

課程大綱

雲端運算的三大關鍵技術

Part 1 : Overview of Cloud Computing Core Technologies

IaaS : 虛擬化平台建置分享

以國網中心雲端平台為例

Part 2 : Introduction to NCHC Cloud WebOS & Ezilla

PaaS : 大量資料分析應用分享

以雲端入侵偵測日誌分析為例

Part 3 : Introduction to Cloud Security & ICAS

SaaS : 大量資料分析與網頁服務整合

以企業內網搜尋引擎為例

Part 4 : Introduction to Crawlzilla



雲端運算的三大關鍵技術

Part 1 : Overview of Cloud Computing Core Technologies

Jazz Wang
Yao-Tsung Wang
jazz@nchc.org.tw



Powered by DRBL

National Definition of Cloud Computing

美國國家標準局 NIST 給雲端運算所下的定義

5 Characteristics

五大基礎特徵

4 Deployment Models

四個佈署模型

3 Service Models

三個服務模式

1. On-demand self-service.

隨需自助服務

2. Broad network access

隨時隨地用任何網路裝置存取

3. Resource pooling

多人共享資源池

4. Rapid elasticity

快速重新佈署靈活度

5. Measured Service

可被監控與量測的服務

4 Deployment Models of Cloud Computing

雲端運算的四種佈署模型

Public Cloud

公用雲端



Microsoft

Google

**Dynamic Resource Provisioning
between public and private cloud**

私有雲端動態根據計算需求
調用公用雲端的資源

Target Market

is **S.M.B.**

主要客戶為
中小企業

Hybrid
Cloud

以大型企業
為主要客戶

**Enterprise is
key market**

Community Cloud

社群雲端

Academia 學術為主



私有雲端

Private Cloud

3 Service Models of Cloud Computing

雲端運算的三種服務模式 (市場區隔)

IaaS

Infrastructure as a Service

架構即服務

PaaS

Platform as a Service

平台即服務

SaaS

Software as a Service

軟體即服務



2 perspectives : Services vs Technologies

您想聽的是「雲端服務」還是「雲端技術」？

Google YouTube e W

amazon
web services™

雲端服務

Microsoft

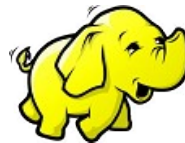
salesforce
SOFTWARE



KVM Xen



libvirt
VIRTUALIZATION API



雲端技術



Cloud computing hype spurs confusion, Gartner says

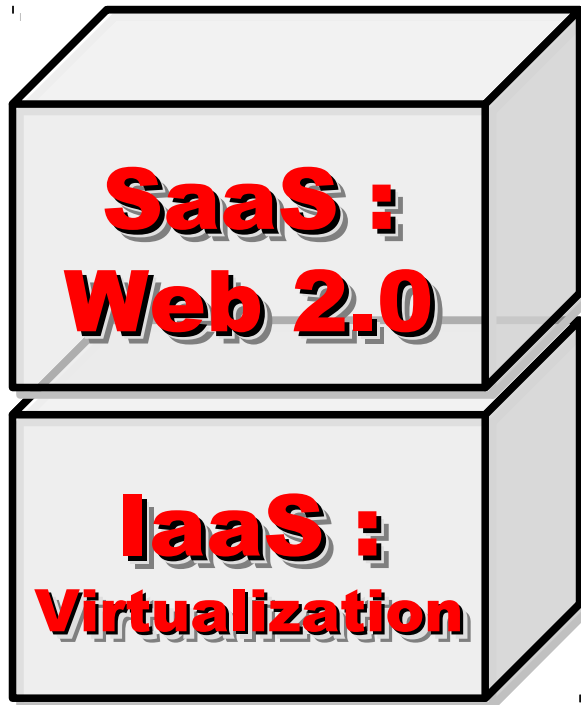
<http://www.computerworld.com/s/article/print/9115904>

淺談雲端運算 (Cloud Computing)

http://www.cc.ntu.edu.tw/chinese/epaper/0008/20090320_8008.htm

Two Type of Cloud Architecture ?

雲端架構的兩大陣營？



想盡辦法誘你用計算跟網路
Computing Intensive



想盡辦法誘你提供資料作分析
Data Intensive



IaaS : 虛擬化平台建置分享

以國網中心雲端平台為例

Part 2 : Introduction to NCHC Cloud WebOS & Ezila

Jazz Wang

Yao-Tsung Wang

jazz@nchc.org.tw



Powered by DRBL

What is Virtualization ??

虛擬化技術有哪些呢??

Application Virtualization

應用程式虛擬化

Desktop Virtualization
Client Virtualization

桌面虛擬化

Presentation Virtualization

顯示虛擬化

OS-level Virtualization

作業系統虛擬化

Network Virtualization

網路虛擬化

Storage Virtualization

儲存虛擬化

資料庫虛擬化

Database Virtualization

資料虛擬化

Data Virtualization

Virtualization and Power Management

虛擬化技術對機房用電管理的幫助

影片：<http://www.youtube.com/watch?v=Nkv0fhu-m2k>

You Tube

VMware Distributed Power Mgmt (DPM)

drummonds1974 6 部影片



1 Server On: ~520 W
3 At Standby: ~60W

0:28 / 2:28 360p

19,484

drummonds1974 於 2008-11-05 上傳

(This is a re-post of a video whose content cannot be displayed on the original page due to an bug identified by YouTube. The original video: <http://www.youtube.com/watch?v=7CbRS0GGuNc>)

18 人喜歡, 0 人不喜歡

藝術家/表演者: Franz Ferdinand

Desktop Virtualization and Thin Client

桌面虛擬化技術對未來辦公環境的改變

影片：<http://www.youtube.com/watch?v=XuYh95y9ROU>



The screenshot shows a YouTube video player interface. The video title is "Using an iPad to connect to your Work Desktop with the iPhone TouchPad I". The channel name is "ShaneTechify" with 9 videos and a subscribe button. The video player shows a person's hand holding an iPhone TouchPad, which is connected to an iPad displaying a web application. The video progress is at 1:34 / 3:08. The video has 45,919 likes and 1 dislike. The video description is "Citrix XenDesktop empowers the remote and mobile workforce on a wide" with a "顯示更多" link.

YouTube

Using an iPad to connect to your Work Desktop with the iPhone TouchPad I

ShaneTechify 9 部影片 訂閱

建議

1:34 / 3:08

45,919

16 人喜歡, 1 人不喜歡

Citrix XenDesktop empowers the remote and mobile workforce on a wide

顯示更多

Cloud WebOS

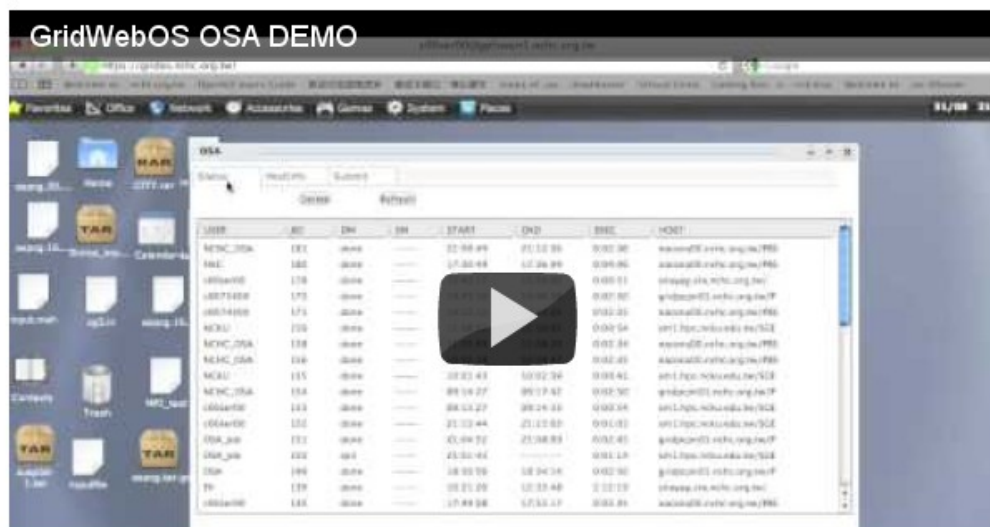
國網中心雲端作業系統

影片：<http://www.youtube.com/watch?v=LZyoG3UAX2U>



Grid WebOS Project

Single Sign-On, CrossPlatform, Widgets



雲端虛擬叢集
系統 線上使用



Cloud WebOS



Visitors
199
11
FLAG

影片：<http://www.youtube.com/watch?v=bw1GpYt7df4>

The screenshot shows the Ezila website's navigation menu with 'Projects' selected. The main content area features a heading for 'Projects' and a sub-heading for '雲端虛擬叢集環境隨選系統 (線上使用手冊)'. Below this is the title 'On Demand Virtual Cluster in Cloud Web-Based OS Environment'. A paragraph describes the Percomp Lab's mission to provide a user-friendly, efficient platform for virtualized environments. A numbered list starts with '1. 虛擬化系統提供者:' followed by text explaining the benefits of virtualization, such as increased system usage and reduced costs.

Home News **Projects** Downloads Documents Presentation Publication Contact us

Grid WebOS Lite Migration On Demand Virtual Cluster

Projects

雲端虛擬叢集環境隨選系統 (線上使用手冊)

On Demand Virtual Cluster in Cloud Web-Based OS Environment

Percomp Lab 致力於提供用者者一個簡易、友善的使用者介面、高效能的使用平台，以服務廣大的使用者。為提高系統使用率及節能省碳，導入虛擬化技術，開發出「雲端虛擬叢集環境隨選系統」，讓每台實體機器利用率提高，並且提供使用者一個完全屬於自己的叢集環境以達到最佳個人化使用平台。不論是虛擬化系統提供者或虛擬化使用者皆可在虛擬化過程中，獲得顯著的益處，因此，積極投入建置虛擬叢集環環的開發。使用虛擬化後，提供者與使用者的優勢，如下所述。

- 1. 虛擬化系統提供者：**

導入虛擬化技術可提高系統使用率及達到節能省碳效果，
讓每台實體機器利用率提高，減少實體機器閒置及添購設備成本。

國立高階網路與計算中心
National Center for High-Performance Computing

雲端虛擬叢集環境隨選系統 線上使用手冊

計算資源

Cloud WebOS

Visitors

	199		2
	11		1

FLAG counter

相關資訊：<http://percomp.nchc.org.tw/index.php/Projects>



PaaS : 大量資料分析應用分享

以雲端入侵偵測日誌分析為例

Part 3 : Introduction to Cloud Security & ICAS

Jazz Wang

Yao-Tsung Wang

jazz@nchc.org.tw



Powered by DRBL

Big Data Analysis : Social Computing & Business Intelligence

「社交運算」與「商業智慧」均仰賴大量資料分析

DIGITIMES 網站內容的著作權為大椽股份有限公司 (DIGITIMES Inc.) 所有, 或其他授權DIGITIMES使用的內容提供者所有。

使用者下載或拷貝網站的內容或服務僅限於供個人、非商業用途之使用, 但不得以任何形式傳輸、重製、散布或提供予公眾。使用人利用時必須遵守著作權法的所有相關規定, 不可變更、發行、播送、轉賣、重製、改作、散布、表演、展示或利用DIGITIMES所屬網站上局部或全部內容及服務以賺取利益。

提升商業分析效果 資料倉儲業提倡資料社交化

2010/10/27 - DIGITIMES 馬培治 / 台北

社交運算(social computing)隨著包括Facebook在內的各式社交網絡服務持續發燒, 也成為企業資訊系統發展的重點之一, 繼IBM、微軟(Microsoft)與甲骨文(Oracle)等大廠提倡在應用軟體功能上支援社交功能之後, 資料倉儲(Data Warehouse)業者Teradata則提倡企業資料分析, 應納入包括社交資訊在內的多元因子, 讓不同資料源間的資料「社交化」(socialization of data), 以增強商業分析效果, 提高掌握用戶行為並輔助商業決策。

Teradata業務發展暨行銷執行副總裁Darryl D. McDonald於25日在自家全球合作夥伴與使用者大會上表示, 除了傳統企業營運資料, 各種可用來擷取資訊的資料源, 如RFID、智慧型裝置、社交網路, 乃至各種感應器, 將會對現今的企業分析帶來龐大的衝擊, 他建議企業可以開始著手思考, 如何將這些新興資料源的資料與傳統商業智慧分析的資料進行整合, 以期從更豐富的資料中, 找出過去商業分析方法看不到的隱性資訊。

McDonald表示, Facebook目前已經擁有超過5億個註冊用戶, 而推特(Tweeter)每天也有超過8,500萬條訊息產生, 若企業能夠將自身的用戶資訊或營運資料與這些龐大的資訊源進行有意義的分析, 將能夠激發在商業分析領域的創新應用。

他以參加Teradata全球合作夥伴暨使用者大會的3,000多名與會者為例進行分析, 發現這些與會者代表的公司總計具有9兆美元的資本額, 以及合計達230萬個線上社交網路服務的人際連結數, McDonald說, 這些資訊代表龐大的商機, 以及可供未來利用在業務推廣、行銷等目的使用。

參考來源：提升商業分析效果 資料倉儲業提倡資料社交化 (2010/10/27)

<http://goo.gl/2GoMo>

中華電信用 Hadoop 技術分析通話明細



郵件、資料庫、防火牆...
輕鬆解決企業 IT 資源需求

[iThome週週為IT人打氣!](#)

[雲端伺服器首選, 半年免費](#)

[企業選平板? 選最相容的!](#)

新聞

新聞專題

即時新聞

新聞簡訊

技術

產品報導

技術專題

IT書訊

IT管理

CIO

IT人物

專欄

新聞總覽

業界動態

訂閱電子報

中華電信用Hadoop技術分析通話明細

文/辜雅菴 2011-06-12



62 人說讚。快免費註冊來查看你的朋友對什麼說讚。

[+ 我要收藏](#)

中華電信利用自行開發的Hadoop大資料運算平臺，找出非結構化資料中的結構性，精簡資料後再置於資料倉儲運算，節省儲存空間

面對資料快速成長以及非結構性資料的增加，中華電信資訊處第四科科長楊秀一表示，中華電信近來利用Hadoop雲端運算技術自行開發了一個專門用來分析非結構化資料的巨量資料 (Big Data) 運算平臺，嘗試在資料進到資料倉儲系統之前，先進行資料的分析與處理以減少資料倉儲的資料量。

近年來行動語音市場趨於飽和，為了掌握用戶特性進行客製化行銷，一份資料要進行分析，就會被多次複製，因此即使用戶增加趨緩，但中華電信擁有的資料量仍快速暴增。

研討會訊息

[Websense TRITON電子郵件資料安全解決方案研討會](#)

[2011 JavaTWO專業技術大會](#)

[+更多研討會](#)

▼ ADVERTISEMENT ▼

Microsoft

Three Core Technologies of Google

Google 的三大關鍵技術

- Google 在一些會議分享他們的三大關鍵技術
- Google shared their design of web-search engine
 - SOSP 2003 :
 - “The Google File System”
 - <http://labs.google.com/papers/gfs.html>
 - OSDI 2004 :
 - “MapReduce : Simplified Data Processing on Large Cluster”
 - <http://labs.google.com/papers/mapreduce.html>
 - OSDI 2006 :
 - “Bigtable: A Distributed Storage System for Structured Data”
 - <http://labs.google.com/papers/bigtable-osdi06.pdf>



Open Source Mapping of Google Core Technologies

Google 三大關鍵技術對應的自由軟體

BigTable

A huge key-value datastore

HBase, Hypertable
Cassandra,

MapReduce

To parallel process data

Hadoop MapReduce API
Sphere MapReduce API, ...

