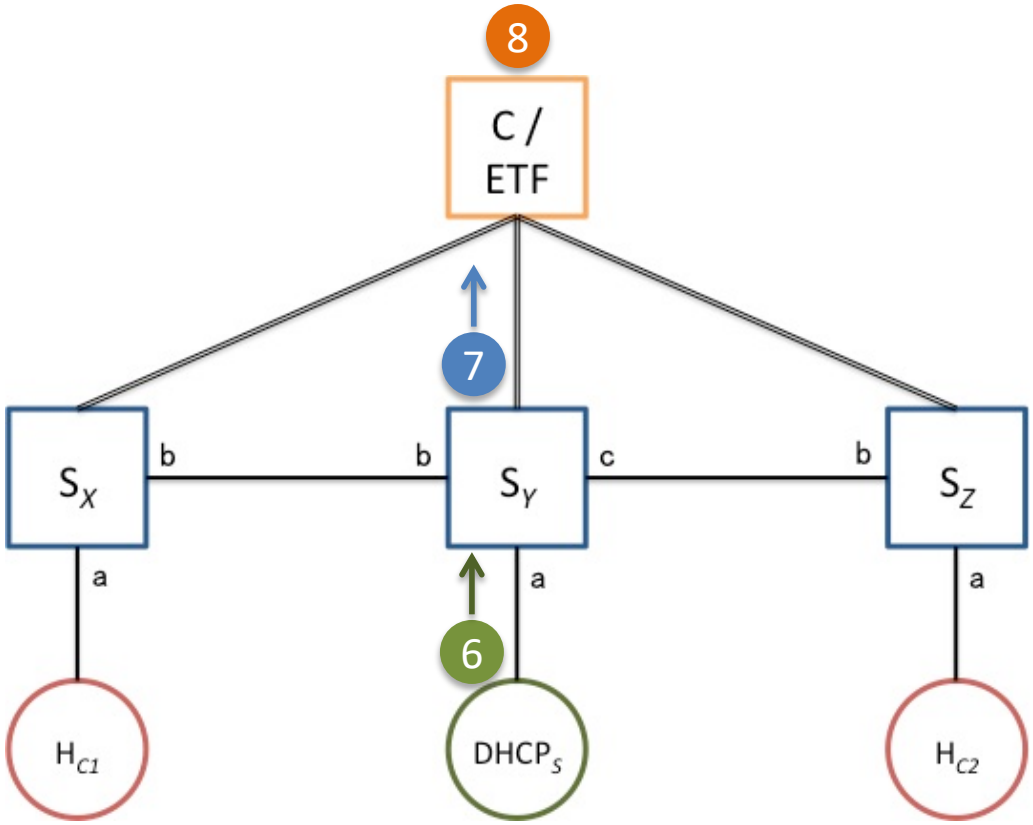
1. HC1 sends DHCP Discover message
2. S_X receives DHCP Discover message and forwards the packet to the ETF
3. ETF does something
   a. Creates an entry for H_{C1} in host_db except IP address
   b. arp_db is kept untouched
   c. Looks up dhcp_db to find the DHCP servers, but no entry is found

Discovering a DHCP Server (2/3)



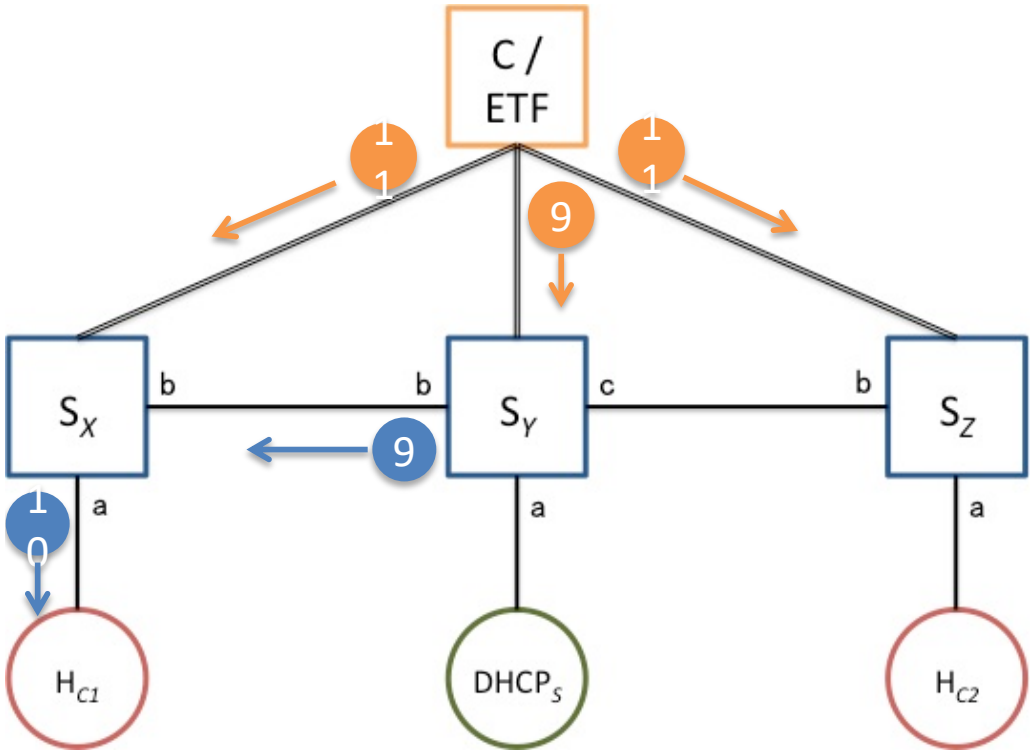
4. ETF sends a packet_out message to each of the switches for flooding the packet
   - This flooding procedure happens only if no DHCP server is known to ETF
5. DHCP server receives DHCP Discover message

