



SDN 交換機核心技術-流量分類以及應用辨識技術

黃能富教授

國立清華大學特聘教授, 資工系教授

捷鑒科技創辦人

威播科技創辦人

E-mail: nfhuang@cs.nthu.edu.tw

Contents

- 1 **Introduction to SDN Networks**
- 2 **Key Issues of SDN Switches**
- 3 **Machine Learning Based Applications Classification**
- 4 **Cloud + MLAC + SDN networks**
- 5 **Demo (Video)**
- 6 **Conclusions**

Today's routers



Routing, management, mobility management, access control, VPNs, ...

Feature

Feature

OS

Custom Hardware

Million of
lines
of source
code

6,000 RFCs

Billions of
gates

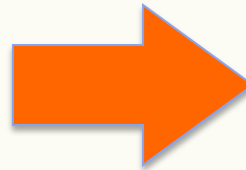
Bloated Power Hungry

- Vertically integrated, complex, closed, proprietary
- Networking industry with “mainframe” mindset

What SDN really is



Vertically integrated
Closed, proprietary
Slow innovation



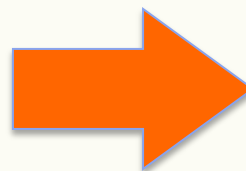
— Open Interface —



— Open Interface —



Horizontal
Open interfaces
Rapid innovation



OpenFlow Switching

Controller

OpenFlow Switch specification

OpenFlow Switch

sw

Secure Channel

hw

Flow Table

OpenFlow
Protocol
SSL

PC

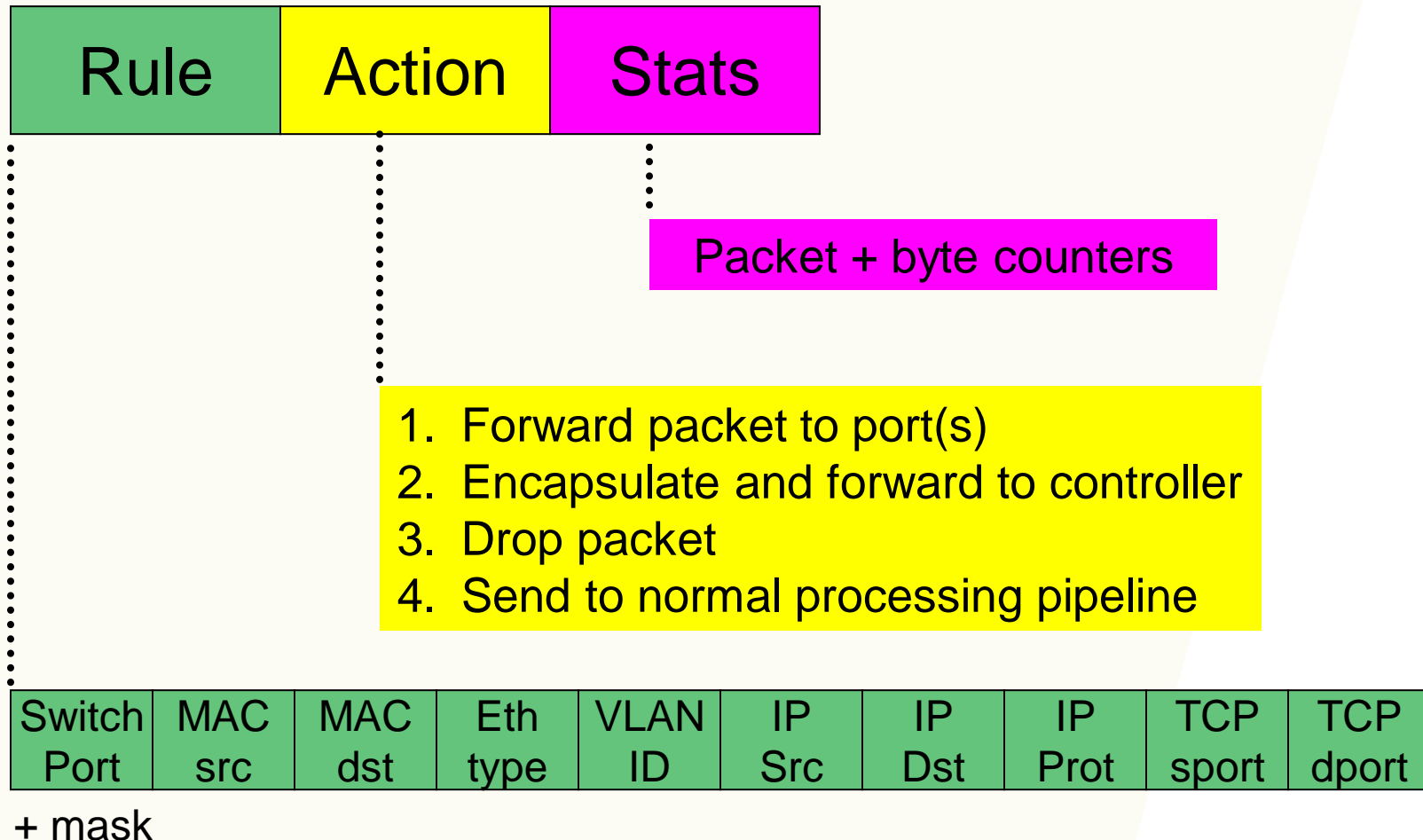


The Stanford Clean Slate Program

<http://cleanslate.stanford.edu>

Flow Table Entry

“Type 0” OpenFlow Switch



Applications for SDN Networks

- ❖ **QoS management**
- ❖ **Security management**
- ❖ **Network management**
- ❖ **BYOD management**

Contents

- 1 Introduction to SDN Networks
- 2 **Key Issues of SDN Switches**
- 3 Machine Learning Based Applications Classification
- 4 Cloud + MLAC + SDN networks
- 5 Demo (Video)
- 6 Conclusions

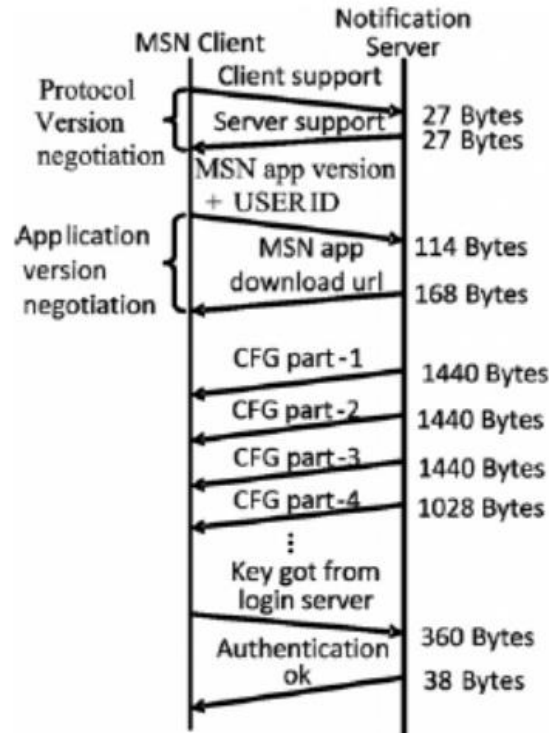
Key issues in SDN switches

- ❖ **Fast and Accurate Flow Classification**
 - Fast – SDN Switching Fabric with TCAM based flow table (TCAM lookup)
 - Accurate – Flow table with accurate flow signatures
- ❖ But **only Layer-4 fields are designed** in the standard flow table of OpenFlow SDN switches
 - Good for port-based applications
 - FTP (TCP/20-21), DNS (UDP/53), Telnet (TCP/23), Http (TCP/80)
 - Not good enough for port-changing applications
 - P2p sharing, On-line Games, Skype, BT, ...

