TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

Some Research Directions of Distributed Data Processing and HPC Group

Assoc. Prof. NGUYEN BINH MINH Ph.D.
School of Information and Communication Technology

Outline

I. Meet our Team

II. Overview of research directions
- Two main directions
- Research assumptions/observations (foundations for our studies)
- Featured papers
- Research direction maps (summary)

III. Research route plans
- For two main directions
- Combination of two directions

IV. Blockchain works

Meet the CORE team

A/Prof. Nguyen Binh Minh
Ph.D. from Slovak Technical University, Slovakia

Dr. Nguyen Huu Duc
Ph.D. from JAIST, Japan

Dr. Do Ba Lam
Ph.D. from Vienna Technical University, Austria

Dr. Dao Thanh Chung
Ph.D. from Tokyo University, Japan
Co-founder of Rikkeisoft

Dr. Michel Toulouse
Ph.D. from Montréal, Canada

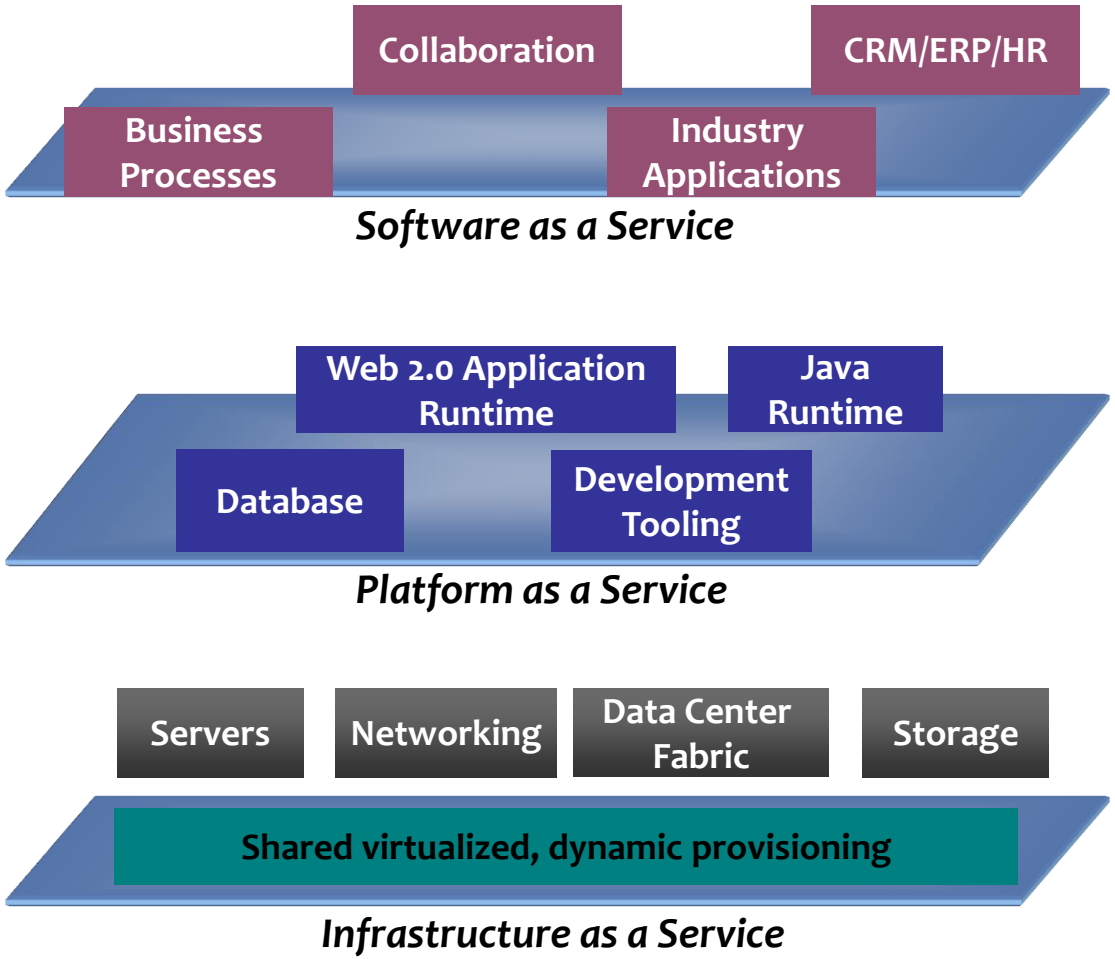
Dr. Trinh Tuan Dat
Ph.D. from Vienna Technical University, Austria

Dr. Tran Vinh Duc
Ph.D. from University of Nice Sophia Antipolis, France

Dr. Vu Van Thieu
Ph.D. from Leiden University, Netherlands

Cornerstone: Cloud Computing

- A Computing Cloud:
  - Set of Services, Inexpensive Computing, Personalized, Simply Access, QoS/SLA
- Types of Cloud Computing:
  - IaaS, PaaS, SaaS
- Service Provisioning in the Cloud:
  - Private, Community, Public, Hybrid Clouds