Discovering a DHCP Server (3/3)



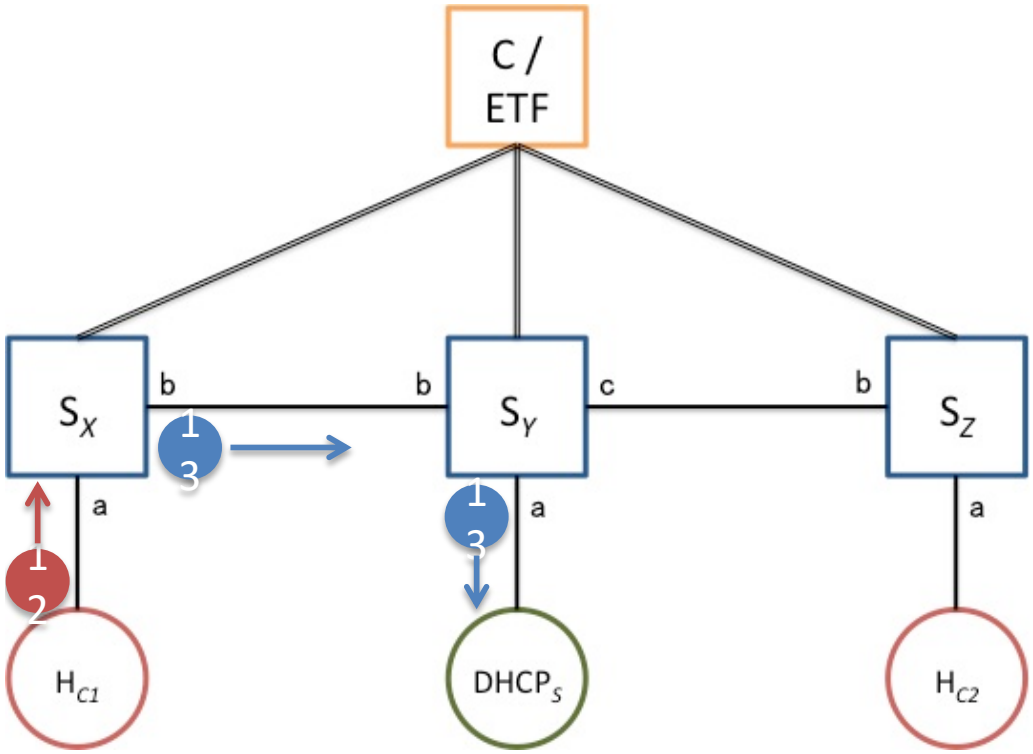
6. DHCPS sends unicast DHCP Offer
7. SY captures the packet and forwards it to ETF
8. ETF discovers DHCP Server and updates dhcp_db, host_db and arp_db to record that DHCP Server is connected to SY on port a

Operating DHCP in Suppress Mode (1/3)



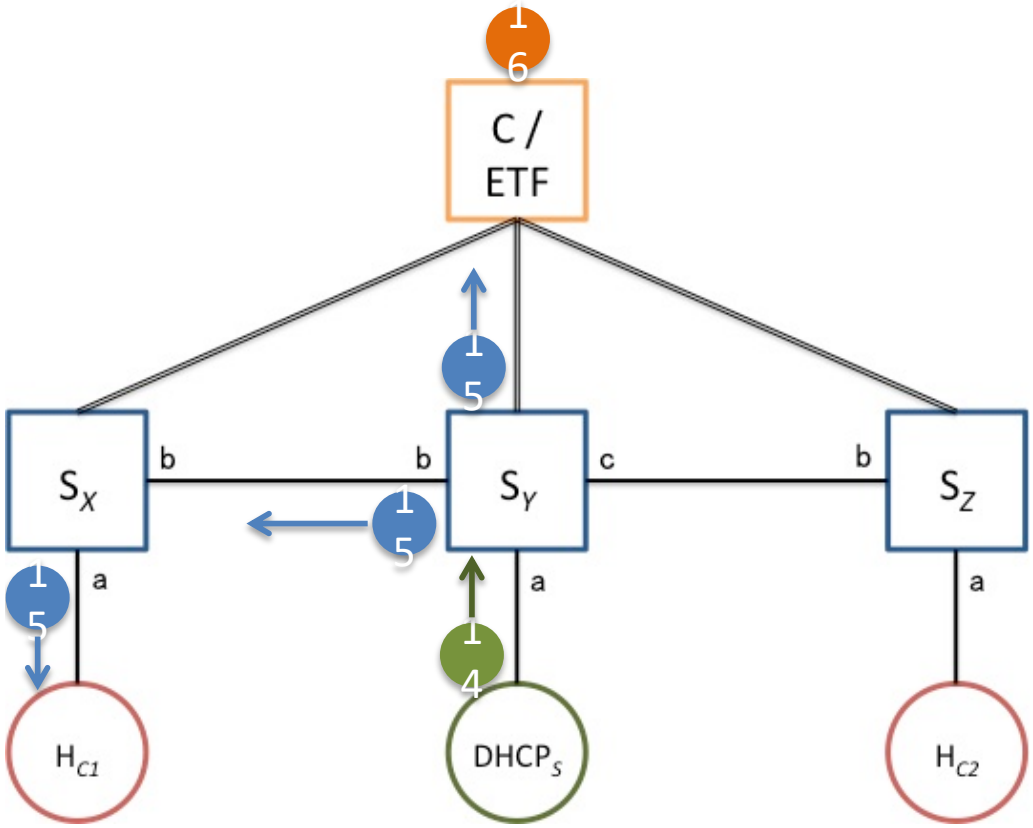
9. ETF gives SY a flow rule to forward DHCP Offer to the port b
10. SX also forwards the message to port a via which HC1 is connected
11. ETF installs proactive flow rule for DHCP Server on all the switches.

