

# 以 Proxmox VE 建置私有雲 虛擬主機服務

NTHU CS - LSALAB  
(周志遠教授)

APNIC 52 Fellows  
牟展佑 [www.mou.tw](http://www.mou.tw)  
[contact@mou.tw](mailto:contact@mou.tw)

# William Mou



- # NTHU SCC
- # New Taipei
- # Community
- # Photographer

I am passionate about computer architecture, operating systems and networking. I show these passions in HPC field and won the championship in ASC20-21 Student Supercomputing Challenge. I'm glad to join here and meet you all!



COSUP



HITCON



SITCON



Super Computer Competition (HPC)



Proxmox VE Overlay Networking



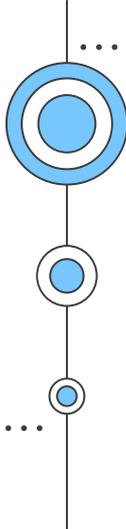
WasmEdgeRuntime

Web Assembly



SiFive

SiFive Intern



# 目錄

壹、Overlay Network 概念與優點

貳、如何在 Mikrotik 建構 Overlay Network

參、Proxmox VE 介紹

肆、私有雲主機服務案例

...



# 需求及挑戰

## 學校實驗室有建立伺服器叢集給研究生使用的需求：

- 伺服器叢集、實驗室 A、實驗室 B 位於不同大樓
- 伺服器需要高自由度，供研究生實驗新的軟硬體

## 在 Naive 的網路及伺服器架構下，使用者有以下困難：

1. 無法透過 SSH 連接內部網路設備
    - A 埠轉發至外部網路(但有可能遭受攻擊)
    - A 設定虛擬私人網路(需要頻繁切換)
  2. 伺服器上臨時架設的網頁難以連接
    - A 請管理員設定埠轉發(較為繁瑣)
    - A 使用 `ssh -R` 進行反向隧道連接
1. 硬體機器資源有限互相爭搶
    - A 使用試算表手動排程(繁瑣)
    - A 未滿載使用容易造成資源浪費
  2. 軟體環境互相干擾
    - A 使用虛擬環境(學習成本高)
    - A 硬碟多系統開機(中斷成本高)

...

# 解決方案

在 Naive 的網路及伺服器架構下，使用者有以下困難：

1. 無法透過 SSH 連接內部網路設備
  - 埠轉發至外部網路(但有可能遭受攻擊)
  - 設定虛擬私人網路(需要頻繁切換)
2. 伺服器上臨時架設的網頁或服務難以連接
  - 請管理員設定埠轉發(較為繁瑣)
  - 使用 `ssh -R` 進行反向隧道連接
1. 硬體機器資源有限互相爭搶
  - 使用試算表手動排程(繁瑣)
  - 未滿載使用容易造成資源浪費
2. 軟體環境互相干擾
  - 使用虛擬環境(學習成本高)
  - 硬碟多系統開機(中斷成本高)

## L3 Overlay Network

- BGP over WireGuard / L2TP over IPSec
- Build with Mikrotik RouterOS
- Need to manage private AS, IPAM, and more

## Proxmox VE Cluster

- PVE Accounts Permission
- Virtual Machine: QEMU / KVM
- Virt-IO / VFIO / PCIe Passthrough
- Virtual Bridge / VLAN

