



# e-Science activities at HCMUT

**Thoại Nam**

[namthoai@hcmut.edu.vn](mailto:namthoai@hcmut.edu.vn)

HPC Lab (<http://www.hpcc.hcmut.edu.vn/>)  
Faculty of Computer Science and Engineering  
HCMC University of Technology  
<http://www.cse.hcmut.edu.vn/>





# Contents

- Motivation
- Problems
- How we do



# Motivation

- HCM city
  - Big plan for Smart city
  - Problems
    - (1) Traffic jam
    - (2) Urban flooding
    - (3) Pollution: waster & air
- Mekong delta
  - Flooding
  - Salinization
  - Healthy

# Traffic in HCMC

## ■ Big data

- Motorbike: 5M
- Car: 500K
- 3.800 roads with the total length of 3.670km
- Realtime

## ■ Complexity

- Many motorbike
- Safety space
- Traffic coming from small routes



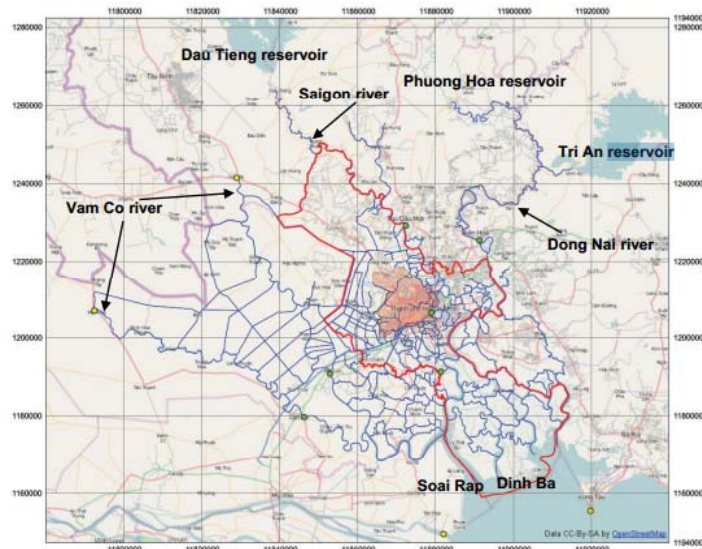
# Urban flooding in HCMC



Street or River?



Airport?

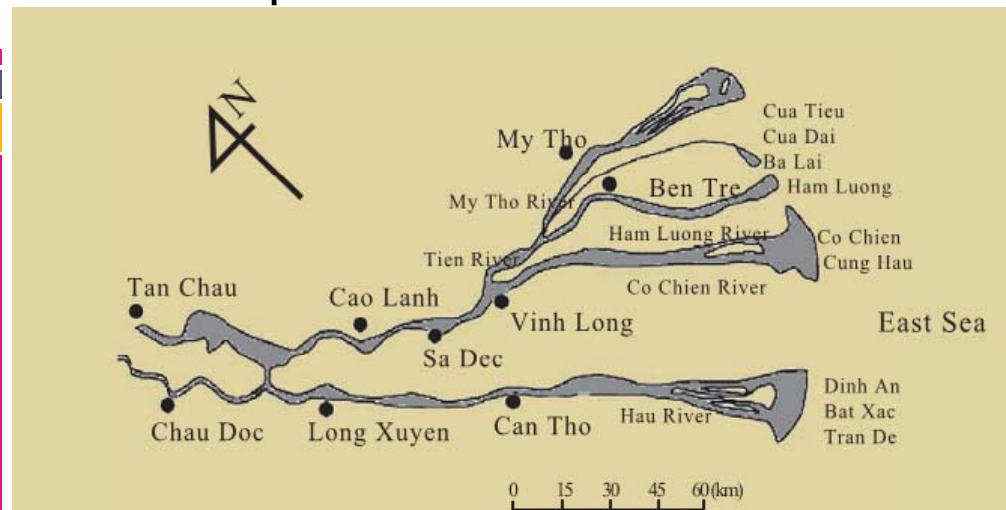


Safety?



# Problems in Mekong Delta

- Flooding
- Salinization
- Bank erosion
- Navigation hazards
- Pets
- Invasive exotic plants
- Healthy risks: water-borne, skin, mosquito-borne diseases

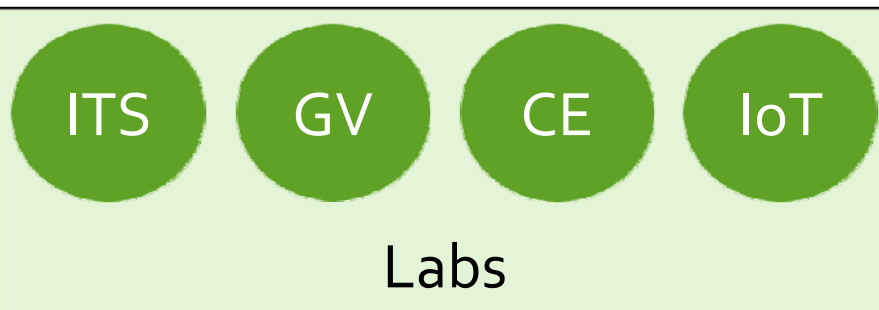


# HCMUT

(1) HCM city: smart city projects  
(2) Mekong delta

Real problems

Traffic jam  
Urban flooding  
Pollution: waster & air



Applications

HPC Lab

Computing  
Infrastructure

# Ecosystem (2012)

1. To foster the advances in diverse fields of Science & Technology
2. To be able to solve complex, realistic problems raised by HCMC and VN

Computational Science and Engineering Center

Institute of Computational Science and Technology

## HPC Center of Excellence

### Intel

1. Intel HPC Lab
2. PR
3. Experts
4. Linking with International partners
5. HPC software, tools
6. Training, Master programs on HPC and CE
7. HPC&CE International Conf.

### HCMUT - VNU

1. Building a HPC center
2. Research funding for HPC and CE
3. Master programs on HPC and CE
4. Human resource
5. Scholarships: outstanding students in HPC and CE
6. HPC&CE International Conf
7. Machine ~**40TFlops** (2014)



# High Performance Computing

## Partners

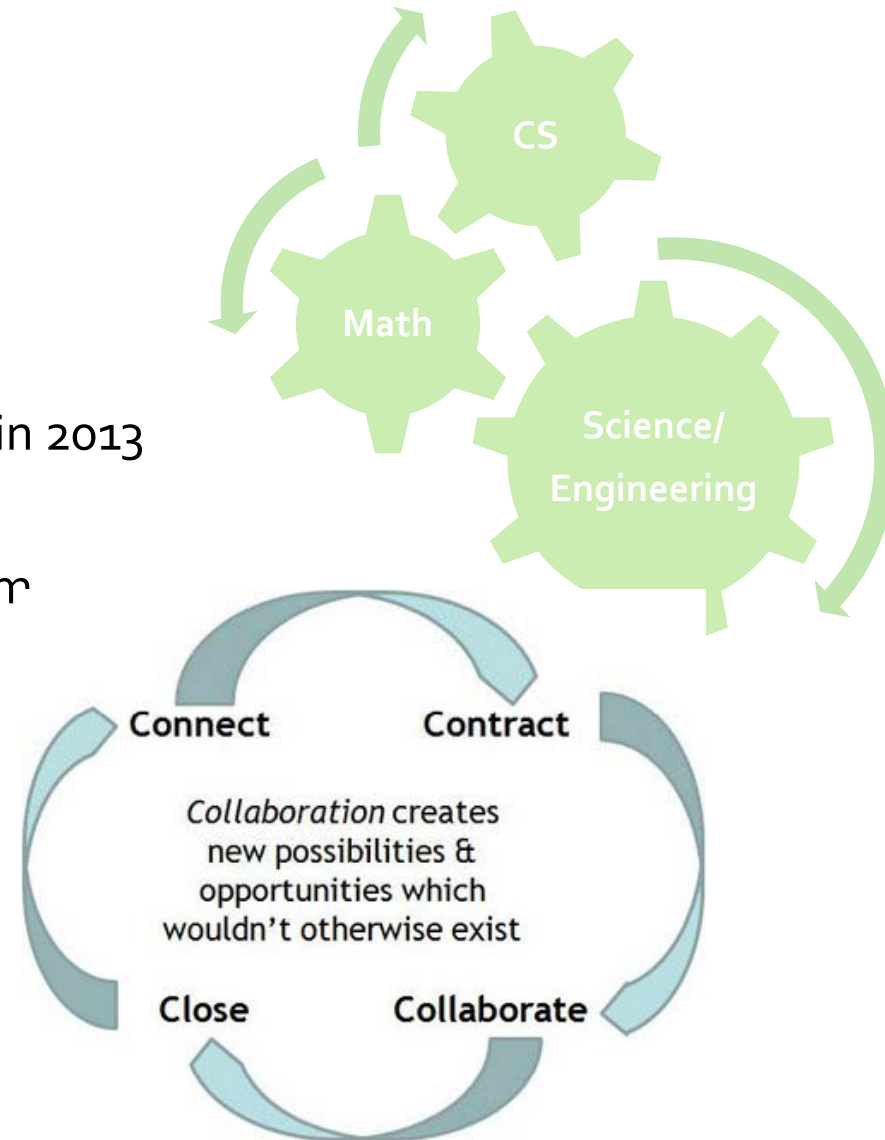
- HCMUT - VNU
- Intel
- Ho Chi Minh city

## Plan: 2012-2022

- HPC research Lab: Set up in 2013
- HPC Center
- Strengthen HPC in Vietnam
- Solving big problems
- Leading in technology

