IP Address Reuse for Multi-tenants Data Center

Version2 Apr19, 2011

Cheng-Chun Tu

Problem Statement

- When a physical data center is used to support multiple virtual data centers (VDC), how does it give each VDC the same private IP address range?
- Why?
  - No network reconfiguration required when a user moves parts of its IT infrastructure between private and public clouds in a hybrid cloud environment

Solution Overview

- To preserve the private IP addresses from each VDC, apply another unique identifier: VDC ID.
- To locate a host, redirect ARP request with VDC id to a centrally managed server
- Forward packets at an all layer 2 network
- To improve performance, apply local cache at physical machine

Architectural Assumptions

- A VDC consists of multiple VMs that reside on multiple PMs
- A Layer-2-only data center network connects all the PMs
- No two user VMs residing on the same PM could be assigned the same private IP address

Addressing

- Each VDC is given a globally unique ID and one or multiple public IP addresses
- Because each VDC is given an identical private IP address range, each VM in a VDC is uniquely identified by the ID of its VDC concatenated by its private IP address
- A physical data center has a set of service nodes, which forms a special VDC (ID = 0), which is given a distinct service address range that is disjoint from the private and the public address range
  - The entire IP address space is decomposed into three ranges: private, public and service

MAC Address Resolution

- ARP queries are transparently intercepted and relayed to a centralized ARP server, which answers all ARP queries
  - An ARP query contains the VDC ID and the IP address of the host whose MAC address is to be resolved
- Each PM maintains a VMmap data structure that records, for each VM on the PM, its VDC ID, IP address and MAC address
- A VM in a VDC can communicate with
  - Another VM X in the same VDC using X's private IP address
  - A VM X in another VDC or a node X on the internet using X's public IP address
  - A service node X using X's service IP address

Classification of Nodes and IP addresses

- IP address space: private, public, service
- Private node includes:
  - VMs running in PM
  - Private IP or public IP

- Service node includes:
  - Dom0 on each PM
  - Service nodes are configured as service IP

Identify Destination's VDC ID

- Key Issue: How to identify the VDC ID of the target host in an ARP query based on the querying host's address?

1. Private → Private

   Use the source's VDCID as the destination's VDCID

2. Private/Service → Service

   Use 0 as the destination's VDCID

3. Service → Private

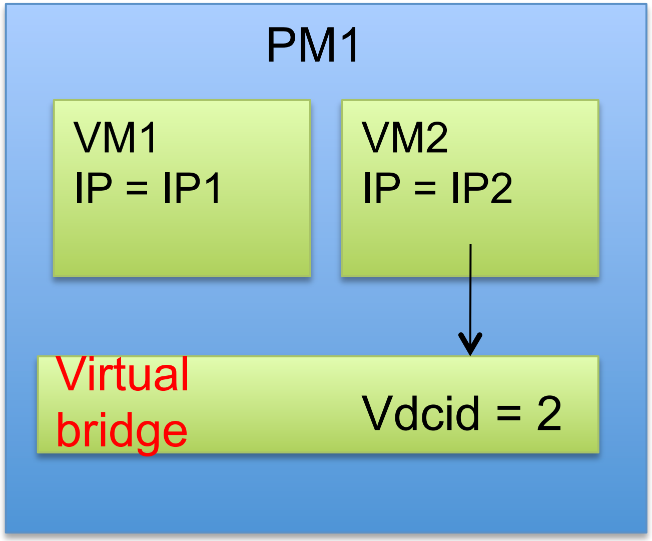
   Case1: Connection initiated from Service node

   Case2: Connection initiated from Private node

   Case3: SLB as service node

Private to Private: case1

Case 1: Destination is on the same PM, same VDC
Example: VM2 wants to talk to VM1, which is on the same PM
With vdcid = 2



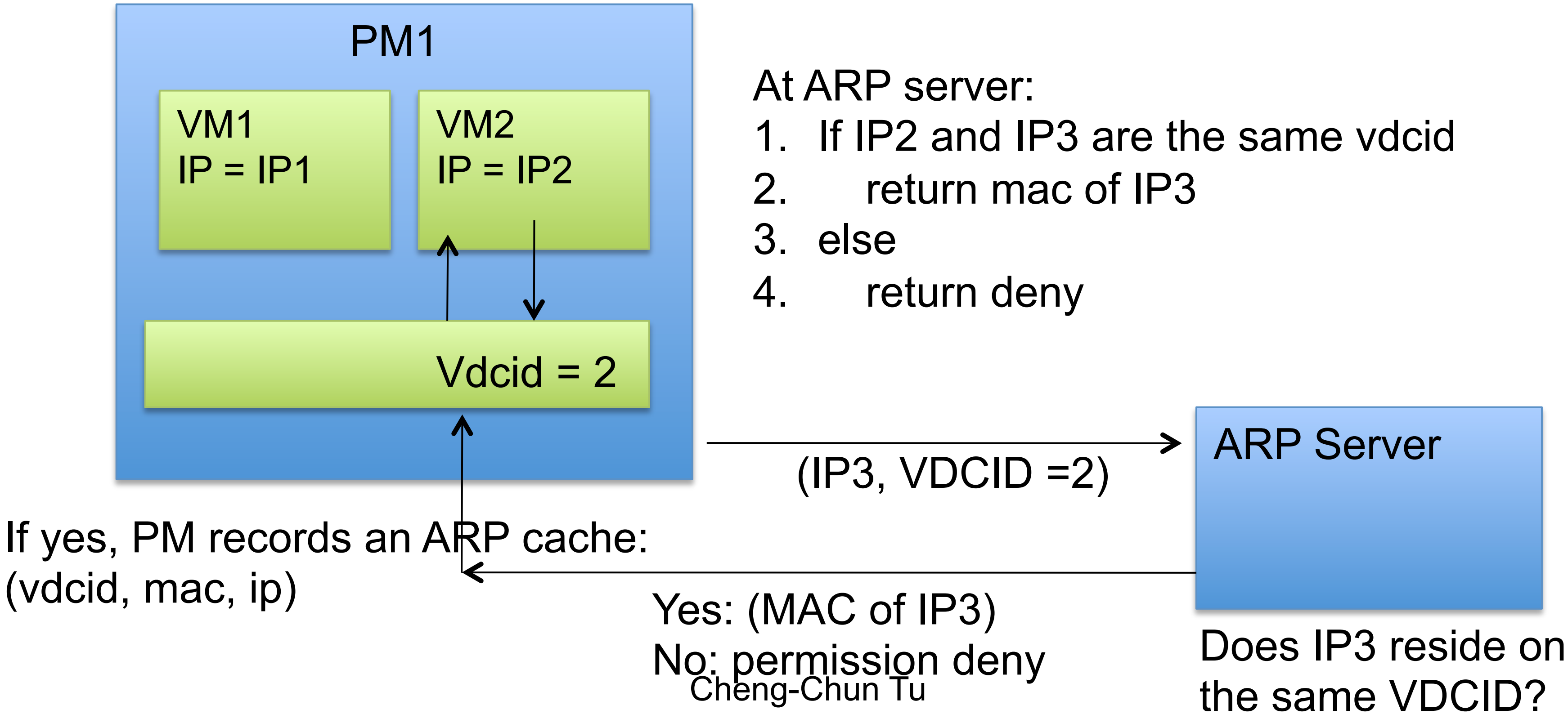
```
1.  if IP1 is on the same PM
2.      if IP1 and IP2 are the same vdcid
3.          transmit
4.  else
5.      redirect ARP to server
```

vm IP	Vdcid	Vm MAC
IP1	2	00:01:02:03..
…	…	…

Cheng-Chun Tu

Private to Private: case2

Case2: Destination is on another PM, within same/diff VDC
ex: VM2 wants to talk to private address IP3, which is on another PM