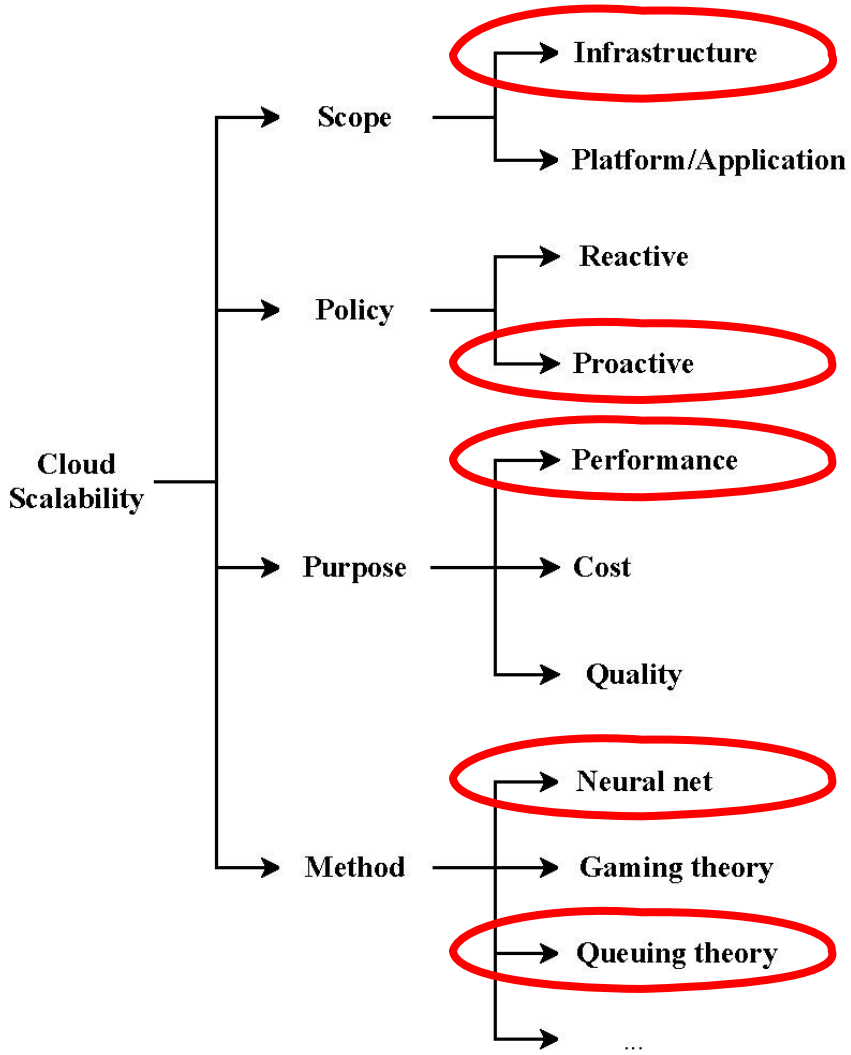


Two Research Directions

1. Proactive resource scalability in cloud systems
2. Interoperability among different clouds

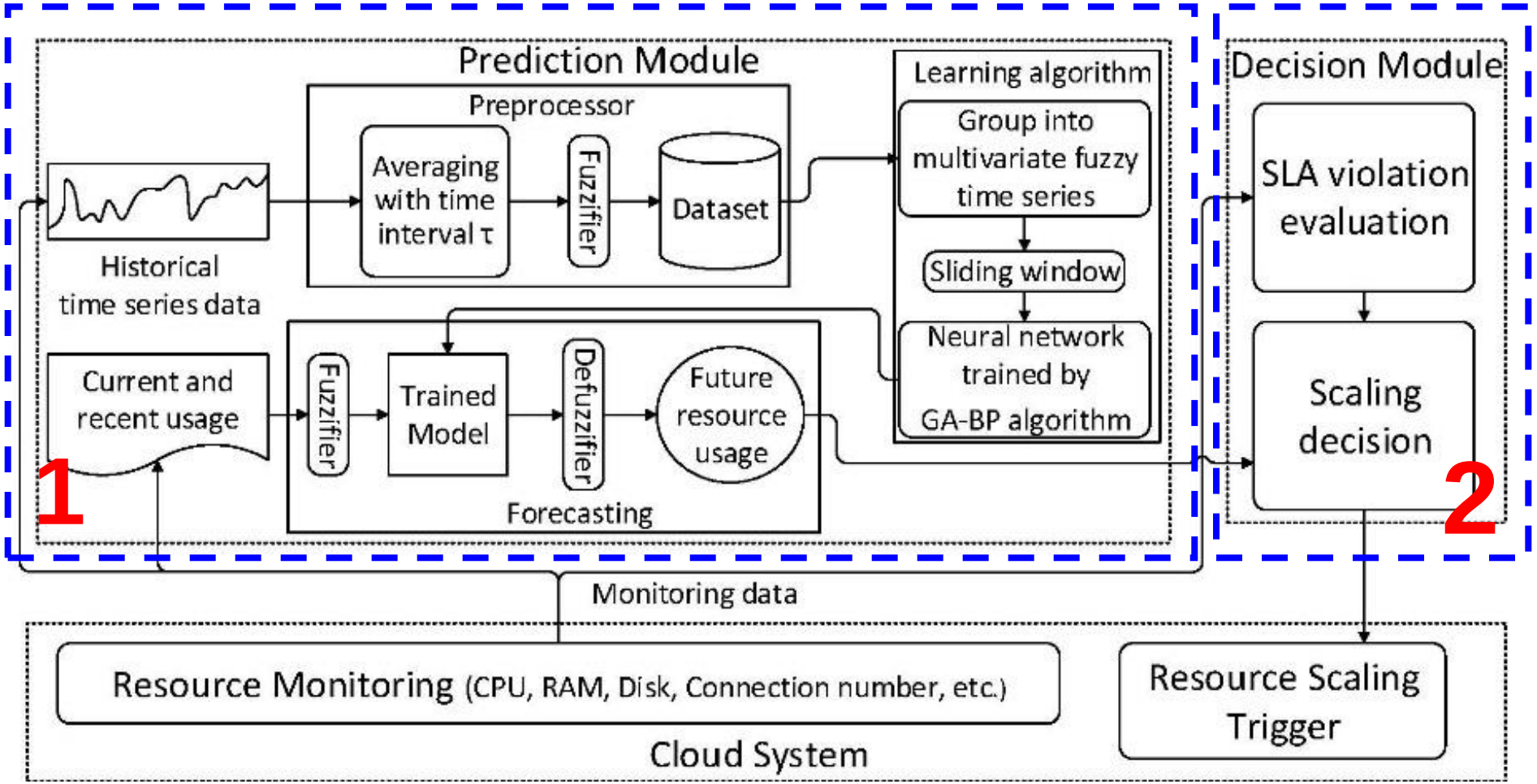
1. Proactive resource scalability

- **Problem statement**
  - Cloud providers offer users a certain auto-scaling mechanism, which mainly operates based on predefined thresholds for resource usage (reactive)
    - ➔Resource provision is often delayed
    - ➔Solution: predict future resource usages to make scaling decision early (proactive)
  - Used techniques: machine learning, gaming theory, …
    - Many existing studies have dealt with the prediction problem with different approaches
    - Not many efforts focus on the decision
  - Two main subproblems: prediction and making scaling decisions
    - Prediction:
      - Time-series analysis
      - Univariate/Multivariate analysis and prediction
      - High fluctuation of values that may lead to low accuracy
      - …
    - Decision: based on SLA, …



Overview proactive scaling resource architecture system

Overview proactive scaling resource architecture system[1]



[1] Tran, Dang, Nhuan Tran, Giang Nguyen, and **Binh Minh Nguyen***. "A proactive cloud scaling model based on fuzzy time series and SLA awareness." *Procedia Computer Science* 108 (2017): 365-374.

Observations/Assumptions

1. Prediction phase:

   (1-1) Periodic data (hourly, daily, monthly, ...)

   (1-2) Data fluctuation (noisy, change quickly by time – second/milisecond)

   (1-3) Multi-metrics (CPU, memory, disk I/O, …)

   (1-4) Learning cost

2. Making decision phase:

   (1-5) Service level agreement (SLA)

   (1-6) Task/job queues

Contributions/meaning (1)

Tran, Dang, Nhuan Tran, Giang Nguyen, and **Binh Minh Nguyen***. "A proactive cloud scaling model based on fuzzy time series and SLA awareness." *Procedia Computer Science* 108 (2017): 365-374. (Scopus)

1. Proposing a novel proactive autoscaling architecture for clouds includes two main components: prediction and scaling decision
2. Proposing a novel prediction approach that *exploits simultaneously multiple monitoring utilization data* such as CPU, memory to forecast the future system usages; ➔ multiple metrics
3. Applying fuzzy time series approach in the prediction model to improve forecast effect; ➔ data fluctuation
4. Proposing a formula to make scaling decisions based on multiple parameters including predictions of multi-resource utilization and SLA violation estimation.

➔ Proactive autoscaling architecture: foundation for our next research works in this direction.