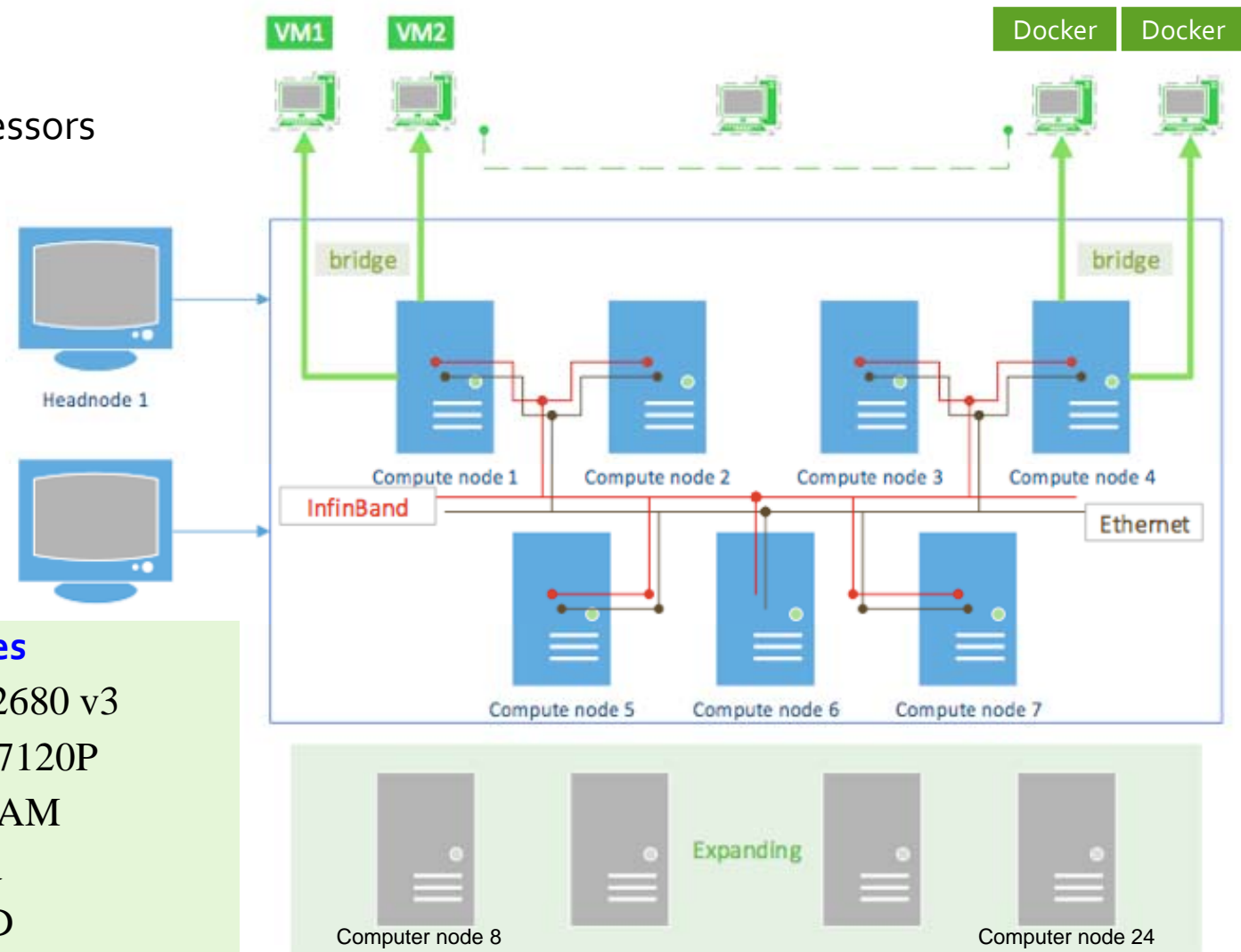
## Applications: 2013-2022

- Traffic analysis
- Urban flooding
- Big data analytics



# SuperNode-XP: 50 TFlops

- Vendors: HPE
- Intel Xeon processors
- Intel Xeon Phi
- Infiniband

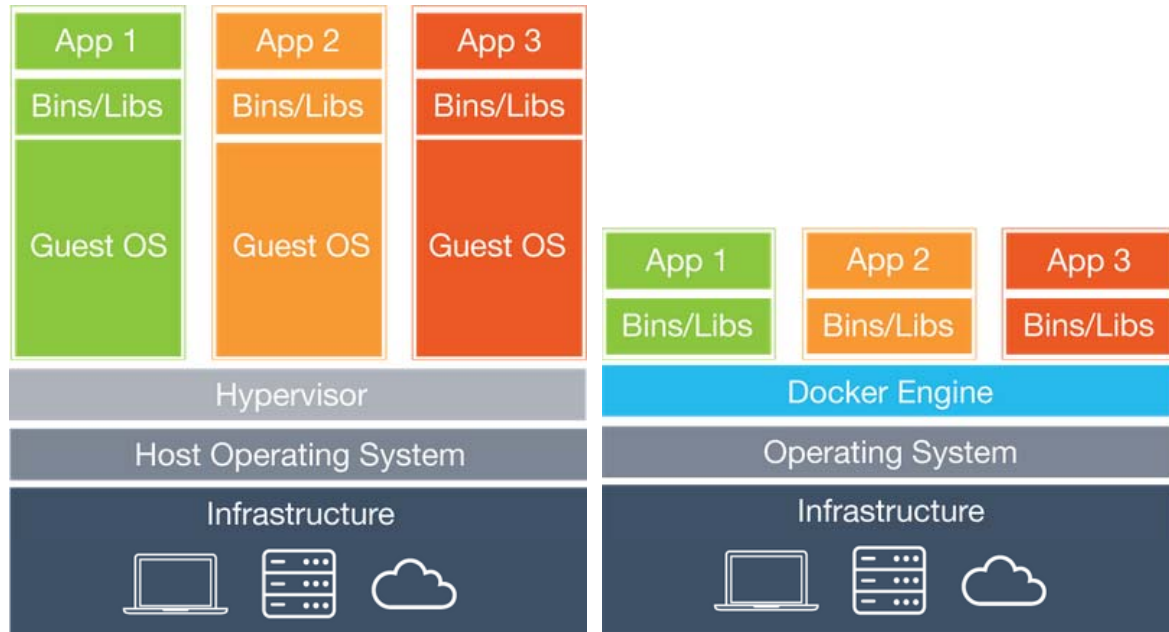


## 24 Computing nodes

- 2 x Intel Xeon E5-2680 v3
- 2 x Intel Xeon Phi 7120P
- 512/256/128 GB RAM
- 50 Gbps Infiniband
- 2 TB hard disk/SSD
- 200 TB Storage

# Computing platforms

Traditional platform	HPC Cloud	
	Virtual machine	Docker
Physical machines Shared resource? Performance Communication Coprocessors	Virtual machines (VMs) Private environment <b>Performance</b> <b>Communication</b> <b>Coprocessors</b>	Dockers Shared resource <b>Performance</b> <b>Communication</b> <b>Coprocessors</b>





# Libraries in SuperNode-XP

- ANSYS
- Cadence
- Open Telemac
- OpenFOAM
- Gromacs
- Gaussian
- BLAST
- Hadoop
- MPI
- Intel@ Parallel Studio XE
- Machine/Deep learning tools

# Performance Evaluation of Environmental Applications using TELEMAC-MASCARET on Virtual Platforms

EWC workshop - The IEEE 12th International Conference on eScience, 23-27 Oct 2016, Baltimore, Maryland, USA

Minh Thanh Chung<sup>1</sup>, Manh-Thin Nguyen<sup>1</sup>, Nhu-Y Nguyen-Huynh<sup>1</sup>,

Nguyen Thong<sup>2</sup>, Nam Thoai<sup>1</sup>

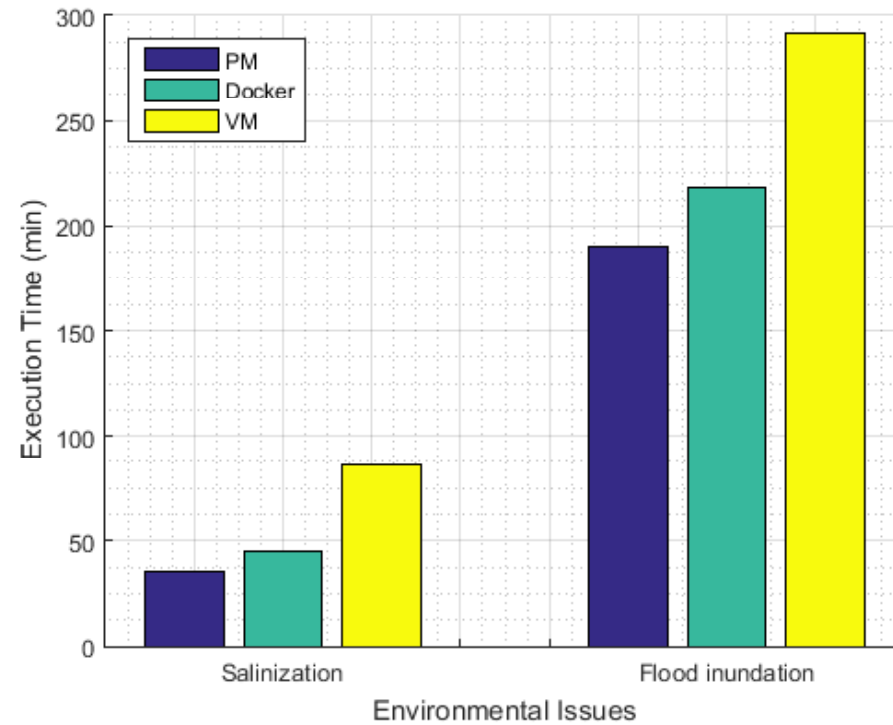
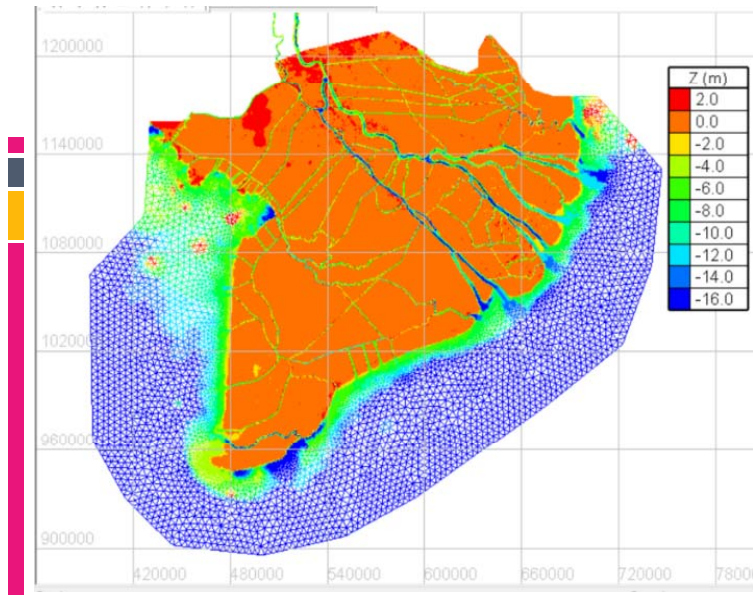
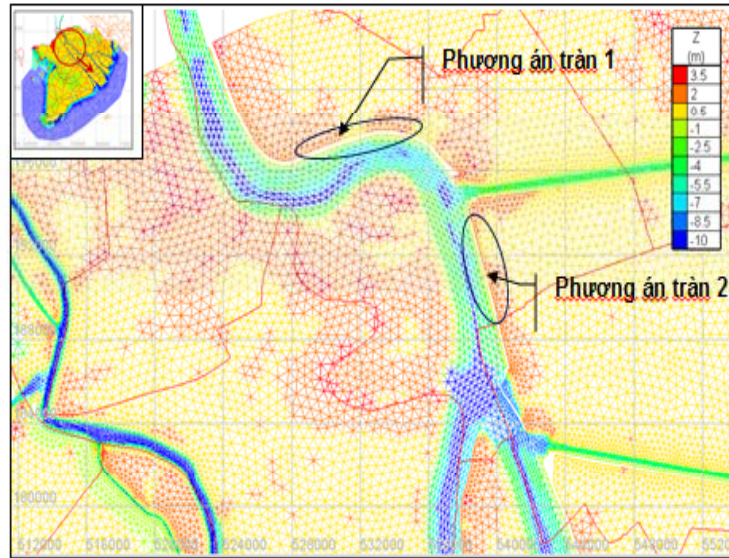
<sup>1</sup>HPC Lab – Faculty of Computer Science and Engineering