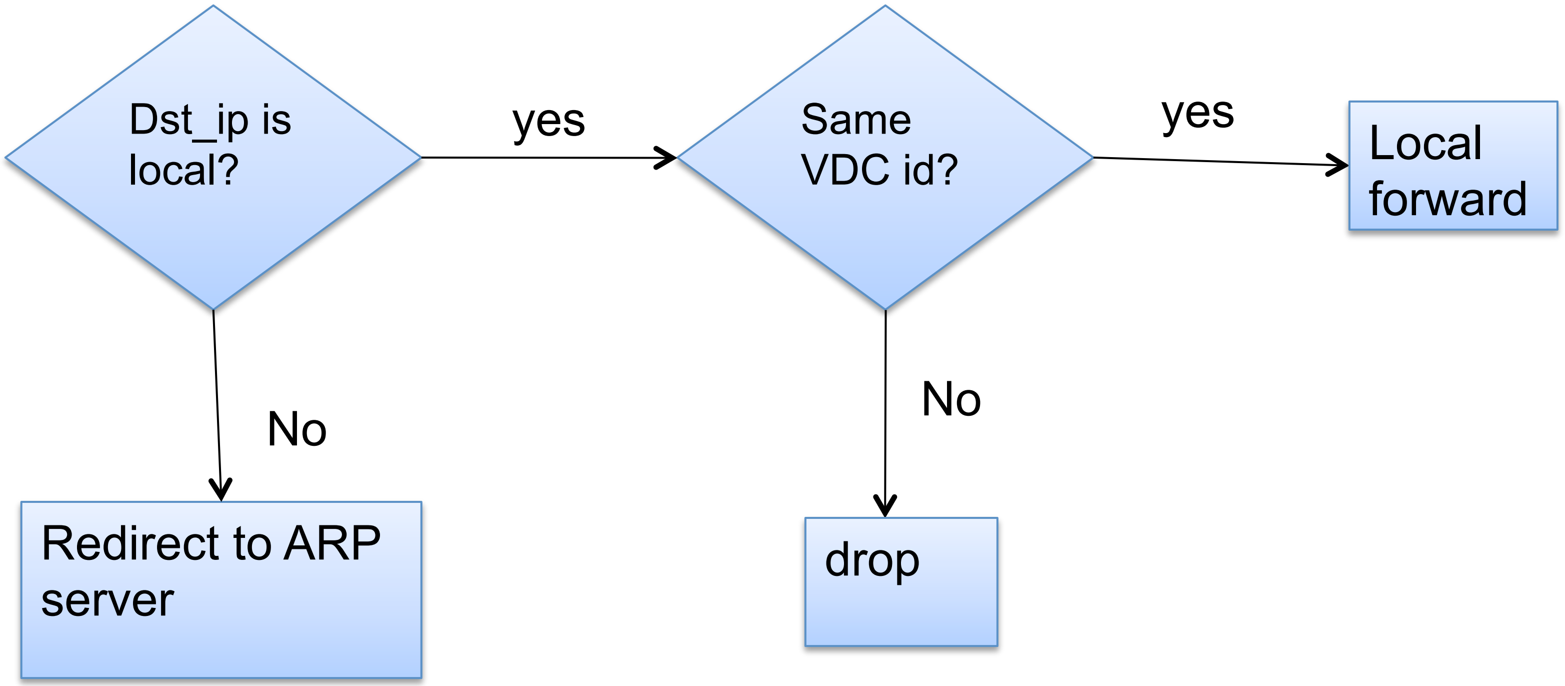
At ARP server:

```
1. If IP2 and IP3 are the same vdcid
2.     return mac of IP3
3. else
4.     return deny
```

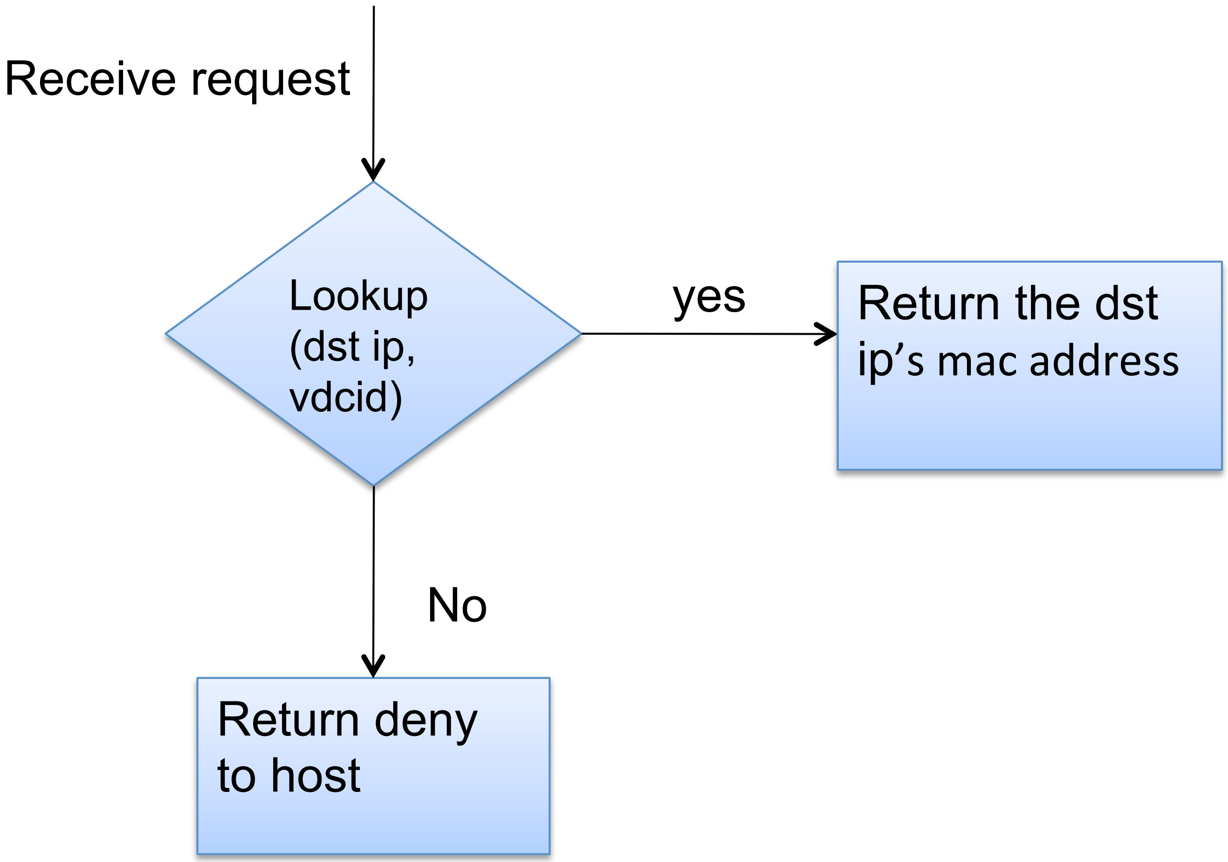
If yes, PM records an ARP cache:
(vdcid, mac, ip)

Does IP3 reside on the same VDCID?