Google File System

To store petabytes of data

Hadoop Distributed File System (HDFS)
Sector Distributed File System

更多不同語言的 MapReduce API 實作：

<http://trac.nchc.org.tw/grid/intertrac/wiki%3Ajazz/09-04-14%23MapReduce>

其他值得觀察的分散式檔案系統：

- IBM GPFS - <http://www-03.ibm.com/systems/software/gpfs/>
- Lustre - <http://www.lustre.org/>
- Ceph - <http://ceph.newdream.net/>

Who Use Hadoop ??

有哪些公司在用 **Hadoop** 這套軟體 ??

- **Yahoo** is the key contributor currently.
- **IBM** and **Google** teach Hadoop in universities ...
- http://www.google.com/intl/en/press/pressrel/20071008_ibm_univ.html
- **The New York Times** used **100 Amazon EC2 instances** and a Hadoop application to process **4TB of raw image TIFF data** (stored in S3) into **11 million finished PDFs** in the space of **24 hours** at a computation cost of about **\$240** (not including bandwidth)
 - from <http://en.wikipedia.org/wiki/Hadoop>
- <http://wiki.apache.org/hadoop/AmazonEC2>
- <http://wiki.apache.org/hadoop/PoweredBy>
 - A9.com
 - ADSDAQ by Contextweb
 - EHarmony
 - Facebook
 - Fox Interactive Media
 - IBM
 - ImageShack
 - ISI
 - Joost
 - Last.fm
 - Powerset
 - The New York Times
 - Rackspace
 - Veoh
 - Metaweb

Hadoop in production run

商業運轉中的 **Hadoop** 應用

- February 19, 2008
- Yahoo! Launches World's Largest Hadoop Production Application
- <http://developer.yahoo.net/blogs/hadoop/2008/02/yahoo-worlds-largest-production-hadoop.html>

Number of links between pages in the index	roughly 1 trillion links
Size of output	over 300 TB, compressed!
Number of cores used to run single Map-Reduce job	over 10,000
Raw disk used in the production cluster	over 5 Petabytes

專家說：雲端每個環節都有安全問題

ZDNet Taiwan - 專家談雲端：每個環節都有安全問題 - 新聞

2010/08/10 19:50:02

專家談雲端：每個環節都有安全問題

ZDNet記者曠文濤／台北報導 雲端的安全問題不是無解，只是不管是雲端服務供應商或者想要建立私有雲的企業用戶，都必須考量到每個環節。

微軟亞太區全球技術支援中心專案經理、同時也是ZDNet專欄作家林宏嘉今（10）日在ZDNet舉行的IT Priorities圓桌論壇中表示，**雲端的安全議題涉及了IaaS、PaaS乃至於SaaS的每個層面**，當然有些問題是原本就存在：例如在討論到IaaS時，就涉及到了**機房的管理**和**硬體設備的可用性**等；但是講到PaaS時，企業用戶倘若要選擇開原碼的作業系統，必須考量到後續的**安全維護**；在SaaS的層次，企業用戶必須確保每一個分區（partition）的安全更新和**資料安全**。

目前正如火如荼建立台灣第一個校園私有雲的台大計算機及資訊網路中心主任孫雅麗則呼應道，Amazon的雲端服務證實了在Hypervisor層有駭客入侵，也就是意味著過去大家在討論如何防範**虛擬機器的資料安全**，但是威脅已經深化到了更下一層。這些問題都有待解決。

「有些問題甚至是來自於內部，舉例而言，MIS可能會把存在記憶體裡的資料倒出來，或者在Hypervisor層就植入了可以蒐集資料的程式，」孫雅麗說。

安全議題是目前台灣企業對雲端持保留態度的最大主因，這也是何以台灣的大型企業對於雲端的想法，還是
仍好建立私有雲。畢竟對用戶而言，資料放在別人家，還和其他企業，甚至具競爭對手「共處」，

先來談談「端的安全」

用雲端
處理資安

**Dealing Security
issues using Cloud**

**Data Security
In the Cloud**

雲內部
的資安管制

**Security Issues
Inside the Cloud**

雲端資料
安全性

端本身
的資安威脅

**Security Threats
to Internet of Things**

以前你只有電腦需要防毒，現在



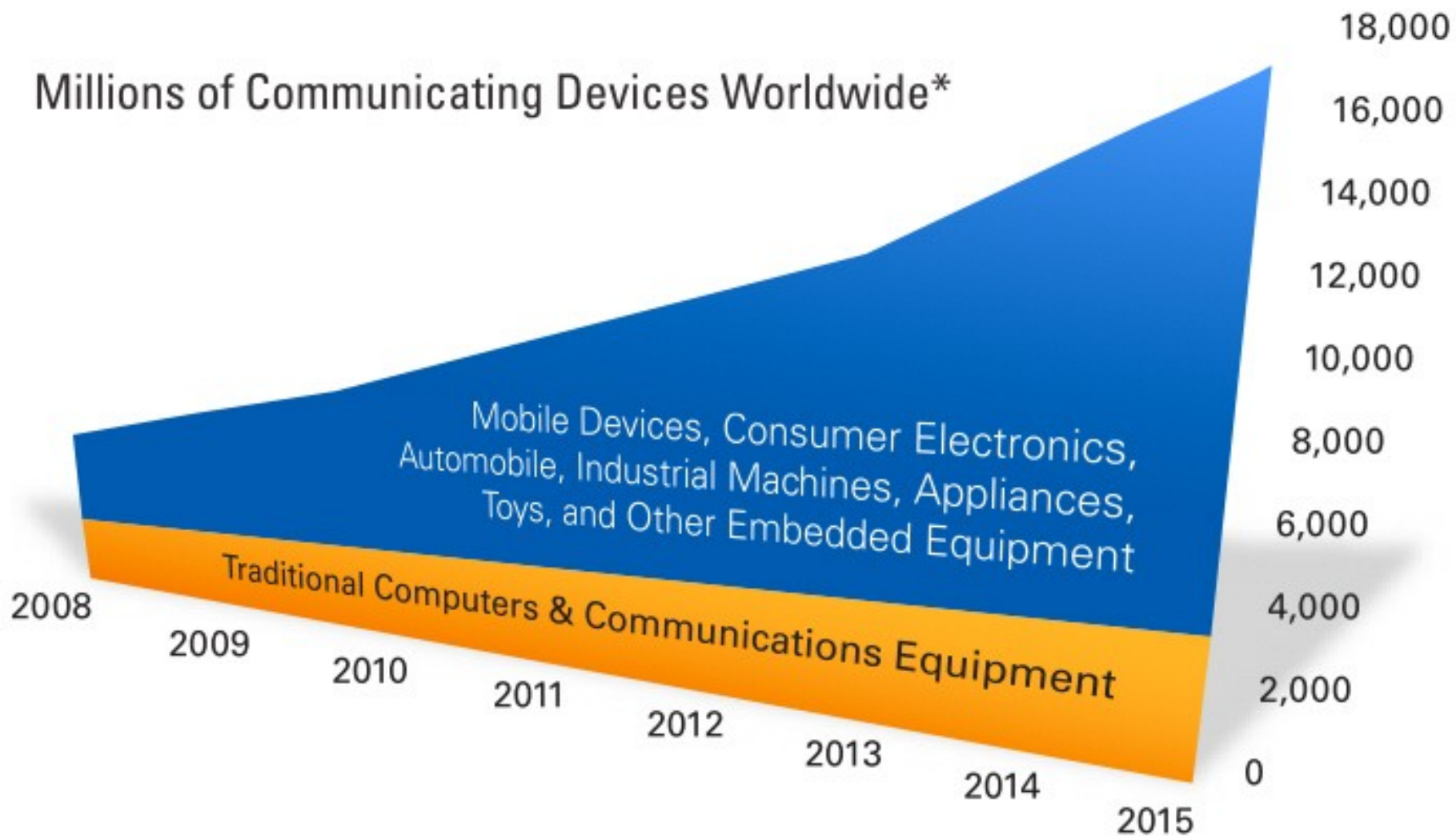
端

symbian OS



多元，中小廠
Diversify，
SMB

全球連網裝置急速成長中



Source: IDC Device Base Model, 2009

*Excludes voice- and SMS-only phones

物聯網的時代來臨

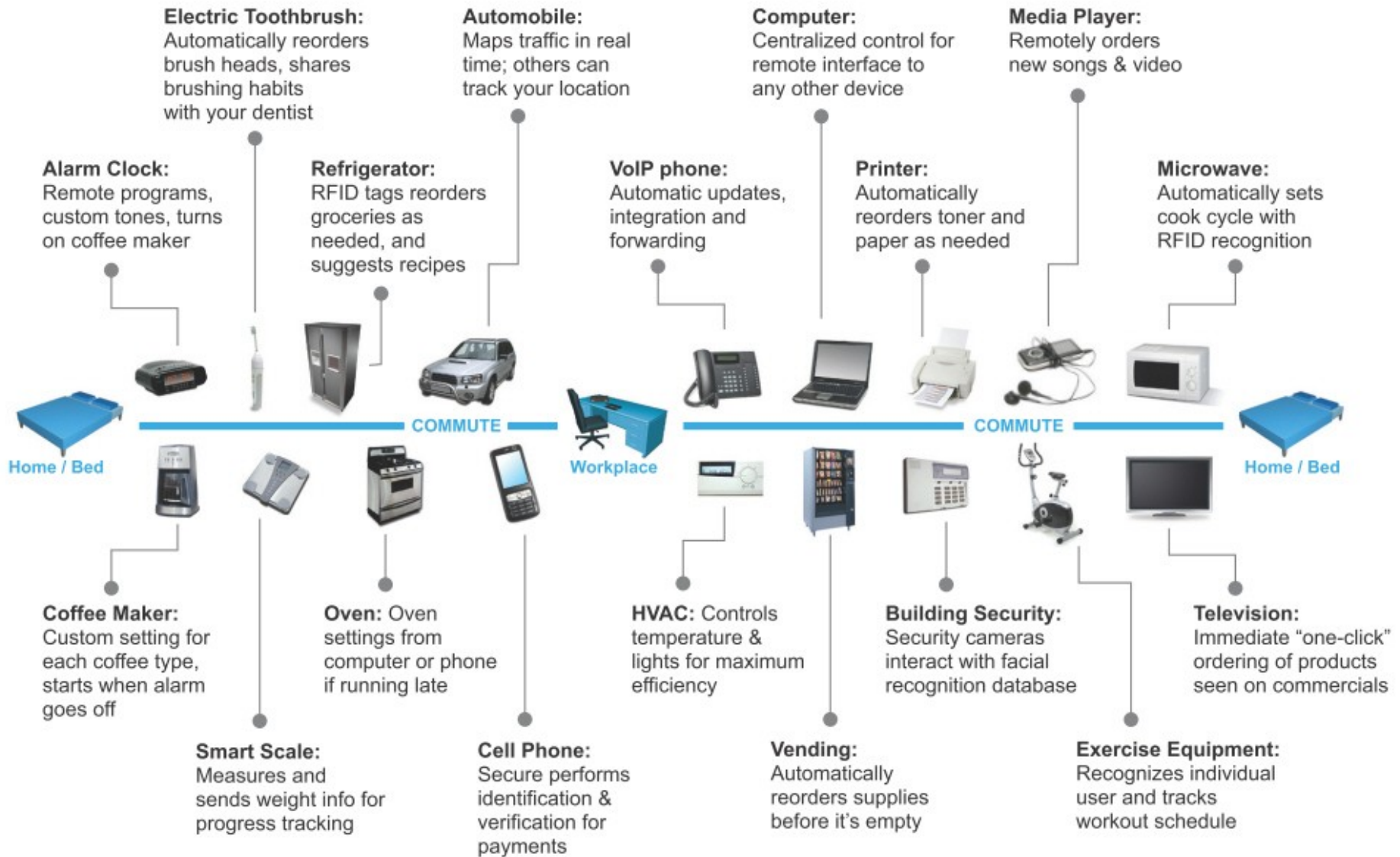
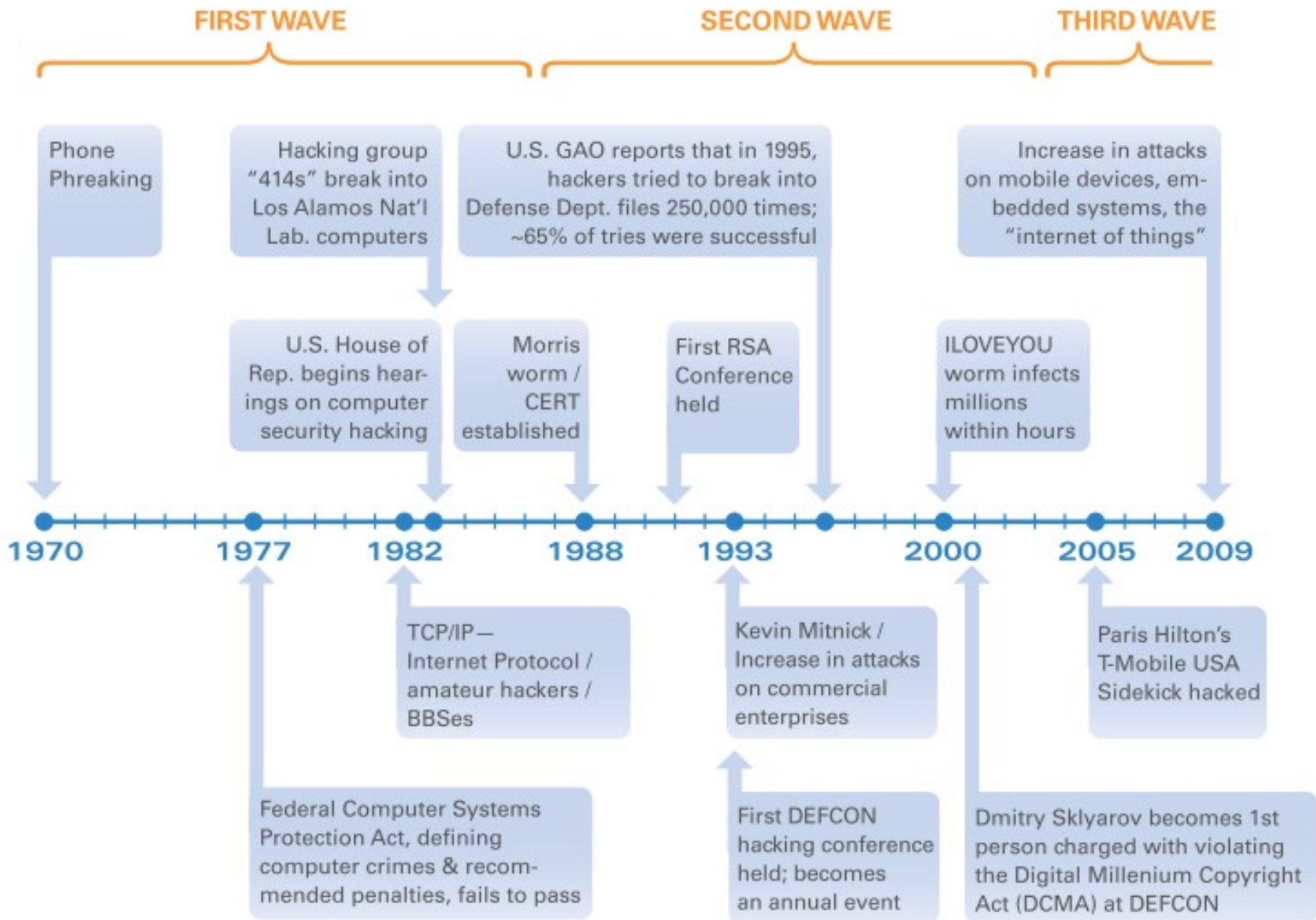