壹、

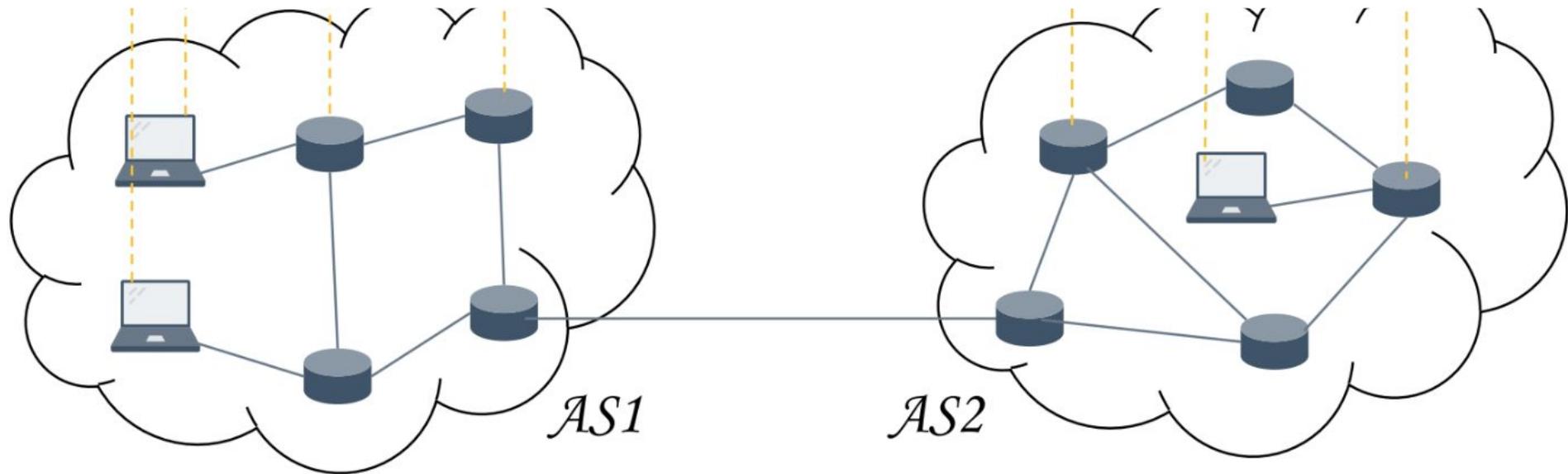
Overlay Network

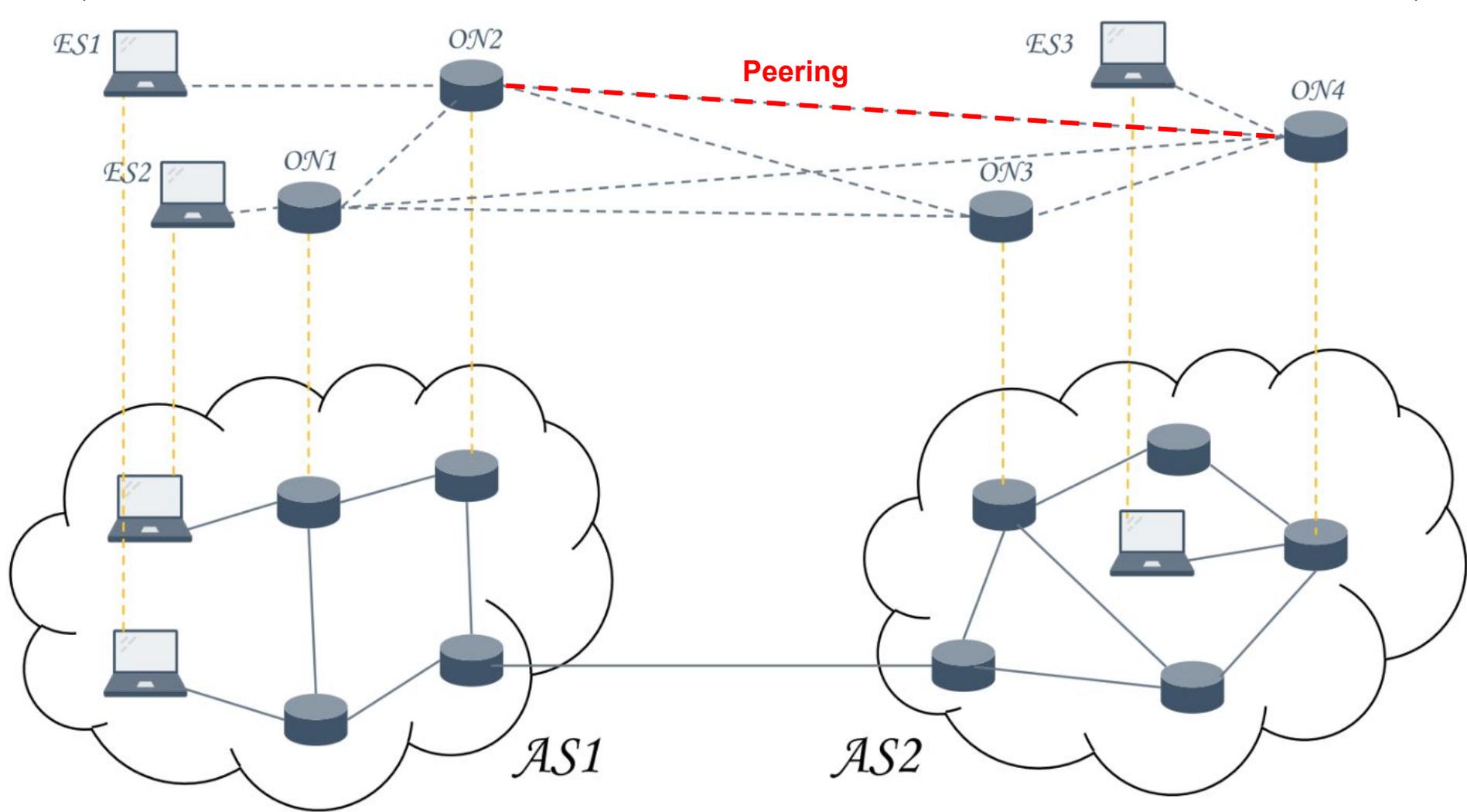
概念與優點

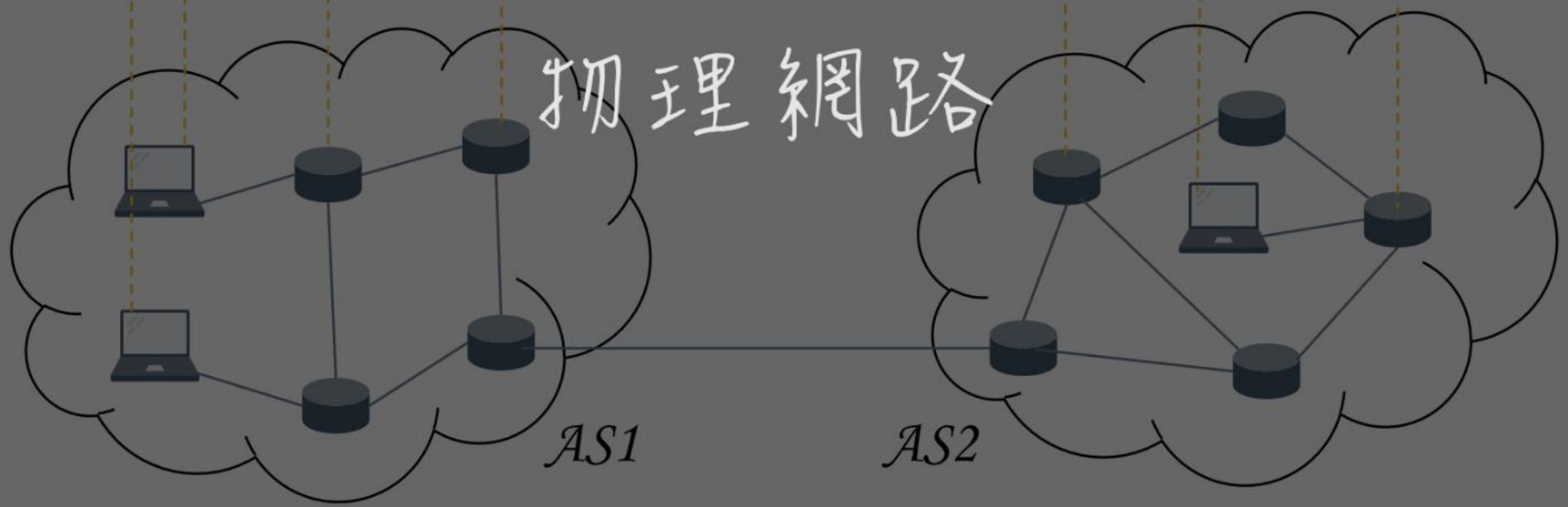
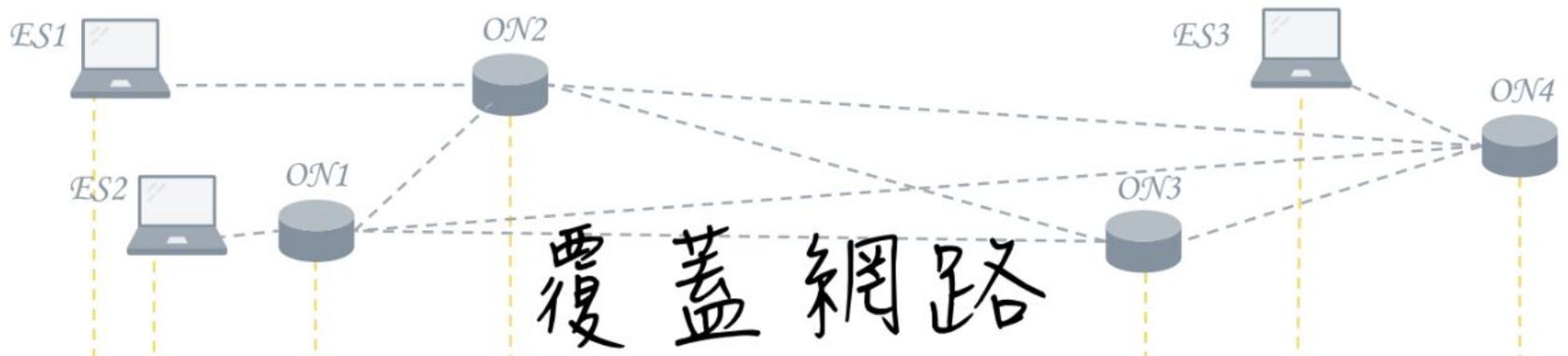
# 覆蓋網路 (Overlay Network)

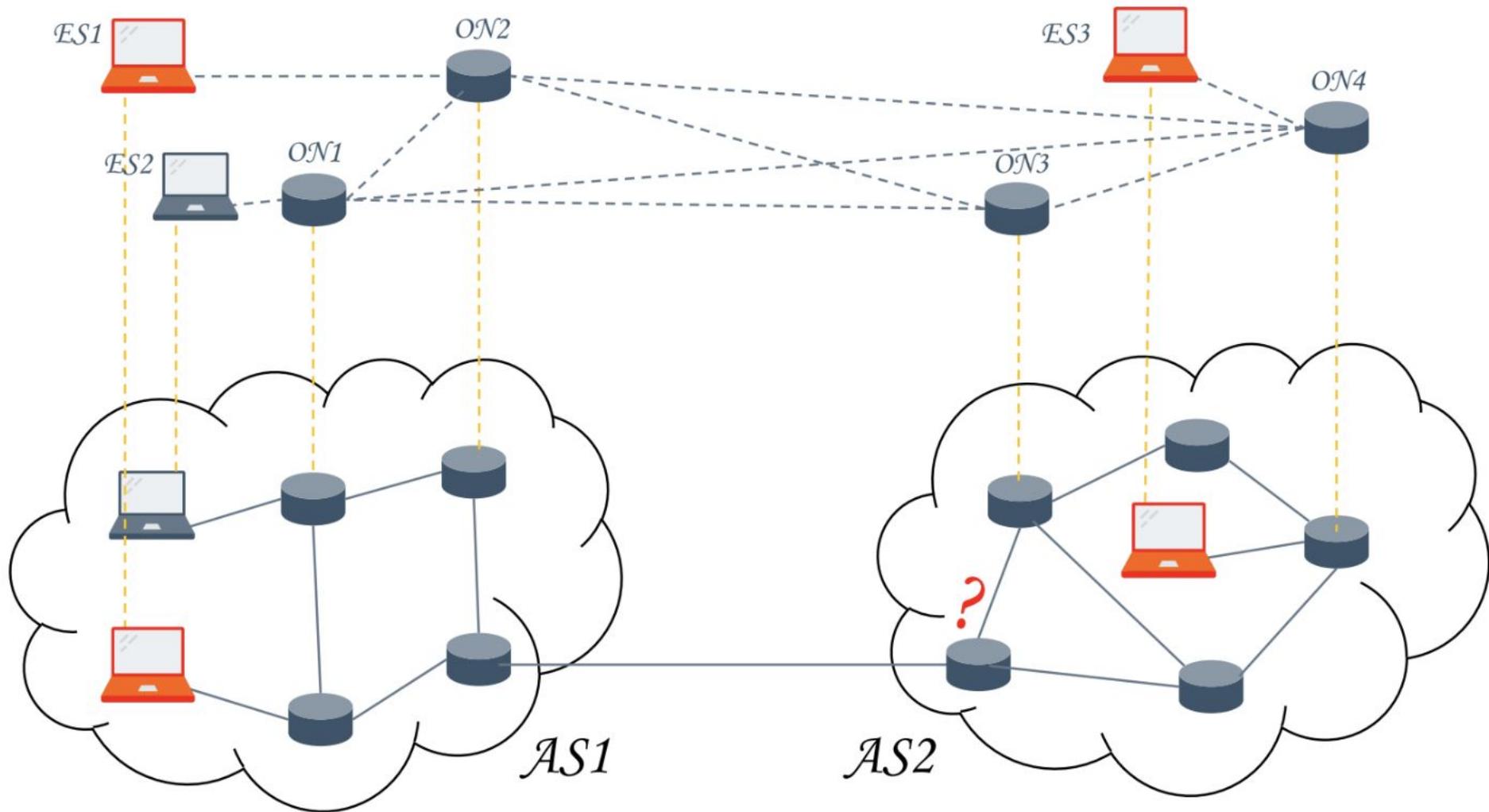
覆蓋網路 (Overlay Network) 是一種在現有網路基礎上建立的虛擬網路架構。覆蓋網路中的節點是通過現有的底層網路 (如網際網路) 相互連接的，並且這些節點之間的連接形成了另一個邏輯上的網路結構。