Flow Chart: Private-to-Private Dom0



Flow Chart: Private-to-Private ARP server


Receive request

Lookup (dst ip, vdcid)

yes

Return the dst ip's mac address

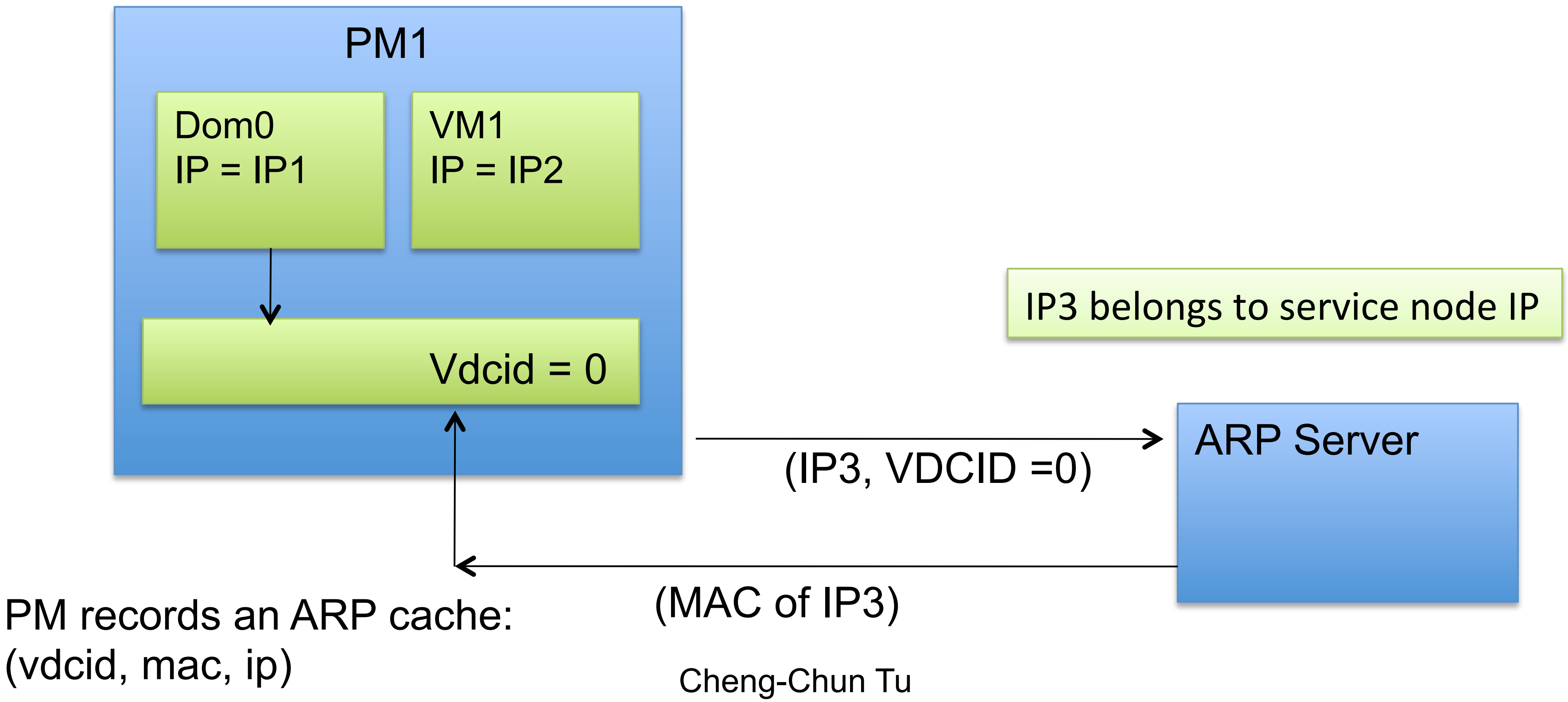
No

Return deny to host

Cheng-Chun Tu

Service node to Service node

Source and destination are both service nodes.
Ex: PM1 Dom0 wants to talk to IP3, which is a service node



PM1

Dom0
IP = IP1

VM1
IP = IP2

Vdcid = 0

IP3 belongs to service node IP

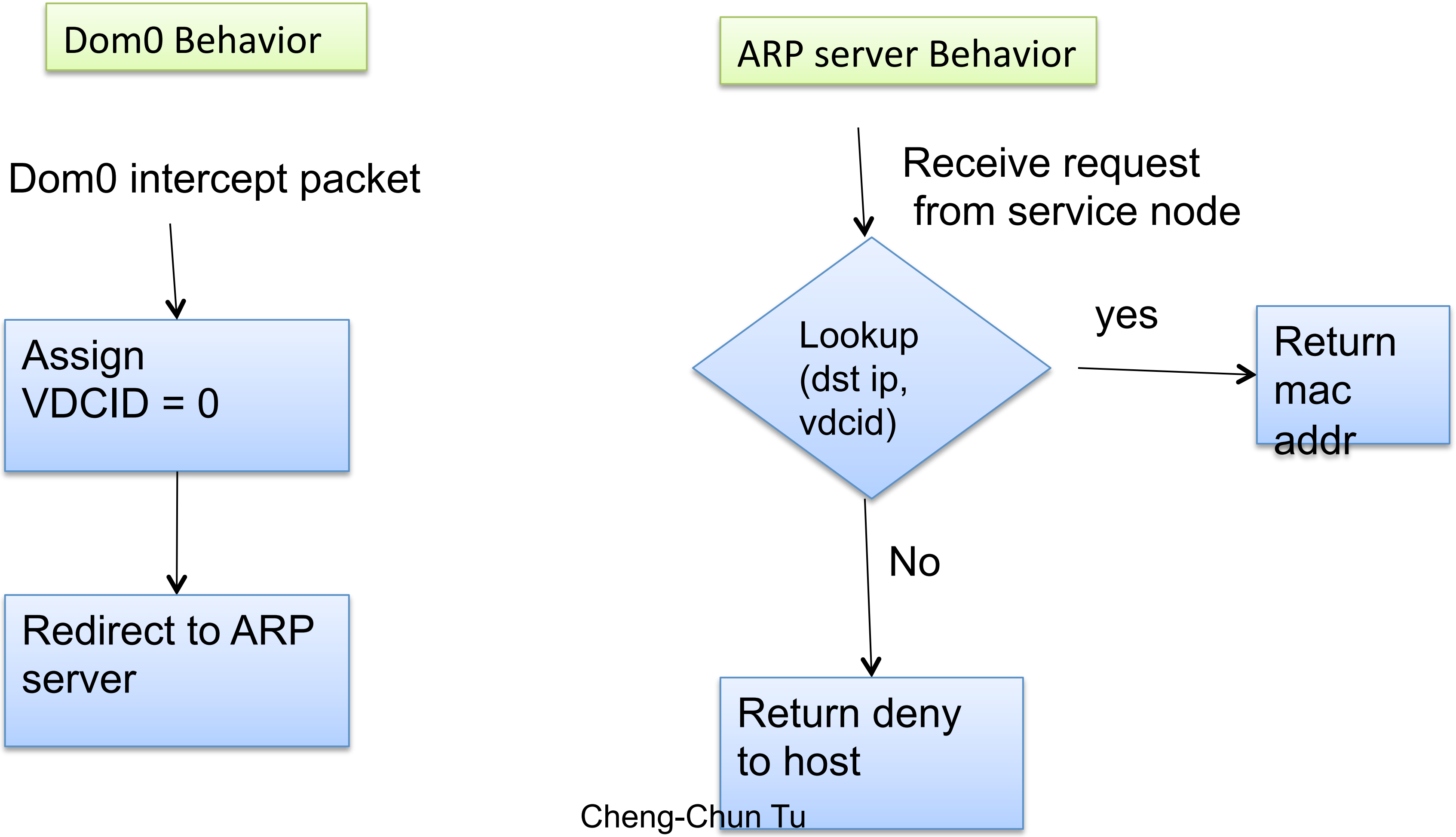
ARP Server

(IP3, VDCID =0)