Figure 3. The Internet of Things

第三波網路入侵對象將鎖定在『物聯網』



圖片來源：Attacks on Mobile and Embedded Systems: Current Trends by Mocana

針對行動裝置的各種資安問題與經驗

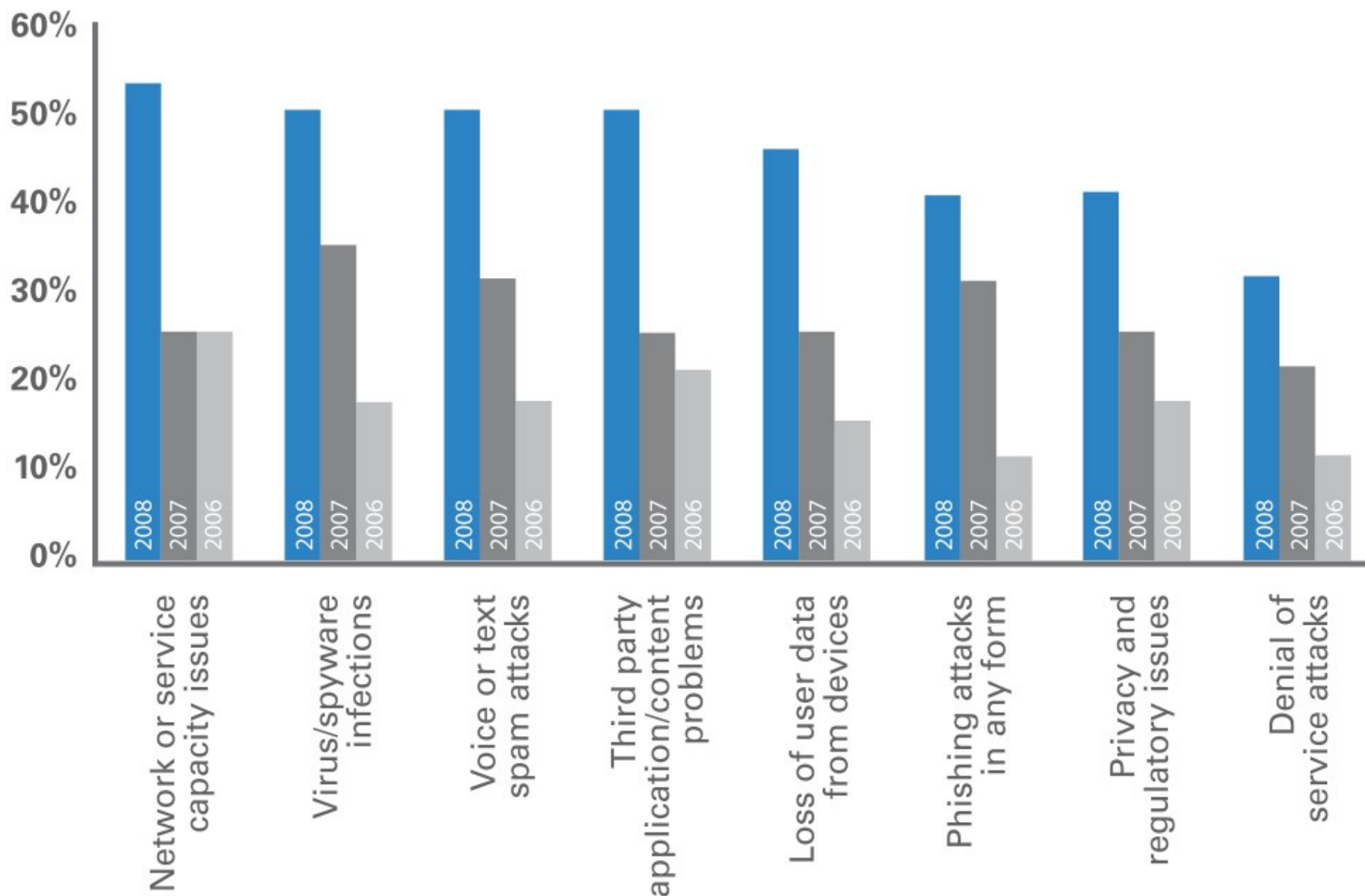
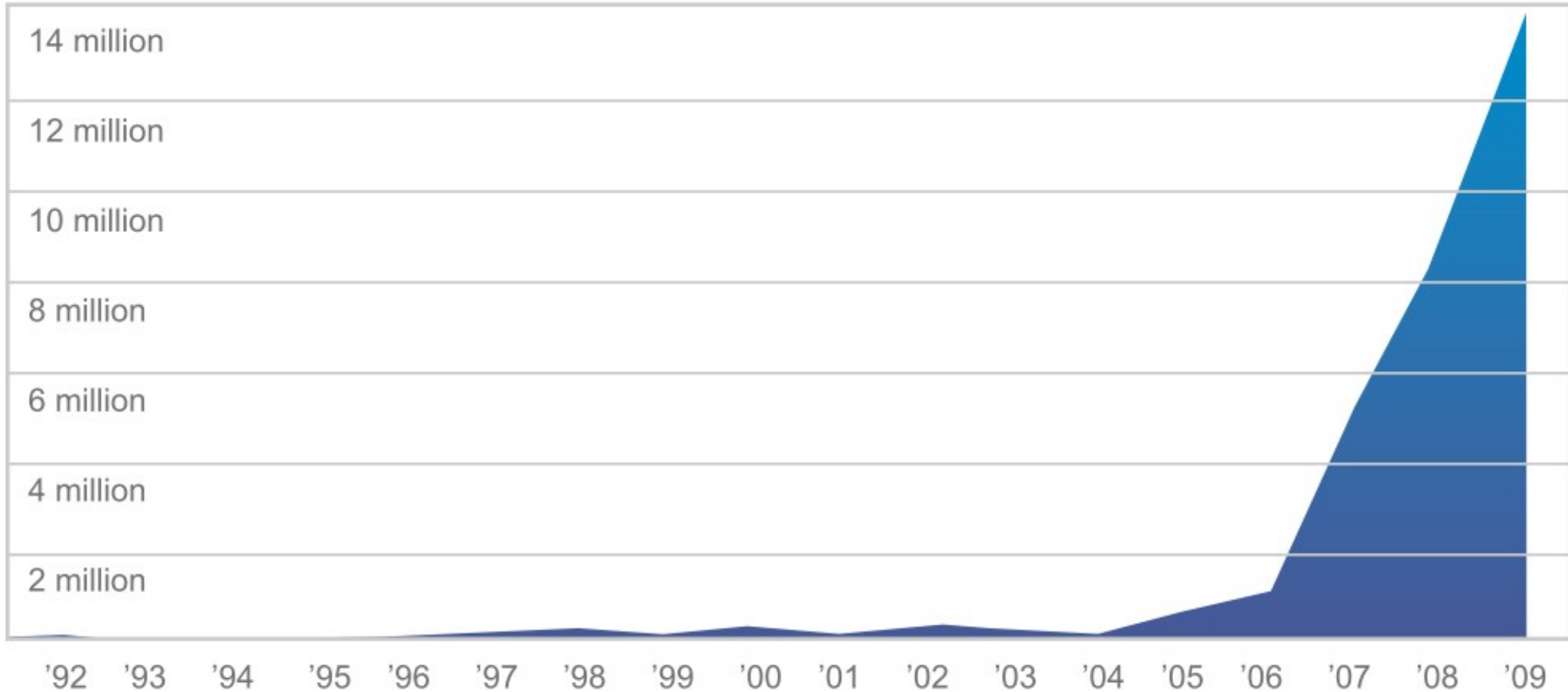


Figure 6. The increase in security issues experienced by mobile device users from 2006 to 2008; % of respondents. McAfee *Mobile Security Report 2009*

圖片來源：[Attacks on Mobile and Embedded Systems: Current Trends by Mocana](#)

網路惡意程式 (Malware) 逐年激增

Malware detected by year



Over 3,000 new "species" of PC malware are released onto the Internet every hour. Now that malware is setting its sights on Device platforms.

Source: AV LABS

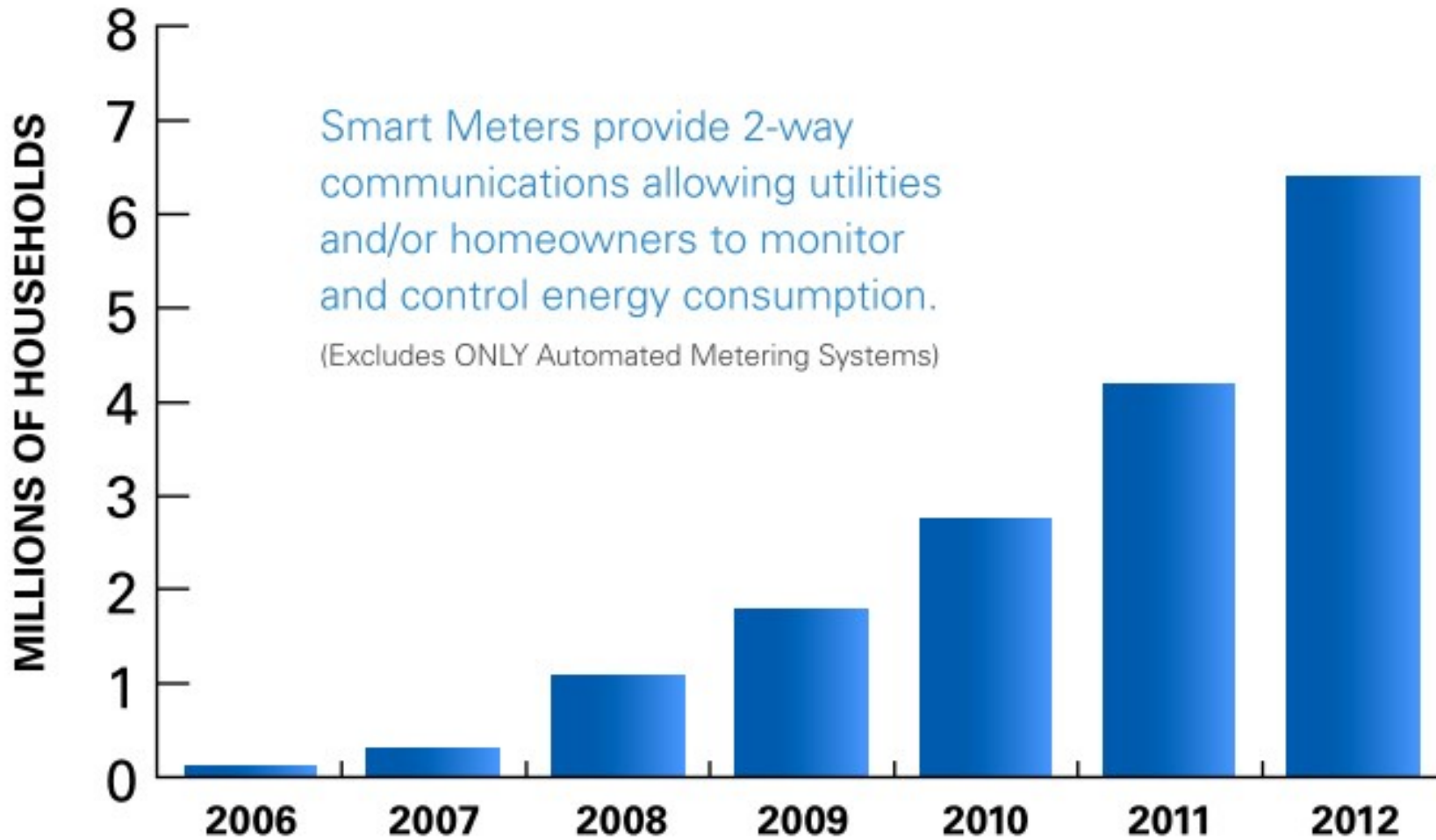
圖片來源：

U.S. Unprepared for Internet Device Flood: Unaddressed Security Problems & Talent Drought Threaten Long-Term Commercial, Government Interests