Operating DHCP in Suppress Mode (2/3)



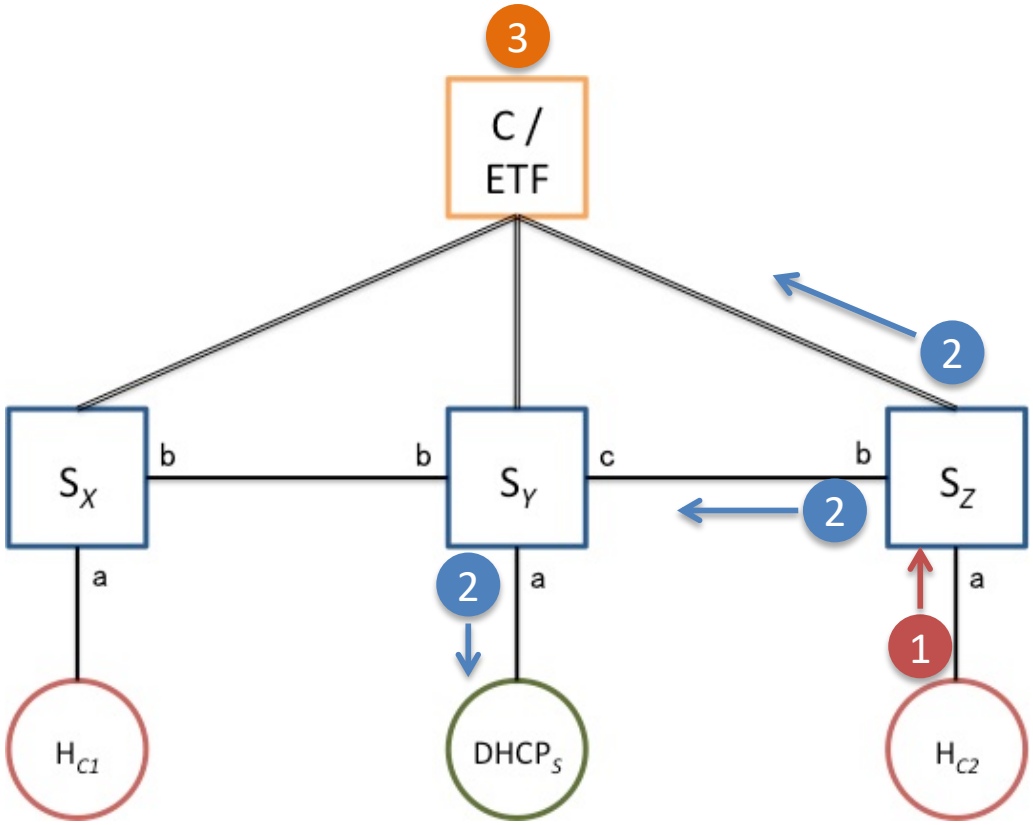
12. HC1 broadcasts DHCP Request to inform all DHCP servers that it takes IP address offered by a specific DHCP server.
13. Now SX and SY have the flow rule for DHCP broadcast packet so it will be forwarded to only the selected port though which the packet can reach DHCP Server

Operating DHCP in Suppress Mode (3/3)