(MAC of IP3)

PM records an ARP cache:
(vdcid, mac, ip)

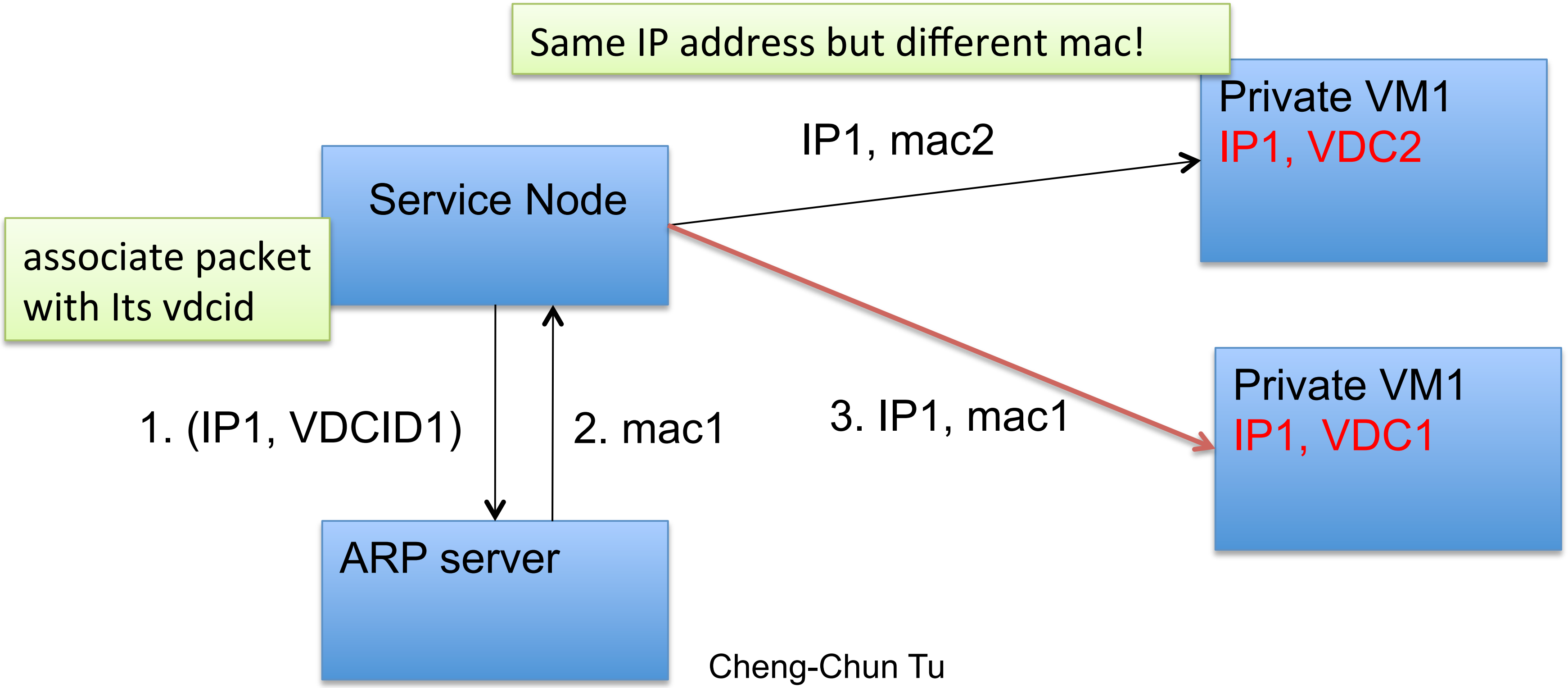
Flow Chart: Service-to-service Dom0 & ARP server



Cheng-Chun Tu

Service to Private: Problem

Two distinct private nodes on different VDC can have identical IP address. Application on the service node, such as ssh, need extra information about vdc id



Same IP address but different mac!

Service Node

associate packet with Its vdcid

IP1, mac2

Private VM1
IP1, VDC2

1. (IP1, VDCID1)

2. mac1

3. IP1, mac1

Private VM1
IP1, VDC1

ARP server

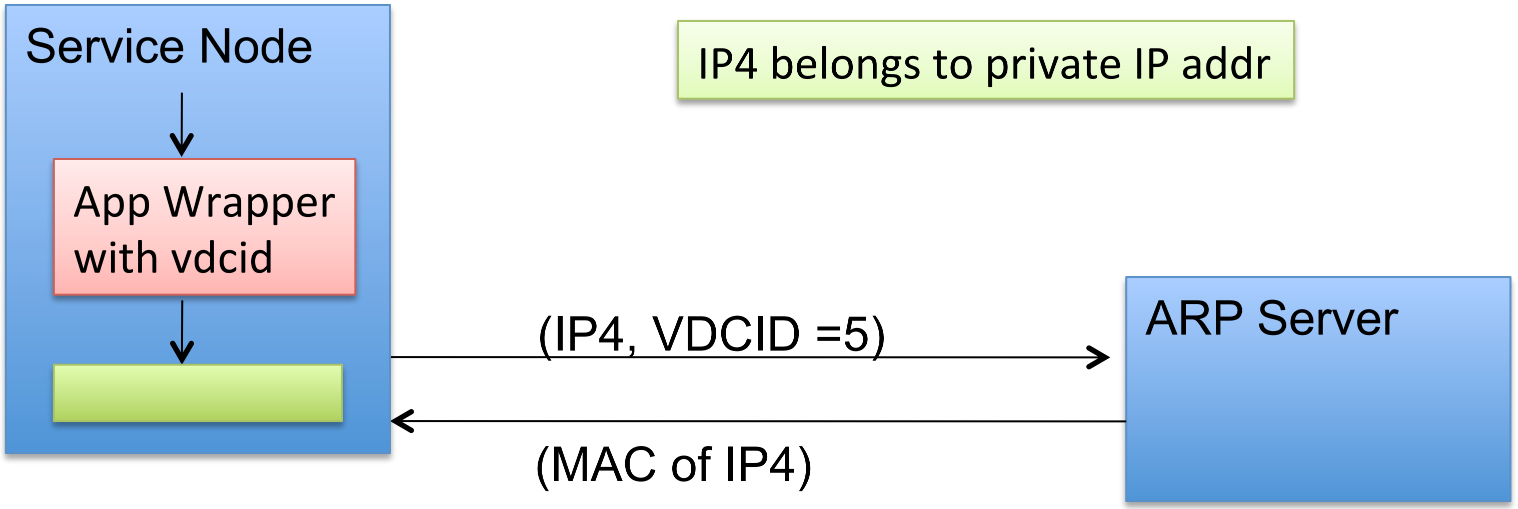
Solution: Application Wrapper embedding vdcid to packet