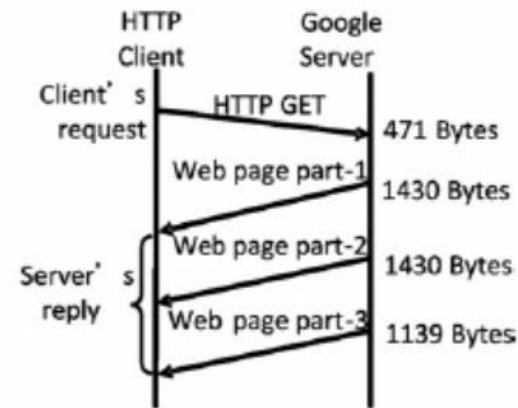
Key issues in SDN switches

- ❖ We can either classify the traffic with **Deep Packet Inspection (DPI)** technology
 - High Accuracy
 - But need to maintain the signatures dynamically
 - Unable to classify the encrypted flow
 - Not practical for cloud based platform
- ❖ Or we can use **Machine Learning** based approach
 - Use “**statistical attributes**” of flows to achieve high accuracy identification
 - Without inspecting the packet payload
 - Practical for cloud based platform as only “**statistical attributes**” are forwarded to the cloud.

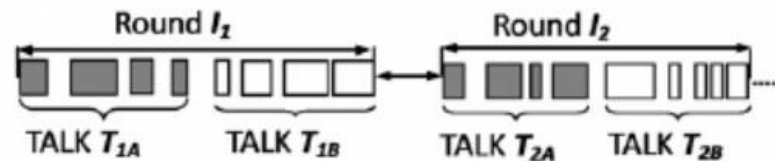
Examples of Statistical Attributes



(a) MSN login



(b) HTTP GET



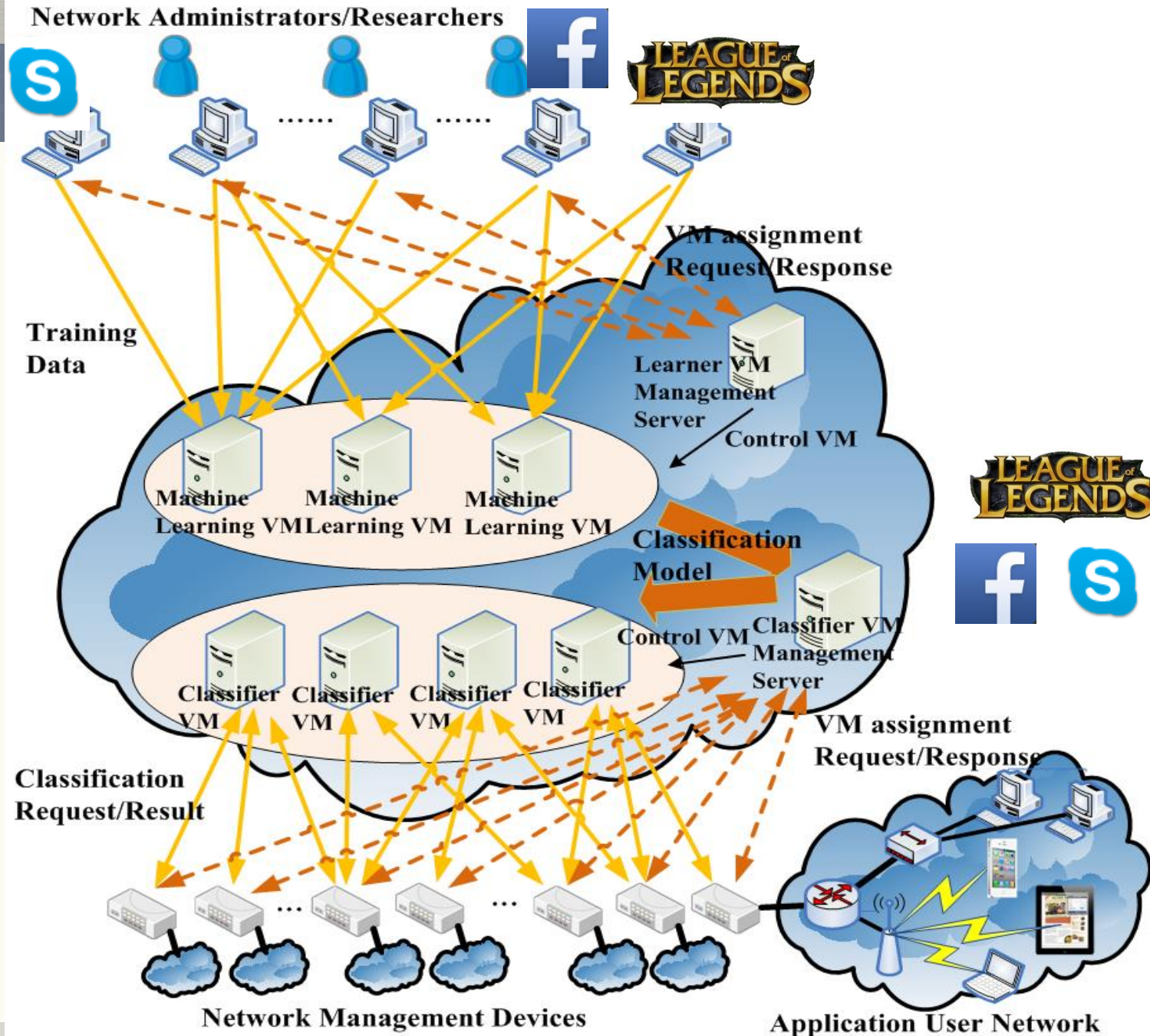
(c)

Contents

- 1 Introduction to SDN Networks
- 2 Key Issues of SDN Switches
- 3 **Machine Learning Based Applications Classification**
- 4 Cloud + MLAC + SDN networks
- 5 Demo (Video)
- 6 Conclusions

A Cloud based Applications Classification Service Platform with Machine Learning Algorithms for SDN Networks

基於機器學習演算法之雲端應用辨識服務平台
與 SDN 網路應用



Architecture of Traffic Classification Service on Cloud

- ❖ **Machine learning virtual machine (LVM)**
 - Running training service
- ❖ **Classifier virtual machine (CVM)**
 - Running classification service
- ❖ **LVM management server**
 - LVM assignment and monitoring
- ❖ **CVM management server**
 - CVM assignment and monitoring
- ❖ **Client** (PCs running the traffic classification client)
 - Computing the statistical attributes of each flow
 - Retrieving the training/classification service.

The used machine learning tools-Weka

❖ **Weka** -Waikato Environment for Knowledge Analysis (running on single PC/NB)

- A JAVA-based open source software suite that collects machine learning algorithms for data mining tasks
- Four kinds of interface



Explore: To see data distribution and apply ML algorithm directly.