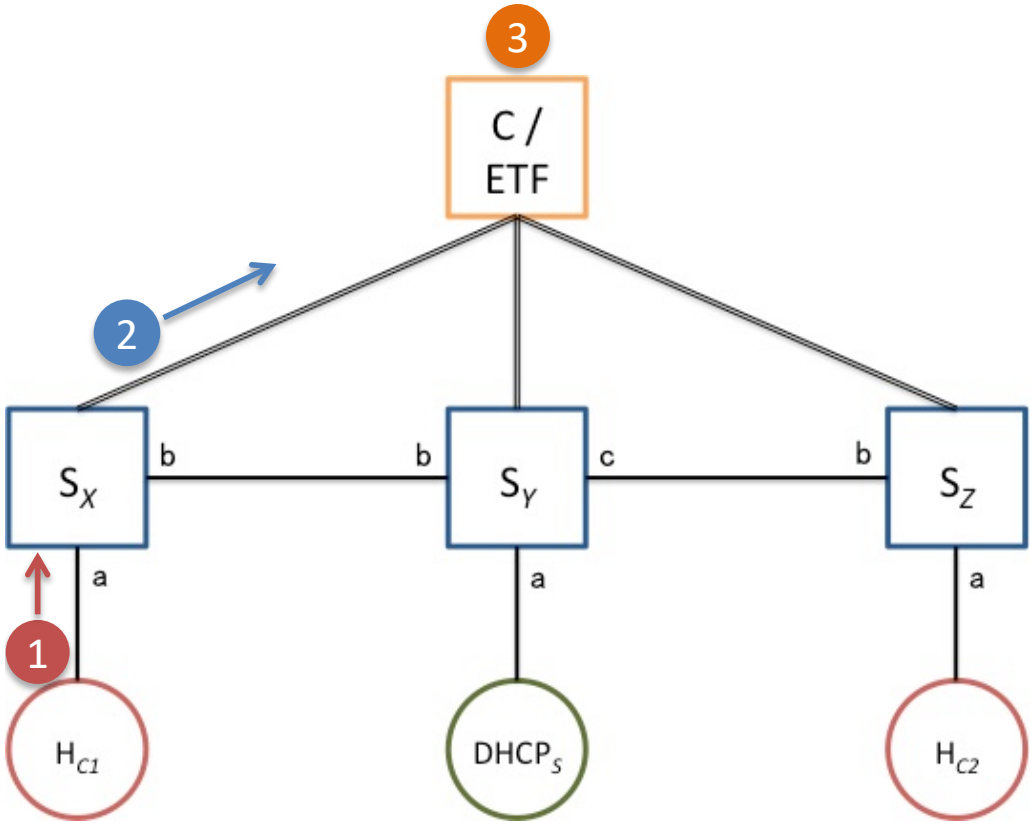
14. DHCP Server sends DHCP Ack to confirm the assignment of IP address to HC 1
15. SY and SX forward the DHCP Ack using the same flow rules that were installed for DHCP Offer
16. ETF updates the IP address of HC1 in host_db and also inserts an ARP entry in arp_db

Operating DHCP for a New Host in Suppress Mode



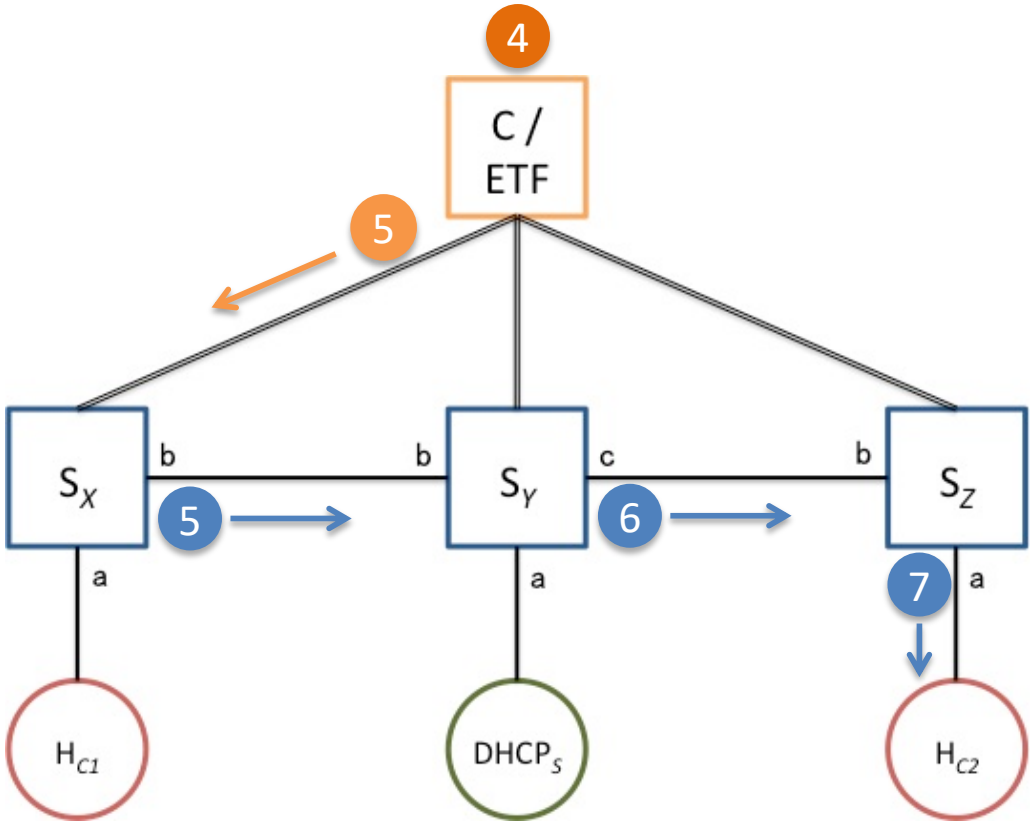
1. A new host HC2 broadcasts a DHCP Discover message
2. DHCP Discover message is sent via the port b on SZ without flooding to any other port and SY forwards it to DHCP Server
3. ETF learns about HC2 using the same process as it learned HC1