- In all layer 2 network with MAC as identifier, same IP addresses doest not matter.
- Assume IPx belongs to vdc1, before a program starts
  - Set vdc1 to the process's environment variable
  - Execute the program
- At kernel: bind packet with vdcid
  - Intercept the packet and lookup vdcid from environment variable using its pid
  - Sent request with this IP and vdcid to ARP server
  - When responds, update the arp cache

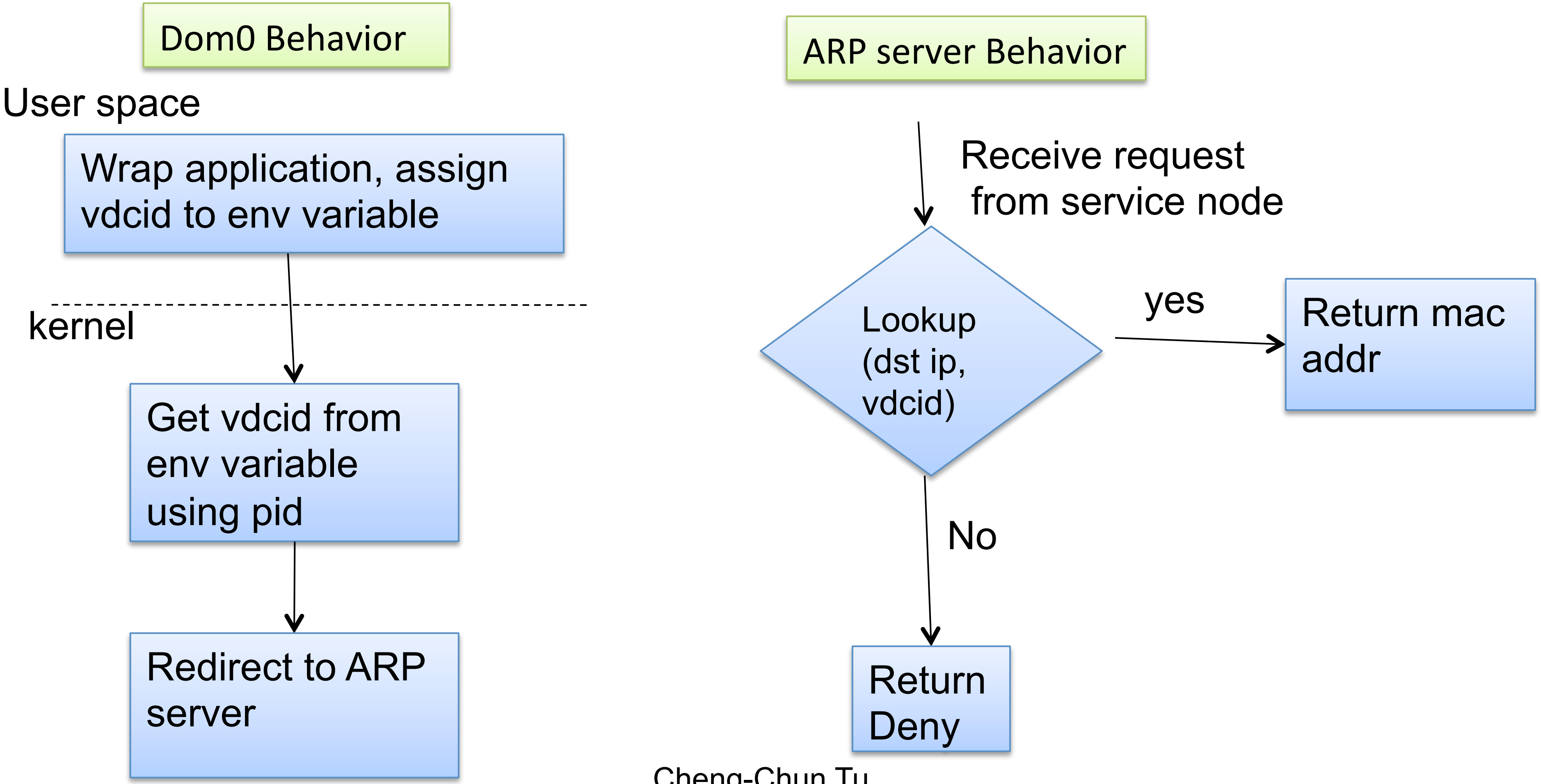
Service to Private example: case1 SSH

Connection initiates from service node.
Ex: Service node wants to talk to IP4, vdcid5, which is a private address

IP4 belongs to private IP addr



Flow Chart: Service-to-private Dom0 & ARP server, case1



Cheng-Chun Tu

Service to Private: Case 2, initiate from private node

1. Connection initiated from private node
2. Service node maintains connection mapp to destination mac

Consult a socket identification map, where a new entry is added when a Private --> Service connection is first established