<sup>2</sup>Faculty of Civil Engineering

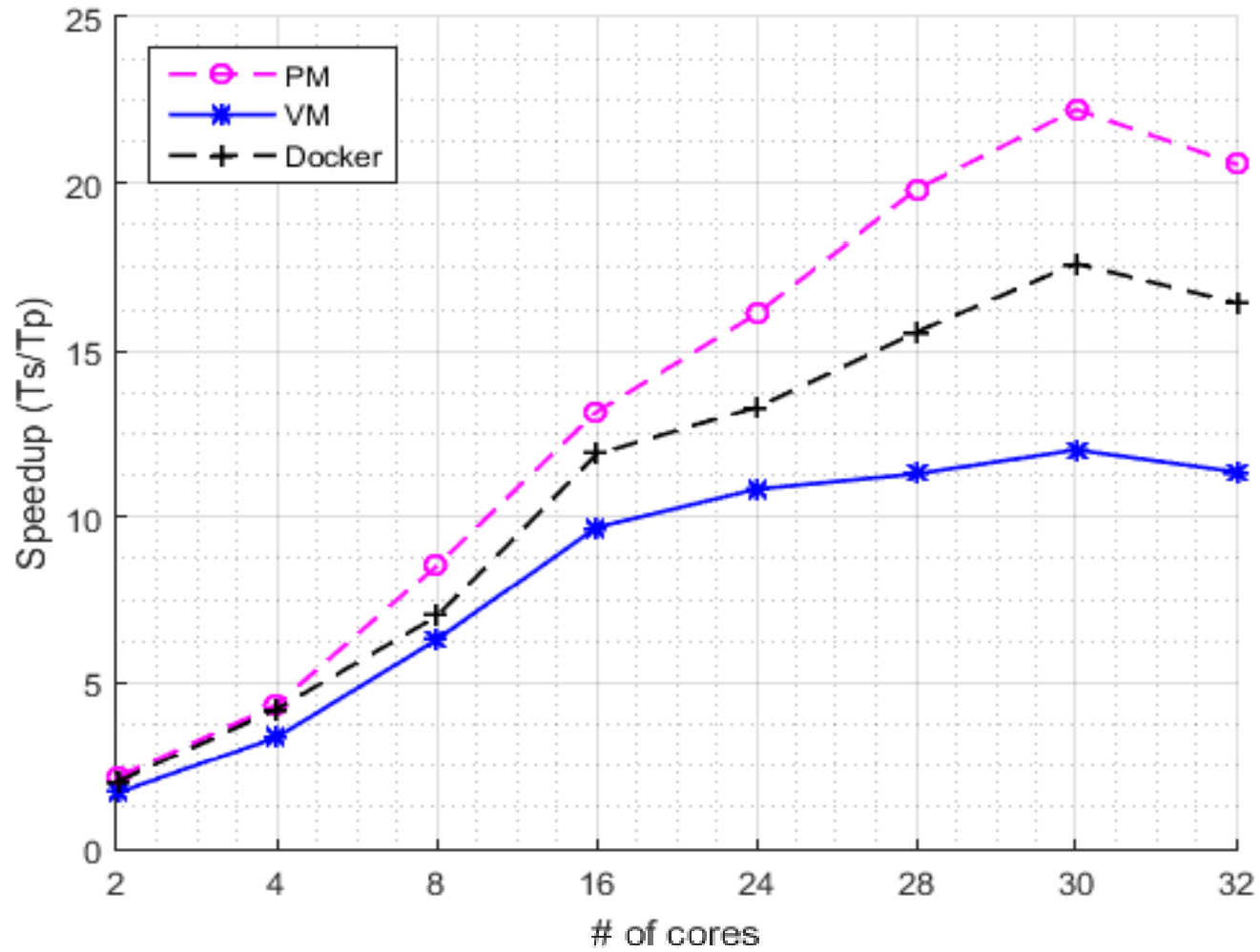
HCMC University of Technology, Vietnam



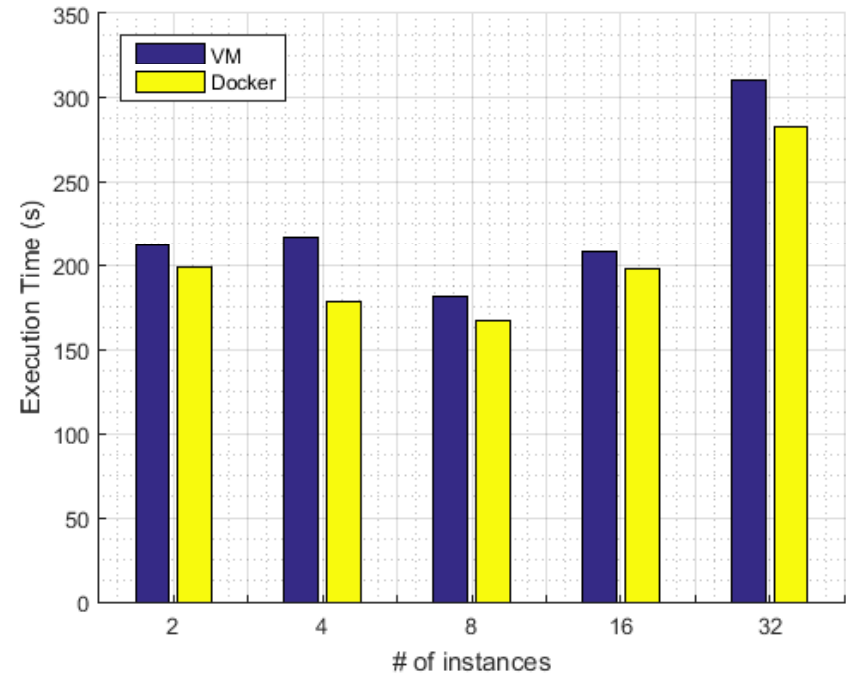
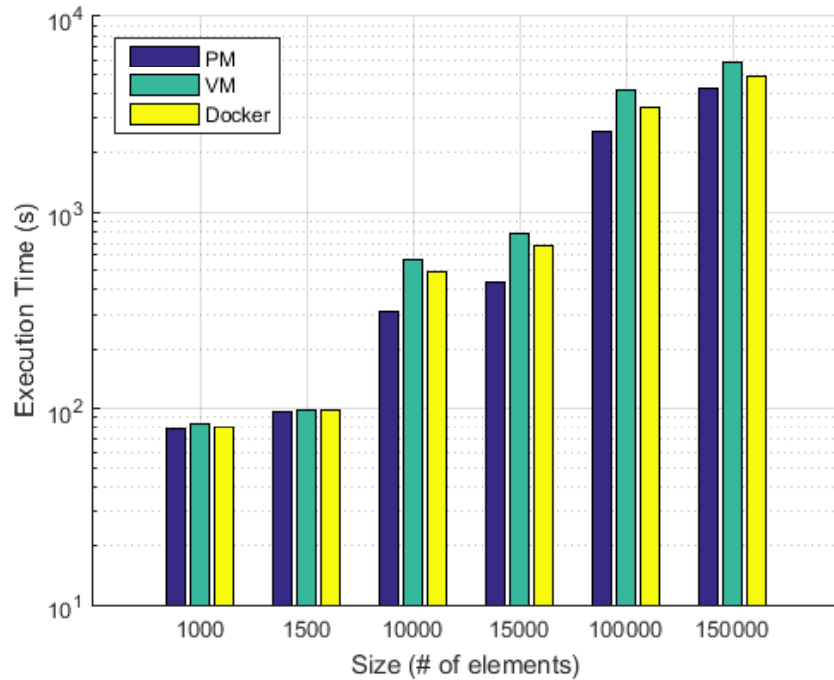
# Real data of Mekong Delta



# Speedup



# Problem size and scalability



# ITS Lab

Pham Tran Vu



## Smart BK Traffic

[traffic.hcmut.edu.vn](http://traffic.hcmut.edu.vn)

Provide the current state of the traffic in Ho Chi Minh City in real-time



## Road Maester

Help the people living in Ho Chi Minh monitor the traffic condition in real-time and avoid bad traffic areas



## Smart Bus

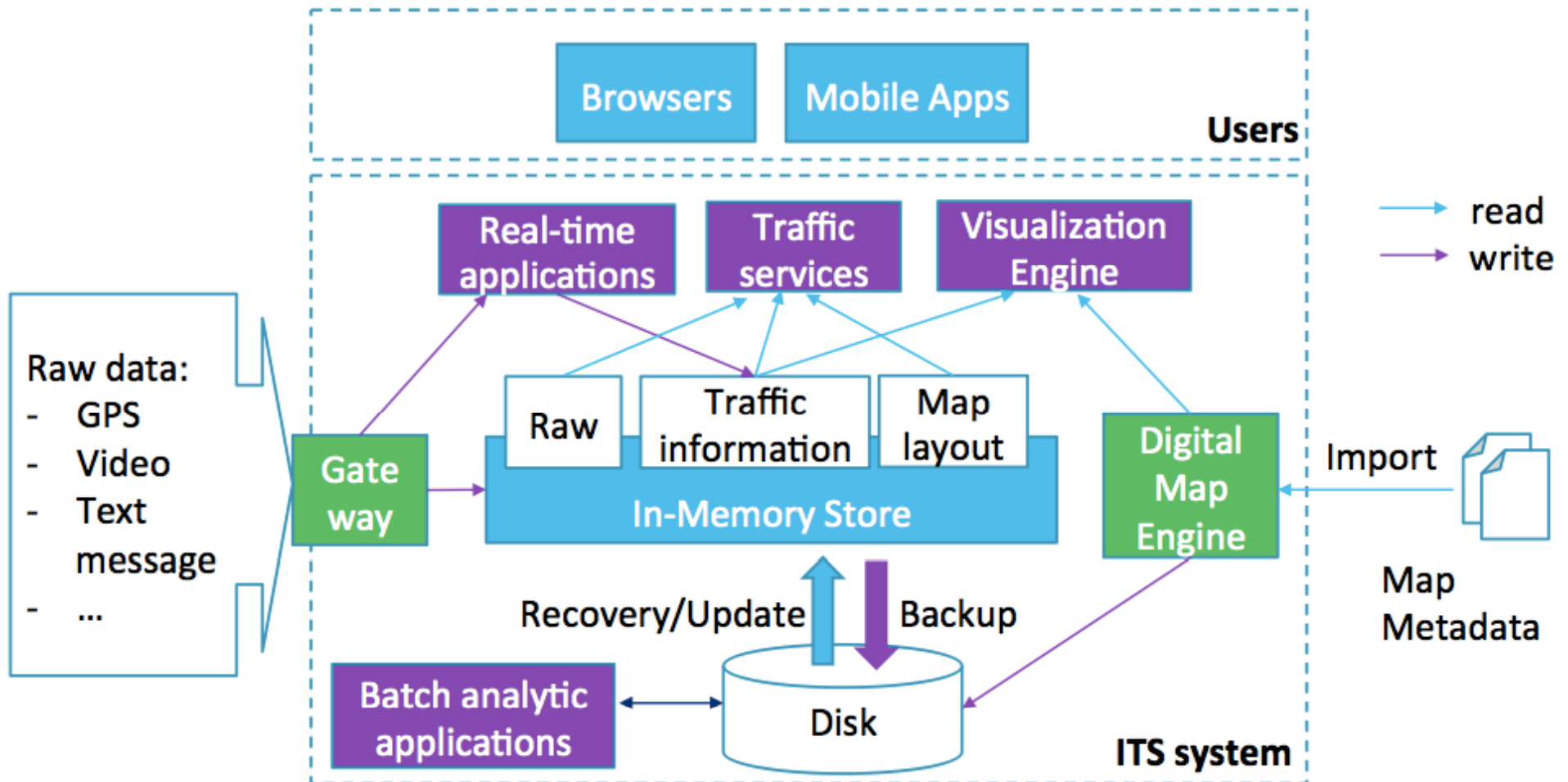
Provide useful information and utilities for bus travellers