By: Kurt Stammberger, CISSP, Adrian Turner and Mat Small, Mocana With: Rich Nass, Sarah Friar, Goldman Sachs

如果你家的智慧電錶被入侵會怎樣？

U.S. Households with Smart Meters



© Copyright 2009 - Parks Associates

圖片來源：

U.S. Unprepared for Internet Device Flood: Unaddressed Security Problems & Talent Drought Threaten Long-Term Commercial, Government Interests
By: Kurt Stammberger, CISSP, Adrian Turner and Mat Small, Mocana With: Rich Nass, Sarah Friar, Goldman Sachs

再來談談「雲的安全」

用雲端
處理資安

**Dealing Security
issues using Cloud**

**Data Security
In the Cloud**

雲內部
的資安管制

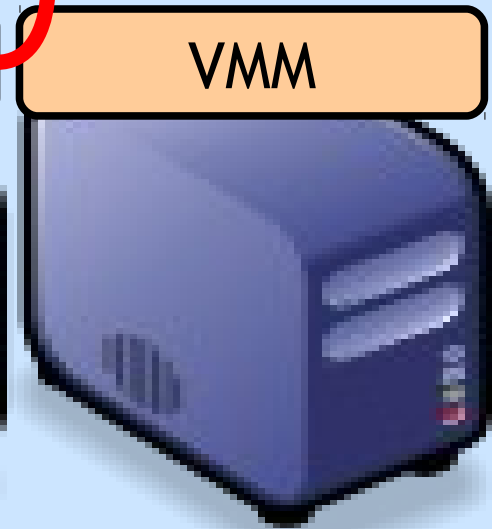
**Security Issues
Inside the Cloud**

雲端資料
安全性

端本身
的資安威脅

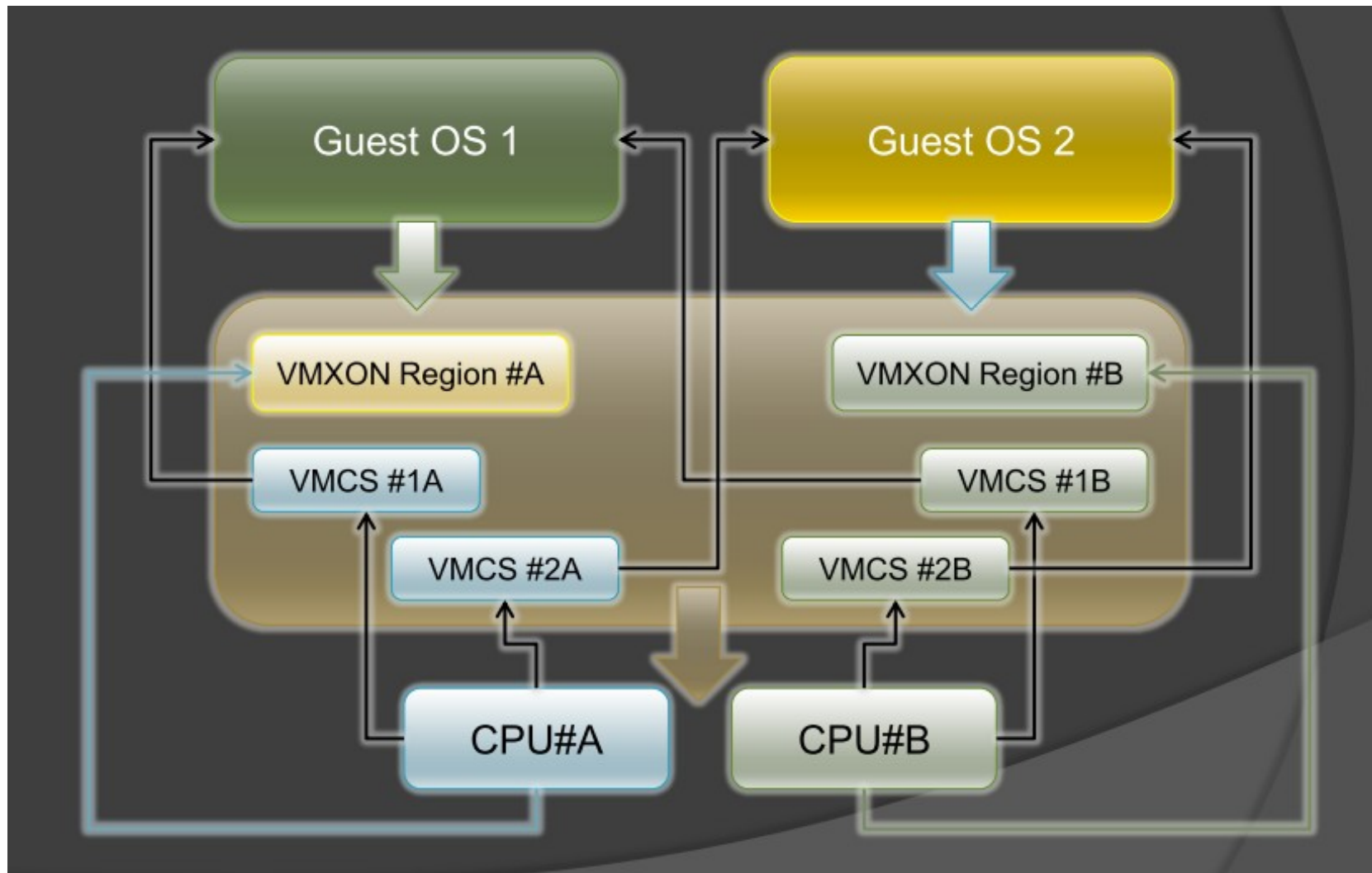
**Security Threats
to Internet of Things**





虛擬化衍生的新興資安問題

透過虛擬機器，竊取鍵盤輸入、植入後門.....



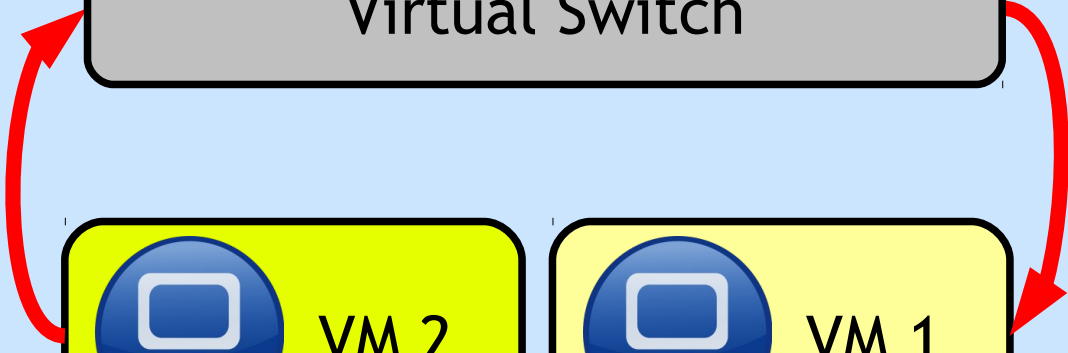
圖片來源： Hacks in Taiwan Conference 2010

http://www.hitcon.org/hit2010/download/6_New%20Battlefield%20For%20Malware%20Game.pdf

王大寶 & PK / Hypervisor - New Battlefield For Malware Game 虛擬機 - 惡意程式攻防的新戰場



Virtual Switch



A yellow rounded rectangle containing a blue computer icon and the text "VM 2".

A yellow rounded rectangle containing a blue computer icon and the text "VM 1".

VMM

A blue server rack icon representing a Virtual Machine Monitor (VMM).

VMM

A blue server rack icon representing a Virtual Machine Monitor (VMM).

VMM

A blue server rack icon representing a Virtual Machine Monitor (VMM).

VMM

A blue server rack icon representing a Virtual Machine Monitor (VMM).

三談「資料安全」

用雲端
處理資安

**Dealing Security
issues using Cloud**

**Data Security
In the Cloud**

雲內部
的資安管制

**Security Issues
Inside the Cloud**

雲端資料
安全性

端本身
的資安威脅

**Security Threats
to Internet of Things**

Ex. 無名照片外流、臉書個資外洩

轟動一時黑澀會妹妹容瑄親密自拍照片外流

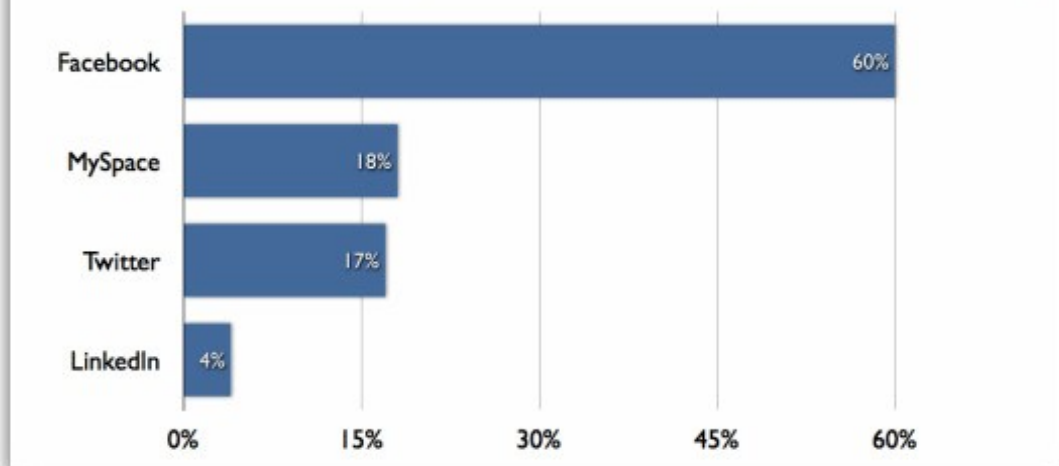


圖片來源：

[Wikileaks and Facebook Privacy / Security: Do we](#)



Which social network do you think poses the biggest risk to security?



圖片來源：

Report Ranks Facebook As Greatest Corporate Security Risk
<http://www.allfacebook.com/facebook-corporate-risk-2010-02>