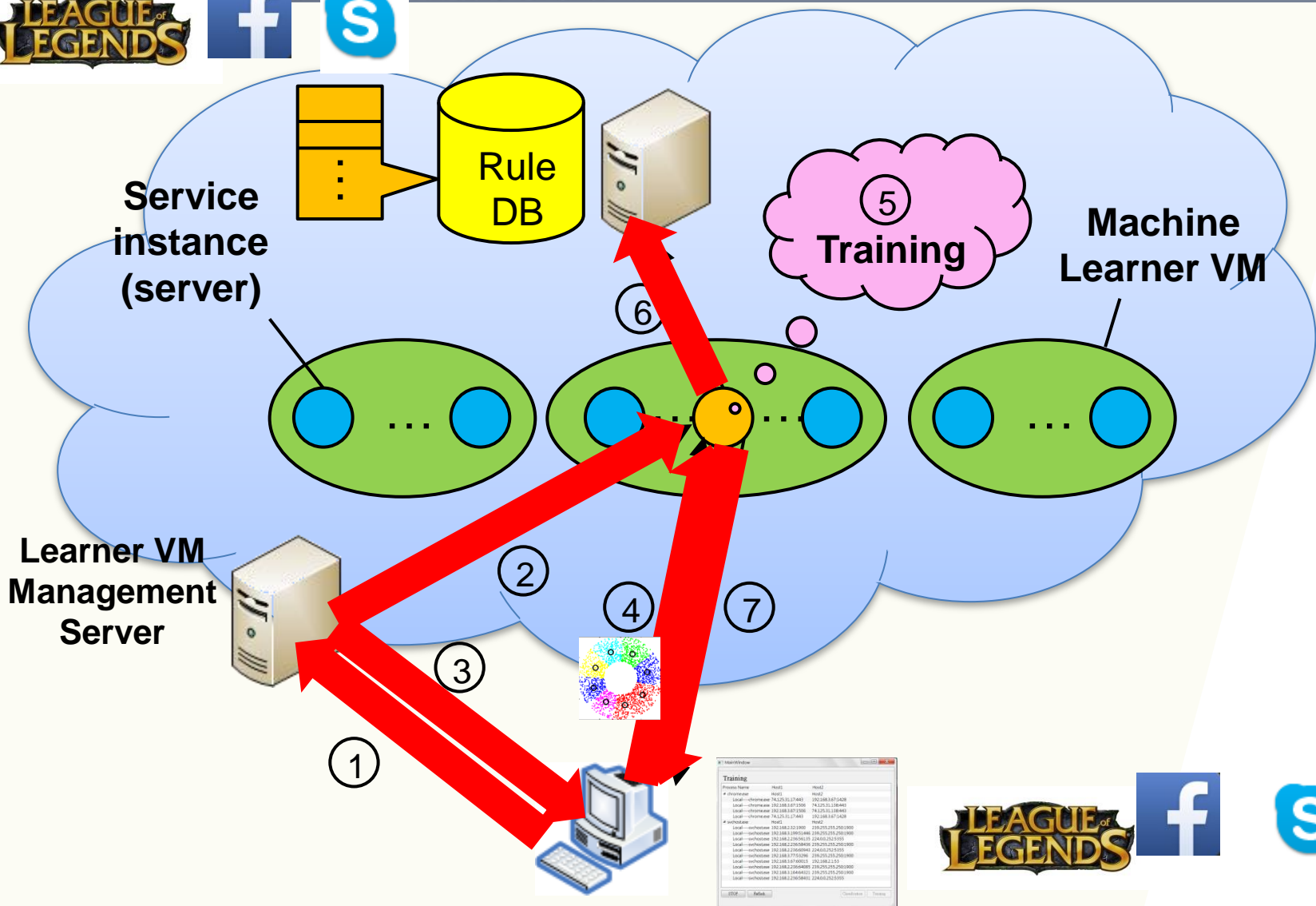
Experimenter: Batch modes for running training-testing evaluation

Knowledge Flow: To combine the different ML algorithms component as graph-based processing flow

CLI: Command Line Interface

- <http://www.cs.waikato.ac.nz/ml/weka/>

The Training Procedure



User's PC running training program

The Training Service

- ❖ **Requesting the training service**
 - **1. Client → LVM management server**

- ❖ **LVM assignment**
 - **2. Checking the loading and status of each service instance on the LVM**
 - **3. IP address & port number of LVM
(LVM Management Server → Client)**

The Training Service

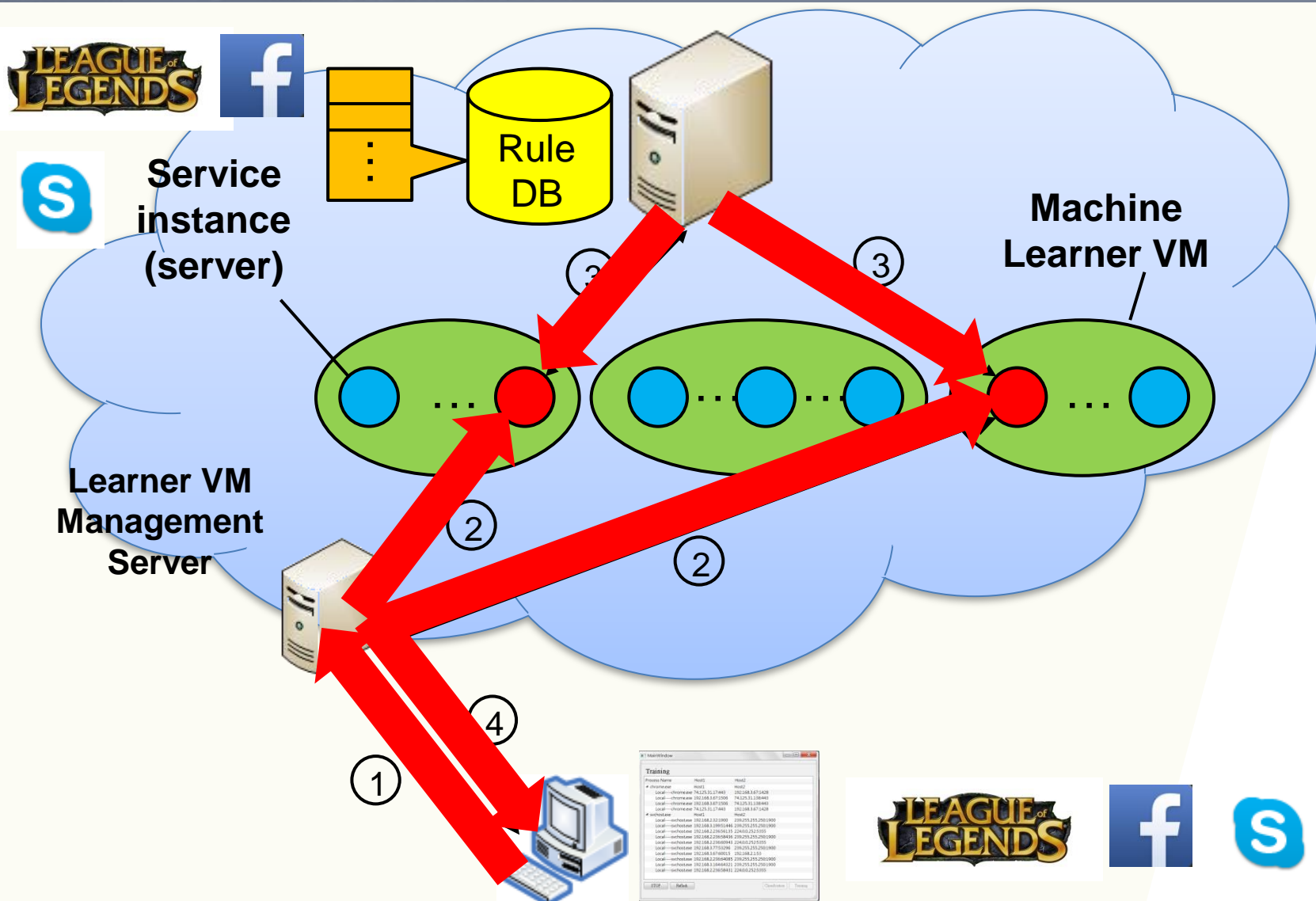
❖ Training

- 4. Training Data (**attributes of each flow to train**): Client → LVM
- 5. Training phase