Operating ARP in Suppress Mode (1/3)



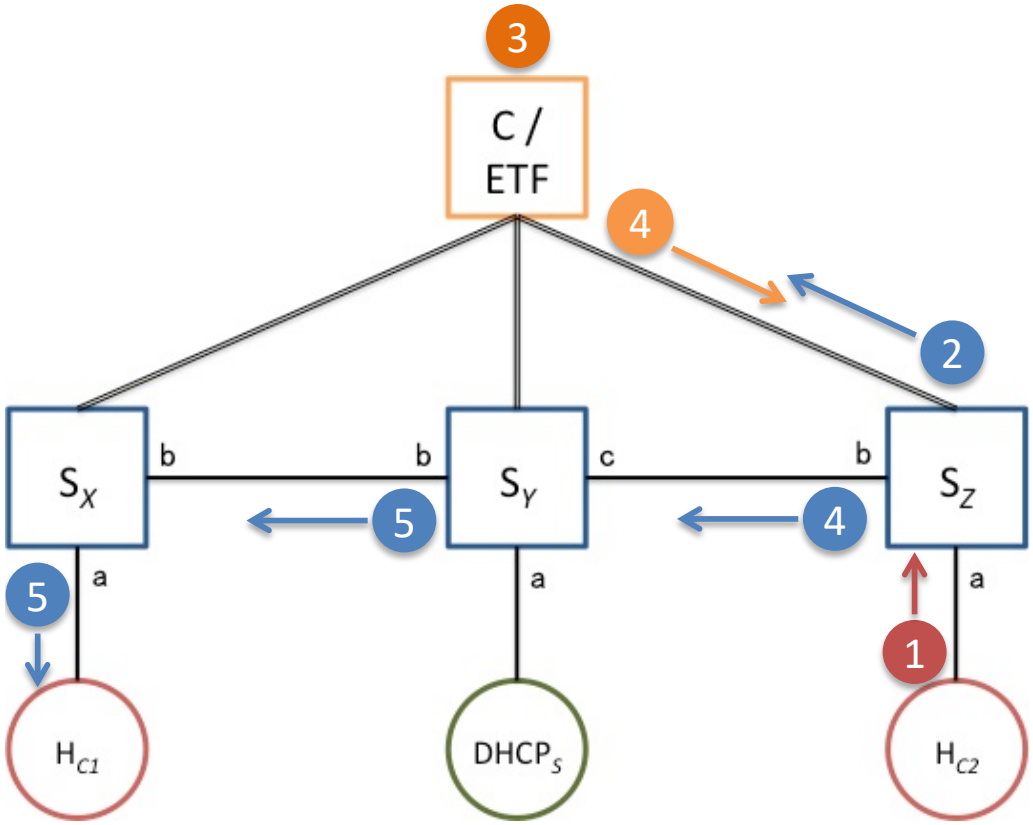
1. HC1 broadcasts ARP Request to look up the MAC address of HC2
2. SX receives the ARP Request and forwards it to ETF
3. ETF does something
   a. Looks up HC1 and HC2, that are the source and the target of ARP Request respectively, in host db and arp db
   b. Gets the MAC address of HC2

Operating ARP in Suppress Mode (2/3)



4. ETF has already learned the MAC address of HC2 in the mac table of each switch
5. ETF gives SX a flow rule to forward such a packet via port b and also to ETF itself
6. SY and SZ also go through the same process as SX
7. HC2 receives the ARP Request

Operating ARP in Suppress Mode (3/3)



1. ARP Reply is a unicast packet back from HC2 to HC1
2. SZ forwards ARP Reply to ETF
3. ETF looks up the mac table of switches to determine the output port
4. ETF gives SZ a flow rule to forward the packet via port b
5. SY and SX will also go through the same process as SZ

When does Flooding happens for ARP?

- ARP works in suppress mode for the end systems which get IP configuration using DHCP

- ETF does not have information of end systems that use static IP addresses
  - When such an static IP address is the target of ARP Request, ETF receives the packet and learns about the source host.
  - ETF sends packet_out message to switches for flooding the packet