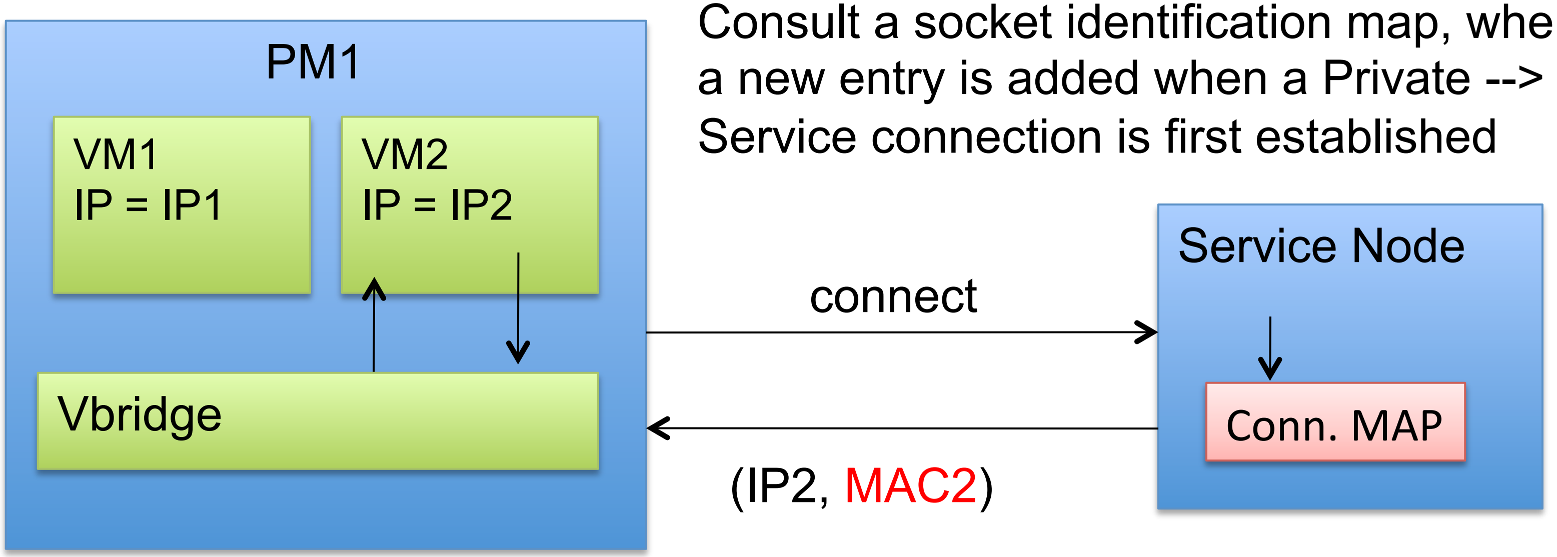
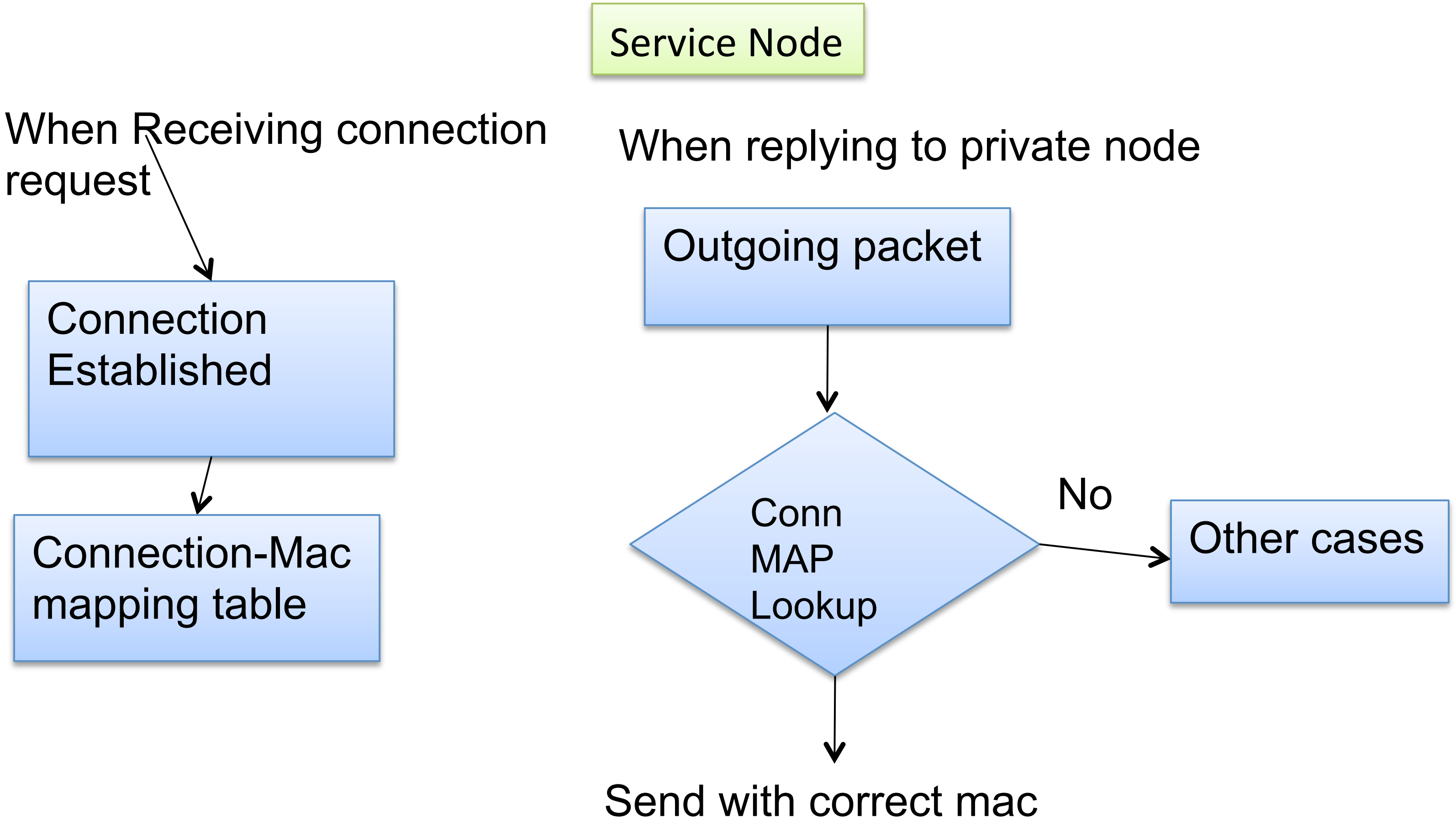


Table entry:
src_ip, src_port, dst_ip, dst_port, mac

Cheng-Chun Tu

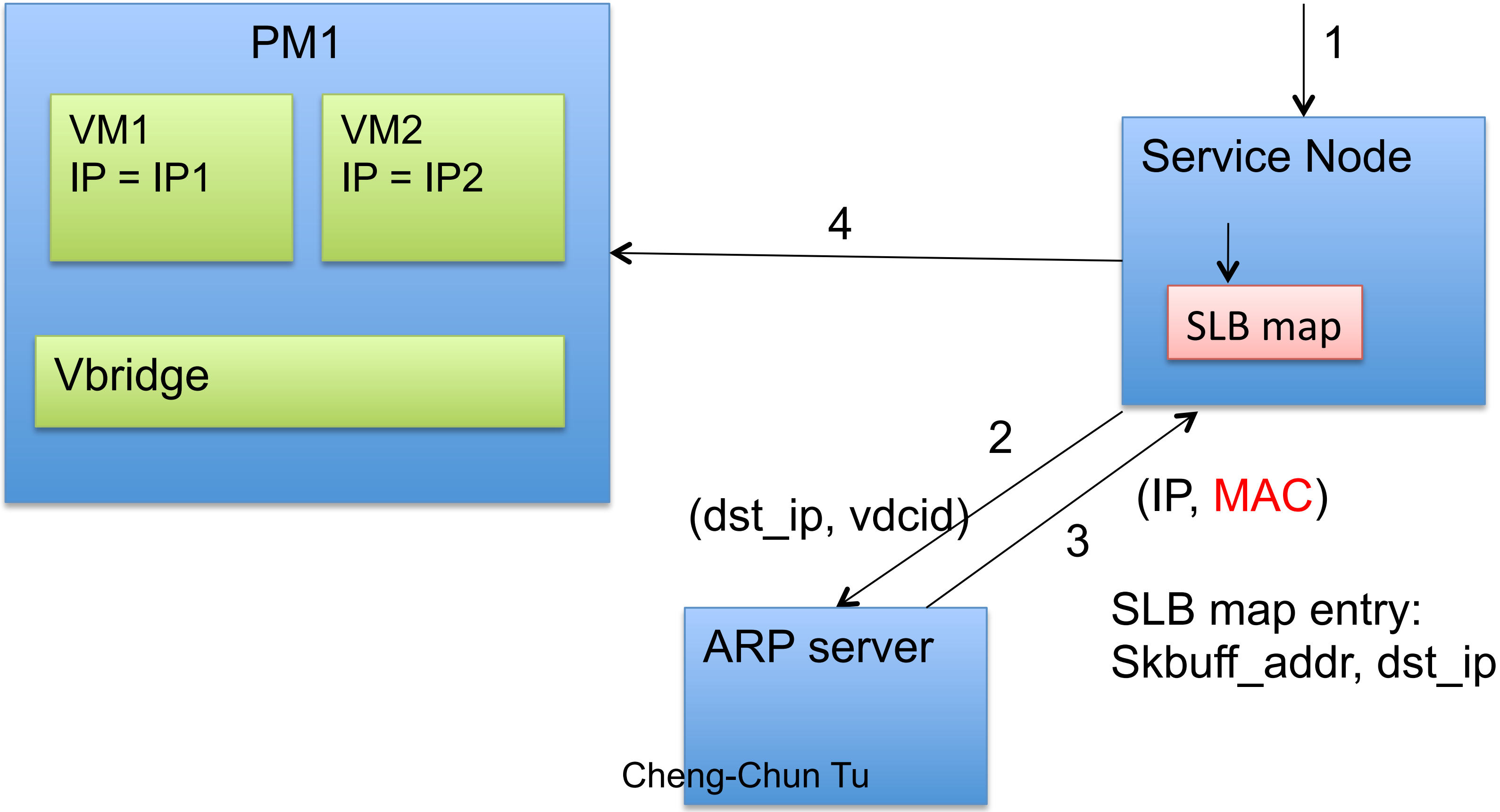
Flow Chart: Service-to-private Dom0 & ARP server, case2