❖ Training Results

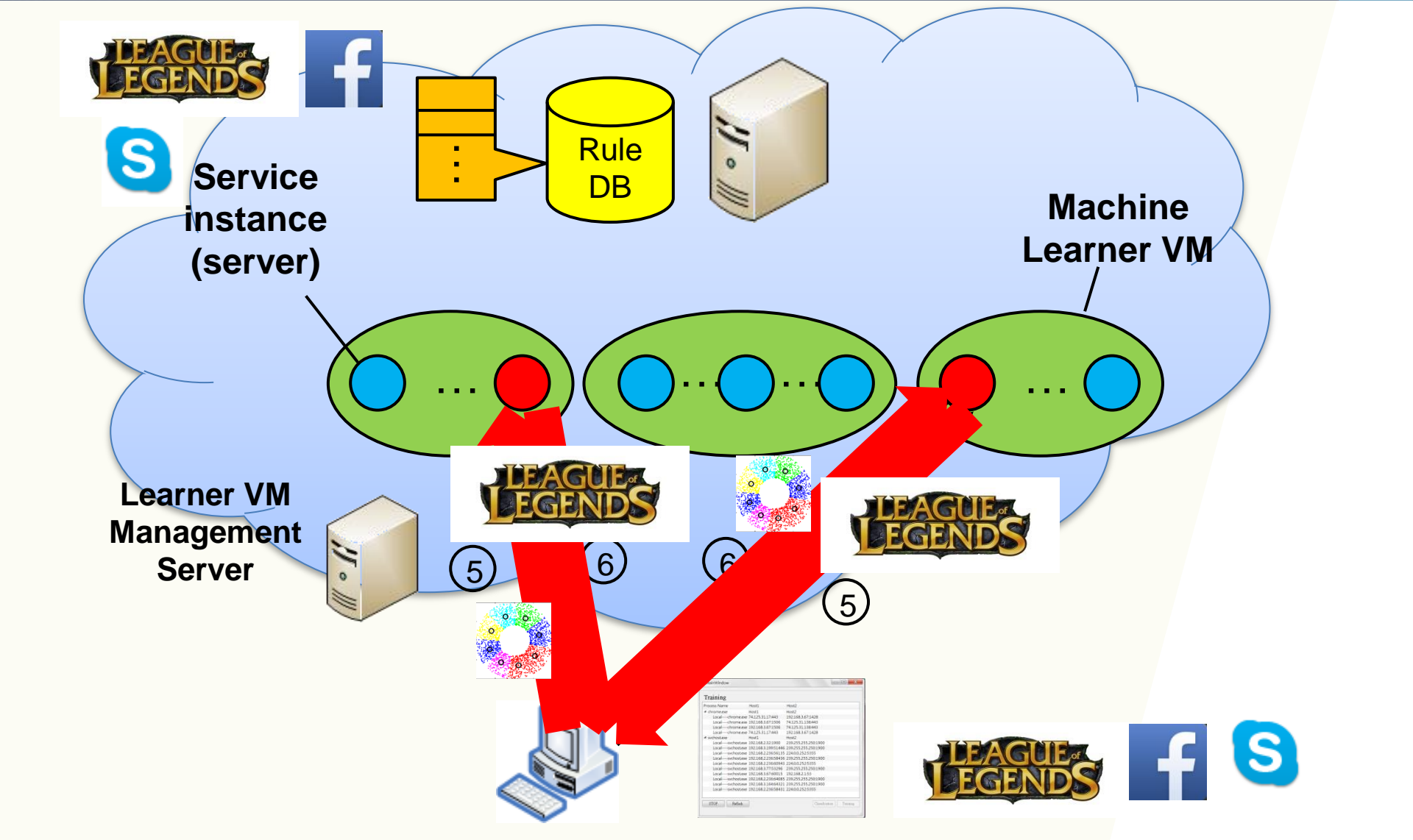
- 6. Storing the rules (LVM → Rule DB)
- 7. Training result (e.g. number of rules.)
- **Attributes: Two way communications, packet size, packet number, packet time interval, download rate, upload rate, ..., etc**

Traffic Classification Procedure



Network Manager's PC with mirrored network traffic

Traffic Classification Procedure (cont.)



The Traffic Classification Procedure

❖ Requesting the classification service

- 1. Client → CVM management server

❖ CVM assignment

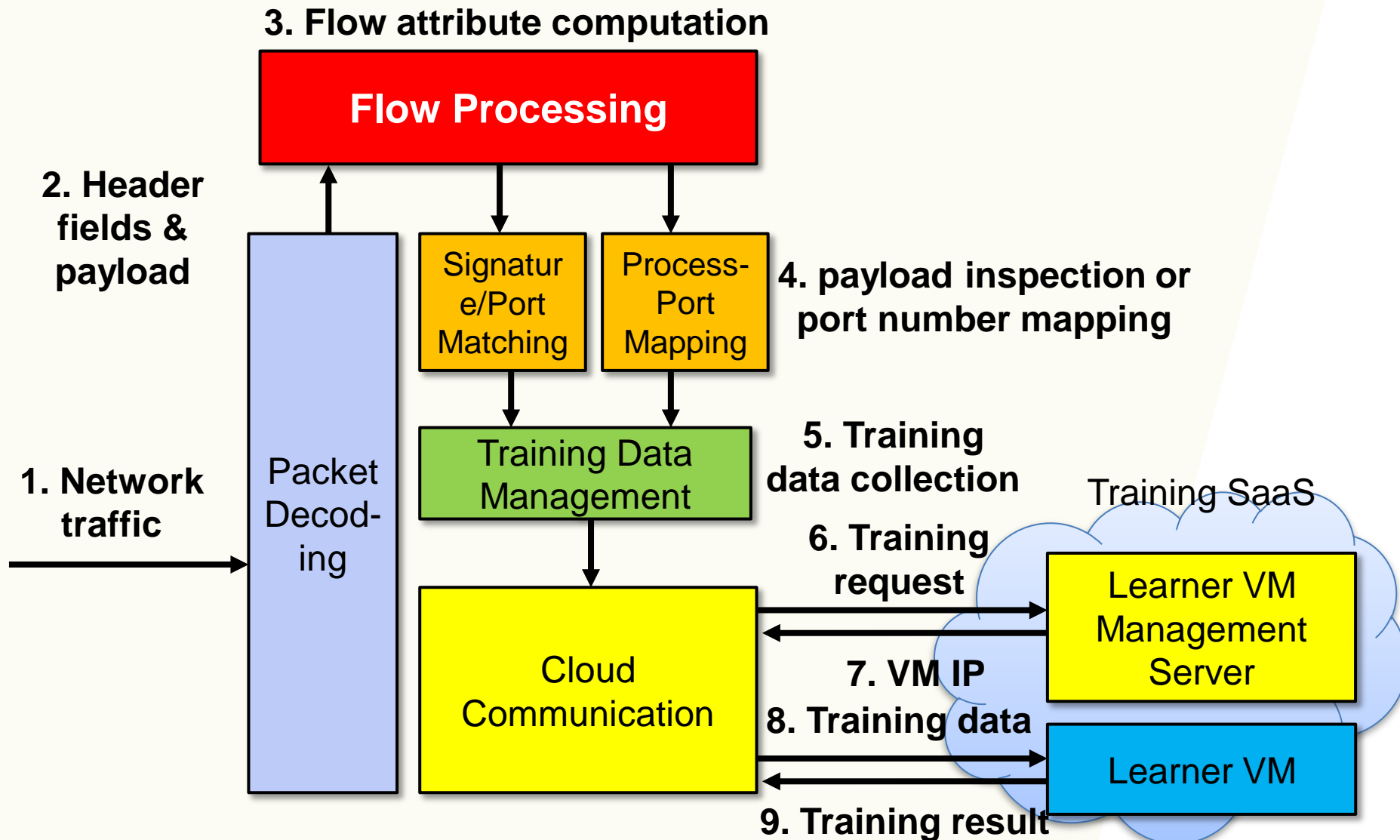
- 2. Checking the loading and status of each service instance on the CVM
- 3. Loading the classification rules from DB.
(Rule DB \leftrightarrow CVM)
- 4. CVMs' IP address & port number
(CVM Management Server → Client)

The Traffic Classification Procedure

❖ Classification

- 5. Sending the **attributes of flows to classify**
 - Client → CVMs
- 6. Receiving the classification result (application id)
 - CVMs → Client