➔ Focus on resolving the observations **(1-2)**, **(1-3)**, and **(1-5) [decreasing fluctuation, multiple metrics, SLA]**

Contributions/Meaning (2)

Nguyen, Thieu, Tu Nguyen, **Binh Minh Nguyen***, and Giang Nguyen. "Efficient Time-Series Forecasting Using Neural Network and Opposition-Based Coral Reefs Optimization." *International Journal of Computational Intelligence Systems* vol. 12, no. 2 (2019): 1144-1161. (SCIE, Q1, IF 2.153)

Motivation: finding simpler prediction model (idea: MLNN + metaheuristics)

1. Proposing a new improvement called opposition-based coral reefs optimization (OCRO), which improves the original CRO algorithm using OBL
2. Combination of CRO and OBL is property applied to MLNN to generate a prediction model ➔ increase accuracy in prediction
3. Carrying out comparisons among the novel proposed prediction model with five well-known ones such as MLNN, GA-MLNN, CRO-MLNN, RNN, and LSTM for time series forecast ➔ OCRO-MLNN gained positive results compared with deep learning method (RNN, LSTM) ➔ simpler prediction model

➔Focus on resolving the observation **(1-4) [decreasing high learning cost]**

Contributions/Meaning (3)

Nguyen, Binh Minh*, Dang Tran, and Giang Nguyen. "Enhancing service capability with multiple finite capacity server queues in cloud data centers." *Cluster Computing* 19, no. 4 (2016): 1747-1767. (SCIE, Q2, IF 1.851)

Motivation: limitation of training data amount ➔ after making scaling decision (prediction-based), system state will be changed ➔ novel data will be generated ➔ scaling decisions are difficult to verify ➔ require a new approach (idea: queuing theory)

1. Three-state model for servers: OFF, MIDDLE and ON
2. Applying the three-state model to a single finite-capacity queue. Demonstrating the model based on the Markov chain and queueing theory
3. Proposing a large-scale architecture for cloud data centers, in which we use multiple finite capacity server queues at the same time to serve arriving jobs.

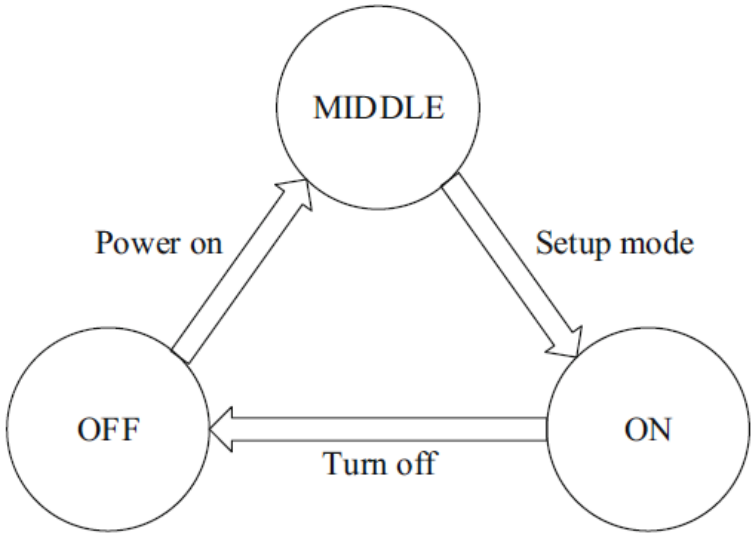
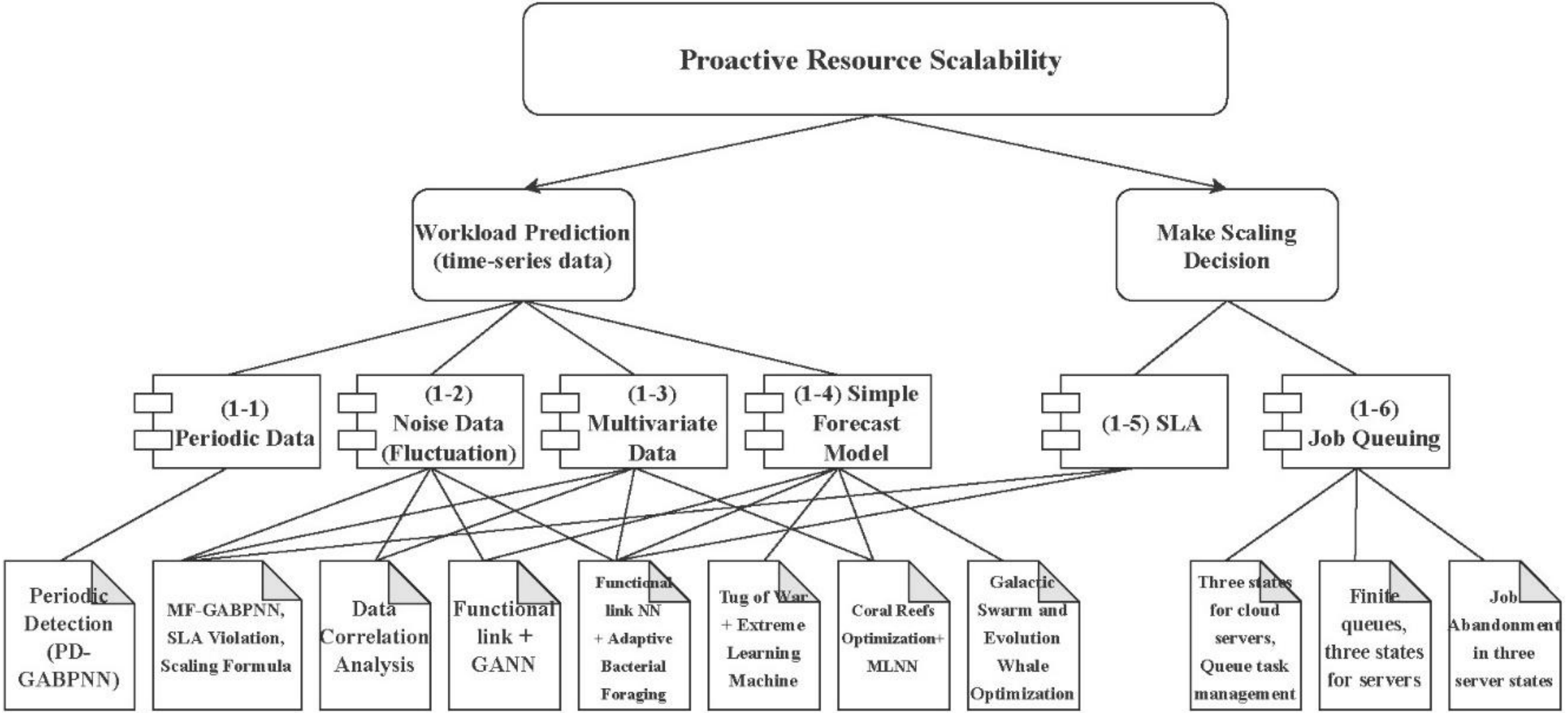