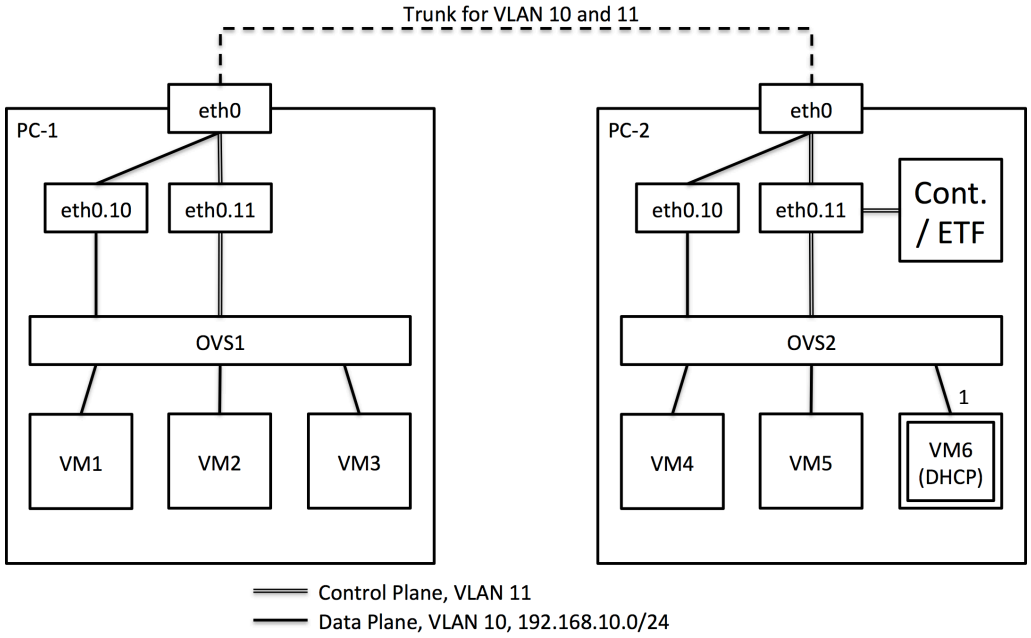
Performance Factors

- Configuration Parameters (such as idle timeout)
  - To maintain the size of flow table and frequency of packet_in
- Delay in processing the broadcast messages
  - Proactive Flow Rule will help
- ARP Request to the non-existent host
  - Rate limiter of ARP messages
- Gratuitous ARP Requests
  - We need IP Conflict Detection and avoidance

Proof of Concept

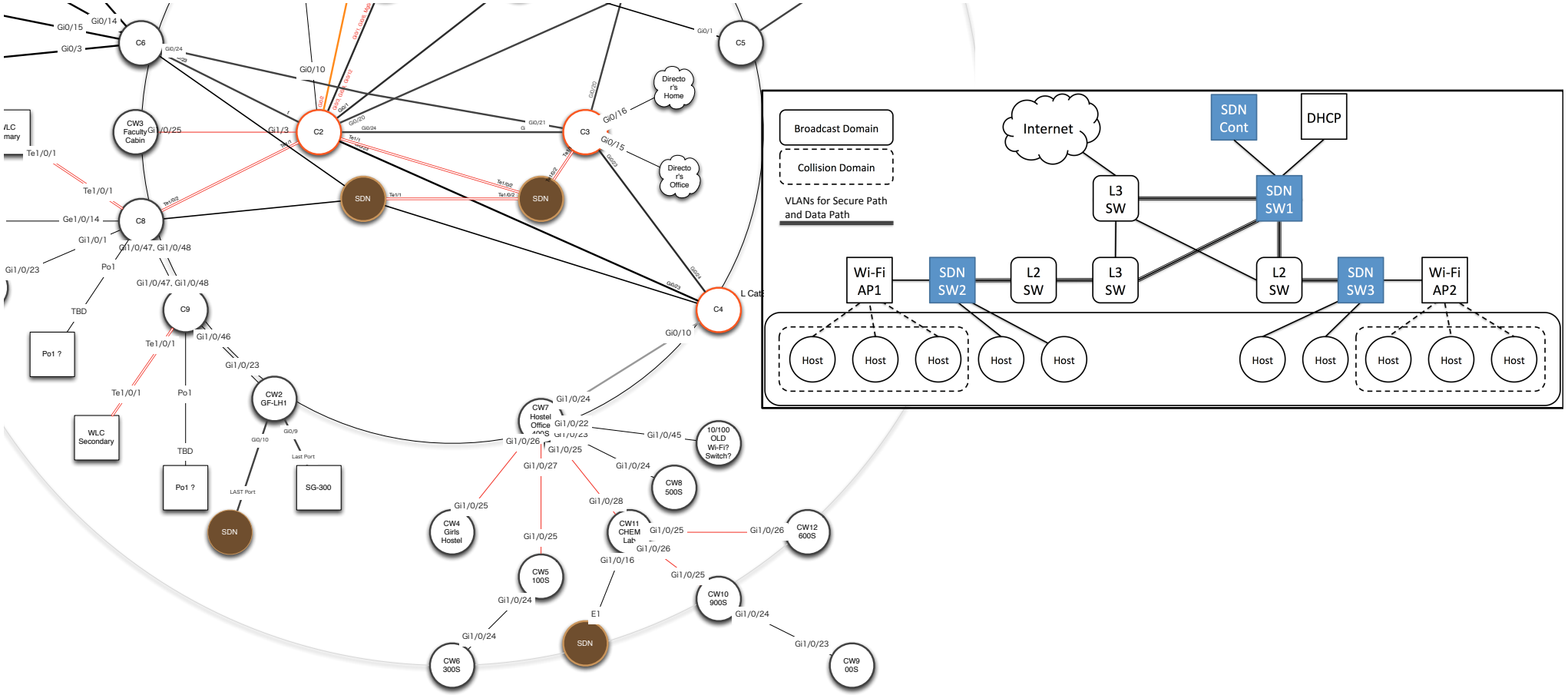
- Suppress Mode DHCP and ARP
- Proactive Flow Rules
  - For ARP
  - Inject the flow rule of return path together with the one of forward path
- Mobility
  - Renew the database once the existing host appears on a different switch / port
- Garbage Collection
  - Expiry of database entry
- Rate limiter
  - ARP request to non-existent host?
  - Currently flooding with managed tx interval