Client Program for Training



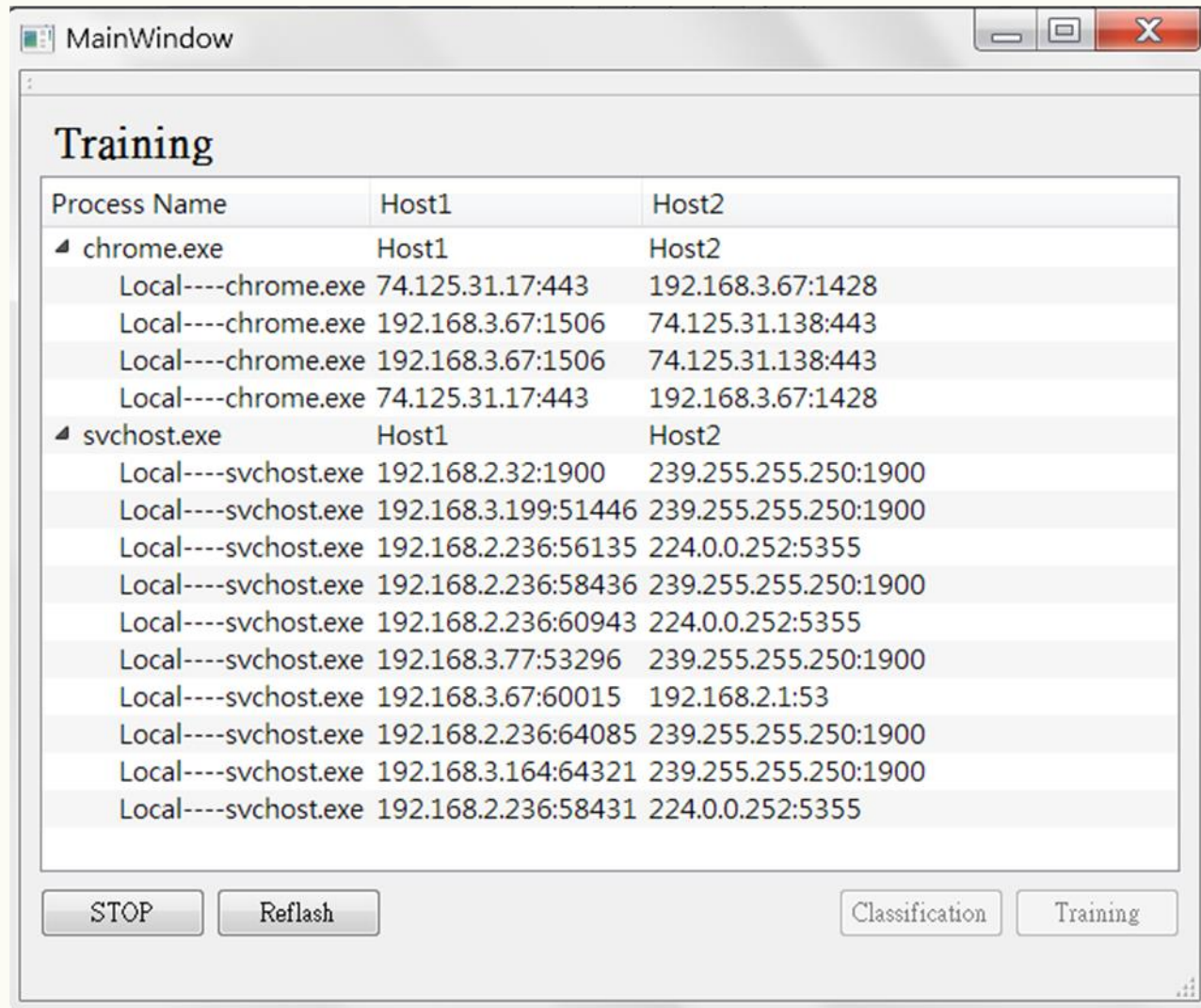
Implementation: Client Program Modules

- ❖ **Packet decoding module**
 - decoding the packet header fields
- ❖ **Flow processing module**
 - computing flow attributes
 - managing the flow table for storing the flow attributes
- ❖ **Signature/port matching and process-port mapping modules (only for training)**
 - To identify the original background truth (Training label).
 - User can choose the signature/port matching or process-port mapping as the module to find the original application class for each flow.

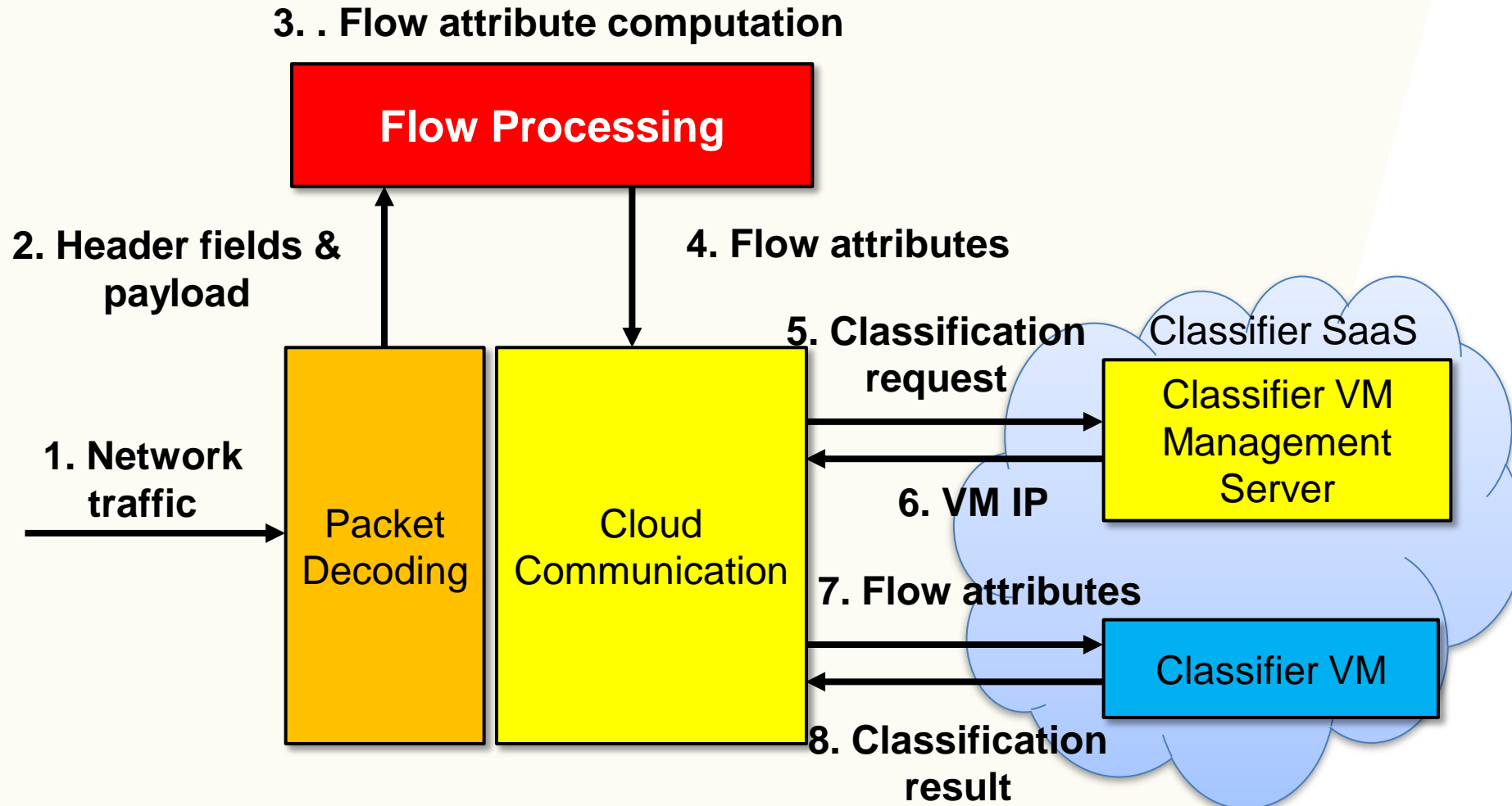
Implementation: Client Program Modules

- ❖ **Training data management module (only for training)**
 - collecting the attributes of flows to form the training data set.
- ❖ **The cloud communication module**
 - Client \leftrightarrow LVM/CVM Management server
 - Client \leftrightarrow LVM/CVM

The client program (Training)