# Challenges

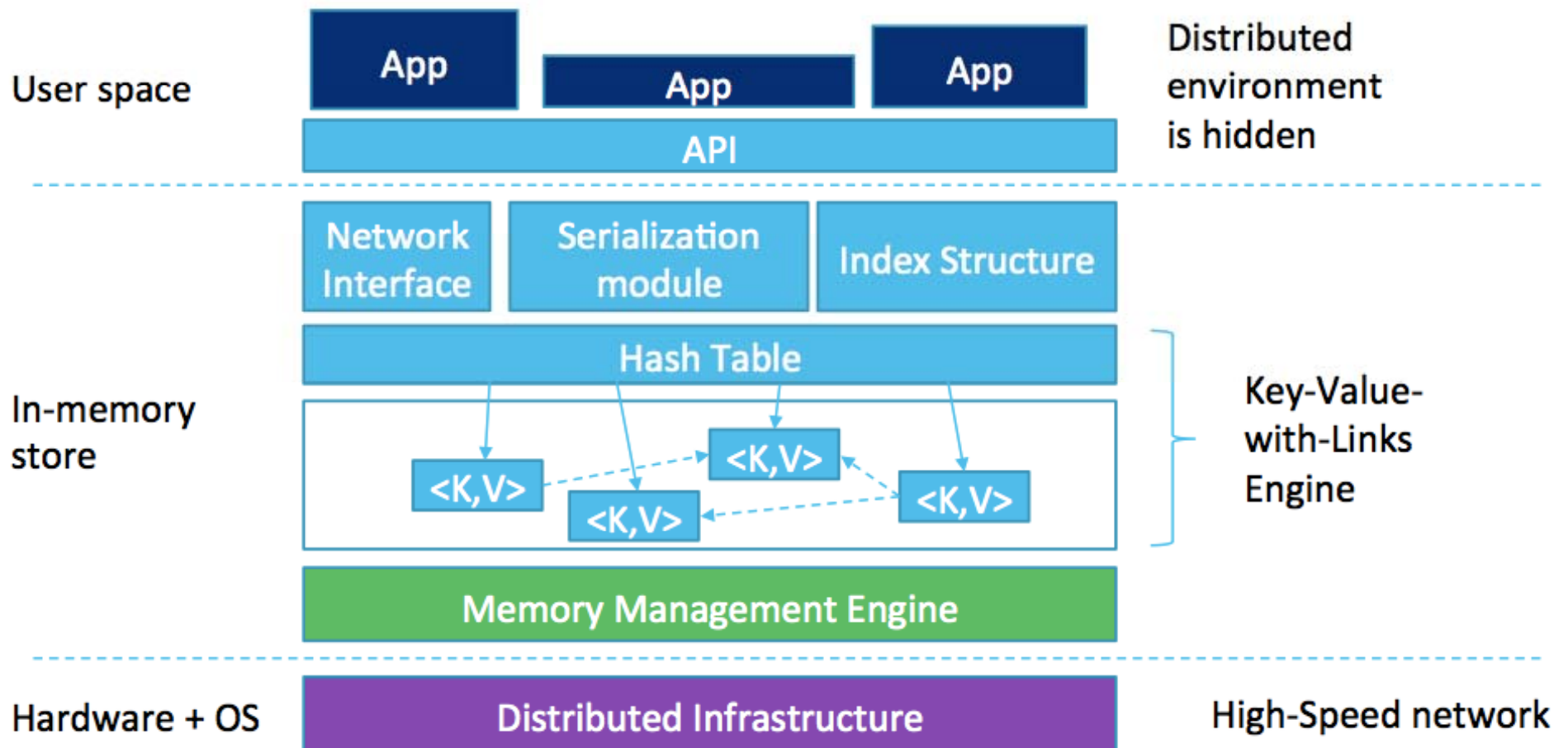
- The volume of sensing data keeps increasing over time.
  - More than 10M records generated per day
- Extracting useful traffic information from a massive amount of data in real-time.
  - The processing must be done within 1 minute
- Low stability of the network infrastructure, The quality and heterogeneity of sensing devices, and the accuracy of the digital map
  - Only ~17% of raw data could be used directly
- Unique (and difficult) traffic problems
  - Motorbike is the most popular mean of transportation
  - Transportation development does not keep up with economic and population growth.



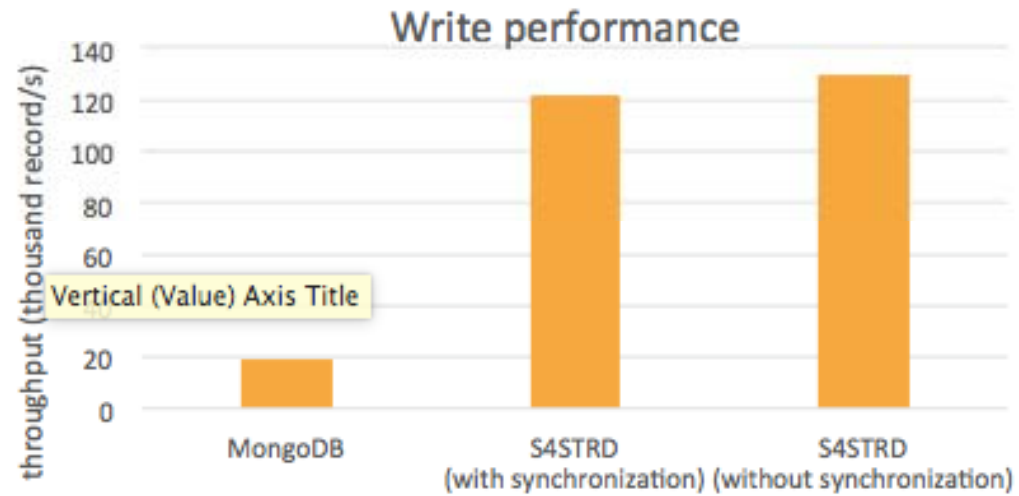
# ITS System Architecture



# || Distributed In-memory Storage System for Big Data



# Performance of Storage System



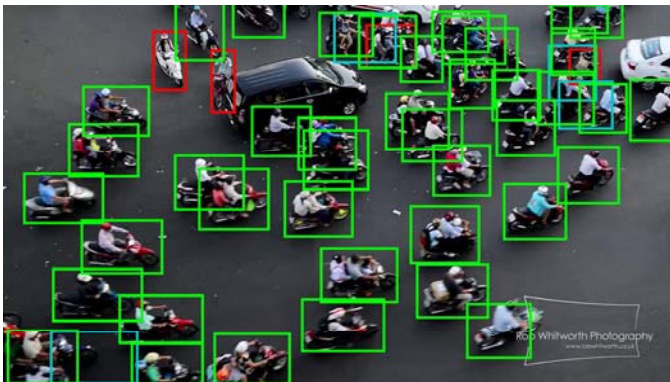
# Graphics & Vision Lab

Le Thanh Sach

## Automatic vehicle detection

- *Deep learning, Machine learning* -

- Object of interest: motorbike, car, bus, etc.
- Achieve an accuracy up to 95% in ideal conditions.
- Can predict correctly if there are around 60 vehicles per image or fewer.
- Can be reused in other problems (human counting, forest density estimation, etc.)



# Graphics & Vision Lab

Le Thanh Sach

## Motorbike count estimation

### - *Machine learning* -

- Counting motorbike by applying regression over global image features (do not know the position of each vehicle).
- Fast (18 FPS) and accurate (93%)
- Can predict accurately up to 100 motorbikes per image.



→ Motorbike count: 63



→ Motorbike count: 87



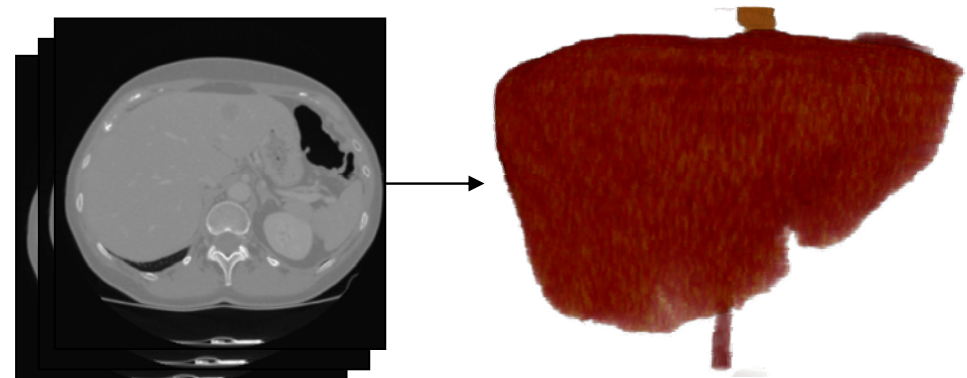
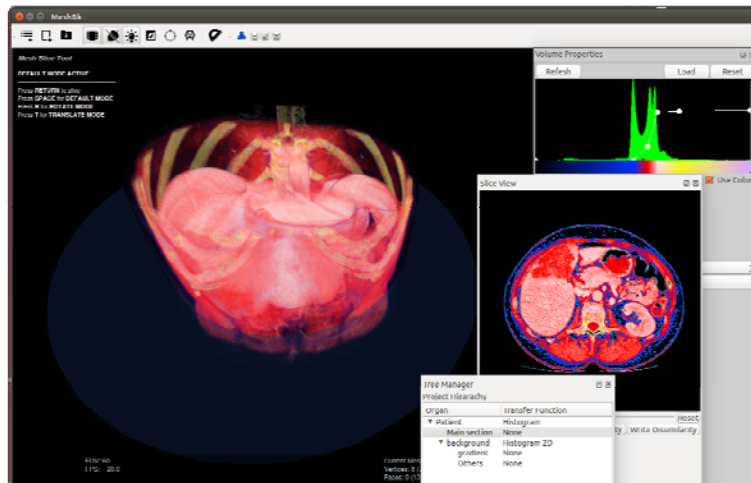
# Graphics & Vision Lab

Le Thanh Sach

## 3D Liver segmentation and visualization

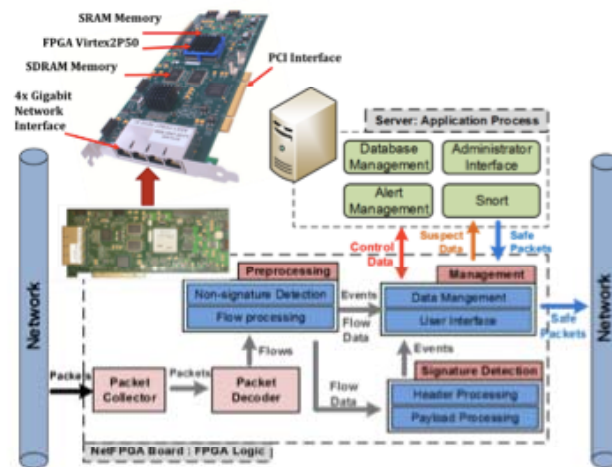
### - Machine learning, 3D Image processing -

- Process stacks of 2D MRI/CT images to generate 3D model of vital organs (bone, liver, kidney, etc.)
- This could help doctors locating anomalies or disease faster.
- There are two workflow: semi automatic and fully automatic. Each has accuracy higher than 80%.