進入今天的主題：用雲端處理傳統資安問題

用雲端
處理資安

**Dealing Security
issues using Cloud**

**Data Security
In the Cloud**

雲內部
的資安管制

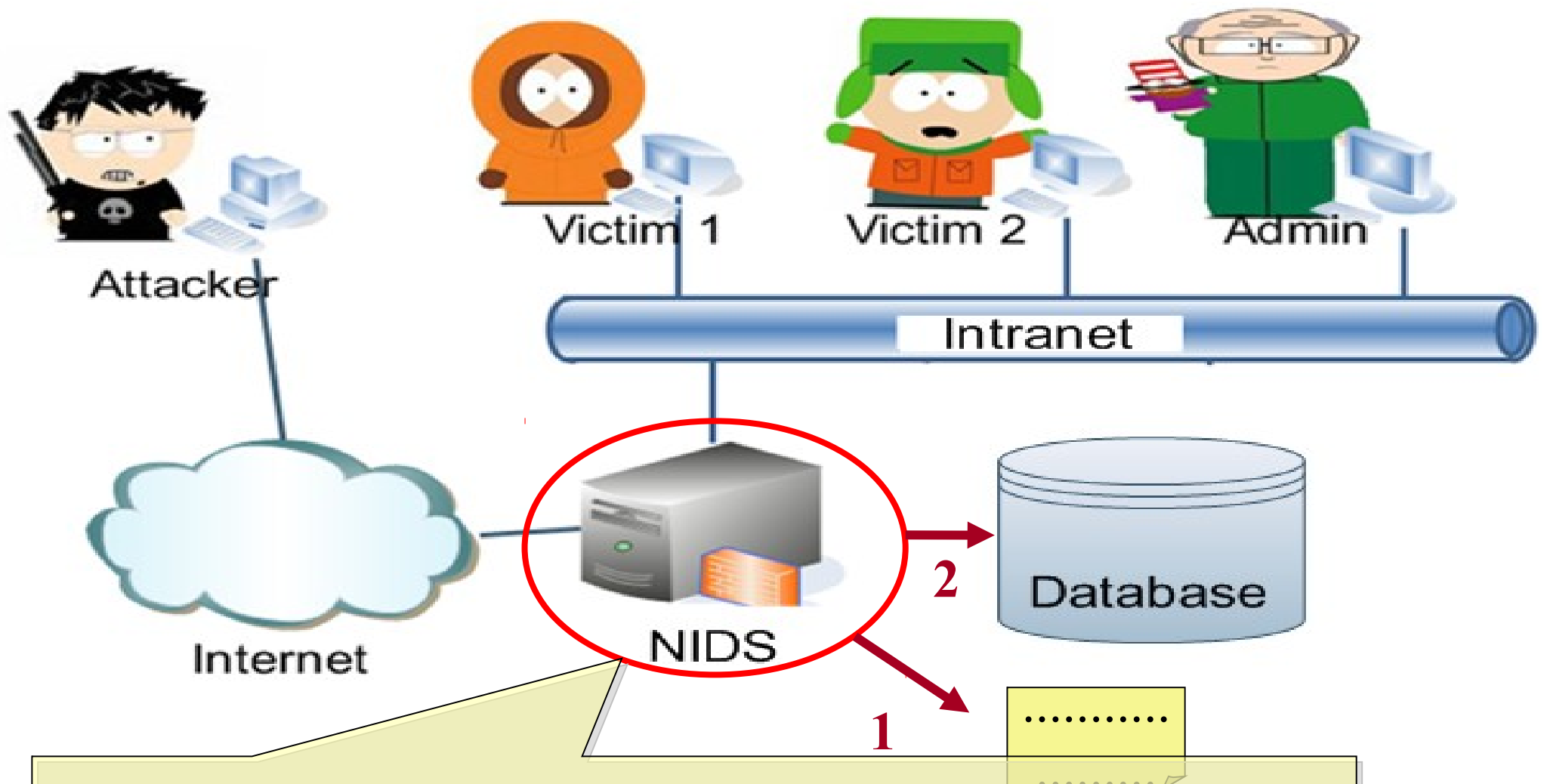
**Security Issues
Inside the Cloud**

雲端資料
安全性

端本身
的資安威脅

**Security Threats
to Internet of Things**

使用入侵偵測系統 (NIDS) 來找出入侵訊息



當入侵偵測系統偵測到網路上有異常封包時，就會產生警訊以告知有攻擊發生。警訊通常有兩種形式：
1. 紀錄成 log 檔 2. 紀錄到資料庫

傳統 NIDS 的警訊型態 (1) 紀錄在日誌檔內

入侵偵測系統所產生警訊日誌檔內一小段內容

```
[**] [1:538:15] NETBIOS SMB IPC$ unicode share access [**]  
[Classification: Generic Protocol Command Decode] [Priority: 3]  
09/04-17:53:56.363811 168.150.177.165:1051 -> 168.150.177.166:139  
TCP TTL:128 TOS:0x0 ID:4000 IpLen:20 DgmLen:138 DF  
***AP*** Seq: 0x2E589B8 Ack: 0x642D47F9 Win: 0x4241 TcpLen: 20
```

```
[**] [1:1917:6] SCAN UPnP service discover attempt [**]  
[Classification: Detection of a Network Scan] [Priority: 3]  
09/04-17:53:56.385573 168.150.177.164:1032 -> 239.255.255.250:1900  
UDP TTL:1 TOS:0x0 ID:80 IpLen:20 DgmLen:161  
Len: 133
```

```
[**] [1:1917:6] SCAN UPnP service discover attempt [**]  
[Classification: Detection of a Network Scan] [Priority: 3]  
09/04-17:53:56.386910 168.150.177.164:1032 -> 239.255.255.250:1900  
UDP TTL:1 TOS:0x0 ID:82 IpLen:20 DgmLen:161  
Len: 133
```

```
[**] [1:1917:6] SCAN UPnP service discover attempt [**]  
[Classification: Detection of a Network Scan] [Priority: 3]  
09/04-17:53:56.388244 168.150.177.164:1032 -> 239.255.255.250:1900  
UDP TTL:1 TOS:0x0 ID:84 IpLen:20 DgmLen:161  
Len: 133
```

```
[**] [1:538:15] NETBIOS SMB IPC$ unicode share access [**]  
[Classification: Generic Protocol Command Decode] [Priority: 3]  
09/04-17:53:56.405923 168.150.177.164:1035 -> 168.150.177.166:139  
TCP TTL:128 TOS:0x0 ID:94 IpLen:20 DgmLen:138 DF  
***AP*** Seq: 0x82073DFF Ack: 0x2468EB82 Win: 0x4241 TcpLen: 20
```

```
[**] [1:1917:6] SCAN UPnP service discover attempt [**]  
[Classification: Detection of a Network Scan] [Priority: 3]  
09/04-17:53:56.417045 168.150.177.164:45461 -> 168.150.177.1:1900  
UDP TTL:1 TOS:0x0 ID:105 IpLen:20 DgmLen:161  
Len: 133
```

```
[**] [1:1917:6] SCAN UPnP service discover attempt [**]  
[Classification: Detection of a Network Scan] [Priority: 3]  
09/04-17:53:56.420759 168.150.177.164:45461 -> 168.150.177.1:1900  
UDP TTL:1 TOS:0x0 ID:117 IpLen:20 DgmLen:160  
Len: 132
```

```
[**] [1:1917:6] SCAN UPnP service discover attempt [**]  
[Classification: Detection of a Network Scan] [Priority: 3]  
09/04-17:53:56.422095 168.150.177.164:45461 -> 168.150.177.1:1900  
UDP TTL:1 TOS:0x0 ID:118 IpLen:20 DgmLen:161  
Len: 133
```

```
[**] [1:2351:10] NETBIOS DCERPC ISystemActivator path overflow attempt little endian  
unicode [**]  
[Classification: Attempted Administrator Privilege Gain] [Priority: 1]  
09/04-17:53:56.442445 198.8.16.1:10179 -> 168.150.177.164:135  
TCP TTL:105 TOS:0x0 ID:49809 IpLen:20 DgmLen:1420 DF  
***A**** Seq: 0xF9589BBF Ack: 0x82CCF5B7 Win: 0xFFFF TcpLen: 20  
[Xref => http://www.microsoft.com/technet/security/bulletin/MS03-026.msp][Xref =>  
http://cgi.nessus.org/plugins/dump.php?id=11808][Xref => http://cve.mitre.org/cgi-bin/cvename.cgi?name=2003-0352][Xref => http://www.securityfocus.com/bid/8205]
```

```
[**] [122:3:0] (portscan) TCP Portsweep [**]  
[Priority: 3]  
09/04-17:53:56.499016 198.8.16.1 -> 168.150.177.166  
PROTO:255 TTL:0 TOS:0x0 ID:1750 IpLen:20 DgmLen:168
```

傳統 NIDS 的警訊型態 (2) 紀錄在資料庫內

以下為利用瀏覽器透過網頁方式呈現警訊資料庫的內容

The screenshot shows a Mozilla browser window displaying the 'Basic Analysis and Security Engine (BASE)' interface. The browser title is 'Basic Analysis and Security Engine (BASE): Query Results - Mozilla'. The page has a menu bar with 'File', 'Edit', 'View', 'Go', 'Bookmarks', 'Tools', 'Window', and 'Help'. Below the menu is a navigation bar with 'Home', 'Search', and 'AG Maintenance' links, and a '[Back]' link on the right. The main content area features a blue header with the text 'Basic Analysis and Security Engine (BASE)'. Below the header, there is a status message: 'Added 0 alert(s) to the Alert cache' and a timestamp: 'Queried DB on : Thu October 14, 2004 22:04:44'. To the left of the main content is a table with criteria: 'Meta Criteria: any', 'IP Criteria: any', 'TCP Criteria: any', and 'Payload Criteria: any'. To the right is a 'Summary Statistics' box containing a bulleted list: 'Sensors', 'Unique Alerts (classifications)', 'Unique addresses: source | destination', 'Unique IP links', 'Source Port: TCP | UDP', 'Destination Port: TCP | UDP', and 'Time profile of alerts'. Below the summary statistics, it says 'Displaying alerts 1-50 of 81 total'. The main part of the page is a table of alert records with columns: ID, Signature, Timestamp, Source Address, Dest. Address, and Layer 4 Proto. The table contains five rows of data, each with a checkbox in the ID column.

<input type="checkbox"/>	ID	< Signature >	< Timestamp >	< Source Address >	< Dest. Address >	< Layer 4 Proto >
<input type="checkbox"/>	#0-(1-84)	[snort] NETBIOS SMB IPC\$ share unicode access	2004-10-08 11:25:41	192.168.1.100:1613	192.168.1.4:139	TCP
<input type="checkbox"/>	#1-(1-83)	[snort] NETBIOS SMB IPC\$ share unicode access	2004-10-08 11:25:31	192.168.1.100:1608	192.168.1.4:139	TCP
<input type="checkbox"/>	#2-(1-82)	[snort] NETBIOS SMB IPC\$ share unicode access	2004-10-08 11:25:05	192.168.1.100:1601	192.168.1.4:139	TCP
<input type="checkbox"/>	#3-(1-80)	[snort] (http_inspect) OVERSIZE CHUNK ENCODING	2004-10-04 22:25:41	192.168.1.4:42164	67.19.245.228:80	TCP
<input type="checkbox"/>	#4-(1-81)	[snort] (http_inspect) OVERSIZE CHUNK ENCODING	2004-10-04 22:25:41	192.168.1.4:42163	67.19.245.228:80	TCP

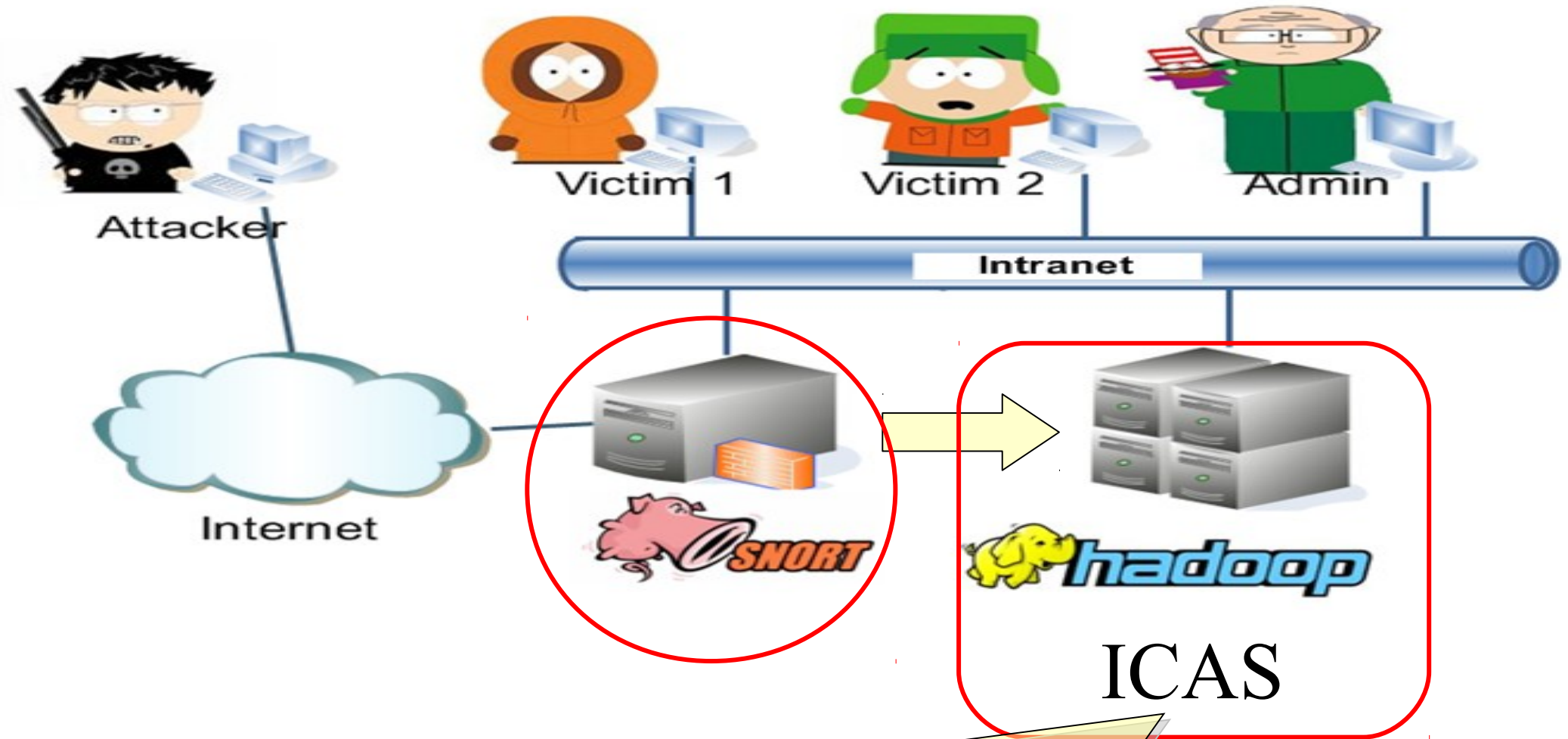
以上作法的缺點

- 警訊僅被『忠實』地被記錄下來，無法顯示彼此間的關聯性，因此系統管理者難以瞭解全部攻擊情形
- 過多的警訊，使得容易忽略重要內容
- 完全依賴單一資料庫，當資料量一大，該台主機的讀寫效率將成爲瓶頸

使用雲端運算的解決方案：ICAS

- ICAS, *IDS Cloud Analysis System*
- 利用雲端運算的特性提供以下好處
 - 對大量資料有高效率
 - 一般主機的叢集
 - 有錯誤容忍
- 分析演算法
 - 整合
 - 關聯

透過 ICAS 協助分析 IDS 的警訊

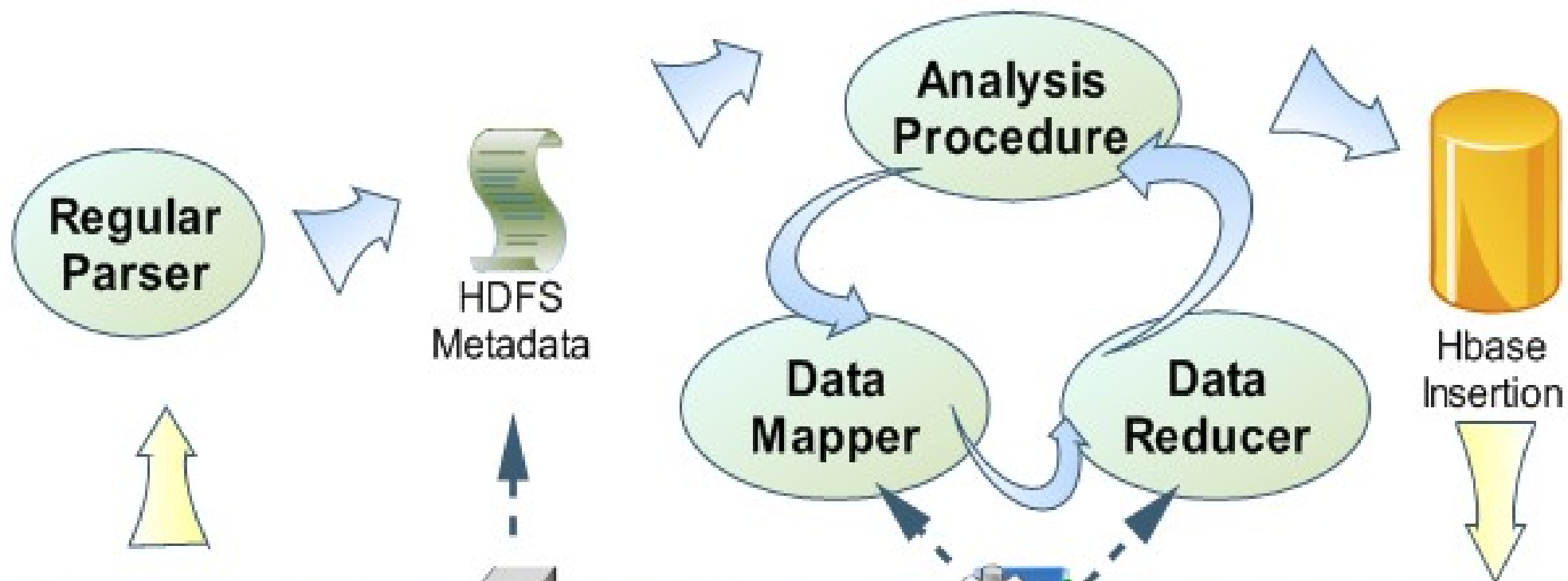


可多個 NIDS 共同產生警訊後，傳送至 ICAS，分析演算法
目前有 ICAS-I 及 ICAS-II

ICAS-I

- 將任意個原始警訊檔上傳到運行 ICAS-I 演算法的 Hadoop 檔案系統空間 (HDFS)
- 利用 Hadoop 的 MapReduce 平台架構所設計的演算法來分析資料
- 分析完後的資料塞入分散式資料庫 HBase 內