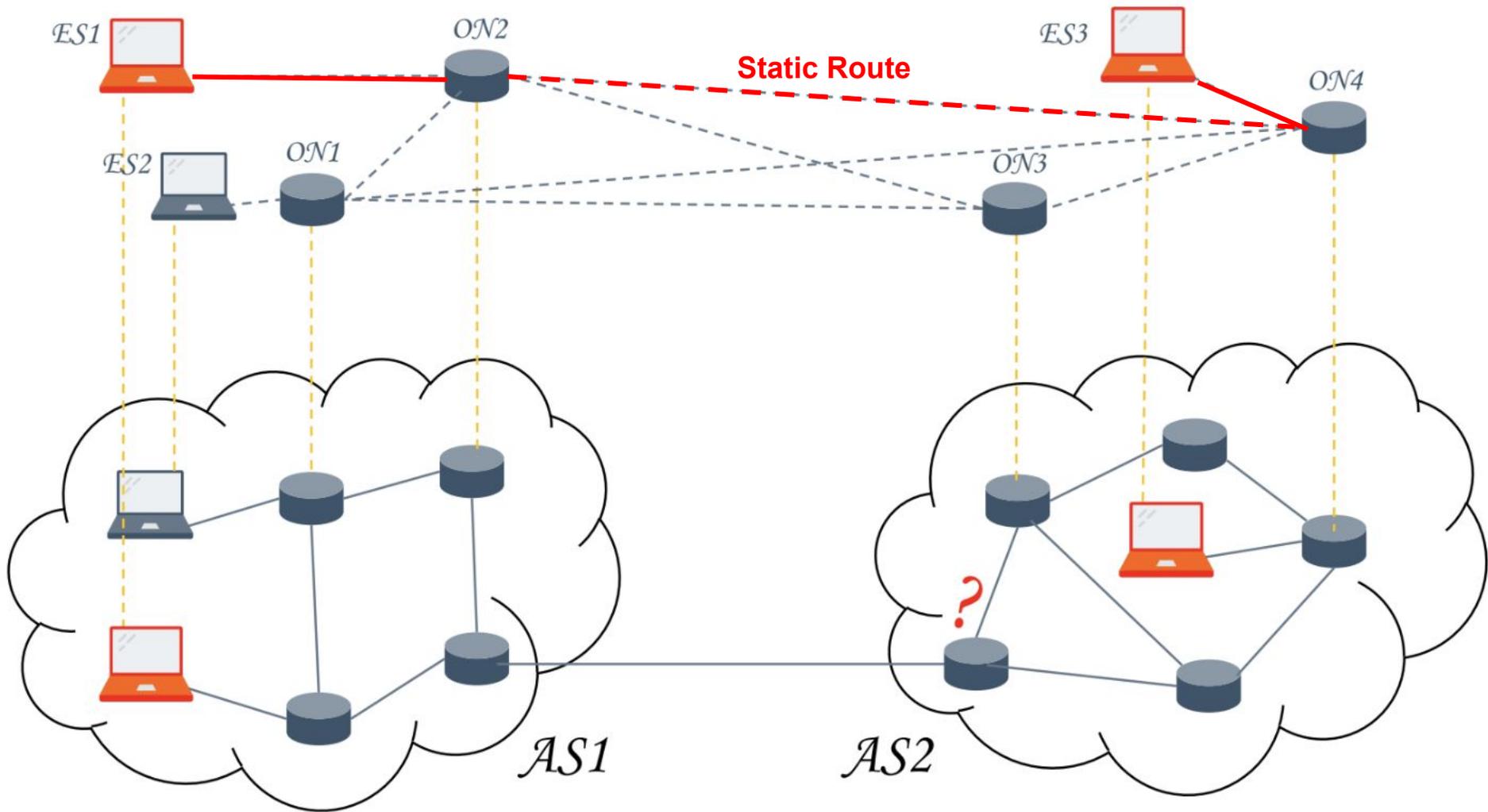
覆蓋網路是在一個物理網路上建立的虛擬網路。

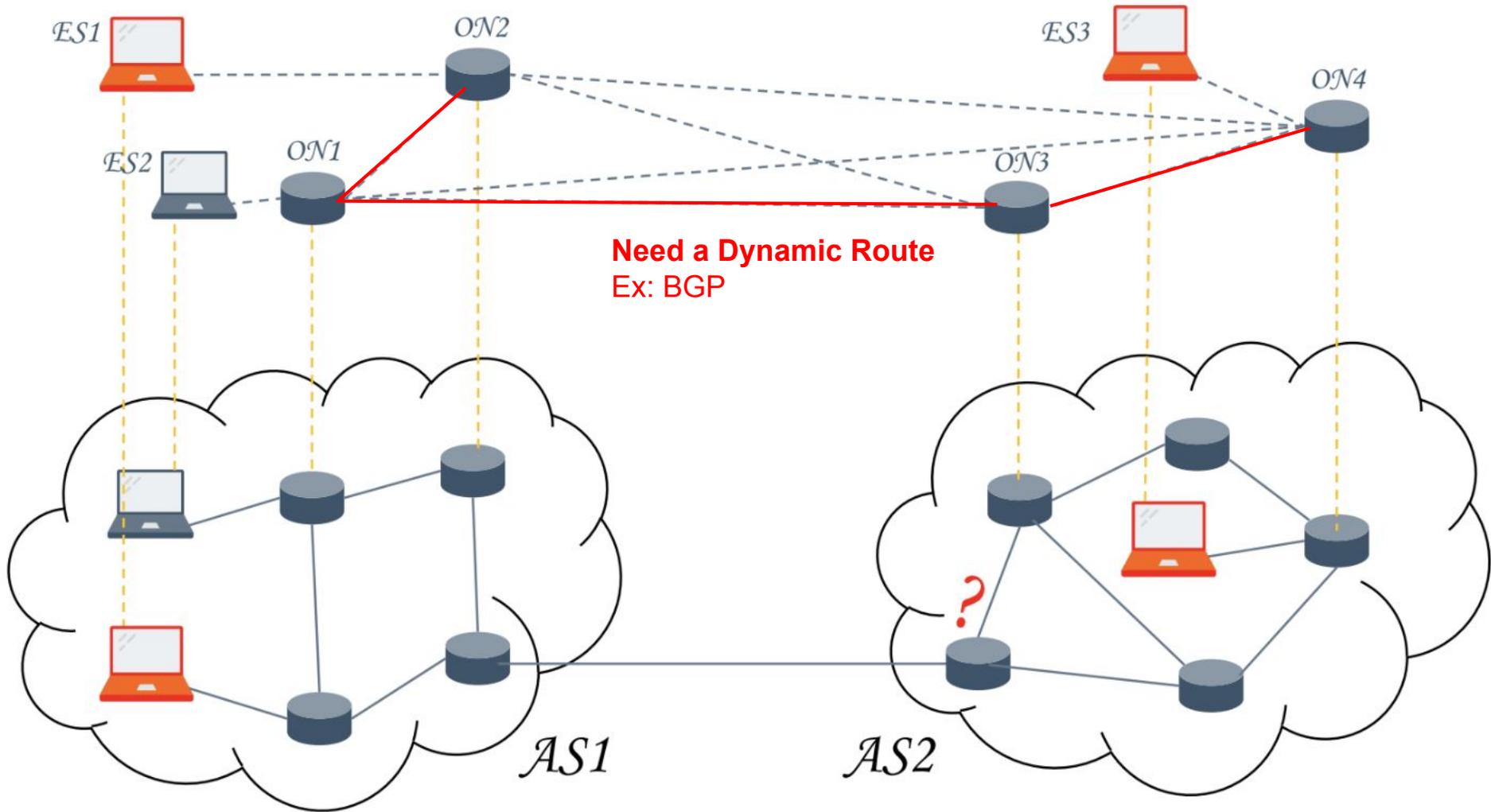


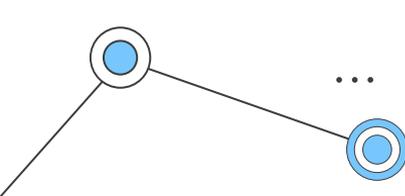




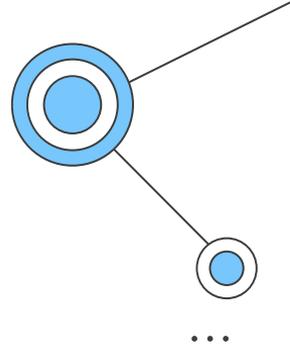


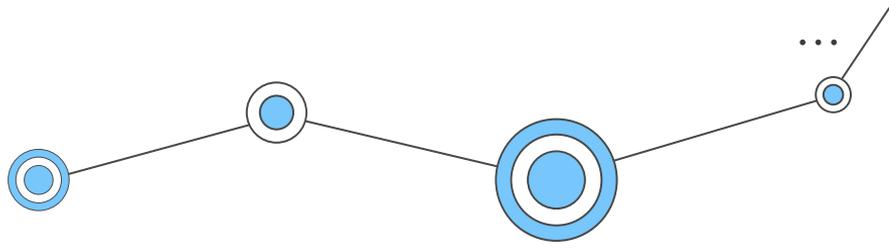






# 覆蓋網路的優勢



1. 靈活性: 如果要新增新的 Site (路由器及伺服器), 可以直接和任一 Site peering 即可上線
  2. 易於管理: 管理員接入 overlay network 後, 可以用符合業務邏輯的方式管理網路
  3. 安全性: 通過 VPN 連入後, 可以訪問網路內的伺服器, 避免暴露 port 等危險操作
  4. 冗餘和效率: overlay 流量可以更輕鬆地根據流量飽和度或網路中斷情況改變路徑
- 

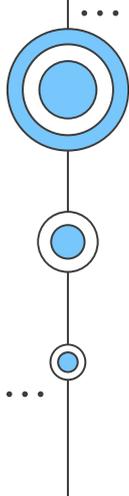
# 使用者體驗

## 對於 User 而言：

- 身處任意 Site 的 Subnet 下 或
  - 連上 VPN 接入 Overlay Network
1. 無法透過 SSH 連接內部網路設備
    - ~~埠轉發至外部網路(但有可能遭受攻擊)~~
    - ~~設定虛擬私人網路(需要頻繁切換)~~
    - 在實驗室內直接連接, 實驗室外一個 VPN 連接全網
  2. 伺服器上臨時架設的網頁或服務難以連接
    - ~~請管理員設定埠轉發(較為繁瑣)~~
    - ~~使用 ssh -R 進行反向隧道連接~~
    - 直接透過 Private IP 連接 ...

貳、

在 RouterOS 建構  
Overlay Network



# Add a WG interface on 2 Routers

MikroTik

Safe Mode Quick Set WebFig Terminal

New Interface

not invalid not running not slave not passthrough

Enabled  Torch  
Reset Traffic Counters

Comment

General

Name

Type WireGuard

MTU

Actual MTU

Listen Port

Private Key

Public Key

Status

Traffic

Cancel Apply OK

MikroTik

Safe Mode Quick Set WebFig Terminal

New Interface

not invalid not running not slave not passthrough

Enabled  Torch  
Reset Traffic Counters

Comment

General

Name

Type WireGuard

MTU

Actual MTU

Listen Port

Private Key

Public Key

Status

Traffic

Cancel Apply OK

MikroTik

Safe Mode Quick Set WebFig Terminal

Address <10.121.4.157/30>

not invalid Remove

Enabled

Comment

Address

Network

Interface

Cancel Apply OK

MikroTik

Safe Mode Quick Set WebFig Terminal

Address <10.121.4.158/30>

not invalid Remove

Enabled

Comment

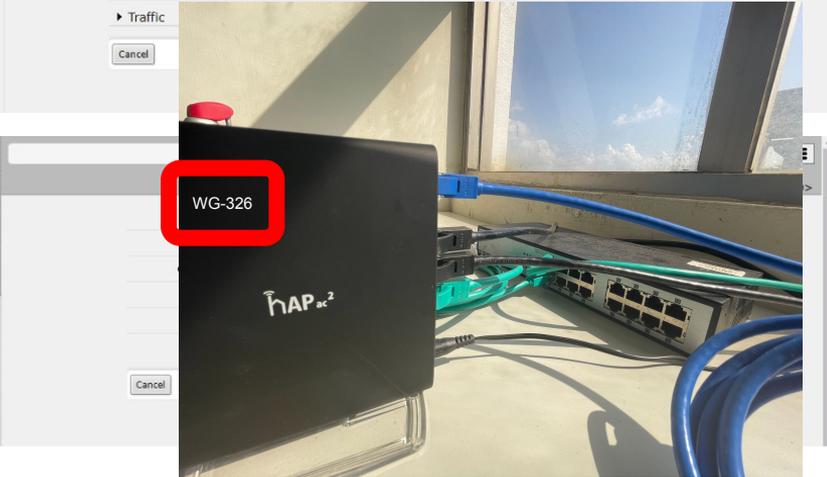
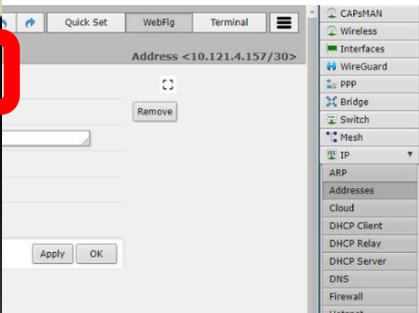
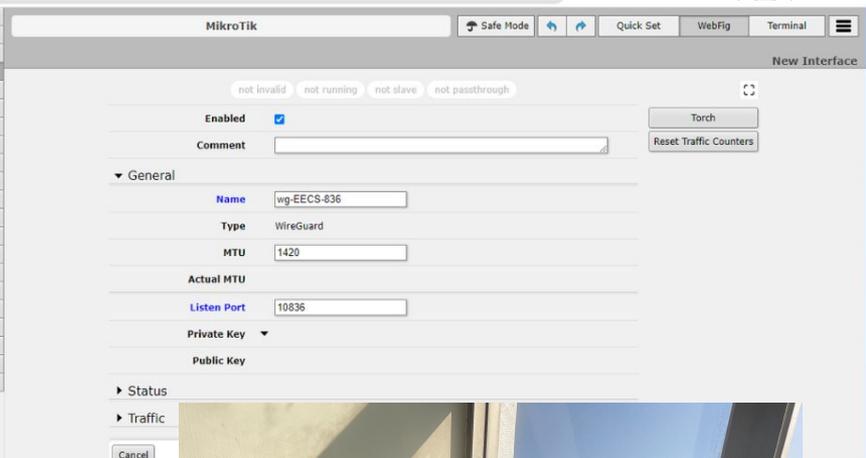
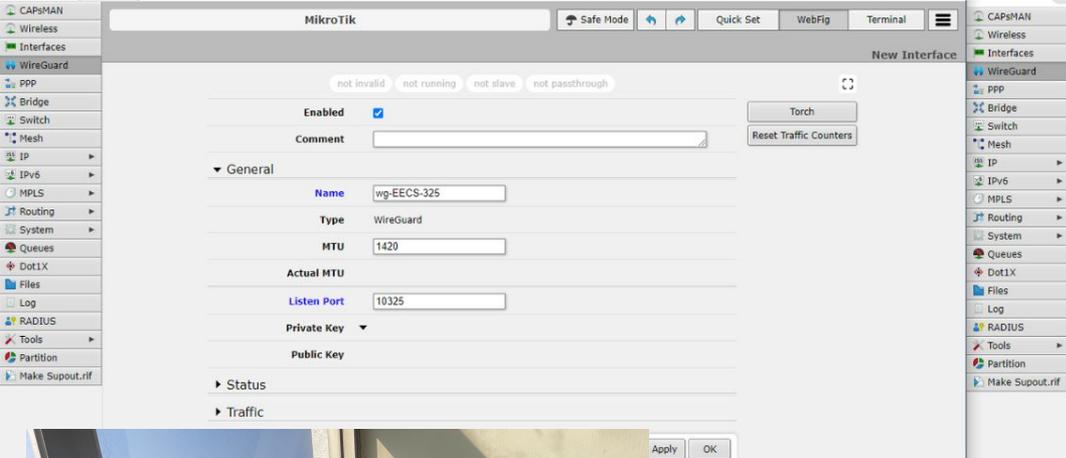
Address