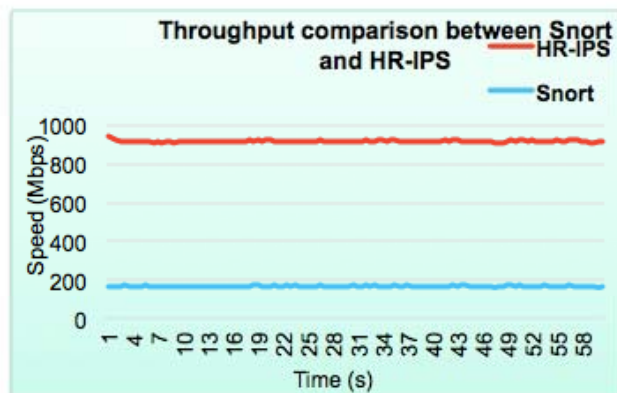
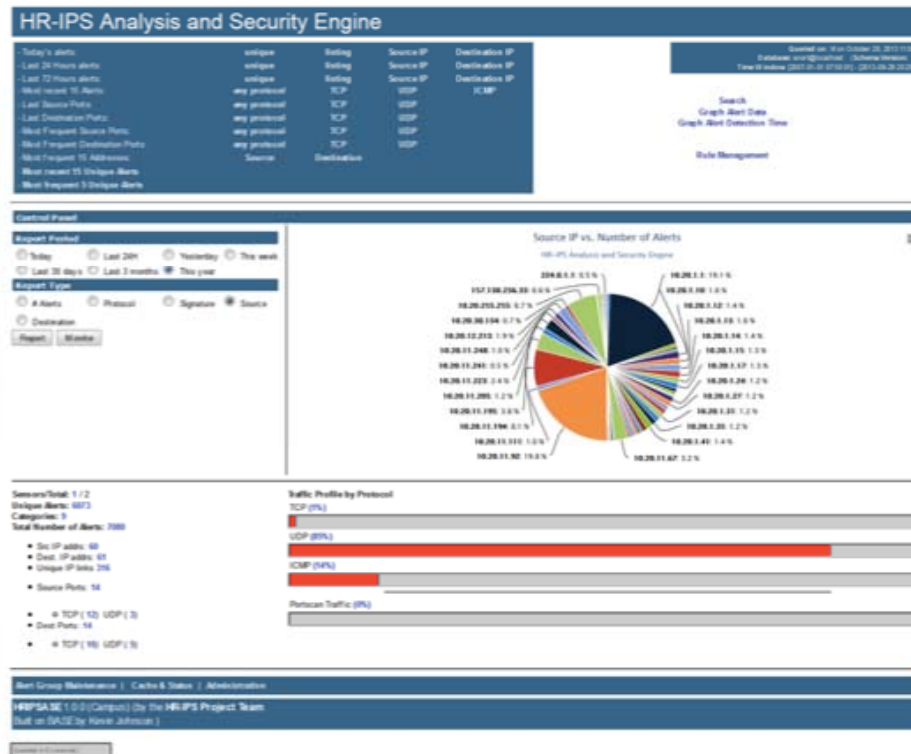


## Applying Reconfigurable Computing for Internet Security

### – HR-IPS: HCMUT Reconfigurable Intrusion Prevention System

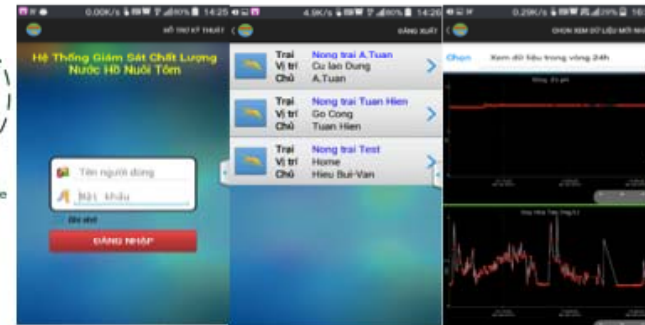
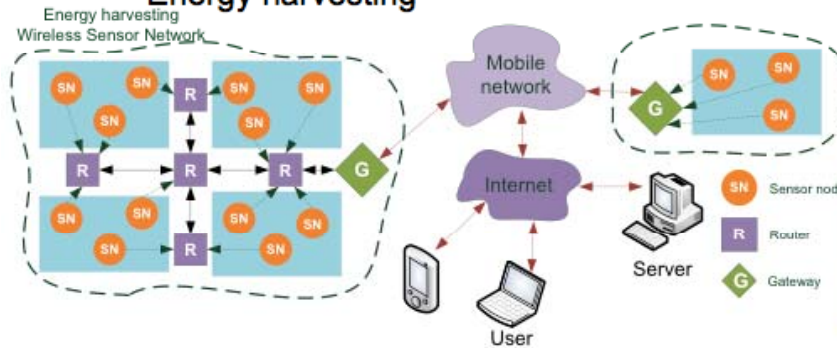


- HR-IPS meets line rate full duplex **Gbps network** link without any drops.
- 5x better** throughput compare to bare software solution (Snort IPS).



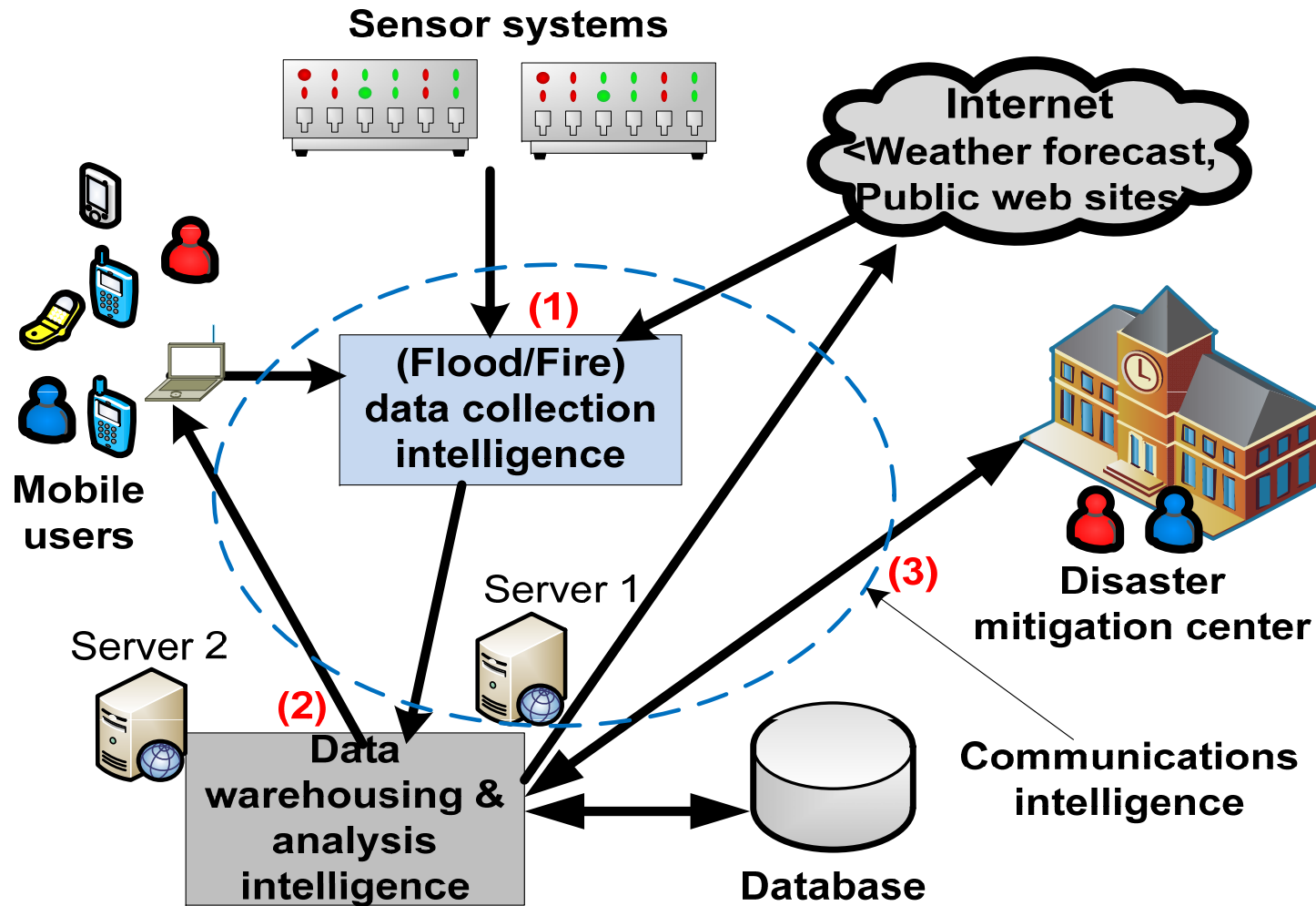
### Low-power wireless water quality monitoring system

- Water quality monitoring system for shrimp farming
- Energy harvesting
- Low-power network protocol
- Data management