Fig. 1 State transition diagram of the physical servers

➔Focus on resolving the observation **(1-6) [queuing]**

Summary of proactive resource scalability



2. Interoperability among different clouds

Problem statement:

There are hundreds of cloud vendors …

Due to
- Business competition
- Lack of official standards

→ **Each vendor imposes its own stack of technologies**

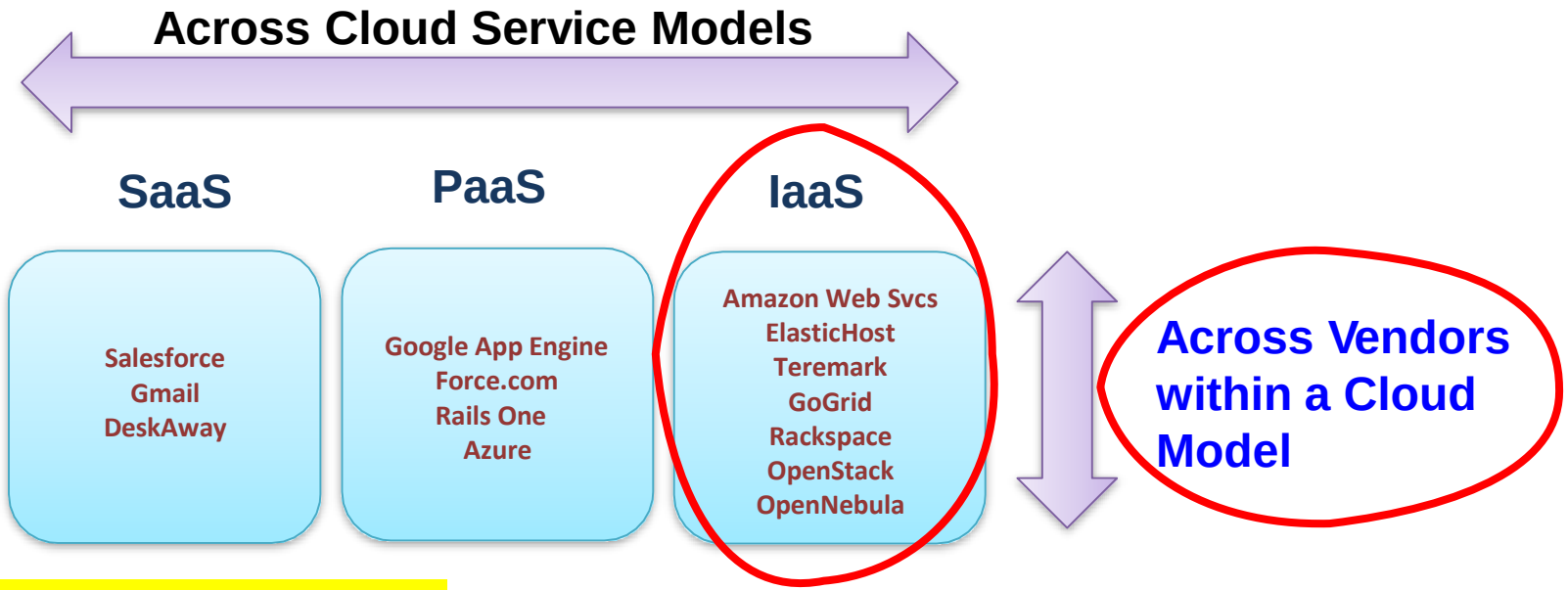
Differences among the stacks: hypervisor, networking infrastructure, data storage facilities, management means, …