Network

Interface

Cancel Apply OK

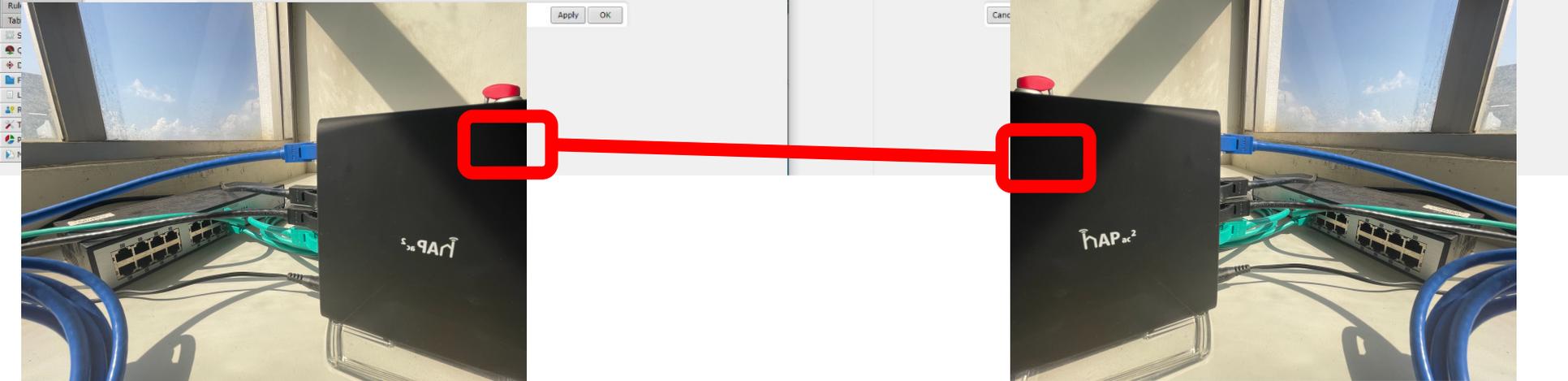
# Add a WG interface on 2 Routers

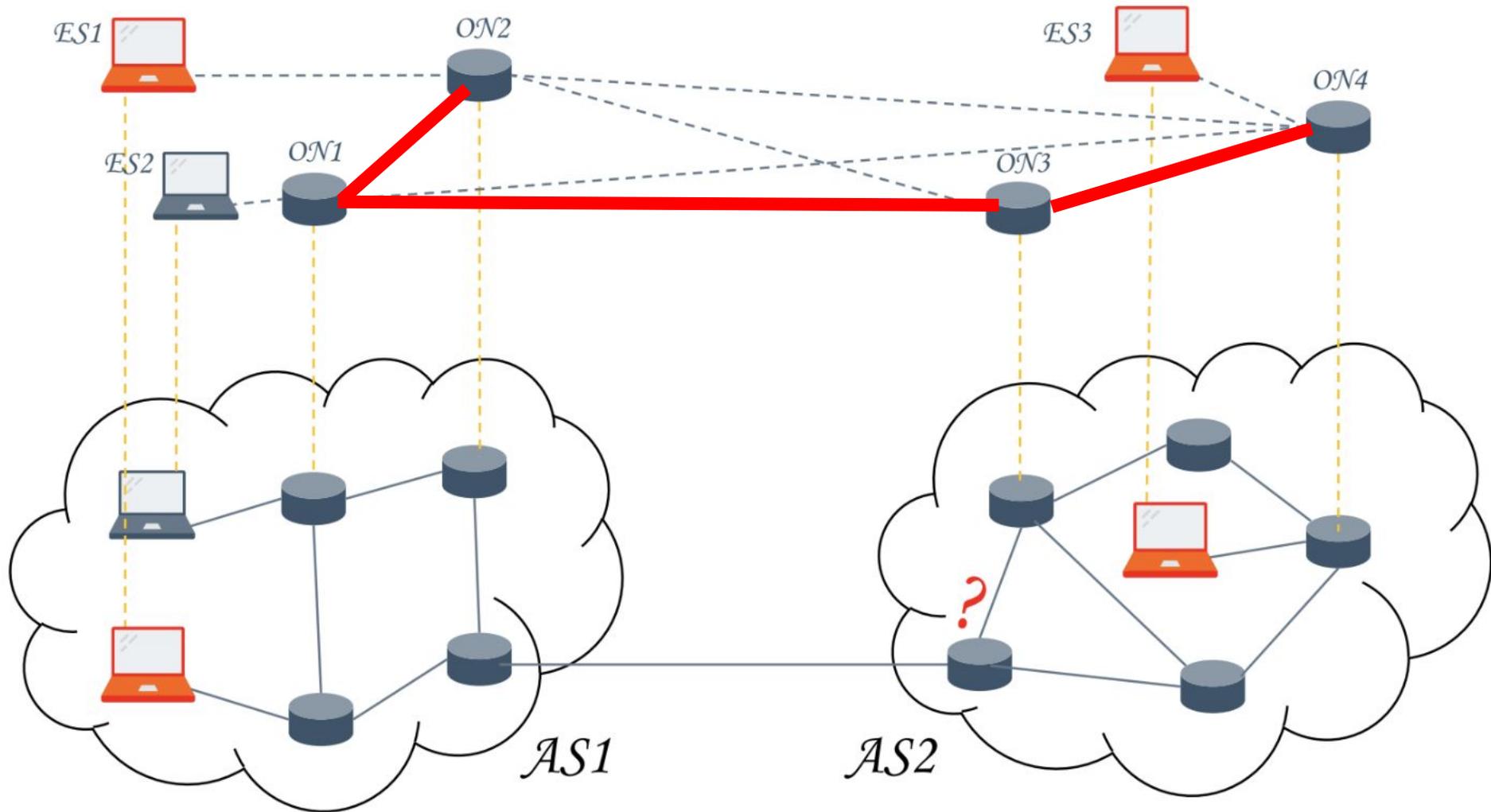


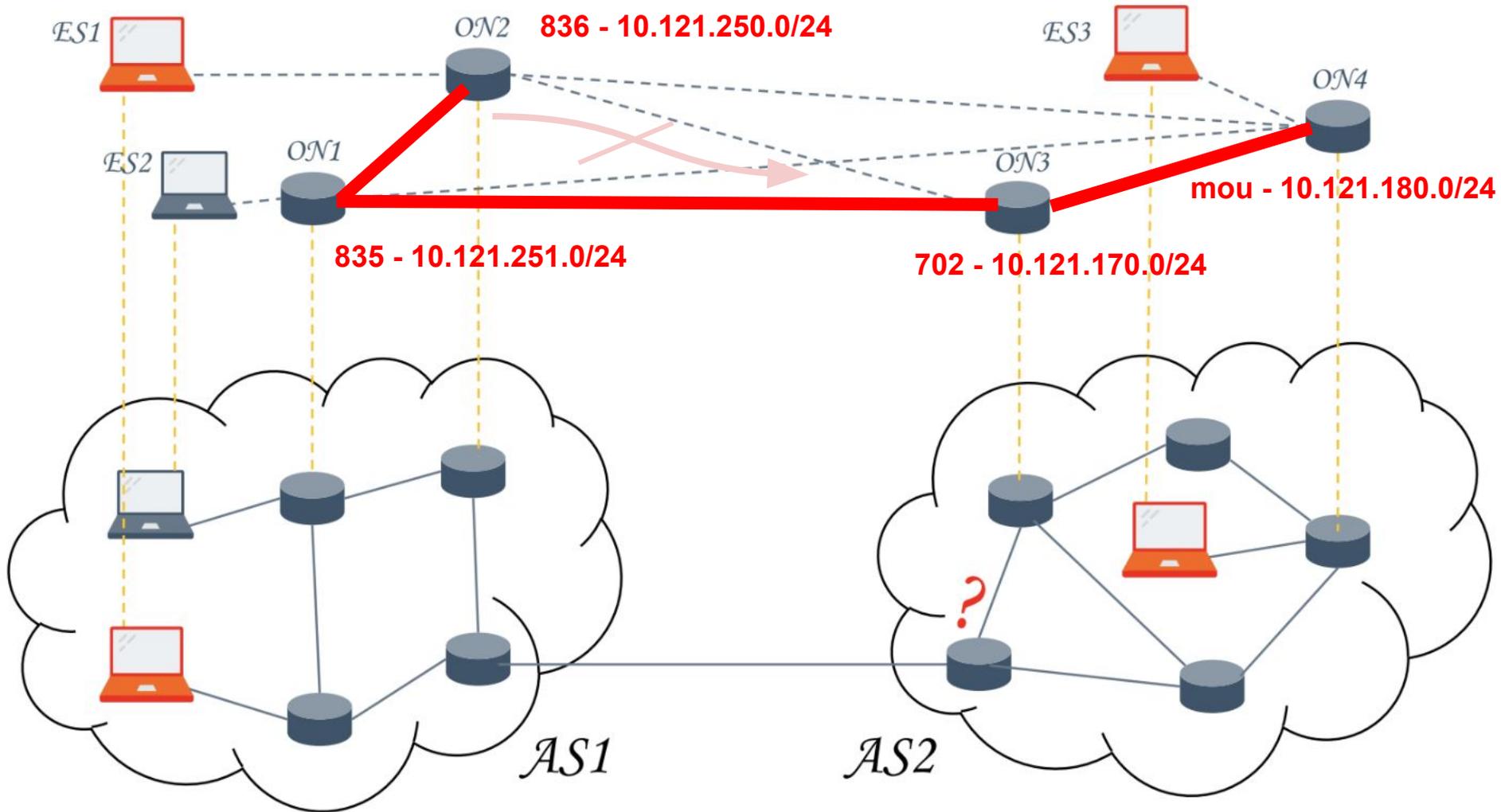
# Add a WG tunnel between 2 Routers

The screenshot shows the MikroTik WinBox interface for configuring a WireGuard Peer. The configuration is for a peer with the public key `5E80bk81AeggCNZ038zTIWfKw9rhRGsqxyvNeWwFFAw`. The interface is set to `wg-EECS-325` and the endpoint is `140.114.91.164` on port `110836`. The allowed address is `0.0.0.0/0`. The persistent keepalive is set to `00:01:00`. The RX and TX rates are `2039.5 MIB` and `3850.7 MIB` respectively. The last handshake was at `00:01:19`.