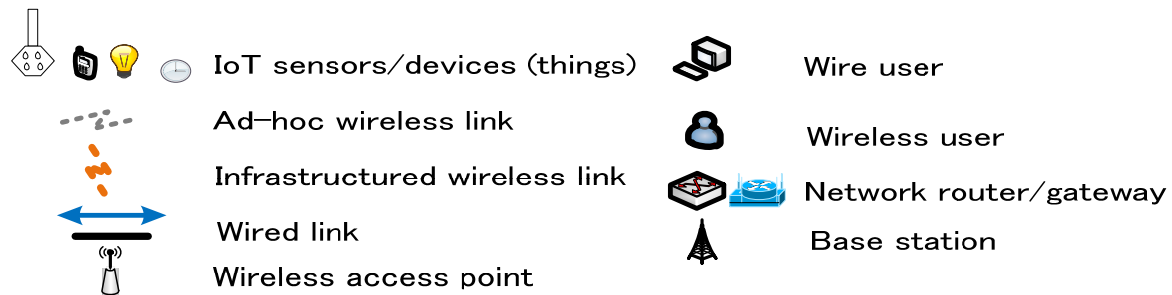
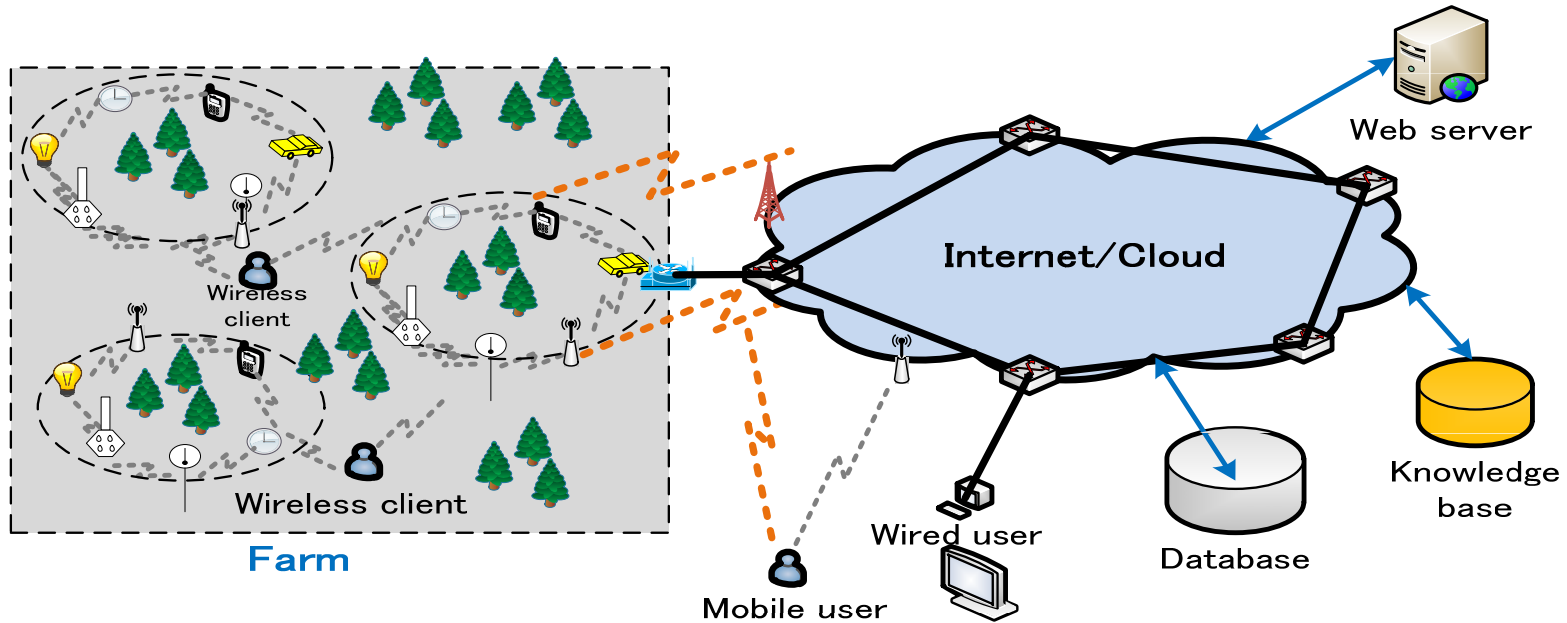
# IoT Lab

Tran Minh Quang

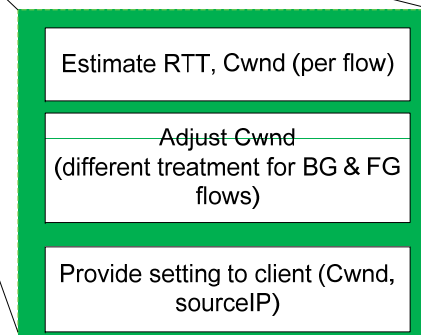
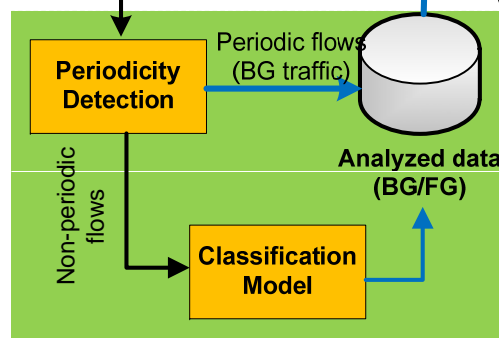
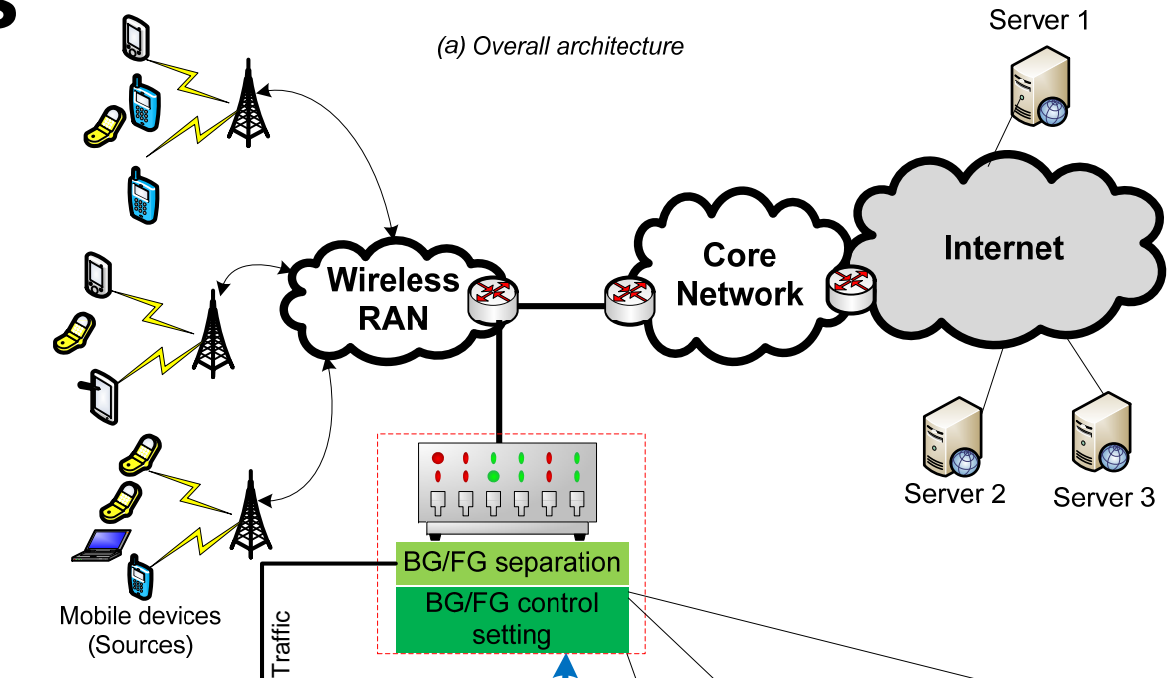


## Urban flooding mitigation

# IoT - Smart farm designs



# Network design and optimization for IoTs



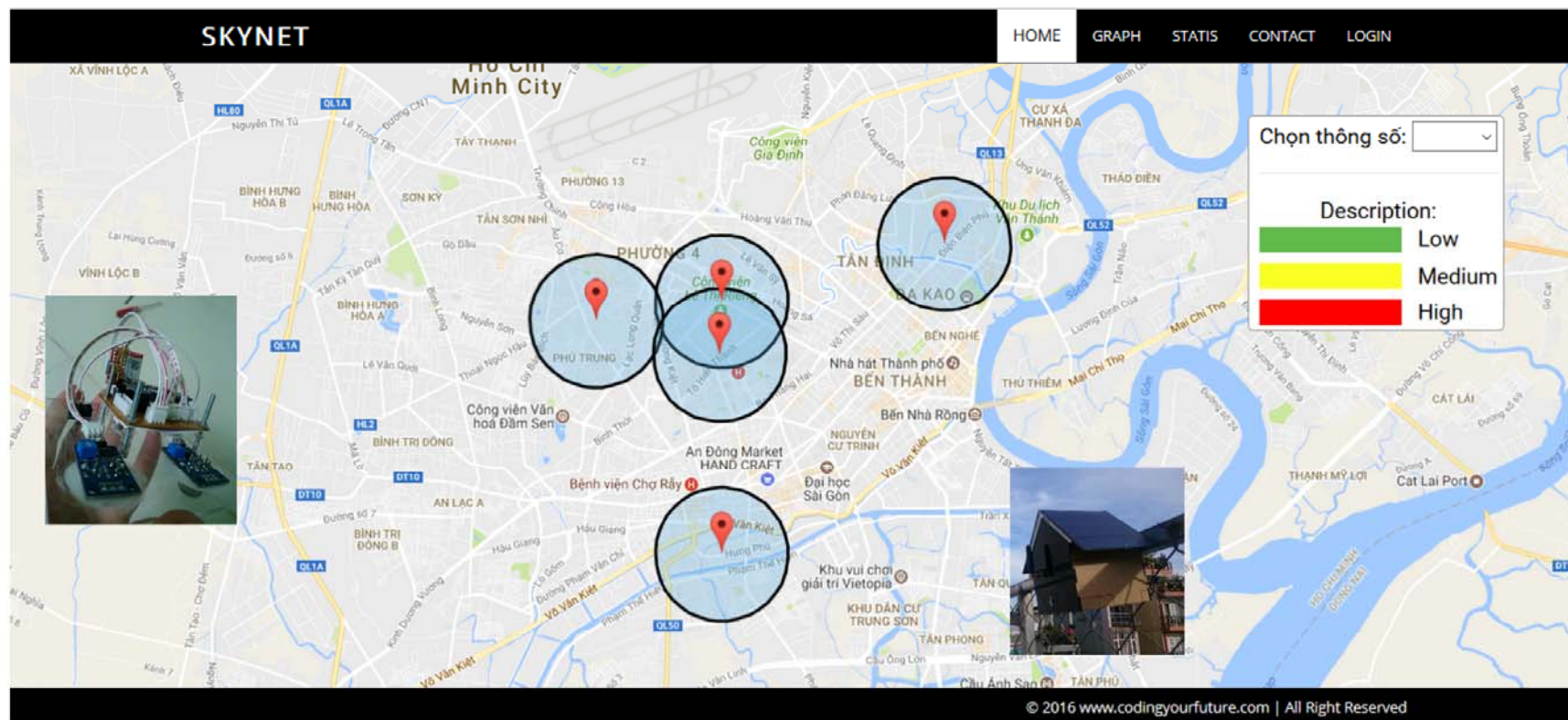
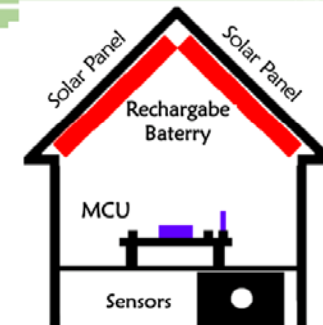


# IoT Lab

Pham Hoang Anh (anhpham@hcmut.edu.vn)

## Air Quality Monitoring System

- CO, Temperature, Humidity, GAS
- Solar Power and Rechargeable Battery.