Cheng-Chun Tu

Service to Private: Case 3 SLB

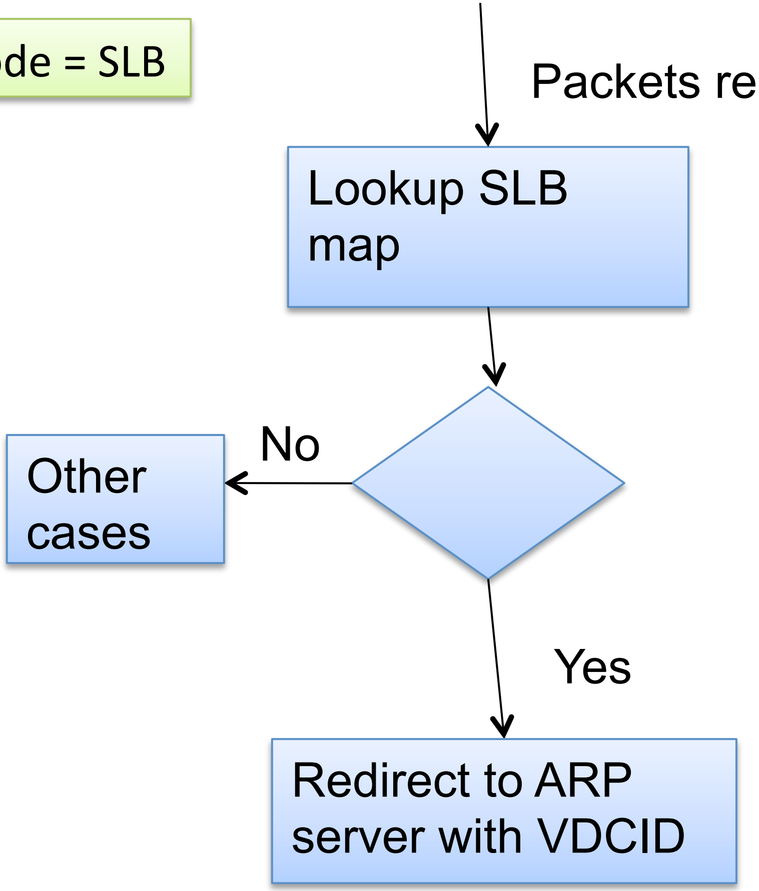
If service = SLB, Look up SLB map to find the destination's VDCID



SLB map entry:
Skbuff_addr, dst_ip, vdcid

Flow Chart: Service-to-private Dom0&ARP serv, case3 SLB

Service node = SLB

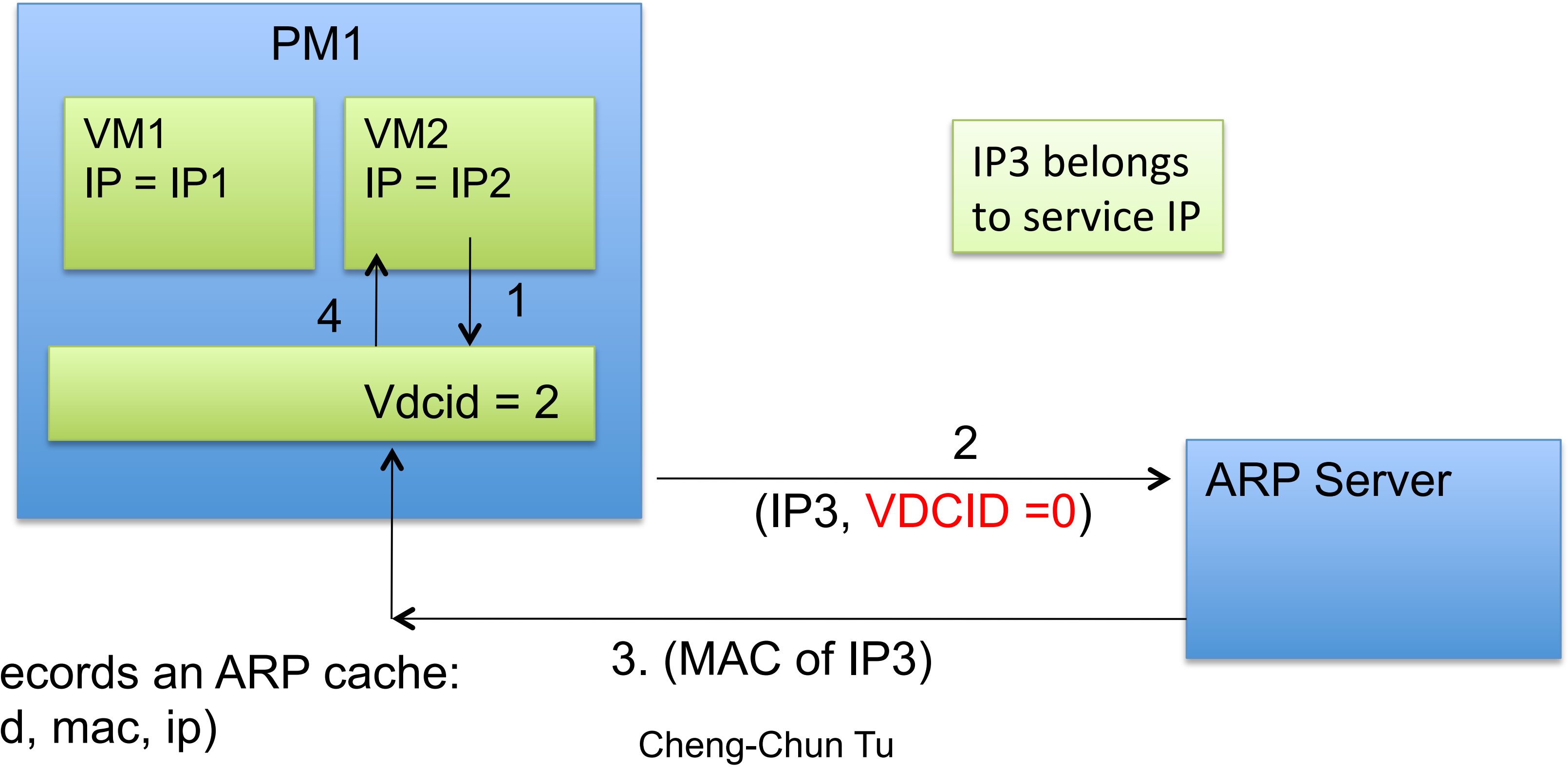


SLB map entry:
Skbuff_addr, dst_ip, vdcid

Cheng-Chun Tu

Private to Service

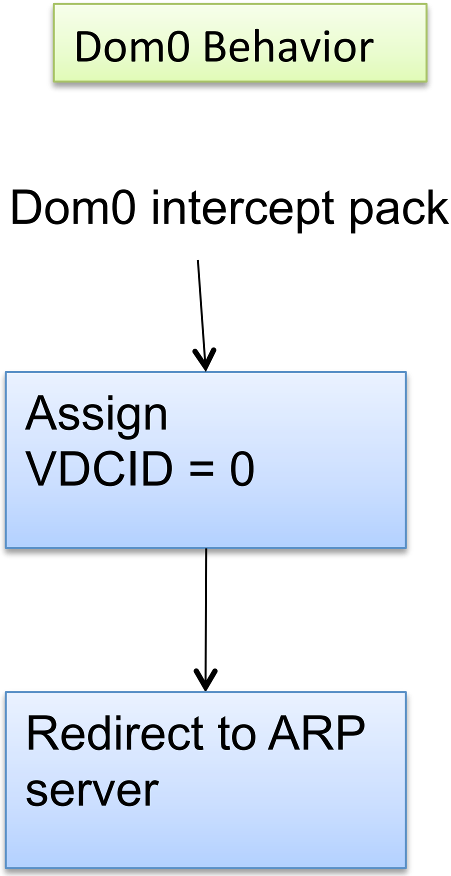
Source nodes with private IP address talk to destination, which nodes are service node or Dom0, Ex: IP3 is a service node