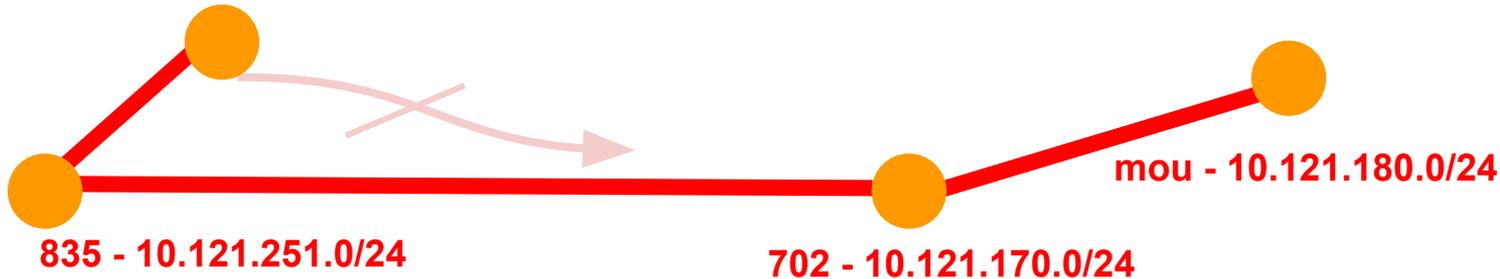
The screenshot shows the MikroTik WinBox interface for configuring a WireGuard Peer. The configuration is for a peer with the public key `F6xd5x2oqywKU+/avhVSttGF3XdytGYzcWbEydSIT0`. The interface is set to `wg-EECS-836` and the endpoint is `140.114.78.12` on port `10325`. The allowed address is `0.0.0.0/0`. The persistent keepalive is set to `00:01:00`. The RX and TX rates are `2642.0 MIB` and `2320.5 MIB` respectively. The last handshake was at `00:01:18`.







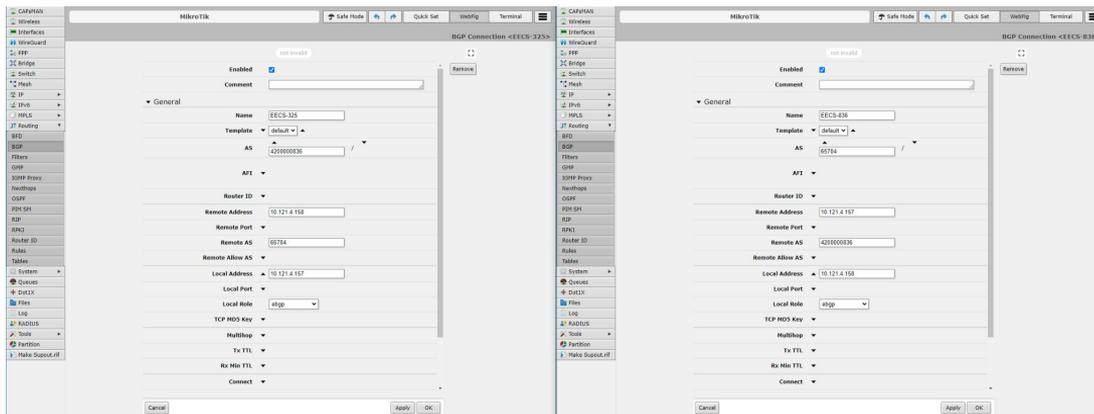
836 - 10.121.250.0/24



mou - 10.121.180.0/24

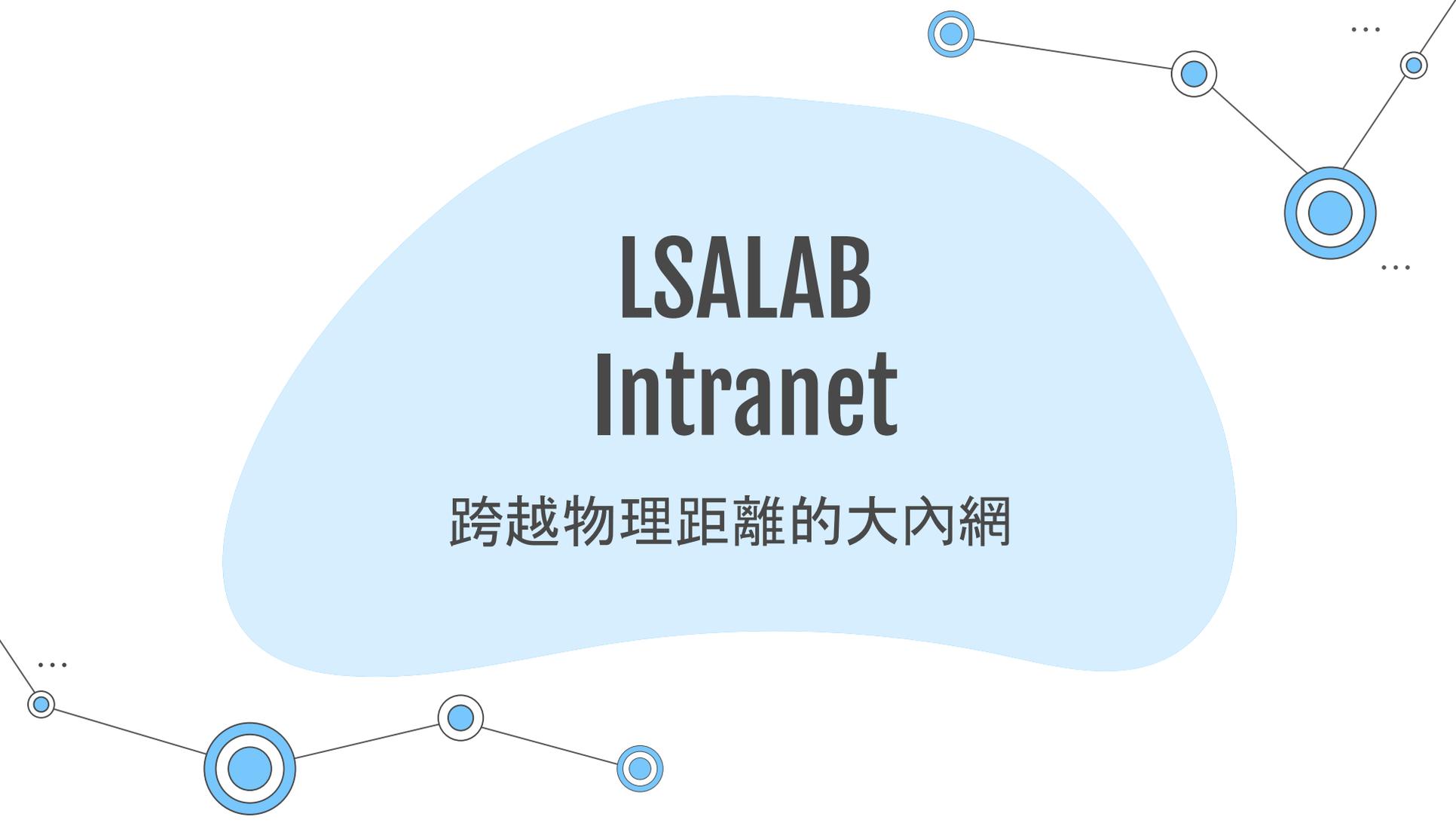
Pool	-	DAC	10.121.250.0/24	%bridge	
Routes	-	DAb	10.121.251.0/24	10.121.4.165	20
SMB	-	Db	10.121.251.0/24	10.121.4.158	20