Client Program for Classification



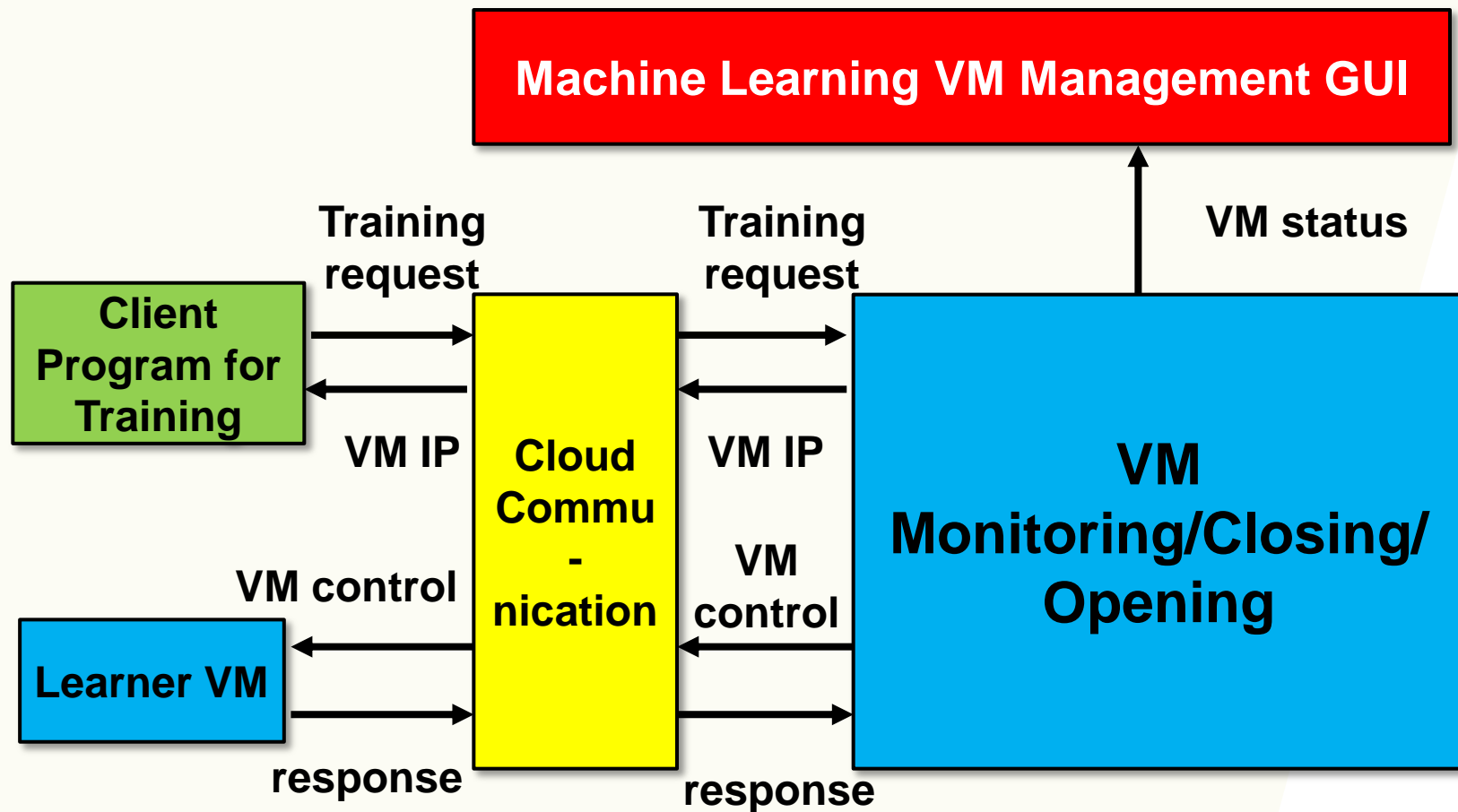
LVM and CVM management server (Cloud)

- ❖ **Machine learning/classifier VM management server graphic user interface (GUI)**
- ❖ **VM monitoring/closing/opening module**
- ❖ **The cloud communication module in the LVM and CVM management server**

Program modules of LVM and CVM (Cloud)

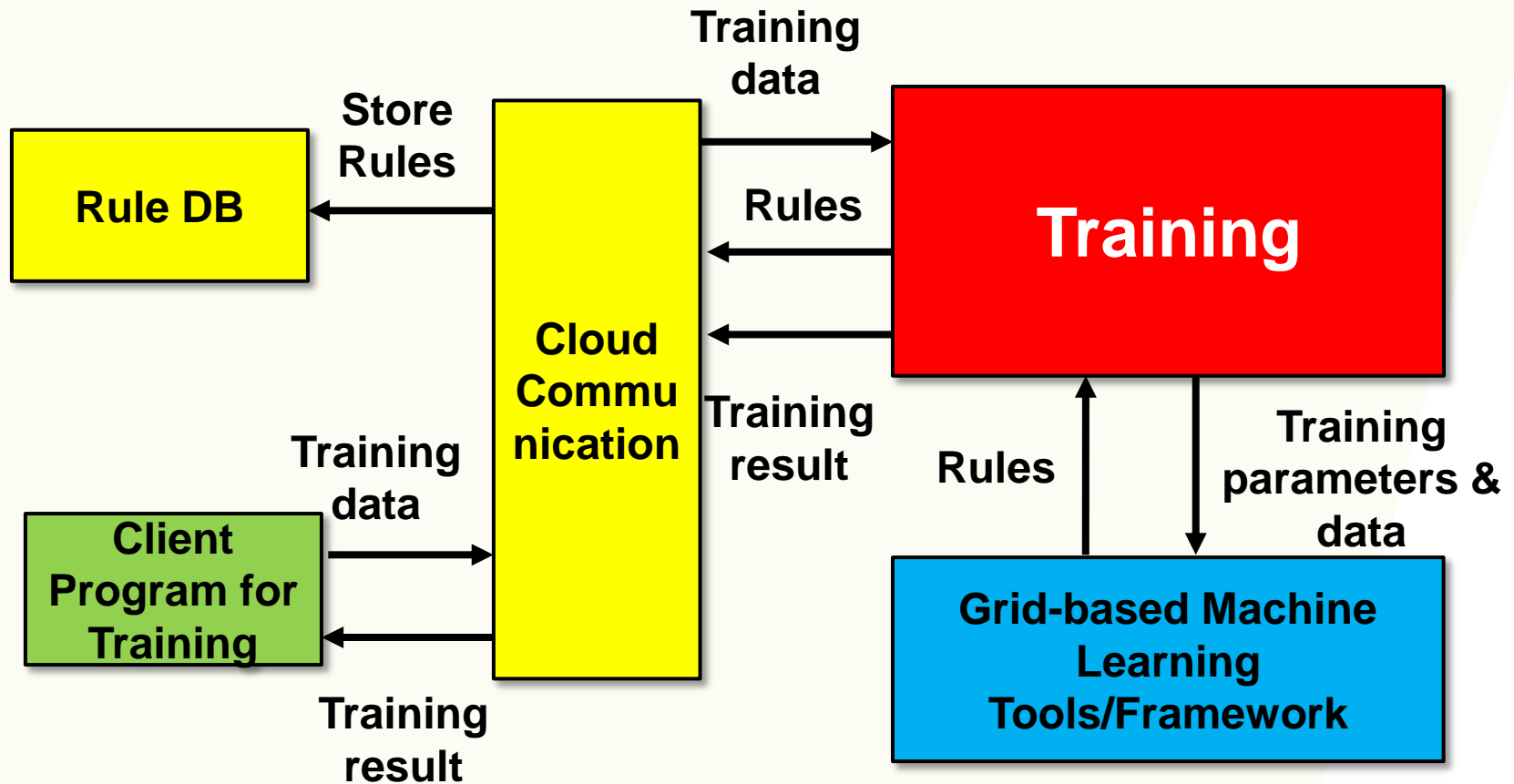
- ❖ **The training/classification modules**
- ❖ **The cloud communication module in the LVM or CVM**
- ❖ **Grid-based machine learning tools/framework**
 - **The implementation of the machine learning algorithms on grid environment.**