Vendor lock-in issues:

- Lock cloud users into services provided by only one vendor!
- Can you transfer data and applications to and from the clouds at the same time?

Interoperability between clouds?

- **Ability to use the cloud services provided by multiple vendors**
  - **Across vendors within a cloud model**
  - Across cloud service models
- **Ability to move data and code** from one cloud to another or back to the enterprise (portability)



Many existing research works: Federation Cloud (EGI-Cloud), European Open Science Cloud, DEEP-HybridDataCloud, TOSCA, OpenStack Interoperability, … with many different approaches: e.g. standardization, API abstraction, ….

Observations/Assumptions

(2-1) Generic functions via programming abstraction layer

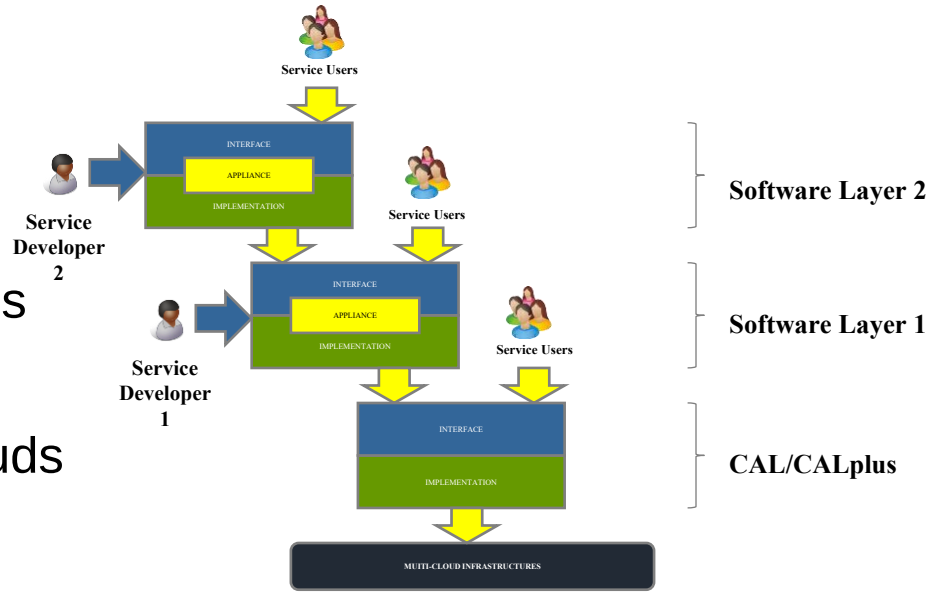
(2-2) Semantic data for resource management

(2-3) Entities identification (resources, objects, services, …)

Contributions/Meaning

Nguyen, Minh Binh*, Viet Tran, and Ladislav Hluchy. "A generic development and deployment framework for cloud computing and distributed applications." *Computing and informatics* vol. 32, no. 3 (2013): 461-485. (SICE, Q3, IF 0.421)

1. Proposing a development and deployment framework for cloud applications based on abstraction programming approach (called CAL)
2. Proposing the use of inheritance for cloud application codes (software layering)
3. Proposing migration of cloud application codes among clouds
4. Publishing CALplus library in Pypi (official Python library)