## 路由協議 BGP, OSPF



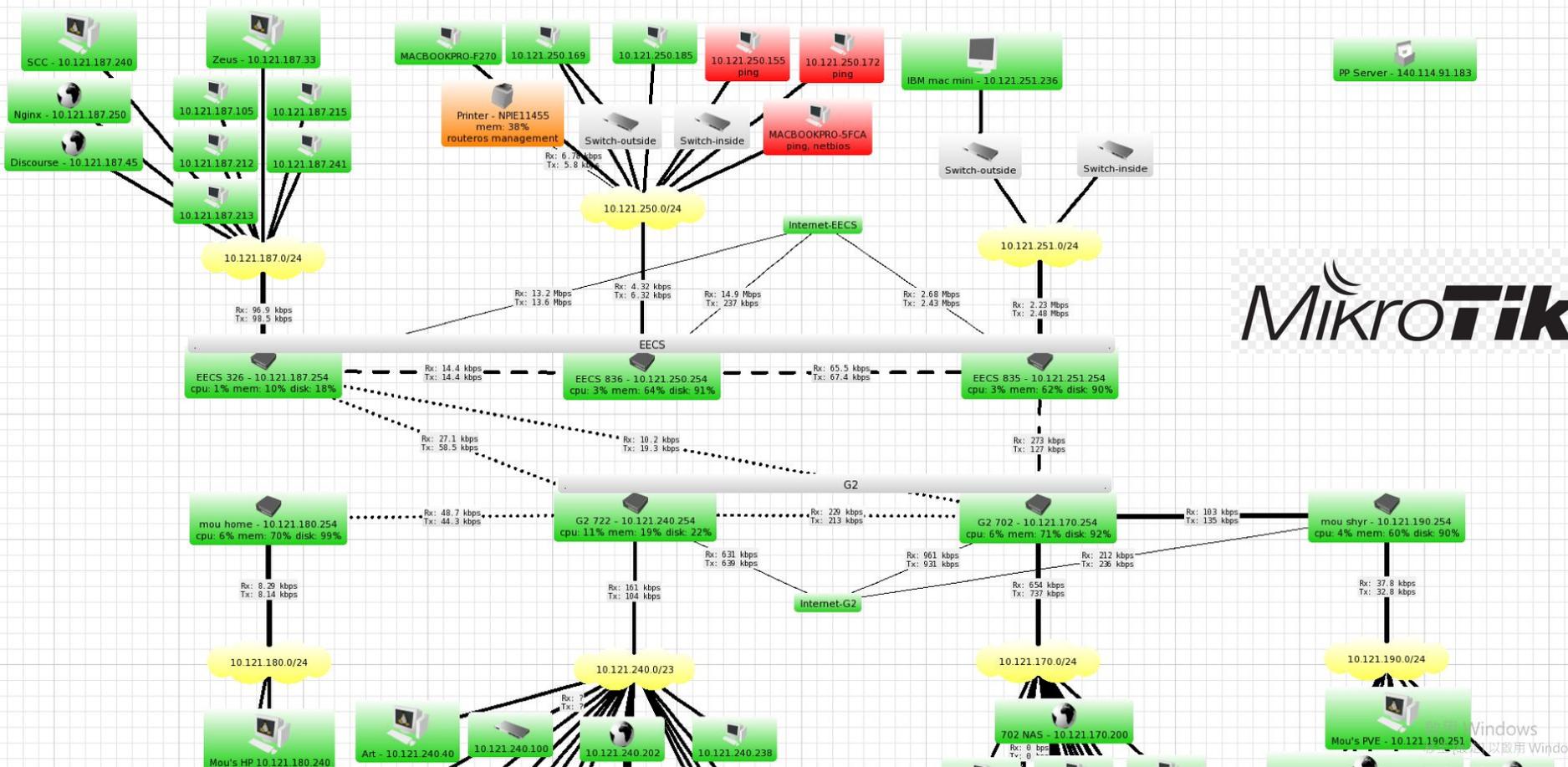
IP
ARP
Addresses
Cloud
DHCP Client
DHCP Relay
DHCP Server
DNS
Firewall
Hotspot
IPsec
Kid Control
Neighbors
Packing
Pool
Routes
SMB
SNMP
SSH
Services

-	Db	10.121.4.156/30	10.121.4.165	20
-	Db	10.121.4.160/30	10.121.4.158	20
-	DAb	10.121.4.160/30	10.121.4.165	20
-	DAC	10.121.4.164/30	%wg-835	
-	Db	10.121.4.164/30	10.121.4.165	20
-	Db	10.121.4.170/32	10.121.4.158	20
-	DAb	10.121.4.170/32	10.121.4.165	20
-	Db	10.121.4.174/32	10.121.4.158	20
-	DAb	10.121.4.174/32	10.121.4.165	20
-	DAb	10.121.4.192/30	10.121.4.165	20
-	Db	10.121.4.192/30	10.121.4.158	20
-	DAb	10.121.170.0/24	10.121.4.165	20
-	Db	10.121.170.0/24	10.121.4.158	20
-	DAb	10.121.171.0/28	10.121.4.165	20
-	Db	10.121.171.0/28	10.121.4.158	20
-	DAb	10.121.171.16/28	10.121.4.165	20
-	Db	10.121.171.16/28	10.121.4.158	20
-	Db	10.121.171.191/32	10.121.4.158	20
-	DAb	10.121.171.191/32	10.121.4.165	20



# LSALAB Intranet

跨越物理距離的大內網



Windows 10 可以改用 Windows 11



# 優點-管理者視角

- **加強對外安全性：**
  - 加密通信:WireGuard 提供端到端加密, 確保跨網絡的數據傳輸安全。
  - 最小化攻擊面:private network 隔離了外部威脅。
- **靈活的網絡連接：**
  - 動態路由:使用BGP, 網絡路由可以自動調整, 以應對節點變化或網絡拓撲變化。
  - 異地連接:地理位置不同的子網也可以輕鬆連接, 適合分佈式團隊或辦公室。
- **簡化網絡管理：**
  - 去中心化設計:WireGuard 為去中心化VPN解決方案, 簡化了傳統VPN 的設置和維護。
  - 易於配置和維護:WireGuard 設計簡潔, 易於配置, 而BGP 支持自動化網絡管理

# 優點-使用者視角

- 遠程工作與協作
- 1. 安全訪問公司內部資源
  - w/o: 員工需要頻繁更換VPN 連接來訪問不同site 的資源, 不便且容易出錯。
  - w/: 員工透過單一VPN 連接就能安全地訪問Intranet 的所有內部資源, 操作簡單且高效。
- 2. 共享文件和網頁程式
  - w/o: 共享文件或連接伺服器時需透過複雜的port forwarding 配置, 增加維護成本。
  - w/: 員工一旦連接到VPN, 即可直接使用內網 IP 和原生 port 訪問共享資源, 簡化流程。
- 3. 分散式設備管理
  - w/o: 不同地點的裝置互連需要透過複雜的網絡配置和NAT, 增加管理難度。
  - w/: 不同地點的裝置可以直接在私有網絡中互聯, 簡化管理和數據收集。
- 4. 安全性和隱私
  - w/o: IoT 裝置的安全性和隱私保護較難保證, 尤其是在公共網絡中。
  - w/: 提供了一個加密且隔離的網絡環境, 增強IoT 裝置的安全性和數據隱私。

# 缺點

- **IT 學習成本及開銷：**
  - IT 需要管理兩個不同的網路層。
  - 這些層必須統一管理，因為overlay 期望的拓撲需要underlay 支援。
- **故障排除：**
  - 與前項相同，需要對兩層debug。
- **網路內部潛在的安全疑慮：**
  - 錯誤的配置造成的負面影響可能會因為overlay 放大。
  - 缺少最小權限概念，不如零信任系統強調安全性。



參、

Proxmox VE

介紹



# Proxmox VE 截圖

PROXMOX Virtual Environment 7.4-16 Search

Documentation Create VM Create CT root@pam

Server View Node 'pve'

Reboot Shutdown Shell Bulk Actions Help

Day (average)

Search Package versions

Summary

Notes

Shell

System

Network

Certificates

DNS

Hosts

Options

Time

Syslog

Updates

Repositories

Firewall

Disks

LVM

LVM-Thin

Directory

ZFS

Ceph

Replication

Task History

Subscription

pve (Uptime: 42 days 19:33:48)

CPU usage 2.38% of 96 CPU(s) IO delay 0.00%

Load average 2.42,2.34,2.26

RAM usage 73.15% (183.86 GiB of 251.35 GiB) KSM sharing 0 B

/ HD space 81.94% (76.97 GiB of 93.93 GiB) SWAP usage N/A

CPU(s) 96 x AMD EPYC 7352 24-Core Processor (2 Sockets)

Kernel Version Linux 5.15.111-1-pve #1 SMP PVE 5.15.111-1 (2023-08-18T08:57Z)

PVE Manager Version pve-manager/7.4-16/039f621

Repository Status ✔ Proxmox VE updates ! Non production-ready repository enabled

CPU usage

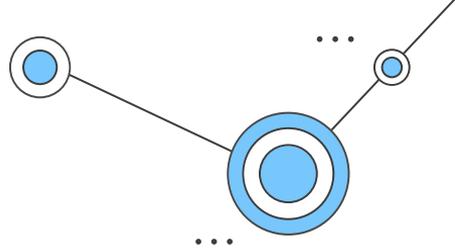
Memory usage

Server load

Network traffic

Start Time	End Time	Node	User name	Description	Status
Jul 18 04:46:21	Jul 18 04:46:23	pve	root@pam	Shell	OK
Jul 18 03:46:32	Jul 18 03:46:35	pve	root@pam	Update package database	OK

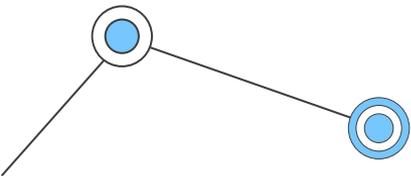
# Proxmox VE 介紹



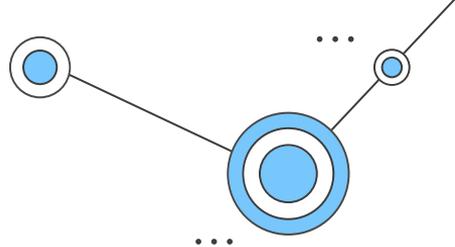
Proxmox VE 是一個開源的虛擬化平台，支援 **虛擬機**和**容器**，讓用戶可以在單一的平台輕鬆管理虛擬化環境。

Proxmox VE 提供友善的 Web 界面，讓用戶能夠輕鬆管理虛擬機、存儲、網絡和使用者帳戶等功能。該平台結合了虛擬機和容器技術，並支援多種虛擬化技術，如 KVM、LXC 等。Proxmox VE 提供豐富的功能和強大的性能，適合企業和個人用戶建立和管理虛擬化環境。

- 全 Web 化管理介面
- KVM 虛擬機與 LXC 容器 Cloud init 功能
- ZFS、Ceph、GlusterFS 等多種檔案系統
- NFS、iSCSI、RBD 等多種連接協定
- 去中心化的叢集機制
- 內建完整備份與還原功能
- 提供虛擬機與容器線上遷移功能
- 支援跨節點虛擬機與容器複寫功能
- 提供高可用性容錯機制
- 繁體中文操作介面



# 傳統伺服器 v.s. Proxmox VE

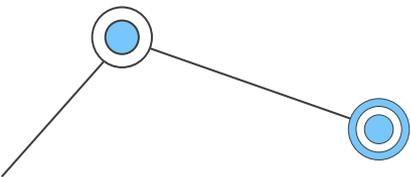


## 傳統伺服器的痛點:

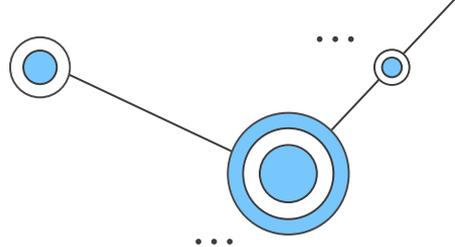
1. 低資源利用率:傳統伺服器通常只運行單一作業系統和應用程式,資源利用率不高。
2. 硬體依賴:使用傳統伺服器時,若要提高性能或可用性,通常需要花費大量成本來升級硬體。
3. 難以管理:管理多個獨立伺服器需要更多的人力和時間,並且可能導致管理上的混亂。
4. 災難恢復困難:傳統伺服器在災難時可能會導致資料丟失和停機時間更長。