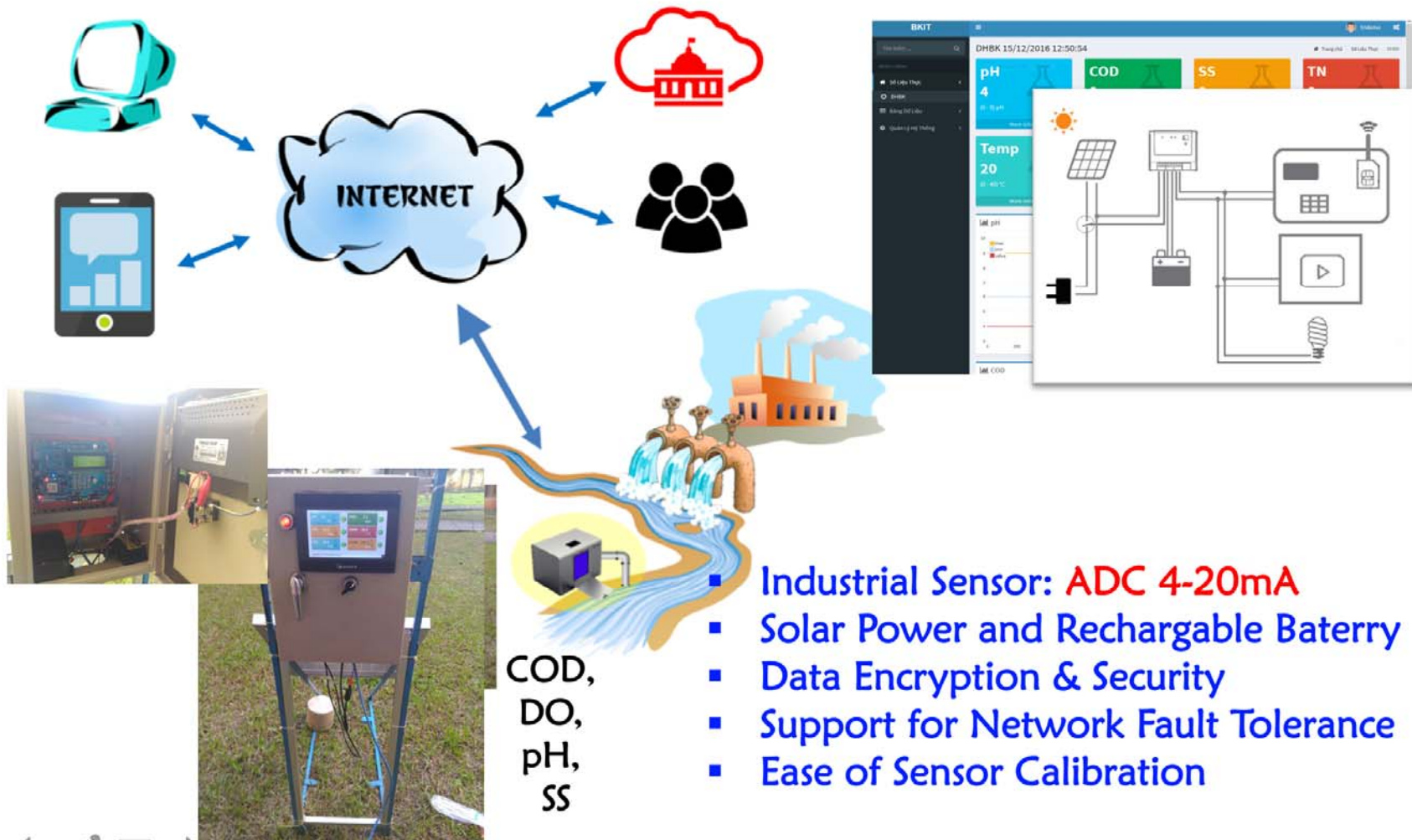




# IoT Lab

Pham Hoang Anh (anhpham@hcmut.edu.vn)

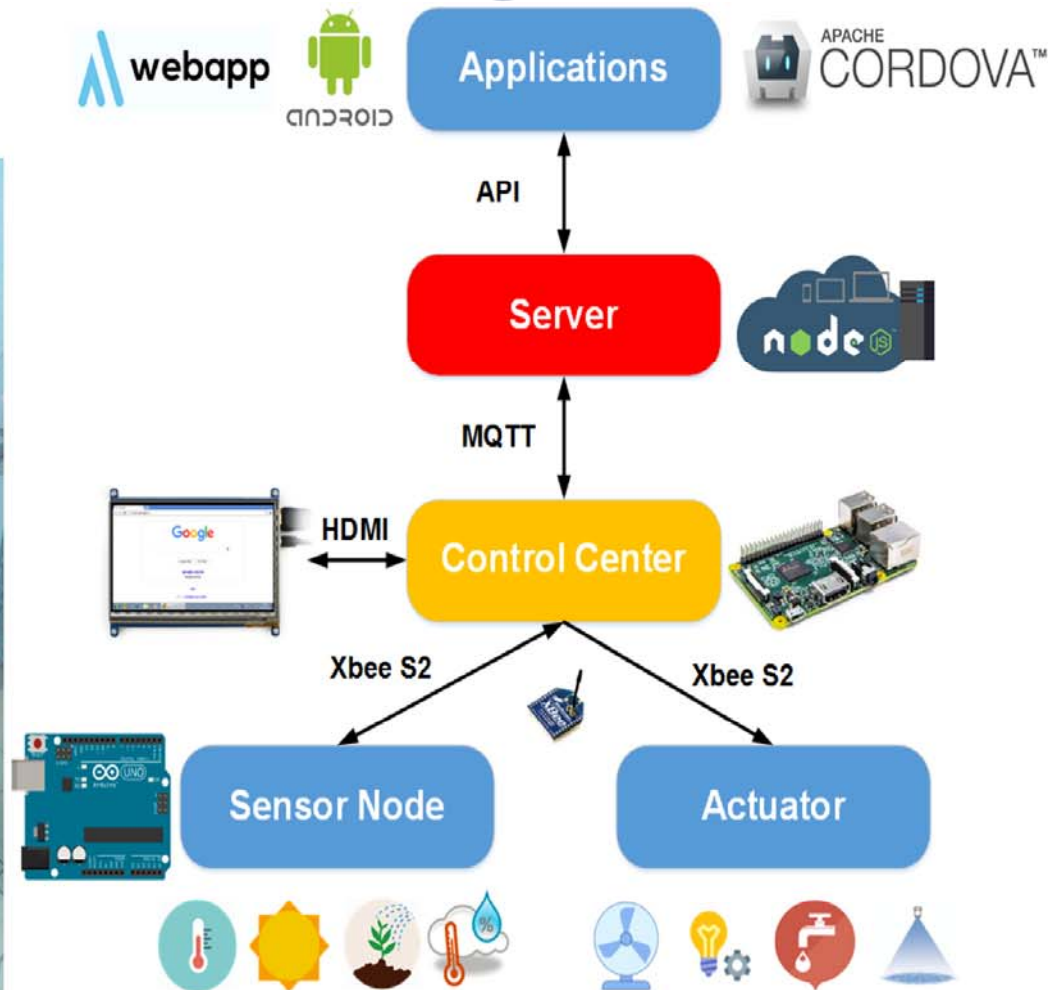
## Waste Water Monitoring System



# IoT Lab

Pham Hoang Anh (anhpham@hcmut.edu.vn)

## Prototype of Smart Farming



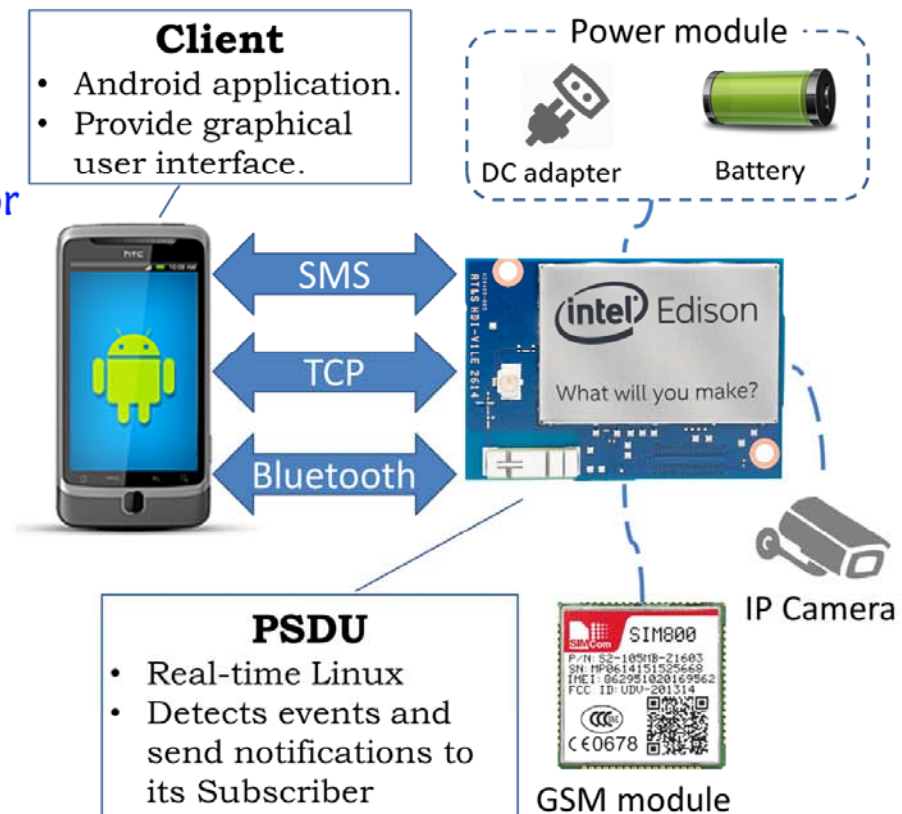
# IoT Lab

Pham Hoang Anh (anhpham@hcmut.edu.vn)

## Framework for Event Management in IoT

- Publish-Subscribe Framework for Event Management in IoT Applications. (**SEATUC 2017**)