ICAS-I 流程圖



**Intrusion
Detectoin
System**

HDFS

JobTracker

hadoop

Cloud Platform

HBASE

Database

ICAS-I 整合後的警訊結果

Destination IP	Attack Signature	Source IP	Destination Port	Source Port	Packet Protocol	Timestamp
Host_1	Trojan	Sip1	80	4077	tcp	T1
Host_1	Trojan	Sip2	80	4077	tcp	T2
Host_1	Trojan	Sip1	443	5002	tcp	T3
Host_2	Trojan	Sip1	443	5002	tcp	T4
Host_3	D.D.O.S	Sip3	53	6007	udp	T5
Host_3	D.D.O.S	Sip4	53	6008	tcp	T5
Host_3	D.D.O.S	Sip5	53	6007	udp	T5
Destination IP	Attack Signature	Source IP	Destination Port	Source Port	Packet Protocol	Timestamp

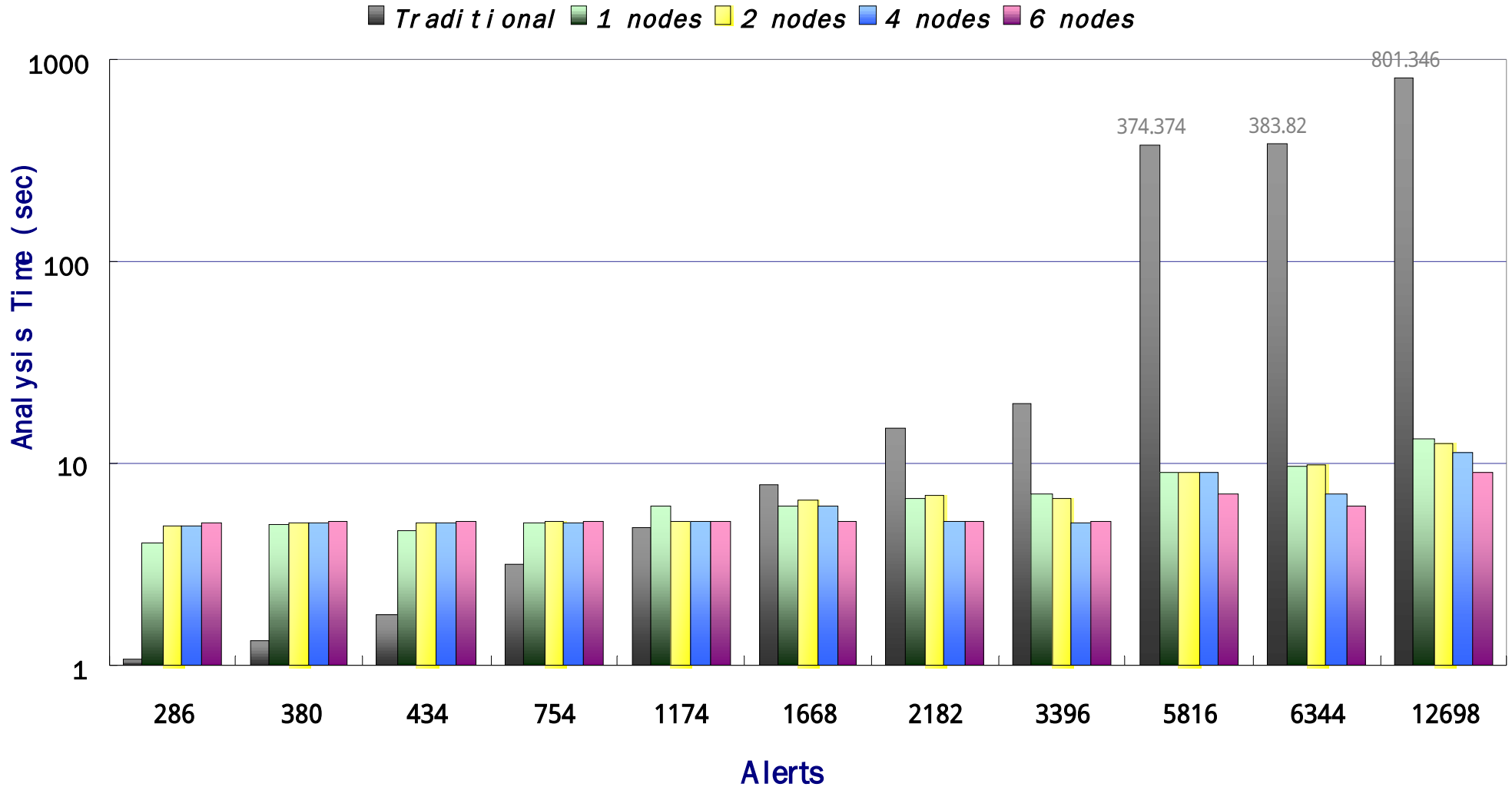
Key		Values				
Host_1	Trojan	Sip1,Sip2	80,443	4077,5002	tcp	T1,T2,T3
Host_2	Trojan	Sip1	443	5002	tcp	T4
Key		Values				

ICAS-I 效能數據的環境

- Machine:
 - CPU : Intel quad-core, Memory : 2 GB,
- OS : Linux : Ubuntu 8.04 server
- Software : version
 - Hadoop : 0.16.4
 - Hbase : 0.1.3
 - Java : 6
- Alerts Data Sets
 - MIT Lincoln Laboratory, Lincoln Lab Data Sets
 - Computer Security group at UC Davis, tcpdump file

ICAS-I 效能分析時間圖

The Consuming Time of Each Number of Data Sets



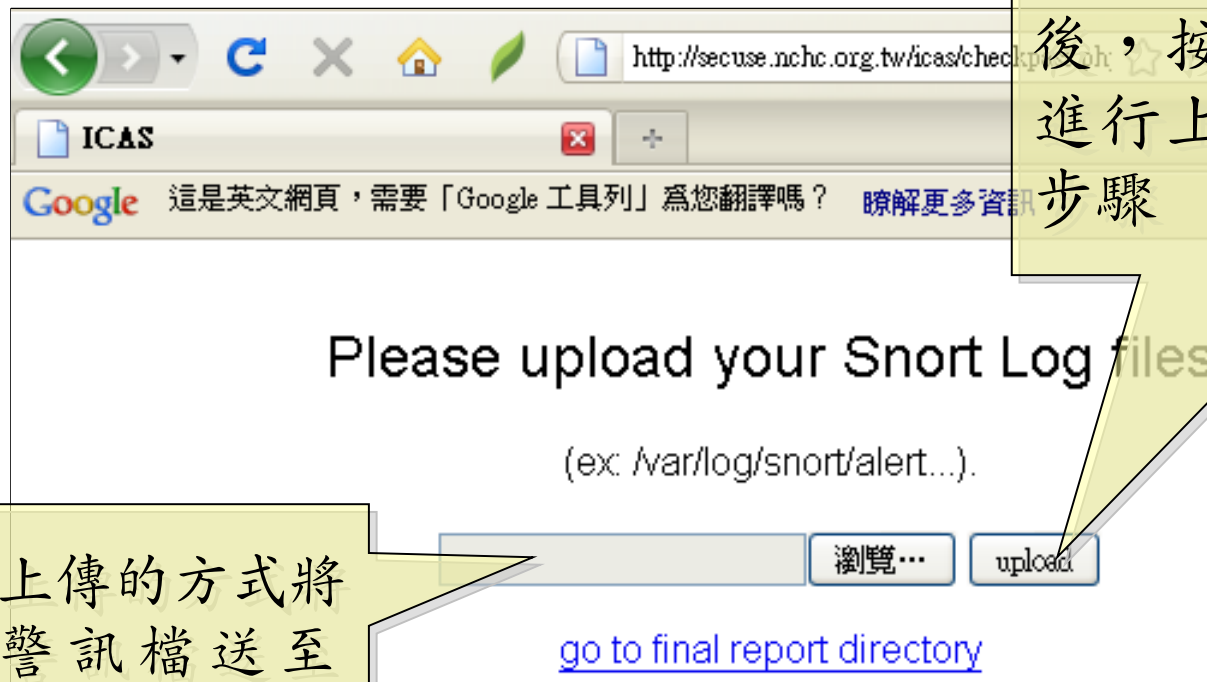
ICAS-I 效能數據表

Throughput Data Overall

Original Alerts	Analysis Time (sec)					Results	Reduction Rate
	Traditional	1 nodes	2 nodes	4 nodes	6 nodes		
286	1.068	4.087	4.869	4.864	5.077	30	89.51%
380	1.333	4.94	5.069	5.067	5.097	11	97.11%
434	1.76	4.61	5.066	5.068	5.09	9	97.93%
754	3.145	5.066	5.079	5.038	5.096	16	97.88%
1174	4.73	6.066	5.093	5.089	5.097	33	97.19%
1668	7.909	6.07	6.56	6.071	5.082	16	99.04%
2182	14.949	6.671	6.95	5.166	5.088	16	99.27%
3396	19.901	7.053	6.654	5.076	5.091	68	98.00%
5816	374.374	9.081	9.076	9.07	7.076	66	98.87%
6344	383.82	9.68	9.872	7.069	6.069	72	98.87%
12698	801.346	13.096	12.367	11.367	9.083	36	99.72%

ICAS-II

- ICAS-I 僅將資料塞入資料庫，然而還是文字的敘述
- ICAS-II 將輸入的任意多個警訊整合成一張警訊關聯圖
- 資料的來源可以透過以下兩種方式上傳到分析平台
 - 系統自動設定以 SCP 傳送到 ICAS 工作目錄
 - 管理者透過 ICAS 網頁上傳

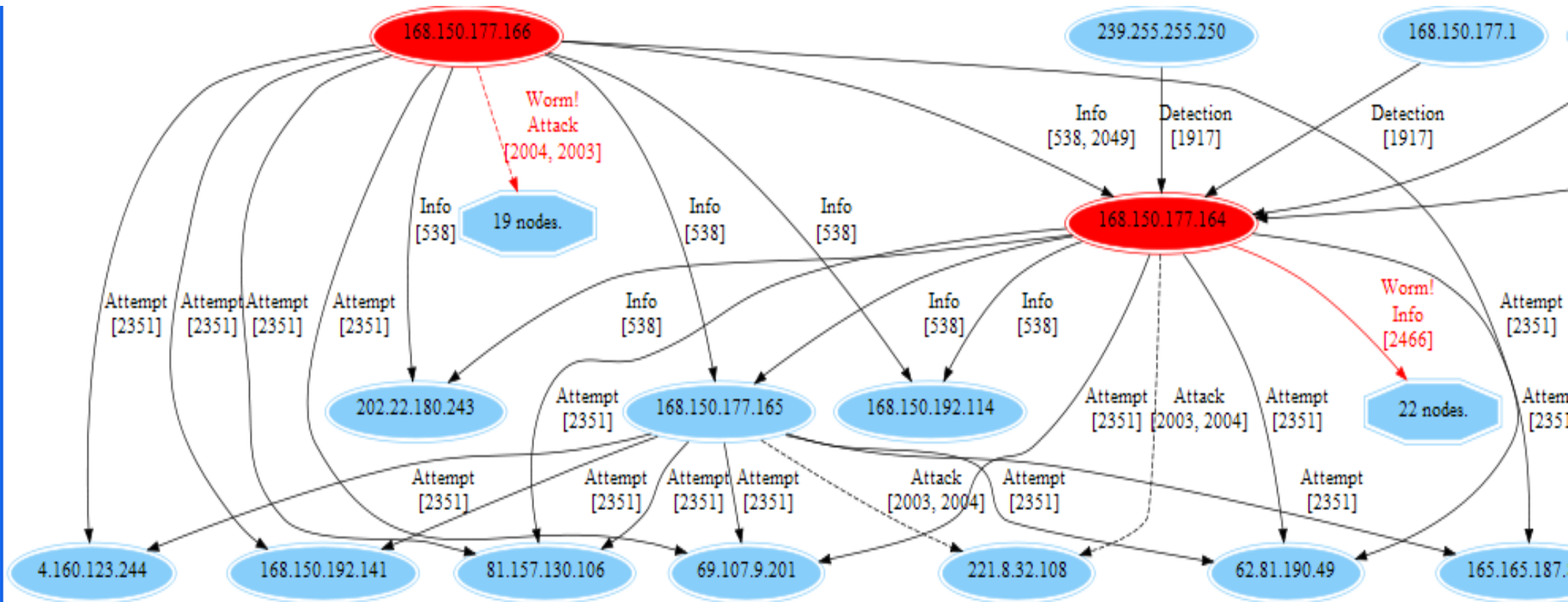


一旦選定需分析的日誌檔後，按下『上傳』，系統進行上傳→分析→繪圖等步驟

透過網頁上傳的方式將 Snort 的警訊檔送至 ICAS 分析

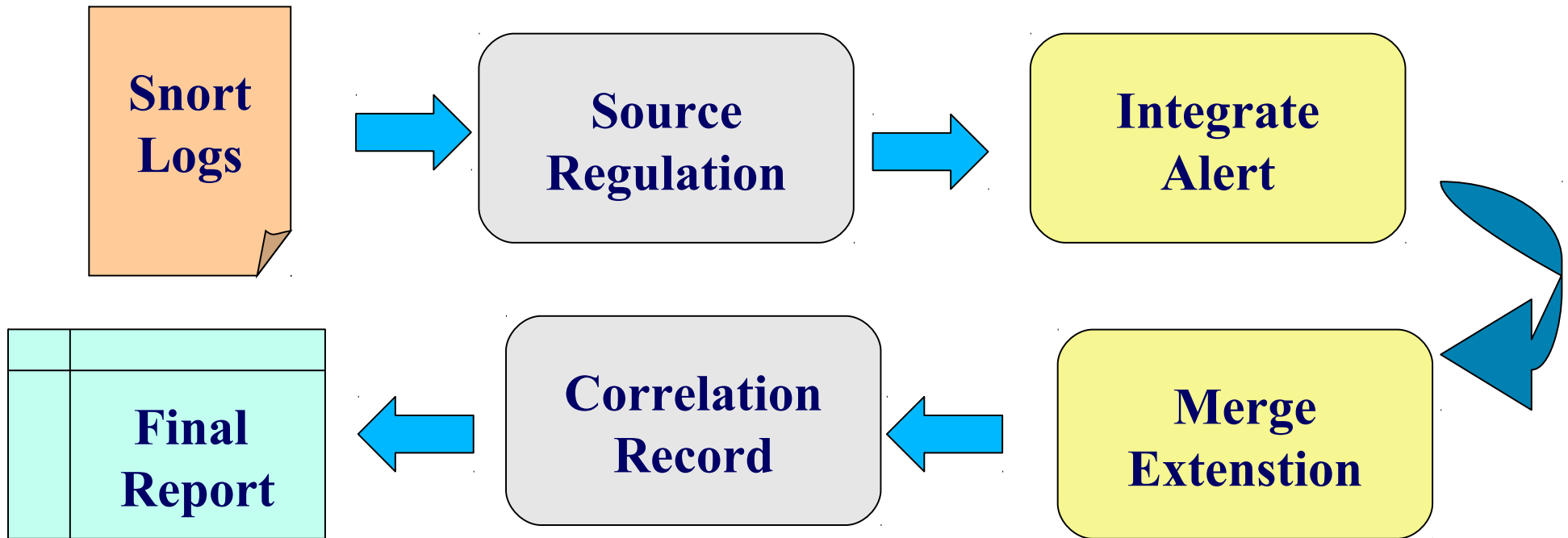
ICAS-II 所產生的報表：警訊關聯圖

- 經過 ICAS-II 分析後，可以得到此警訊關聯圖。
- 圖中橢圓形代表節點，箭頭及線上文字代表攻擊方向與攻擊方法。
- 標為紅色則是經過系統分析之後，被判定有攻擊行為的節點與方法。
- 此圖說明 IP 168.150.177.166 與 168.150.177.164 有進行蠕蟲的攻擊行為