## Proxmox VE 優勢:

1. 虛擬化技術:可以讓您在單一物理伺服器上運行多個虛擬機器,提高伺服器資源的利用率。
2. 管理與監控:提供直觀的 Web 介面,讓您輕鬆地管理和監控所有虛擬機器。
3. 高可用性:支援儲存和虛擬機器的冗餘,可以確保系統的高可用性。
4. 彈性擴展:可以輕鬆地修改伺服器資源,而無需進行更動硬體。
5. 成本較低:軟體本身免費使用,入門門檻低,易於部署中小規模 Cluster。



# 在一個 OS 上虛擬化另一個 OS



## 問題：

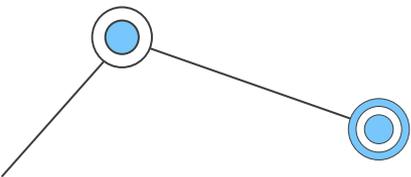
當你想要在一個主機操作系統 (Host OS) 上虛擬化另一個操作系統時，會出現以下問題：

- OS 認為自己負責管理所有硬體，這可能導致與其他虛擬化的 OS 發生衝突。

## 解決方案：

為了避免操作系統之間的衝突，必須妥善切分他們的資源：

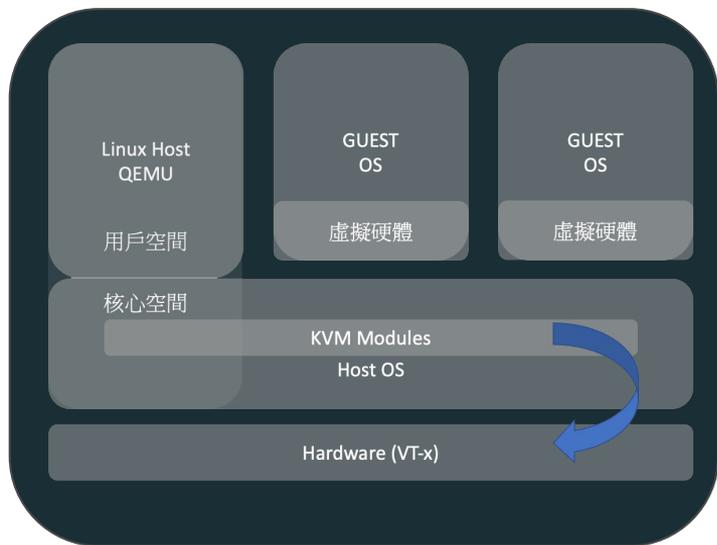
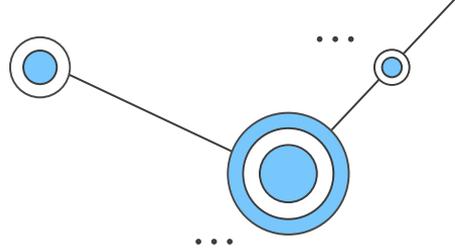
- CPU: 包括寄存器的管理
- Memory: MMU 的切換
- IO Device: 確保各自的設備不互相干擾
- Interrupt: 妥善處理中斷請求
- Timer



# 虛擬化技術介紹

	全虛擬化： 利用二進位翻譯	硬體輔助虛擬化 ：VT-x / VT-d 輔助	半虛擬化： 作業系統協助
實現技術	Binary Translation 和直接執行	遇到特權指令轉到 root 執行	Hypercall
Guest 兼容性	無修改 Guest 兼容性高	無修改 Guest 兼容性高	需修改 Guest 僅適用開源 OS
性能	差	僅切換模式的開銷	接近於物理機
應用廠商	VMware (Workstation) Virtual PC QEMU	VMware ESXi Microsoft Hyper-V Xen 3.0 KVM	Xen

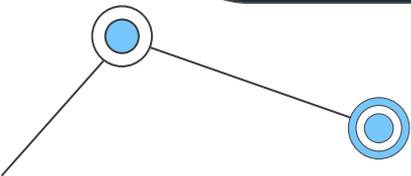
# KVM-QEMU 架構



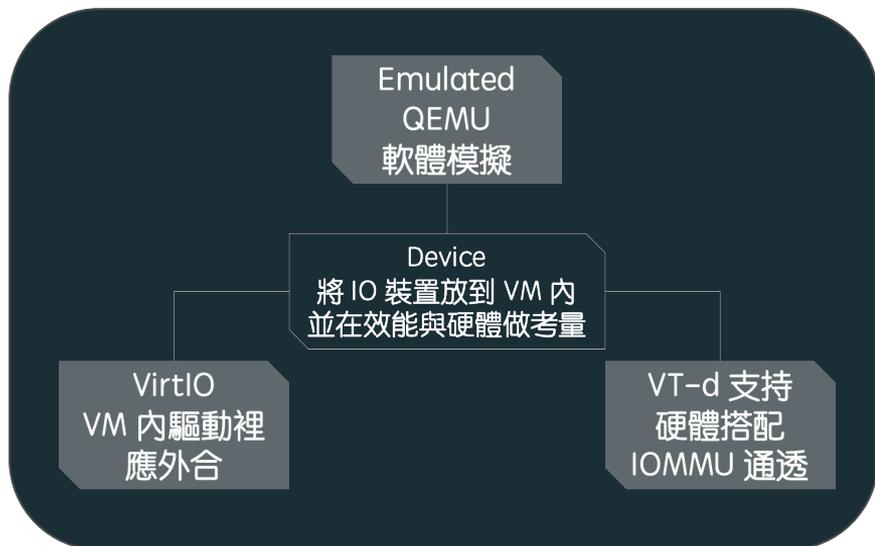
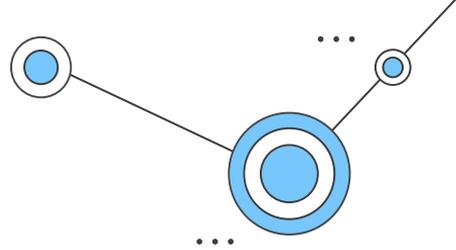
KVM (Kernel-based Virtual Machine) 是一個基於 Linux Kernel 的虛擬化解決方案，結合了 Linux Kernel 的模組化和 QEMU 的虛擬化功能。KVM 允許將 Linux Kernel 轉換為一個類似於 Hypervisor 的角色，使得主機系統可以運行多個虛擬機器。

QEMU 是一個開源的虛擬機器監視器和仿真器，它可以模擬多種硬體環境，包括處理器、記憶體和設備。KVM 通過與 QEMU 結合，實現了高性能的虛擬化解決方案，同時提供了完整的虛擬化功能，如動態記憶體管理、網路訪問控制等。

KVM-QEMU 架構的組合優勢在於其高效率和穩定性，使得用戶可以輕鬆建立和管理虛擬機器，同時實現資源的最佳化利用。



# PVE Device I/O 虛擬化



## 1. Emulated QEMU(軟體模擬)

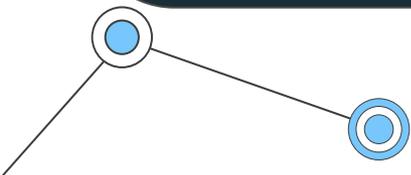
透過 QEMU 模擬硬體裝置來提供虛擬機使用  
軟體模擬會影響性能

## 2. VirtIO(VM 內驅動裡應外合)

使用 VirtIO 驅動程序來提升虛擬機內 I/O 效能  
需要在虛擬機內安裝對應的 VirtIO 驅動程序

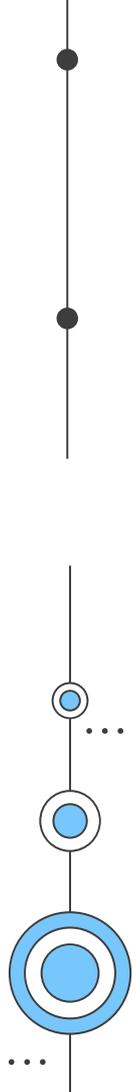
## 3. VT-d 支持(硬體搭配 IOMMU 通透)

利用 VT-d 技術將實體硬體直接分配給虛擬機  
需要硬體支持 IOMMU 技術





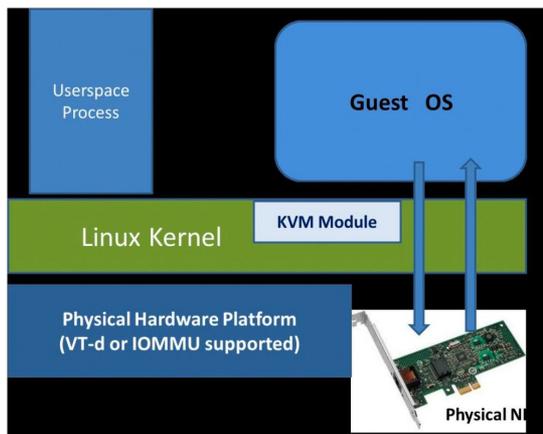
肆、  
私有雲主機  
服務案例



# 案例介紹 1

## 狀況：

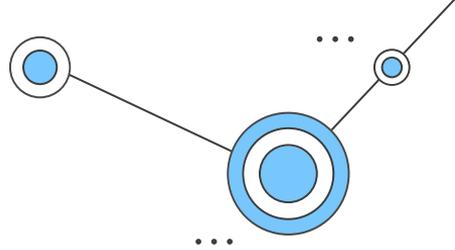
若學生想進行AI 研究，學生 A 及 B 各需要 2 張運算卡，但僅有一台伺服器。



## 解決方案

在 Proxmox VE 環境下，可建立 2 台虛擬機，並分別將 2 張計算卡使用 PCIe Passthrough 的方式傳入虛擬機中，兩個學生便擁有獨立的環境進行實驗。