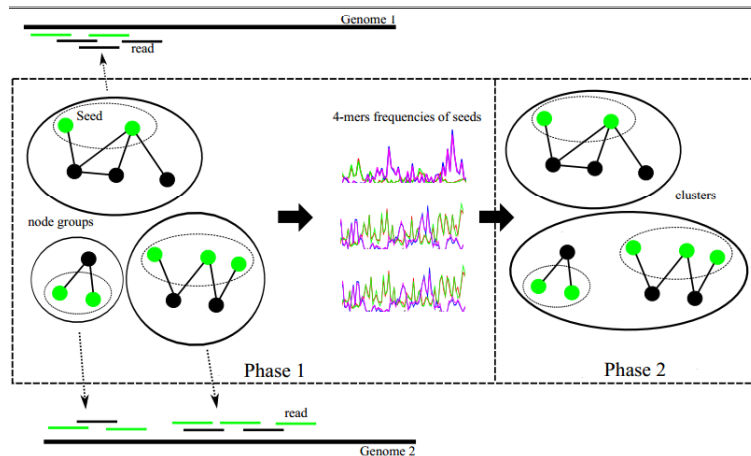
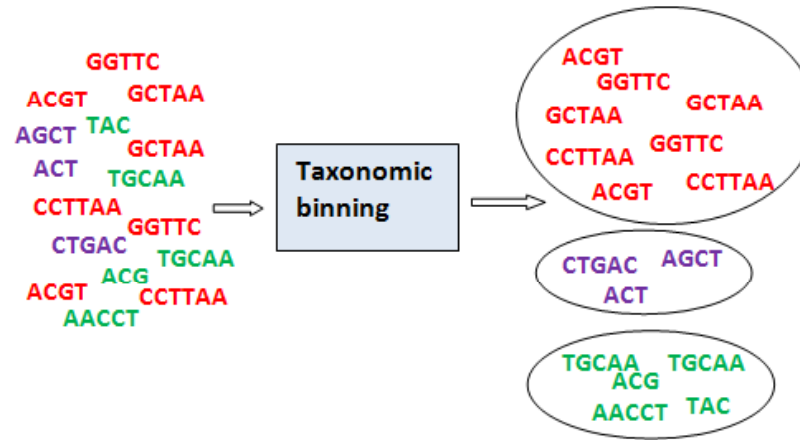
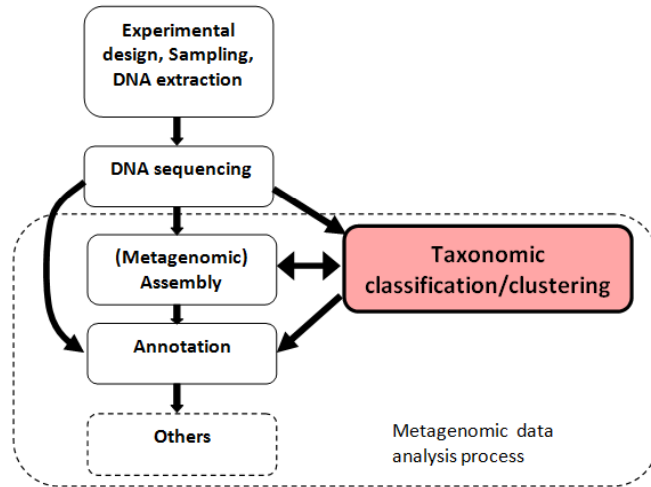
- Demo Application
  - Power Signal Detection Unit for Home Security Application



# Metagenomic data analysis

## BiMeta: Clustering of metagenomic reads

Tran Van Hoai



edu.vn/bioinfo/metapro/BiMeta.html

24 cuốn sách sẽ giúp

### Bioinformatics

- Home
- MetaAB
- BiMeta
- SeMeta
- Publications

### BiMeta

BiMeta is a new software for binning metagenomic reads. It supports both single-end and paired-end reads. The software is implemented in C++.

#### Download source code here:

- + Version 1.1: BiMeta1.1.tar.gz (updated date: 08/03/2015)
- + Version 1.2: BiMeta1.2.tar.gz (updated date: 16/01/2015) (Support an additional function to separate output file into different files of different species)

#### Publication:

Vinh, L. V., Lang, T. V., Binh, L. T., & Hoai, T. V. (2015). A two-phase binning algorithm using 1-mer frequency on groups of non-overlapping reads. *Algorithms for Molecular Biology*, (1), 2.

#### Datasets:

The application was tested with various datasets. The datasets can be download here:

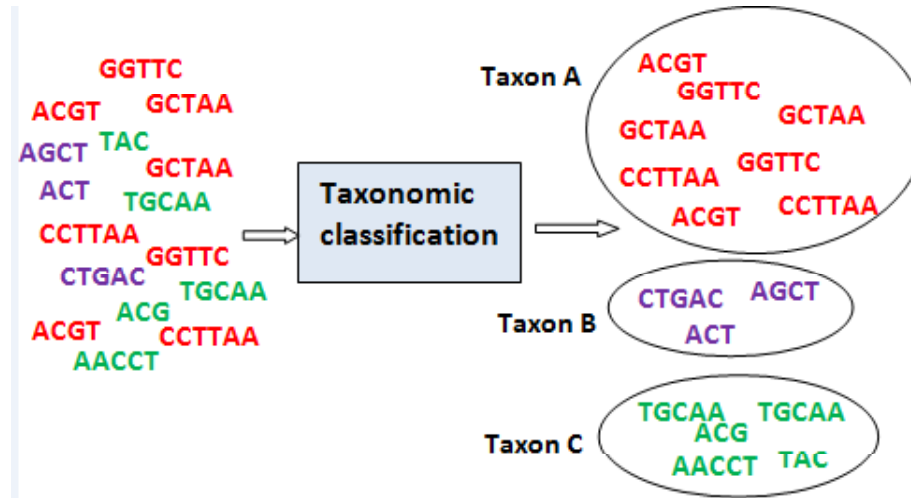
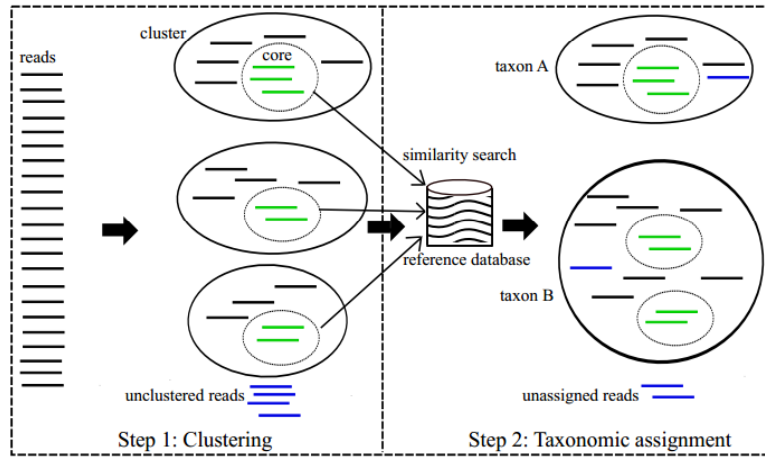
- + Short read datasets: S1.fna.gz, S2.fna.gz, S3.fna.gz, S4.fna.gz, S5.fna.gz, S6.fna.gz, S7.fna.gz, S8.fna.gz, S9.fna.gz, S10.fna.gz, L1.fna.gz, L2.fna.gz, L3.fna.gz, L4.fna.gz, L5.fna.gz, L6.fna.gz
- + Long read datasets: R1.fna.gz, R2.fna.gz, R3.fna.gz, R4.fna.gz, R5.fna.gz, R6.fna.gz, R7.fna.gz, R8.fna.gz, R9.fna.gz
- + Acid mine drainage: <http://www.ncbi.nlm.nih.gov/books/NBK6860/>



# Metagenomic data analysis

## SeMeta/ParSeMeta: Classification of metagenomic reads

Tran Van Hoai



[edu.vn/bioinfo/metapro/SeMeta.html](http://edu.vn/bioinfo/metapro/SeMeta.html)

24 cuốn sách sẽ giúp

### Bioinformatics



- Home
- MetaAB
- BiMeta
- SeMeta
- Publications

### SeMeta

SeMeta is a new software for taxonomic assignment of metagenomic reads. It supports both single-end and paired-end reads. The software is implemented in C++

#### Source codes:

Download source code here:

Version 1.0 (release 02/10/2015): [SeMeta1.0.tar.gz](#)

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

The application was tested with various datasets. The datasets can be downloaded here:

+ Dataset 1: [ds1.fna.gz](#)

+ Dataset 2: [ds2.fna.gz](#)

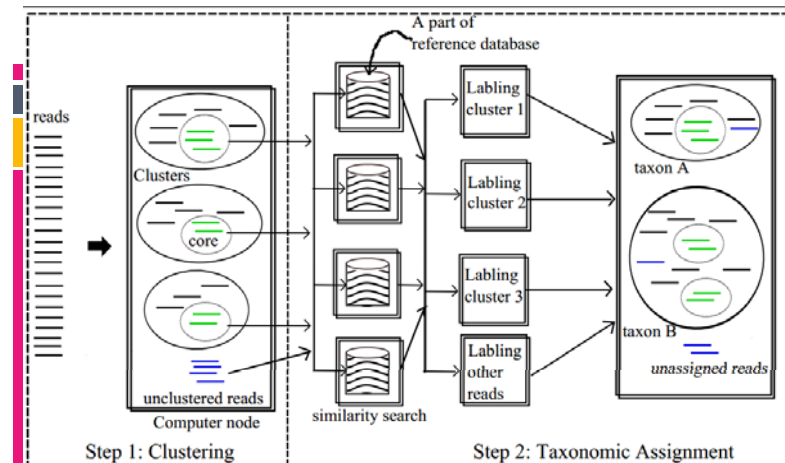
+ Dataset 3: [ds3.tar.gz](#)

+ Acid mine drainage: <http://www.ncbi.nlm.nih.gov/books/NDK6060/>

+ Human gut metagenome: <http://public.genomics.org.cn/BGI/gutmeta/>

#### How to compile the source code, and prepare input files:

Please read the README.txt file in the source code folder.





# HCMUT support

- Open Computing Platform at HPC Lab – HCMUT  
<http://www.hpcc.hcmut.edu.vn>
- Short course training in IoT , Big Data analytics, HPC
- Master program in “Data Science”, 2018



# ACOMP & FDSE 2017

- The 4<sup>th</sup> International Conference on Future Data and Security Engineering 2017
  - Homepage: <http://cse.hcmut.edu.vn/fdse2017/>
  - FDSE 2017 (LNCS - Springer Verlag)
  - Ho Chi Minh City, Vietnam, November 29 to December 1, 2017
- International Conference on Advanced Computing and Applications
  - Homepage: <http://cse.hcmut.edu.vn/acomp2017/>
  - ACOMP 2017 (IEEE)
  - Ho Chi Minh City, Vietnam, November 29 to December 1, 2017
- The FDSE and ACOMP proceedings are indexed by DBLP and some other major indexing systems
- Since 2015, FDSE Proceedings have also been indexed by Scopus and listed in Conference Proceeding Citation Index (CPCI) of Thomson Reuters.
- FDSE 2017, ACOMP2017 and RCCIE2017 are co-located



# Summary

- There are still very critical issues that need to be solved in Vietnam
- Computing resource & data collection are problems
- Collaborations
  - Big data analytics/mining
  - Sharing computing resources
  - Data Science, IoT & HPC in undergraduate/postgraduate programs



# Thank you!

More information:

[namthoai@hcmut.edu.vn](mailto:namthoai@hcmut.edu.vn)

<http://www.cse.hcmut.edu.vn/>

<http://www.hpcc.hcmut.edu.vn/>