Learning VM Management Server

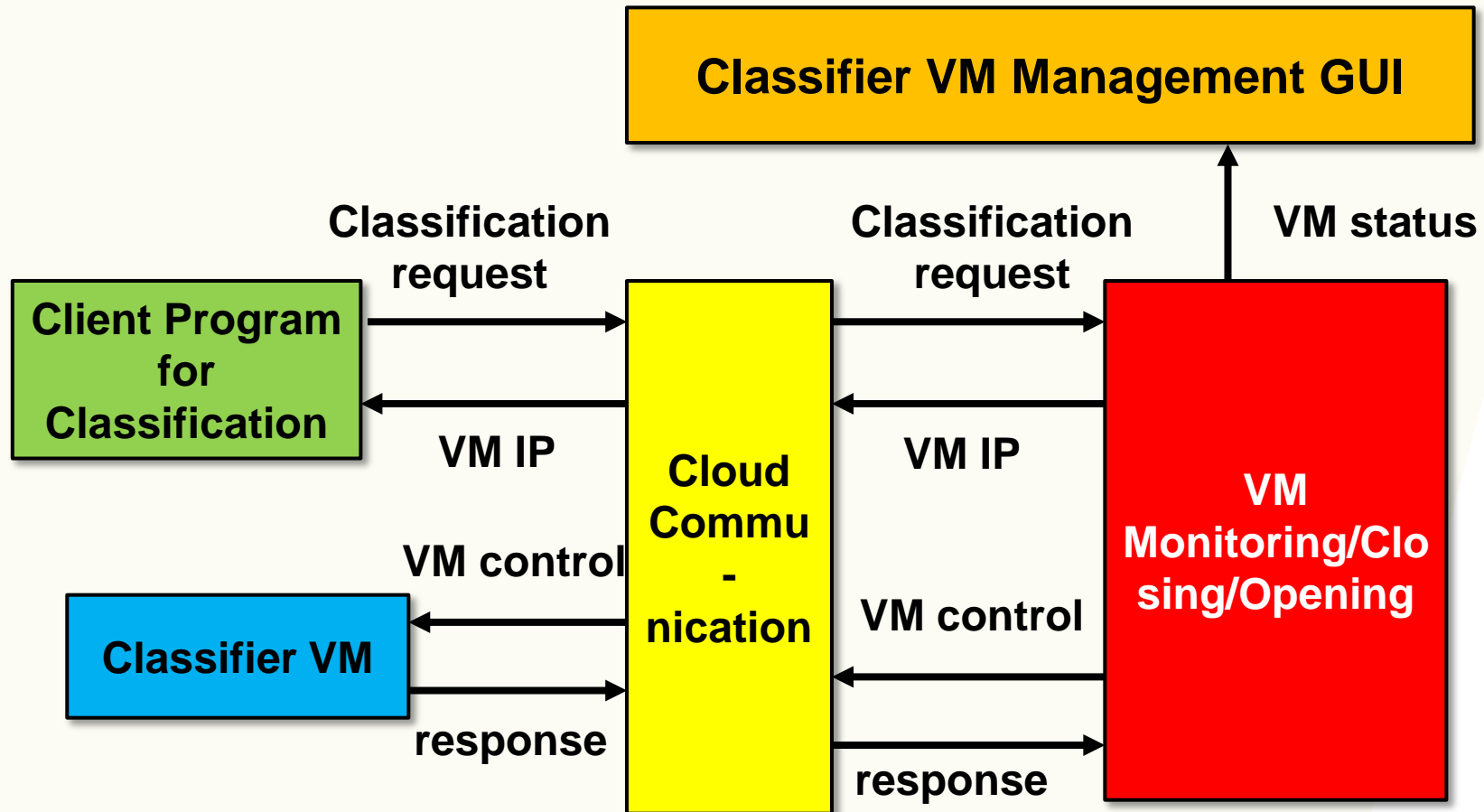


Learning VM Management Server

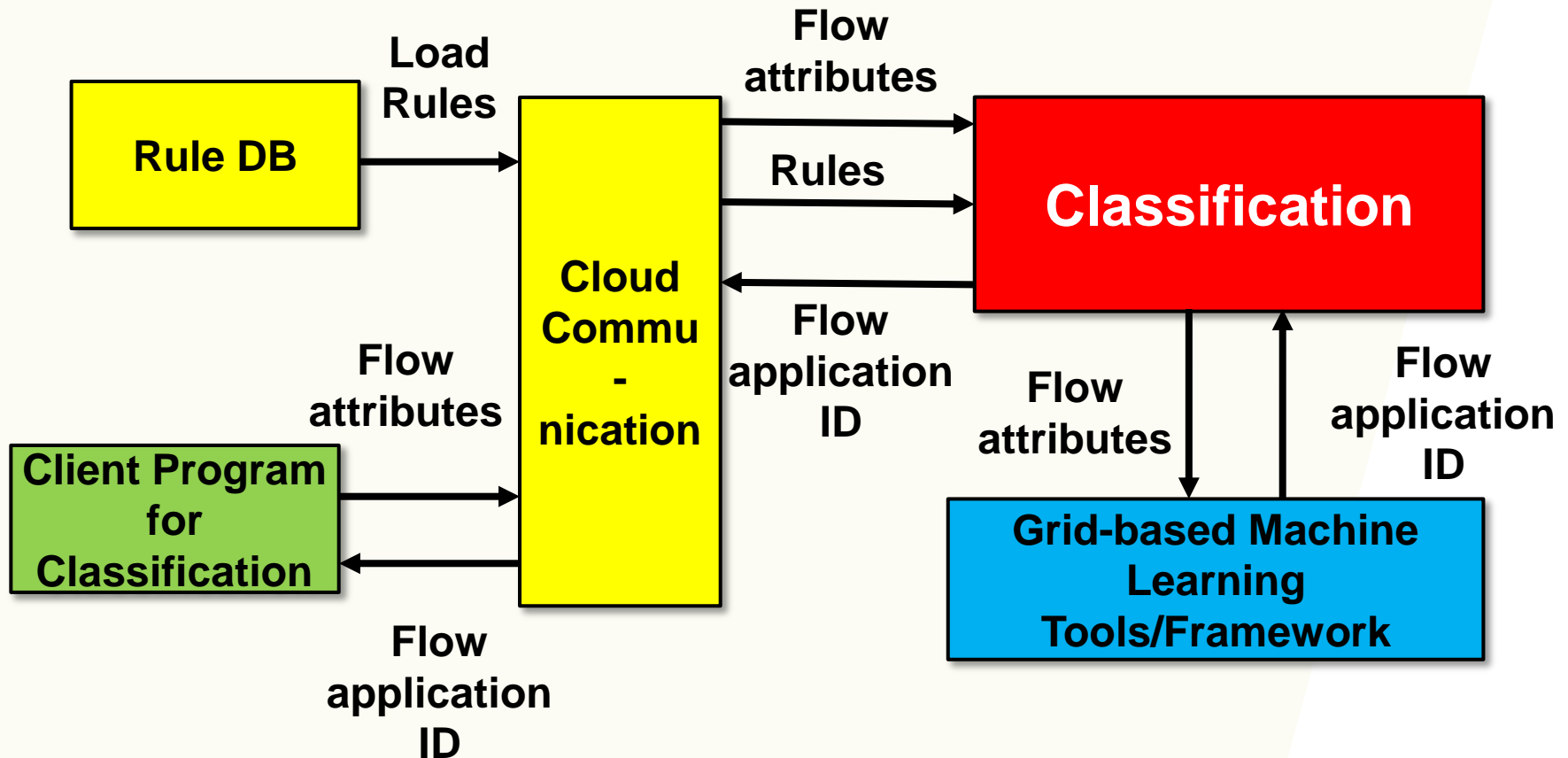
Learning VM



Classifier VM Management Server



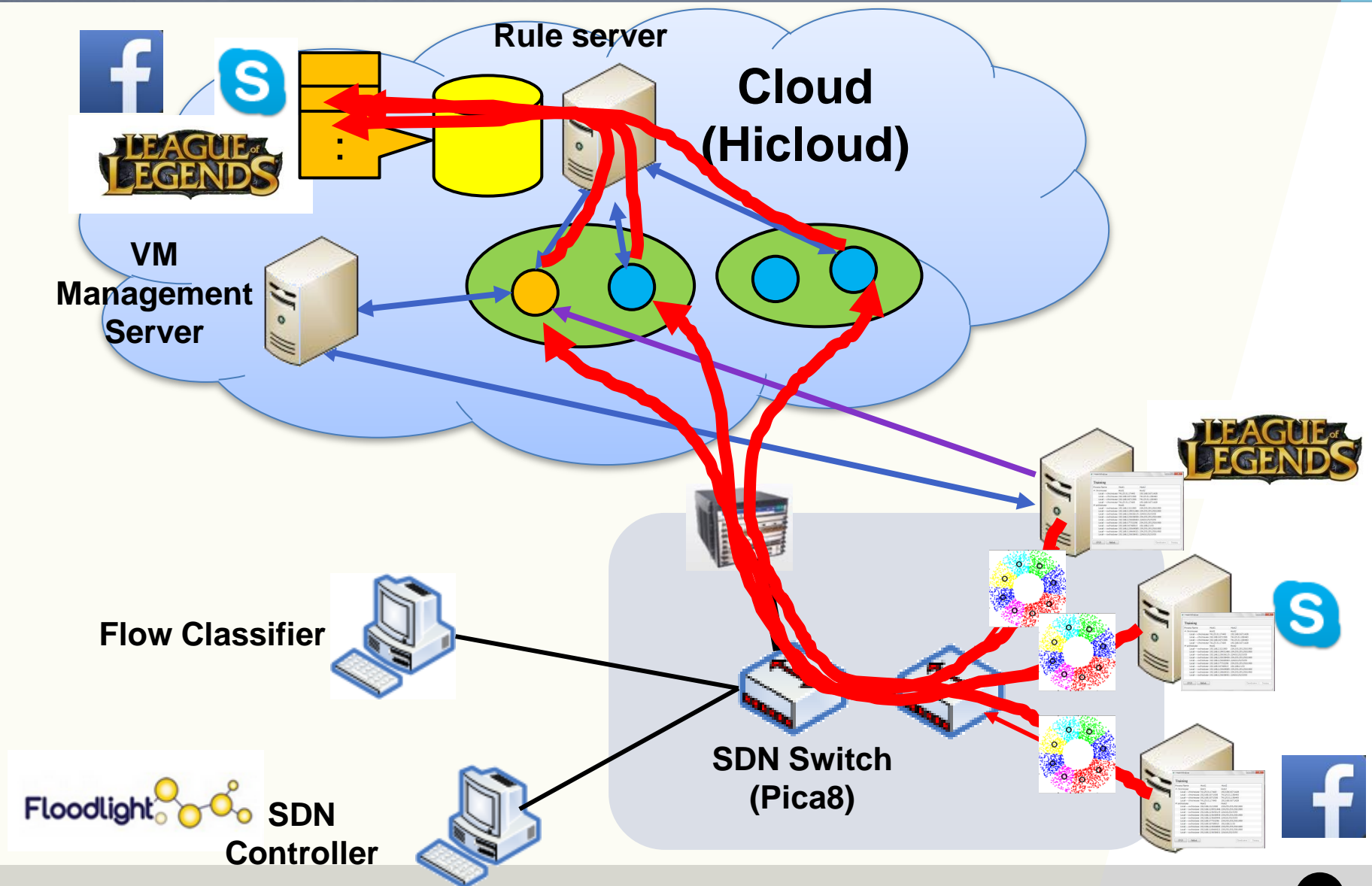
Classifier VM



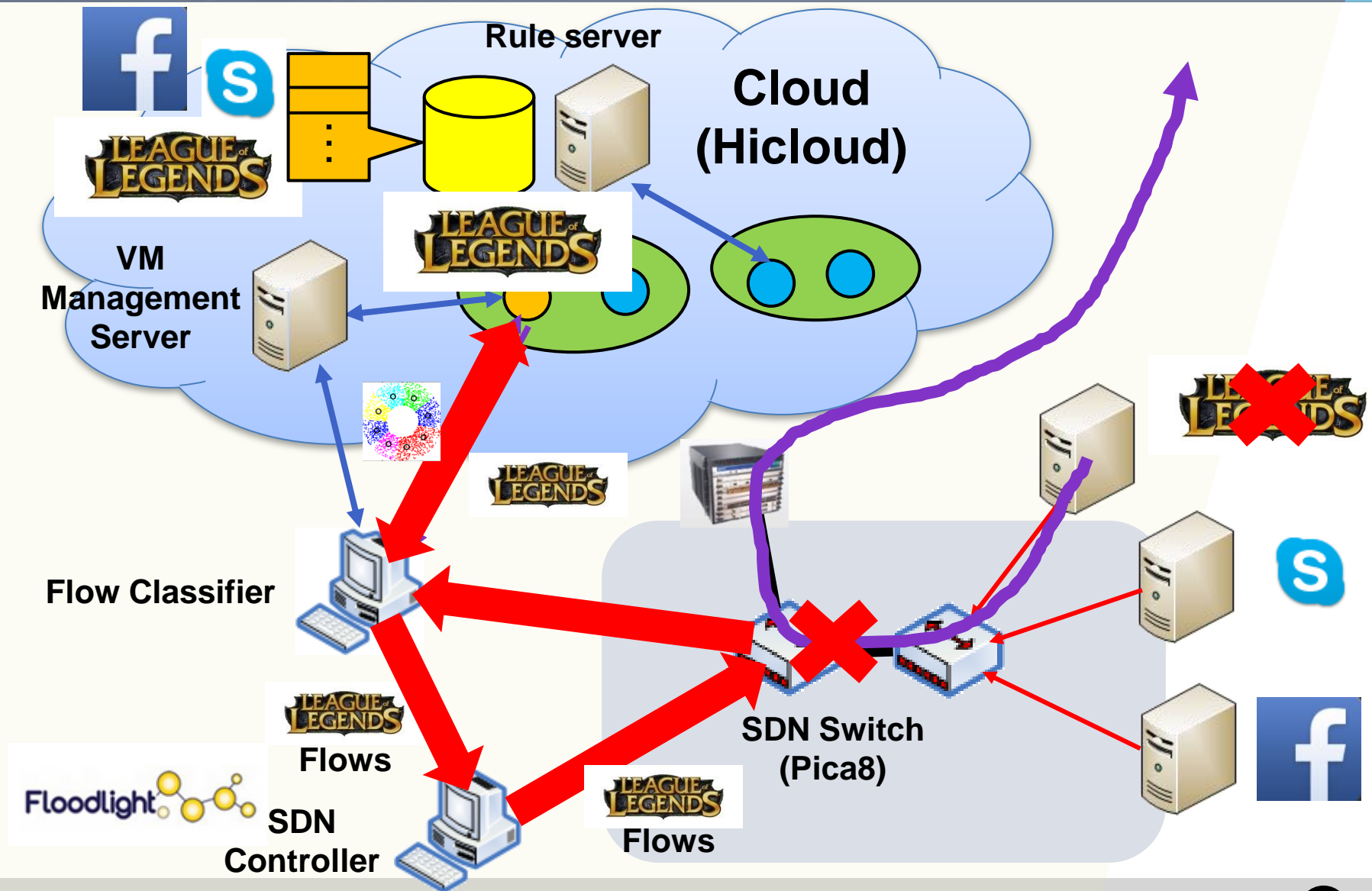
Contents

- 1 Introduction to SDN Networks
- 2 Key Issues of SDN Switches
- 3 Machine Learning Based Applications Classification
- 4 **Cloud + MLAC + SDN networks**
- 5 Demo (Video)
- 6 Conclusions

Cloud + MLAC + SDN (Training)



Cloud + MLAC + SDN (Classification/Management)



Contents

- 1 Introduction to SDN Networks
- 2 Key Issues of SDN Switches
- 3 Machine Learning Based Applications Classification
- 4 Cloud + MLAC + SDN networks
- 5 Demo (Video)
- 6 Conclusions

Demo of Applications Training (Video)

Demo of Applications Classification and Management with SDN Switch (Video)

Conclusions and future works

- ❖ The key issues of SDN networks include **fast and accurate flow classification schemes**
- ❖ The current SDN switches only support layer-4 flow table.
- ❖ A new service framework of applications classification for SDN networks
 - Cloud based for scalability, reliability
 - Machine Learning based for light communications
 - Global collected application attributes (statistical signatures) database.
 - Early stage application management
- ❖ Currently, **this platform is still only a prototype**, more training clients need to be installed (the client training software will be released later) to learn the worldwide applications as complete as possible.
- ❖ The accurate of the training and classification of this platform (currently around 80-98% of accuracy) needs also to be enhanced by tuning the machine learning model.

未來展望

- ❖ 建立全球最完整之網路應用軟體統計特徵資料庫
 - 開放式, 眾志成城 (more PCs to run applications)
- ❖ 與雲端平台業者組織合作
 - 佈建此資料庫 (for millions of applications)
 - 可動態擴增之雲端應用軟體**訓練**虛擬機器 (scalable)
 - 可動態擴增之雲端應用軟體**辨識**虛擬機器 (scalable)
- ❖ 提供全球網路應用軟體即時辨識服務 (SaaS)
 - 搭配 SDN 網路提供加值服務
 - QoS, Security, Network Management, ..., etc
- ❖ 臺灣學術網路或校園網路示範建置



Thank you for your attentions!