## 案例介紹 2

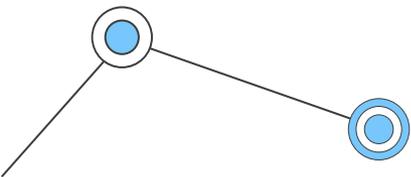


### 狀況:

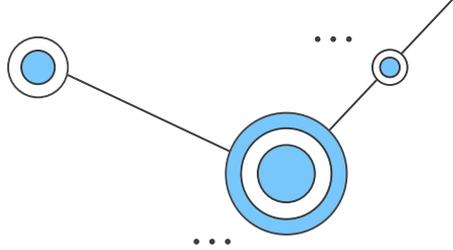
使用者需要測試新的硬體的軟體兼容性或性能, 如網路卡。

### 解決方案

在 Proxmox VE 環境下, 可用建立多台虛擬機, 並分別將網卡 passthrough 進去 VM, 可以在僅有一台實體機的情況下測試網卡運作。



# 案例介紹 3

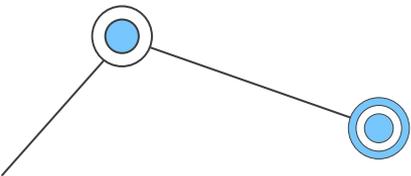


## 狀況：

使用者需要大量伺服器模擬大型系統，例如需要多組 4 nodes cluster。

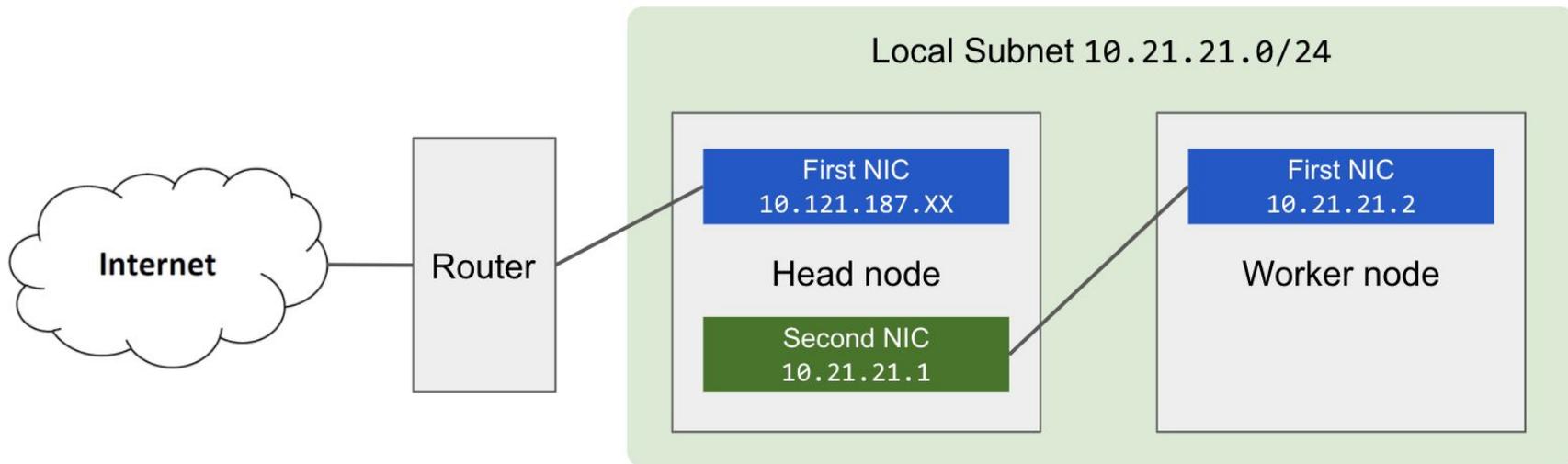
## 解決方案

在 Proxmox VE 環境下，可用腳本建立多台虛擬機，使用 cloud-init 等方式快速部署多台虛擬機，並使用 VLAN 等功能切分 L2 網路，供使用者測試。

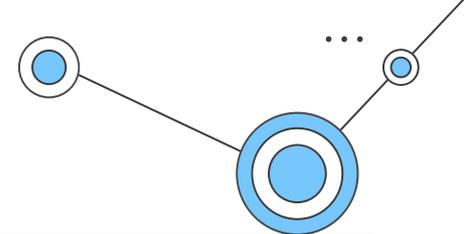


# Cluster Topology

- There are two node each group.
- Head node is the login node. It's responsible for hold userdata (/home), NIS server and forward IP.



# Batch scripts



training-camp-pveenv Public

Edit Pins Watch 1 Fork 1 Star 1

main 3 Branches 0 Tags

Go to file Add file <> Code

About

setting vm env in PVE

- Readme
- Activity
- Custom properties
- 1 star
- 1 watching
- 1 fork

Report repository

Releases

No releases published  
[Create a new release](#)

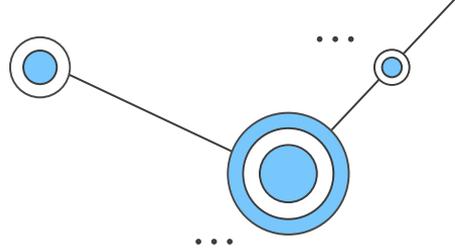
Packages

William-Mou Merge pull request #1 from seadog007/patch-1 52941ec · 3 years ago 12 Commits

README.md	Fixed typo	3 years ago
adduser-sc.sh	Init	4 years ago
create-sc.sh	Update create-sc.sh	3 years ago
deluser-sc.sh	Init	4 years ago
remove-all-vm-sc.sh	Add start stop	4 years ago
start-all-vm-sc.sh	Add start stop	4 years ago
update.sh	Add Updated.sh	3 years ago
usbip.sh	Create usbip.sh	3 years ago



# 案例介紹 4

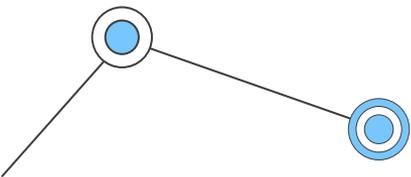


## 狀況:

使用者在實驗室需要連入某VM 進行網頁開發。

## 解決方案

在 Proxmox VE 搭配 Overlay Network 環境下, 即便 Proxmox VE 的 VM 屬於 X subnet, User 所在實驗室配發的是 B Subnet, 也能透過 BGP 的路由輕鬆連線到 Proxmox VE 虛擬機進行開發。



# 解決方案

在 Naive 的網路及伺服器架構下，使用者有以下困難：

1. 無法透過 SSH 連接內部網路設備
  - 埠轉發至外部網路(但有可能遭受攻擊)
  - 設定虛擬私人網路(需要頻繁切換)
2. 伺服器上臨時架設的網頁或服務難以連接
  - 請管理員設定埠轉發(較為繁瑣)
  - 使用 `ssh -R` 進行反向隧道連接
1. 硬體機器資源有限互相爭搶
  - 使用試算表手動排程(繁瑣)
  - 未滿載使用容易造成資源浪費
2. 軟體環境互相干擾
  - 使用虛擬環境(學習成本高)
  - 硬碟多系統開機(中斷成本高)

## L3 Overlay Network

- BGP over WireGuard
- Build with Mikrotik RouterOS
- Need to manage private AS, IPAM, and more ...

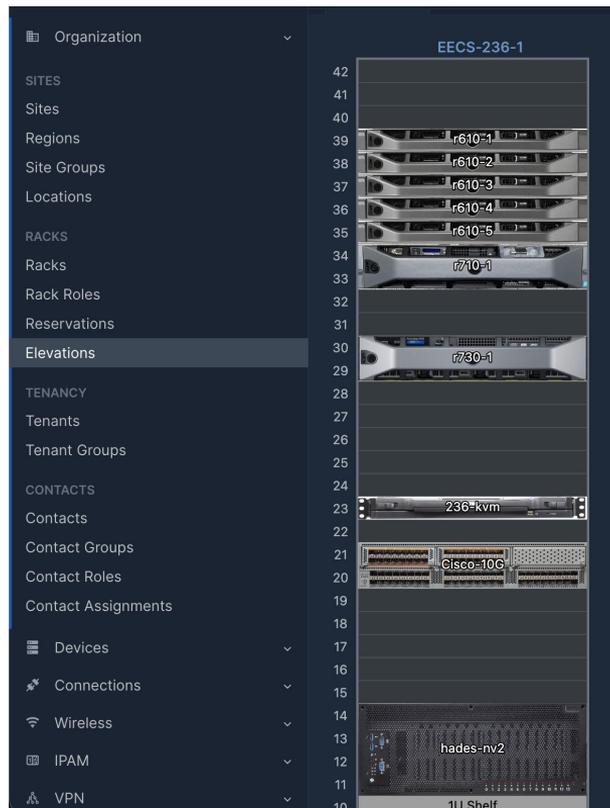
## Proxmox VE Cluster

- PVE Accounts Permission
- Virtual Machine: QEMU / KVM
- Virt-IO / VFIO / PCIe Passthrough
- Virtual Bridge / VLAN

# Future Work

在現有系統中新增加以下服務

1. [Netbox](#)
2. VXLAN
3. SSO + LDAP
4. Self host service



# Education

- DN42

dn42 

## A dynamic interconnected VPN.

dn42 can be used to learn networking and to connect private networks, such as hackerspaces or community networks. But above all, experimenting with routing in dn42 is fun!

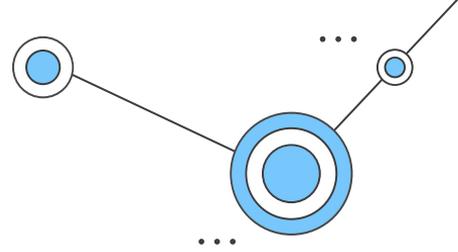
[Read the Wiki](#)

<https://dn42.us/>

- STUIX



<https://stuiX.io/>



# Resource & 感謝

- Overlay-network [Link](#) co-work 張意鴻
- PVE 虛擬化——給我一個邁出 Windows 的理由！ [Link](#)
- 在家機器學習？用虛擬化技術實現個人 AI 環境配置 [Link](#)

Jan Ho 的網絡世界 網絡知識教學網站

IP Address Version 4 (IPv4) 網際網路協定位址

26 October, 2016

## 目錄

前言

Network ID

Subnet Mask 子網路遮罩

Supernet

VLSM 可變長子網路遮罩

Private IP Address

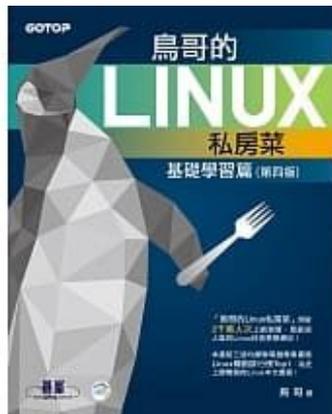
Private IP Address

VLSM 可變長子網路遮罩

Subnet Mask

Supernet 可變長子網路遮罩

Network ID



seadog007 (seadog007)

IT邦新手 5 級 · 點數 176

台灣數位串流有限公司維運暨開發工程師

個人背景	0	22
發問	發問	文章



鐵人檔案

• 第 12 屆 IT 邦幫忙鐵人賽

Security

從 0 開始的 Web Security 系列

參賽天數 22 天 | 共 22 篇文章 | 299 人訂閱 | 團隊 Seal and Friends

每篇天數 22 天 | 共 22 篇文章 | 299 人訂閱 | 團隊 Seal and Friends

從 0 開始的 Web Security 系列

Security

# Thanks!

Do you have any questions?

contact@mou.tw

+886 920 992 660

[www.mou.tw](http://www.mou.tw)



**CREDITS:** This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), infographics & images by [Freepik](#) and illustrations by [Stories](#)