ICAS-II 的分析流程

- Hadoop v 0.20



ICAS-II 結論

- ICAS-II 可經過警訊的來源、目的、攻擊事件綜合分析
 - 提供巨觀攻擊關聯圖來瞭解攻擊事件的始末
 - 自動透過標記顏色的方法將較高危險的事件呈現出來。
- ICAS-II 尚在整合關聯式資料庫，因此未進行數據量測

ICAS 總結

- 雲端運算處理資料格式相似且資料量大的情況下，能展現其效益
- 提供高容錯率、低獨占系統資源、多工作同時執行等能力
- 可搭配其他軟體作即時的警訊資料呈現， ICAS 可補充分析後資料的部份
- 未來工作
 - 整合多種資料來源平台
 - 產生更詳細與人性化的分析資料



打造內網搜尋引擎：Nutch & Crawlzilla

Part 4 : Introduction to Nutch and Crawlzilla

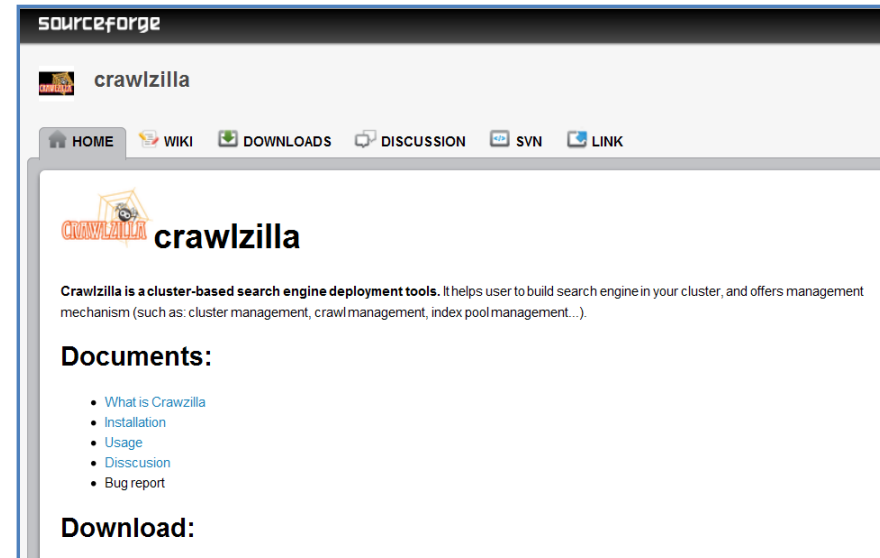
Jazz Wang
Yao-Tsung Wang
jazz@nchc.org.tw



Powered by DRBL

Crawlzilla ?

- 建構專屬於己的搜尋引擎
自由軟體專案
 - 快速找到資訊所在
 - 保障資料機敏性
 - 提供索引庫統計資訊
 - 會操作滑鼠就會使用



- 專案網站

- 中文：<http://crawlzilla.info/>
- 英文：<http://sf.net/p/crawlzilla>



Crawlzilla @

建立防火牆內的搜尋

- 現有的公開搜尋引擎無法、也不可穿透防火牆，搜尋內部網路的資料

在正確的資料內搜尋

- 減少廣告、不必要的內容、不當的資訊

破除資料庫內搜尋的限制

- 使用者上傳的附件檔（ doc, ppt, pdf... ） 、或超連結到站外的網站資訊

Crawlzilla !

Admin

```

正在重建相依關係
正在讀取狀態資料... 完成
正在讀取延伸狀態檔案
初始化套件狀態... 完成
沒有套件將會被安裝、升級或移除。
0 個套件升級, 0 個新安裝, 0 個將移除且 13 個不會升級。
需要下載 0B 的解壓縮檔案。解裝後將用去 0B。
正在編輯延伸狀態訊息... 完成
正在讀取套件清單... 完成
正在重建相依關係
正在讀取狀態資料... 完成
正在讀取延伸狀態檔案
初始化套件狀態... 完成

系統有 Sun Java 1.6 以上版本
系統已有 ssh。
系統已有 ssh Server (sshd)。
系統已有 dialog。
歡迎使用 Crawlzilla, 此安裝程序會為您新建一個 crawler 帳號並協助您設定密碼。
請輸入欲設定的 crawler 密碼:
password:
請再輸入一次確認密碼:
password:

Master 網域 IP 位址為: 146.110.138.186
Master 的 MAC 為: 08:00:27:99:4d:09
請確認上述的安裝資訊: 1. 正確 2. 不正確
    
```

```

[ Crawlzilla 管理介面 ] -by NCHC =
[ 管理功能選項 ]
請選擇:
luster_status 檢查 Cluster 狀態
fast_manage 快速啟動/關閉所有服務及 Tomcat
cluster_setup 設定 datanode & tasktracker
server_setup 設定 namenode & jobtracker
tomcat_switch 啟動/停止/重新啟動 Tomcat
tomcat_port 更改 Tomcat port
lang_switch 更換語言
client_install Client 安裝步驟
exit 結束
    
```

建立搜尋環境、選擇佈署叢集

IT

Crawl-建立搜尋引擎

▼ Crawl 爬取設定

索引庫名稱:

輸入欲爬取的網址(可多行):

爬取深度設定:

▶ 排程設定(Option)

▼ 基本資訊

索引庫名稱: narl_3
 搜尋引擎連結位置: /home/crawler/crawlzilla/user/admin/IDB/narl_3/index
 搜尋引擎狀態: OK
 爬取深度: 3
 建立時間: 20110503-12:15:19
 執行時間: 0:55:53
 超始連結: http://www.narl.org.tw/tw/

索引庫內容 - narl_3

▶ 資料總覽:

▼ 被搜尋分析到的網址:

Order	Contents	Counts	Order	Contents	Co
0	site:www.narl.org.tw	248	1	site:www.stb.org.tw	20
2	site:www.nspo.org.tw	3	3	site:conf.ncrec.org.tw	3
4	site:www.niac.narl.org.tw	2	5	site:i-one.org.tw	2
6	site:www.cic.narl.org.tw	1	7	site:www.mirdc.org.tw	1
8	site:web1.nsc.gov.tw	1	9	site:www.itri.org.tw	1
	site:www.ttfri.narl.org.tw	1	11	site:www.itrc.narl.org.tw	1

建立索引庫、瀏覽索引庫統計資訊

User

Crawlzilla 管理介面

搜索

[首頁](#) | [關於](#) | [公告](#) | [新聞](#) | [服務](#) | [支援](#) | [聯繫](#) | [隱私](#) | [安全](#) | [法律](#) | [其他](#)

國家高速網路與計算中心
 National Center for High-Performance Computing
 Better HPC Better Living

科技貢獻獎

Hits 1-11 (out of about 11 total matching pages):

[國網中心公告系統](#)
 ... 關「行政院傑出科技貢獻獎實施要點」、「行政院 ... 理行政院傑出科技貢 ...
https://intra.nchc.org.tw/HCMS/itr/inform_info.php?post=1302156081 (cached)

[ISO文件::管理規範專區](#)
 ... 驗研究院傑出科技貢獻獎作業要點 TOP ...
<http://iso.nchc.org.tw/document/> (cached) (explain) (anchors)

[重要記事::國研院國網中心](#)
 ... 年度行政院傑出科技貢獻獎 2007年6月 國 ... 年度行政院傑出科技貢 ...
<http://www.nchc.org.tw/tw/about/history.php> (cached) (explain) (anchors)

享受搜尋效益

技術的突破性

- **Crawlzilla** 被打造成企業或個人都可以輕鬆擁有專屬的搜尋引擎，也是目前沒有任何軟體 / 搜尋引擎可以取代的。
- 以自由軟體為巨人的肩膀，讓使用者有使用、複製、修改與再散播的自由
- 化繁為簡，透過簡明的介面完成建構需複雜資訊技術的搜尋引擎，是本專案最大的突破

技術的突破性

使用最新的視覺化網頁介面：**Web 2.0**

- AJAX 技術, W3C standard

整合最熱門的雲端運算演算法：**MapReduce**

- Google like
- 高效率、高容錯、高平行化

依循最穩固的程式開發架構：**Model-View-Control**

- 單元開發、程式再利用
- 全球化

2010 開放原始碼創新應用開發大賽 職業組冠軍



來自世界各地的下載

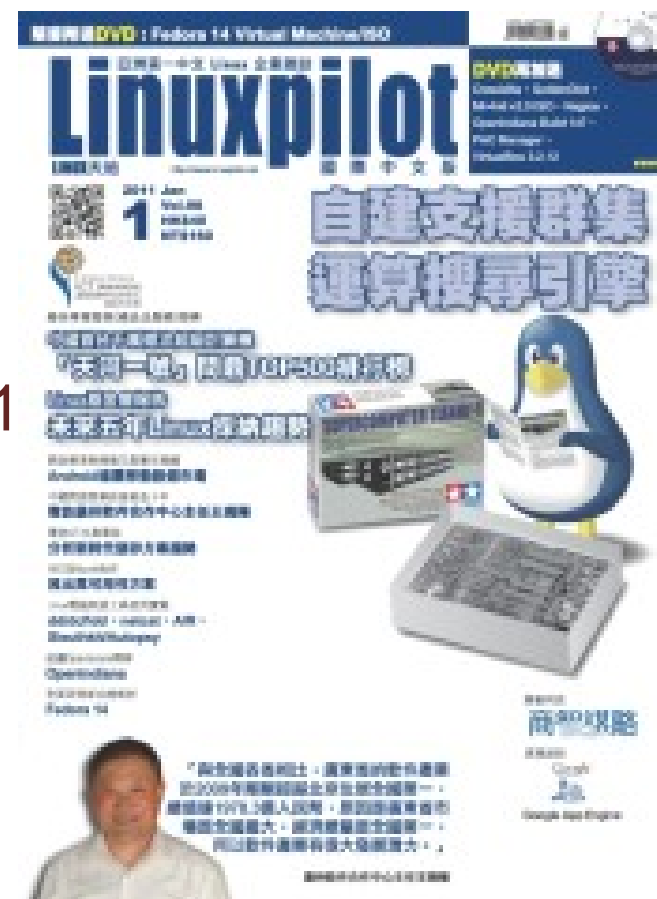
Visits ▾



- 來自 33 個國家， 1,397 次下載量
- 來自 53 個國家， 4479 次造訪紀錄
- 統計區間： 2010-08-17~2011-03-25

媒體報導

- 專案文獻：
 - TANET 2010：快速佈署叢集式搜尋引擎 - Crawlzilla
 - INTENSIVE 2011：Crawlzilla - A Toolkit for Deploying Cluster Search Engine Quickly and Easily
- 媒體報導：
 - 自建支援群集運算搜尋引擎
 - LinuxPilot Taiwan Vol.98 2011/01/13
 - 開放原始碼創新應用開發大賽
 - 創新發現誌 2011/01/25 Vol.20110
 - Three little zillas from Taiwan - iTWire 201



特色

雲端運算

- **Map Reduce** 演算法，分散工作量，整合運算結果

全系列Linux

- 單機、叢集上的任何**Linux** 套件版本，並自動解決軟體相依的問題

雲端介面

- 只需使用瀏覽器

統計管理

- 搜尋選項、瀏覽統計資料庫、叢集狀態

全文索引

- 全文索引引擎，並且能分析各種檔案格式（**html, txt, pdf, doc, ppt...**等）。

在地化MIT

- 提供完整的中英文操作設定，

多庫並存

- 可多個搜尋引擎庫同時上線使用

與其他國際知名的自由軟體比較

	Spidr	Nutch	Crawlzilla V 0.3	Crawlzilla V 1.0
安裝方式	Rube 套件 安裝	配置設定 檔	提供自動安裝 程式	提供自動安裝 程式
爬取網頁	O	O	O	O
分析內容	X	O	O	O
搜尋庫資訊	X	X	O	O
操作介面	指令	指令	Web-UI	Web-UI
中文最佳化	X	X	O	O
多人帳號, 排程機制	X	X	X	O

應用實例

- 嘉義縣網中心
 - 將用於課堂教材的統籌搜尋，如校園資訊、教育部成語字典、..等網站為搜尋字庫的基礎，提供學生正確有益的關鍵字結果
- 慈濟 - 資訊處
 - 將使用 **crawlzilla** 來對所有內部的文件，提供統一的搜尋服務，提供更快更便捷的方式找到資料
- 東海大學高效能實驗室
 - 結合雲端分散儲存與 **Nutch** 搜尋引擎之影音網站



應用實例

NCHC- 內網首頁

院部數位服務

- ▮ HRMS人資系統
- ▮ 電子表單
- ▮ 電子郵件服務
- ▮ 公文系統
- ▮ 工時系統



科技貢獻獎

簡介 常見問題

Search [help](#)

Hits **1-11** (out of about 11 total matching pages):