➔CAL approach: foundation for our next research works in this direction

➔Focus on resolving the observation **(2-1) [abstraction]**

Some other works on interoperability research direction

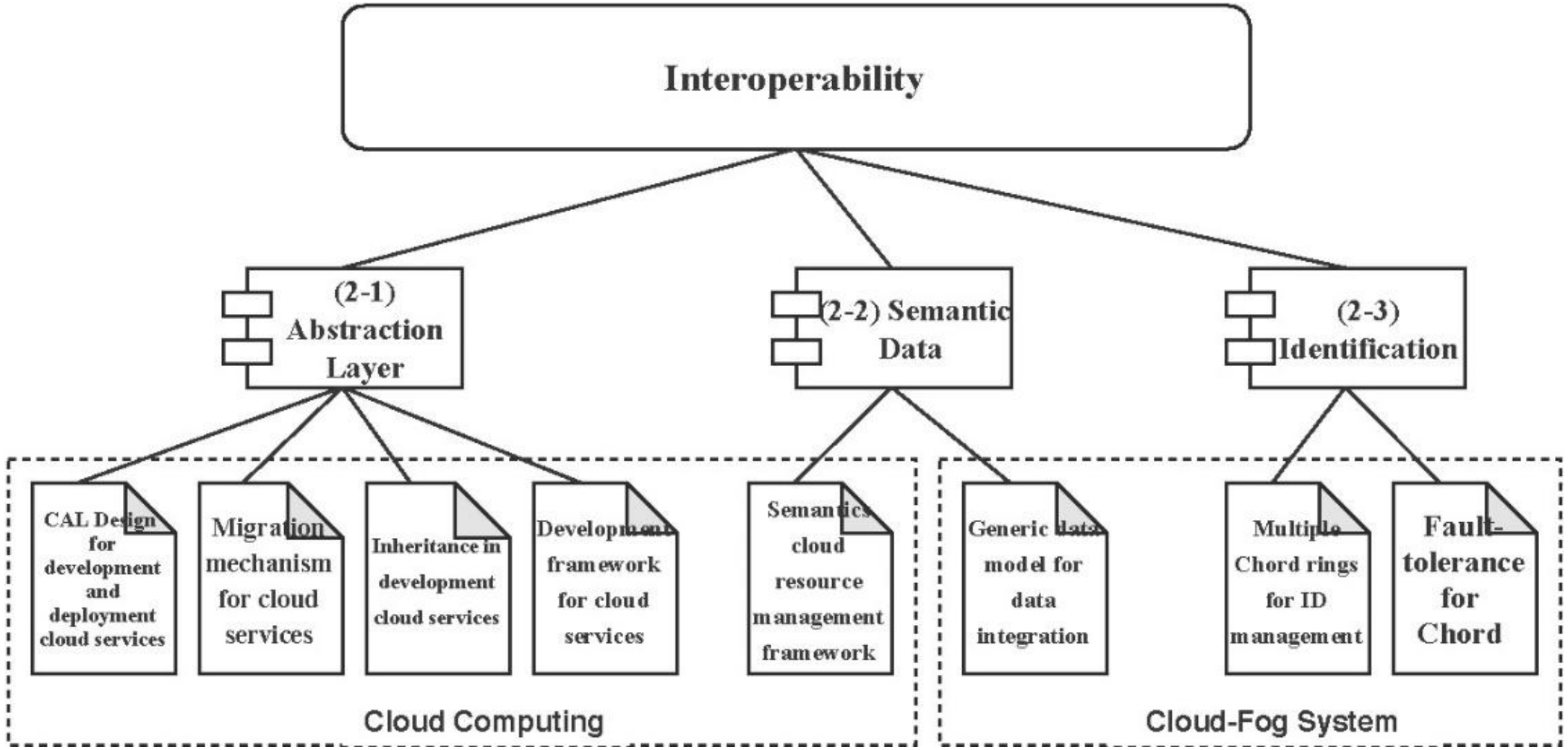
- Nguyen, Binh Minh and Dao, Quang Minh. Towards a Semantic Model of Resource in Cloud Environment. In proceeding of *5th International Symposium on Information and Communication Technology (SOICT)*, ACM, Hanoi, 2014, p. 271-279, ISBN: 978-1-4503-2930-9.
- Nguyen, Binh Minh; Phan, Huan; Ha, Quang Duong; and Nguyen, Giang. An Information-centric Approach for Slice Monitoring from Edge Devices to Clouds. In *Procedia Computer Science*, Elsevier, *9th International Conference on Ambient Systems, Networks and Technologies (ANT)*, Porto, 2018, Vol. 130, 2018, p. 326-335, ISSN 1877-0509.

➔ Semantic data model for resource management (2-2)

- Nguyen, Binh Minh; Hoang, Quoc Hong-Nhat; Hluchý, Ladislav; Vu, Tuyet Trinh and Le, Hieu. Multiple Peer Chord Rings Approach for Device Discovery in IoT Environment. In *Procedia Computer Science*, Elsevier, *14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC)*, Leuven, 2017, vol. 110, p.125-134, ISSN 1877-0509.
- Nguyen, Dan; Hoang, Nhat; Nguyen, Binh Minh*; and Tran, Viet. PSPChord - A Novel Fault Tolerance Approach for P2P Overlay Network. In *Lecture Notes in Computer Science (LNCS)*, Springer, *3rd International Conference on Smart Computing and Communication (SmartCom)*, Tokyo, vol. 11344, 2018, p. 386-396, ISBN 978-3-030-05754-1, ISSN 0302-9743.