PM records an ARP cache: (vdcid, mac, ip)

Cheng-Chun Tu

Flow Chart: Private-to-Service

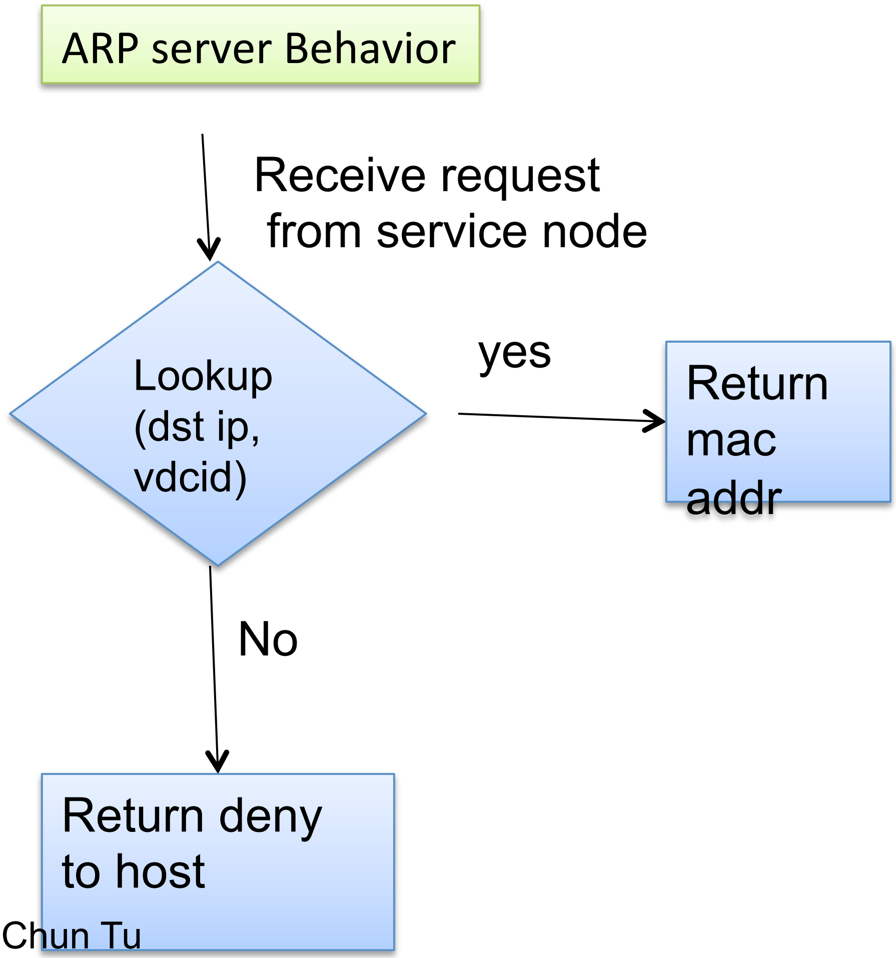


Dom0 Behavior

Dom0 intercept packet

Assign VDCID = 0

Redirect to ARP server



ARP server Behavior

Receive request from service node

Lookup (dst ip, vdcid)

yes

Return mac addr

No

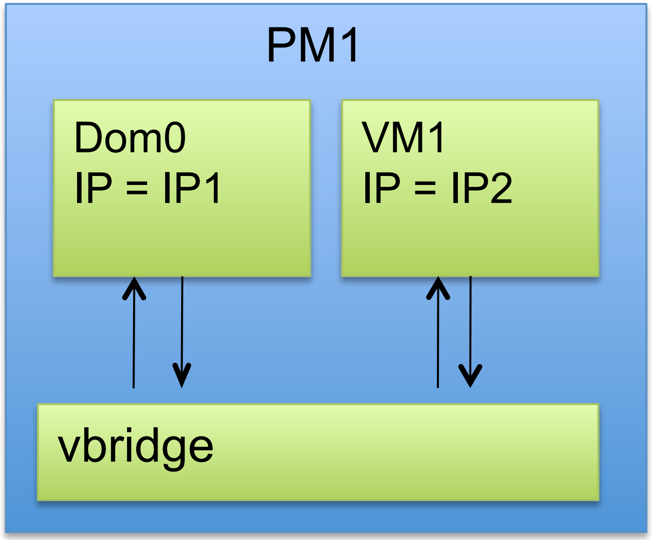
Return deny to host

Same as Service to Service

Cheng-Chun Tu

Local PM: Private to Service and Service to Private

Private node (VM1) and service node (Dom0) locate in the same PM



1. No two user VMs residing on the same PM could be assigned the same private IP address

2. ARP is not redirected, VM1 and Dom0 directly talk to each other.

Cheng-Chun Tu

Security Group

- Problem statement:
  - A large flat network lacks the ability of host authentication
  - prone to ARP spoofing, Man in the middle attack
- Solution:
  - Using VDCID to group a set of VMs as mutually trusted hosts
  - Hosts across VDC have limited ability to attack
  - Centralized ARP server guarantees isolation and no arp spoofing

Summary

- Allowing each VDC to own its private IP address space greatly simplifies the migration from private to public clouds
- Layer-2 forwarding enables reuse of same IP addresses
- Including a destination host's VDC ID in every ARP query
- Key is to determine the VDC ID of a target host in an ARP query based on its querying host's address

Claims

- A layer 2 network system for address reuse comprising:
  - A directory service
  - A Plurality of tenants, each with their own private IP address space (VM)
  - A Plurality of hosts (SN + CN dom0) assigned a set of IP addresses for shared resources
- A directory service is configured to be queried:
  - (vdc id, IP address) -> mac address
- A method for determine the VDCID of outgoing packets
  - Consists of 4 cases
- A method for isolating multiple groups of host (tenants) with same IP address space
  - Using VDC id + IP address as unique identifier
- A method for legacy protocols to work under the same IP address but goes to different hosts

Claims

- Divide IP address space into three categories: public, private, and service
- Using Generalized and centralized ARP: VDCID + IP --> MAC to preserve IP addresses
- A method of enforcing Inter-VDC isolation by checking destination's IP and VDC id
- A method of host authentication and preventing ARP spoofing attack at virtual data center

Claims

- Methods of identifying vdcid: Private to Private, Service to Service, Private to Service, Service to Private communication mode
- Method of attaching vdcid to environment variable at user space and retrieve it by observing packets current context at kernel

END