國網中心公告系統

... 關「行政院傑出**科技貢獻獎**實施要點」、「行政院 ... 理行政院傑出**科技貢** ...
https://intra.nchc.org.tw/HCMS/itr/inform_info.php?post=1302156081 ([cached](#))

ISO文件::管理規範專區

... 驗研究院傑出**科技貢獻獎**作業要點 TOP ...
<http://iso.nchc.org.tw/document/> ([cached](#)) ([explain](#)) ([anchors](#))

重要記事::國研院國網中心

... 年度行政院傑出**科技貢獻獎** 2007年6月 國 ... 年度行政院傑出**科技貢** ...
<http://www.nchc.org.tw/tw/about/history.php> ([cached](#)) ([explain](#)) ([anchors](#))

Crawlzilla 效益

- 節省建置的成本
 - 某商業版的搜尋引擎費用為 USD \$18000 (NTD \$57 萬) ，年費和客製費用另計，且不提供程式碼。
- 節省資料搜尋的時間
- 基於 Apache License 2.0 ，讓企業可客製化成自由軟體或商業獨家軟體
- 透過 開放源碼的搜尋引擎 Crawlzilla 激發未來更多學術和商業價值



資料異機同步儲存的機制：DropBox

Part 5 : Introduction to DropBox

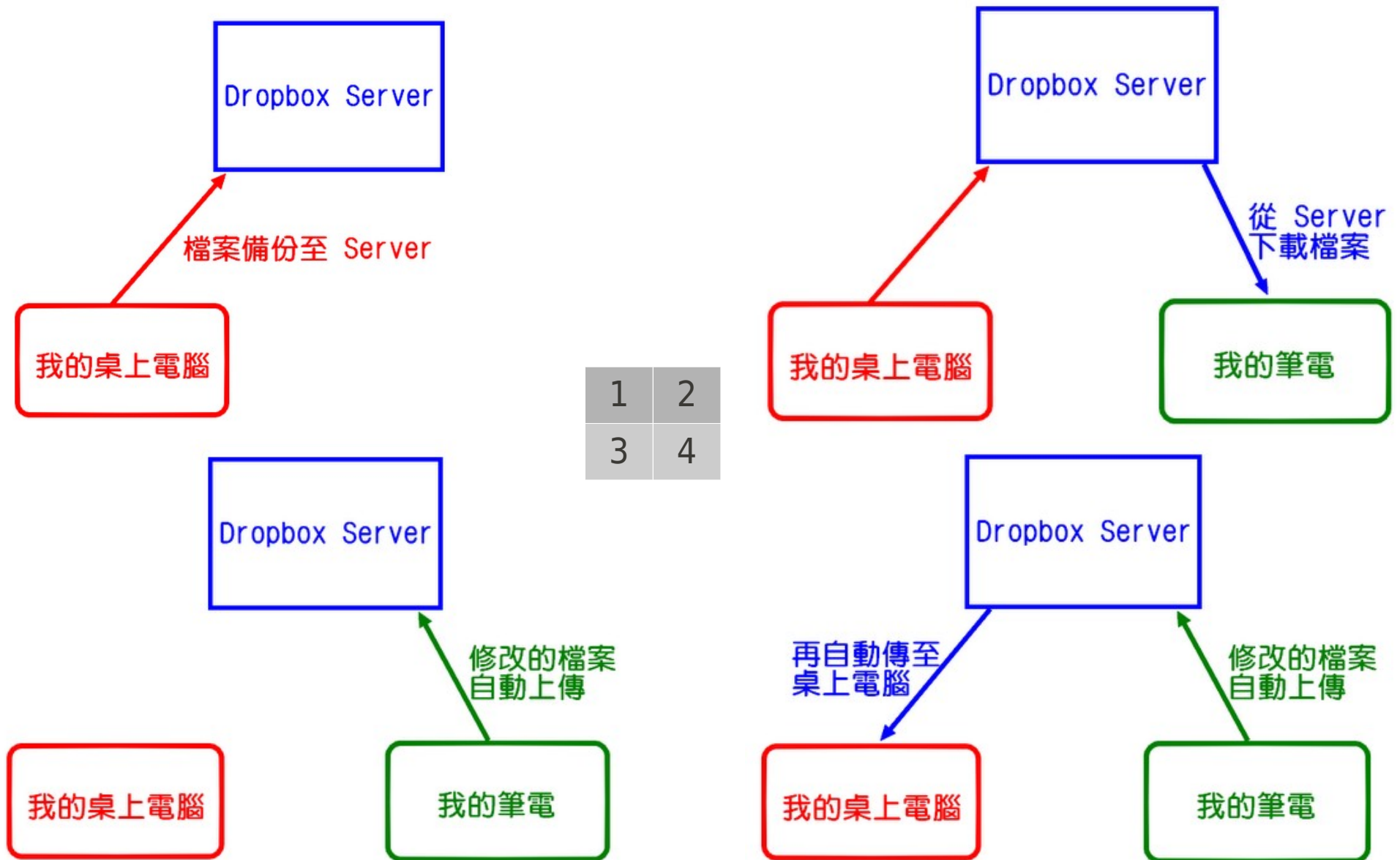
Jazz Wang
Yao-Tsung Wang
jazz@nchc.org.tw



Powered by DRBL

DropBox 可以自動同步多台電腦的資料

What is DropBox ?



如何申請 DropBox ?

How to apply DropBox accounts ?

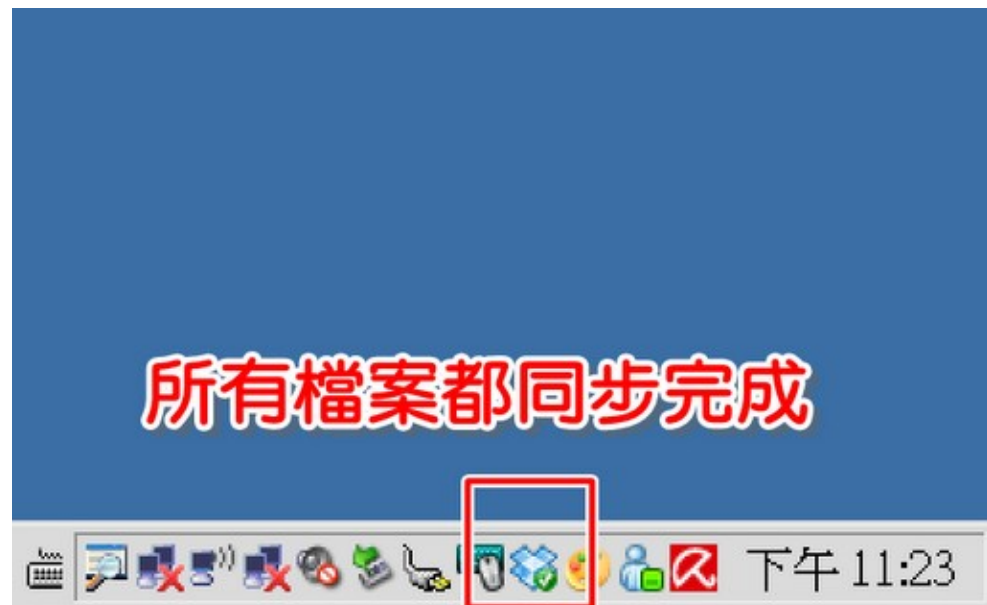
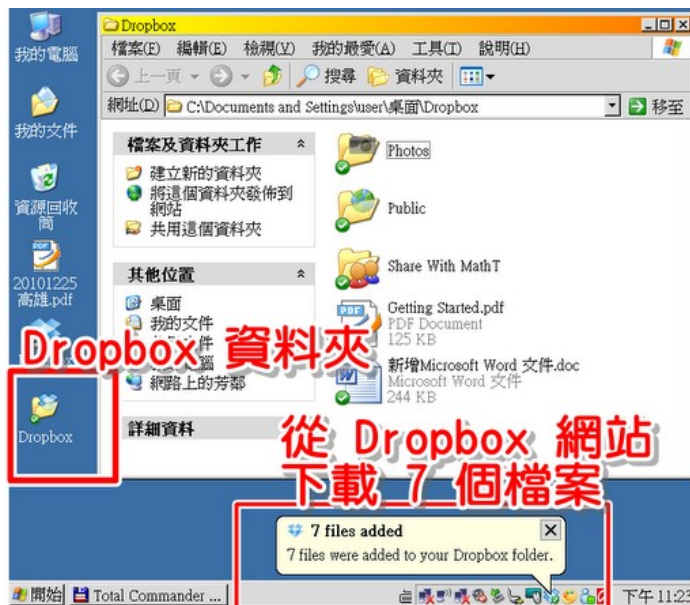
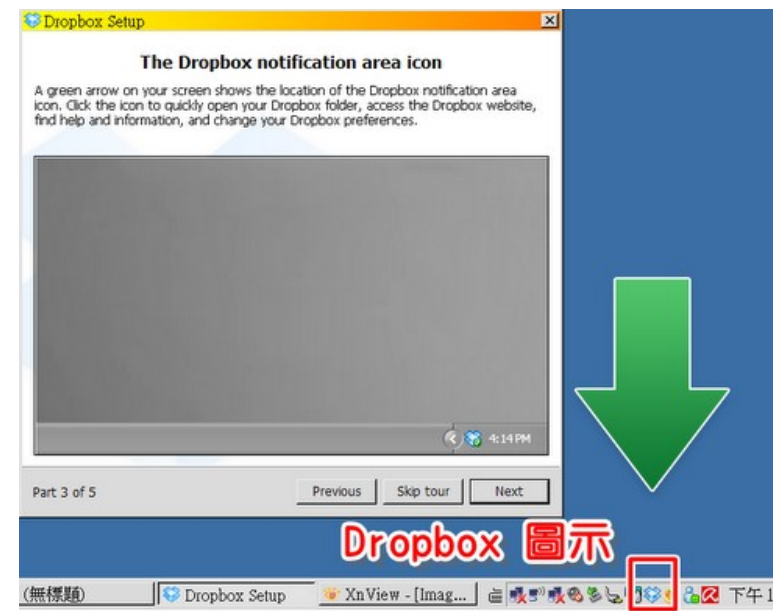
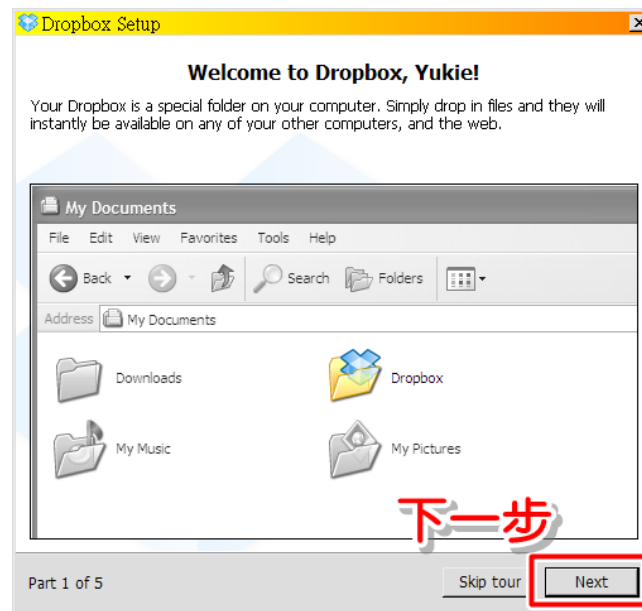
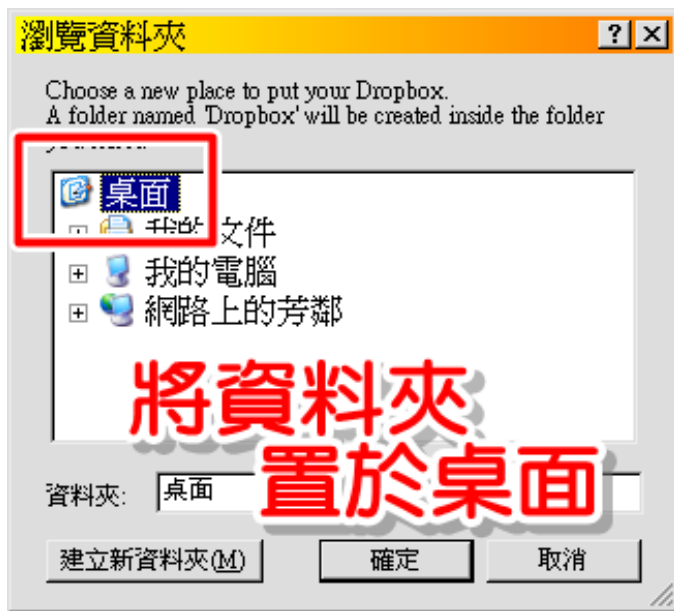
The image displays six sequential screenshots of the Dropbox Setup wizard, illustrating the steps to create and configure a new account. Each screenshot includes a red box highlighting a key element and a red text annotation in Chinese.

- Step 1: Welcome to Dropbox Setup**
This wizard will install Dropbox on your computer. Click Install to start the installation.
Annotation: 安裝 Dropbox (Install Dropbox)
- Step 2: Register new account**
Dropbox 註冊新帳號
 I don't have a Dropbox account
 I already have a Dropbox account
Annotation: 註冊新帳號 (Register new account)
- Step 3: Create your Dropbox**
Create your Dropbox
First name:
Last name:
Email:
Password:
Verify password:
Computer name:
If proceeding, you agree to our [Terms of Service](#).
Annotation: Email 就是帳號 (Email is the account)
- Step 4: Log in to Dropbox**
Log in to Dropbox
Email:
Password:
[Forgot password?](#)
Computer name:
Annotation: 以註冊的 Email 登入 (Log in with registered email)
- Step 5: Choose setup type**
Choose setup type
 Typical (recommended)
Set up Dropbox with normal settings.
 Advanced
Choose your Dropbox's location and which folders will be synced.
Annotation: 調整個人設定 (Adjust personal settings)
- Step 6: Advanced setup - Dropbox location**
Advanced setup - Dropbox location
 Install the Dropbox folder in the 'My Documents' folder.
 I want to choose where to put my Dropbox.

Annotation: 選擇要放 Dropbox 資料夾的位置 (Choose the location for the Dropbox folder)

如何設定與使用 DropBox ?

How to setup and use DropBox ?





Questions?

Slides - <http://trac.nchc.org.tw/cloud>

Jazz Wang
Yao-Tsung Wang
jazz@nchc.org.tw



Powered by DRBL