- Softwared-based Implementation
  - SDN Controller: Trema
  - SDN Switch: Open vSwitch
  - OpenFlow: Version 1.0



SDN Deployment in IIT Hyderabad

- SDN switches are deployed at core and edges of the campus network
  - Overlaying SDN switches on traditional LAN
- IITH network is ready for testing developed codes
- Hardware: HP 3800-24G-Poe+-2SFP+, HP 5400zl with 8 x 10Gbe Ports



Deployment in Campus LAN using Real SDN Switches (1/2)

- SDN Controller

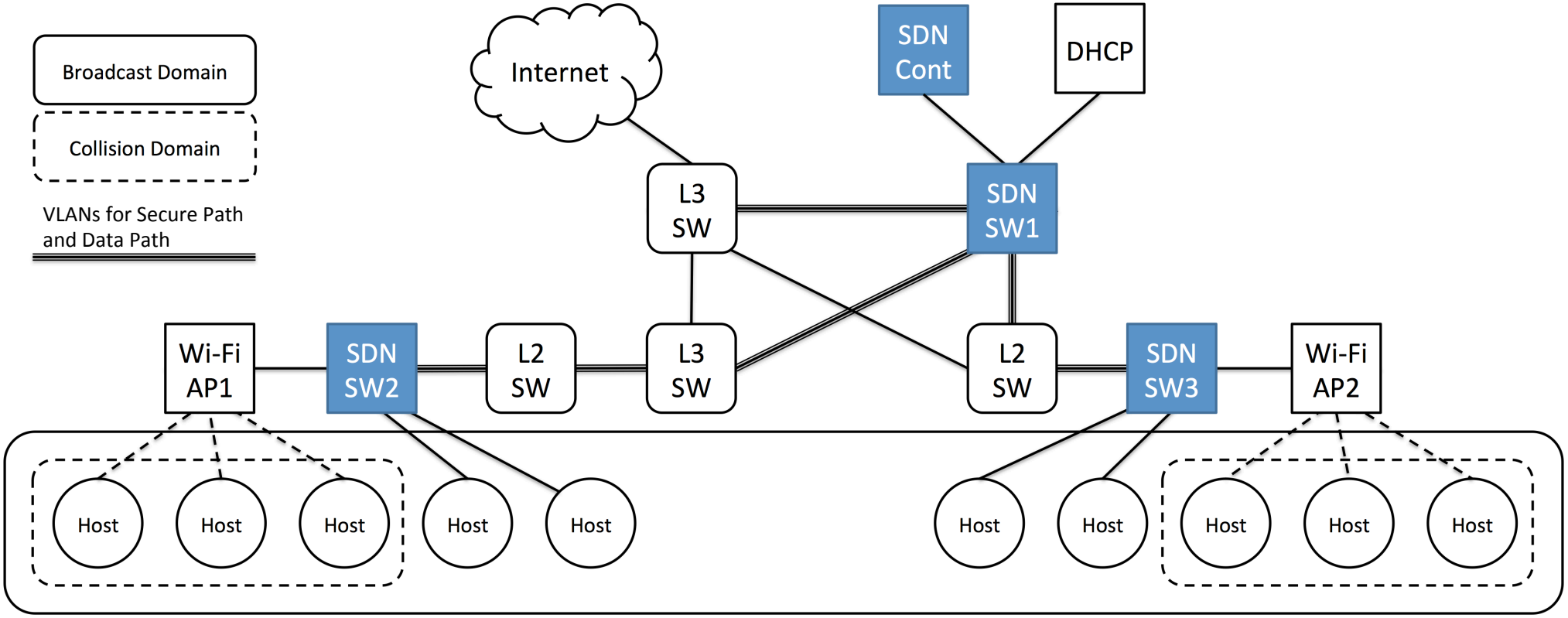
Specification	
Operating System	Ubuntu 12.04 LTS
CPU	Intel Core i7 L620 2.0 GHz
RAM	8GB
NIC	Intel 82577LM Gigabit Ethernet Network Connection
SDN Controller	Floodlight 0.90

- SDN Switch

Specification	
Model	HP J9573A 3800-24G-Poe+-2SFP+
Firmware	KA.15.13.0005

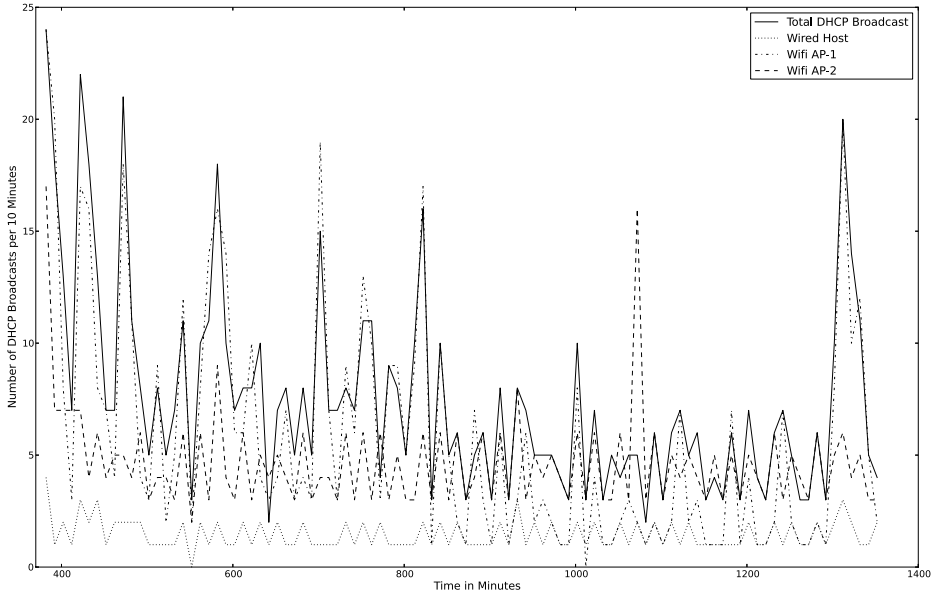
Deployment in Campus LAN using Real SDN Switches (2/2)

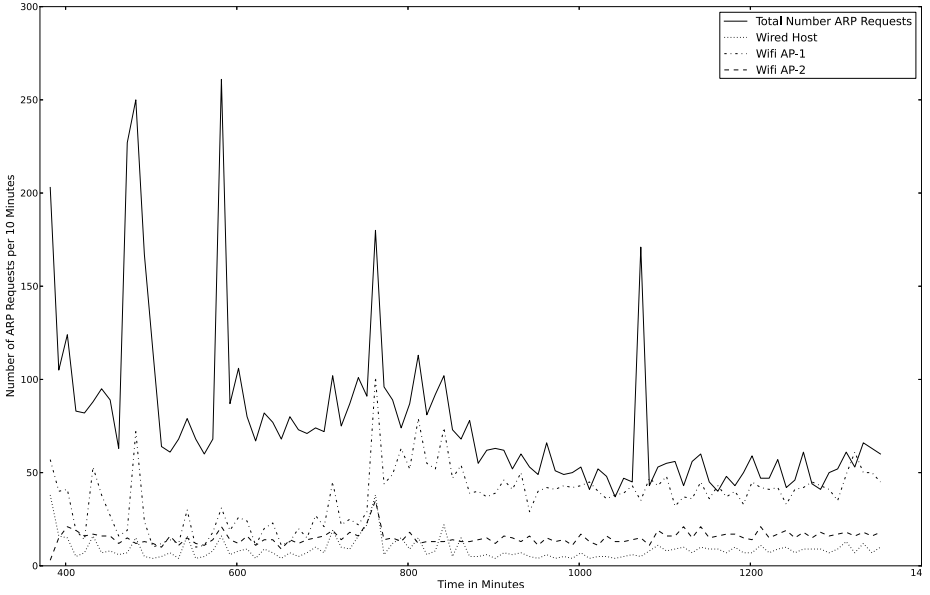
- Two VLANs for SDN (Control and Data) across the campus network
- SDN-enabled CSE Lab. served by Wi-Fi and Wired



Evaluation

- Broadcast of DHCP and ARP was reduced by more than 80%. Wired network works ideally, but Wi-Fi network showed less performance than expectation.
  - Testing result exhibited good isolation of broadcast.
  - Queries for non-existent host were flooded in Wi-Fi
  - Smaller number of hosts. If we have more hosts, then suppress rate will improve.





DHCP / ARP Suppression Rate

Capture Point	Total Number	Avg. PPM	Max. PPM	Suppress Rate
DHCP Server	740	0.76	8	N/A
Wired Host	143	0.147	2	80.6%
Wi-Fi Host-1	577	0.59	10	22.0%
Wi-Fi Host-2	457	0.47	13	38.0%

Table: Overall Performance for Suppressing DHCP

Capture Point	Total Number	Avg. PPM	Max. PPM	Suppress Rate
SDN Controller	7502	7.72	147	N/A
Wired Host	873	0.90	21	88.3 %
Wi-Fi Host-1	3605	3.71	55	51.9 %
Wi-Fi Host-2	1467	1.51	19	80.4 %

Table: Suppression of ARP Broadcast Messages

Related Work on Floodless LAN

- Backward compatibility, transparency, deployment

- [Myers] Replaces ARP and DHCP with alternative directory service and integrating it on switches.
- [EtherProxy] Proxies ARP and DHCP near the end host on a bridging device.
- [SEATLLE] Converts ARP and DHCP to unicast-based services. MAC address indicates switch id and host id.
- [MOOSE] Introduces separation of switch id and host id in MAC address. Enhanced Lookup replaces ARP and DHCP using proxy-based approach.
- [FDSM] Floodless Service Discovery Model compatible with MOOSE.

Conclusion and Future Work

- ETF suppress the DHCP/ARP broadcast
- Other directory services can also be suppressed but we need to look into the message format carefully
- ETF deployment for 10+ Wi-Fi AP is being planned in the new campus of IITH
- Are switches performing good? We have to see them more… (If flow rule is handled in RAM, i.e. software, it means slow.)