➔ Identification (2-3)

Summary of Interoperability in Cloud/distributed systems



Continuous research routes

1. Proactive resource scalability
   - Vertical routes:
     - Proposing other assumptions: e.g., abnormal peak (1-7), uncertainty (1-8), streaming data (1-9), …
     - Applying/improving methods, techniques for solving those assumptions
   - Horizontal route: resource allocations in distributed environment/systems: cloud-fog (some papers already has been published in the way)
2. Interoperability
   - Cloud-fog environment (with multiple clouds and IoT platforms – heterogeneity)
   - Cross-chain (usage of multiple blockchains – V-chain platform)

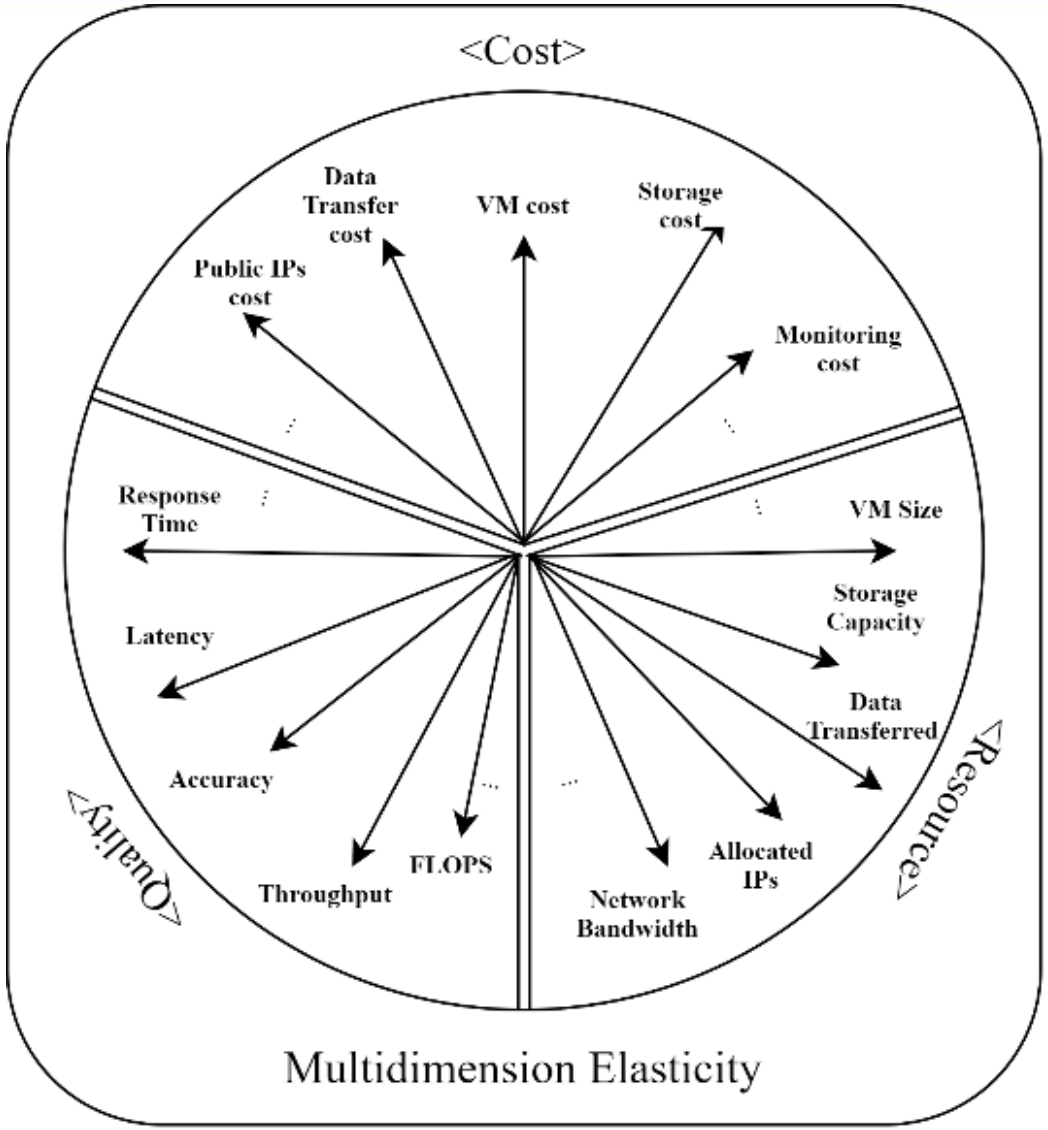
Future research routes – Combination of two directions

- Multi-dimensions scalability/elasticity: resource, cost, quality in distributed systems

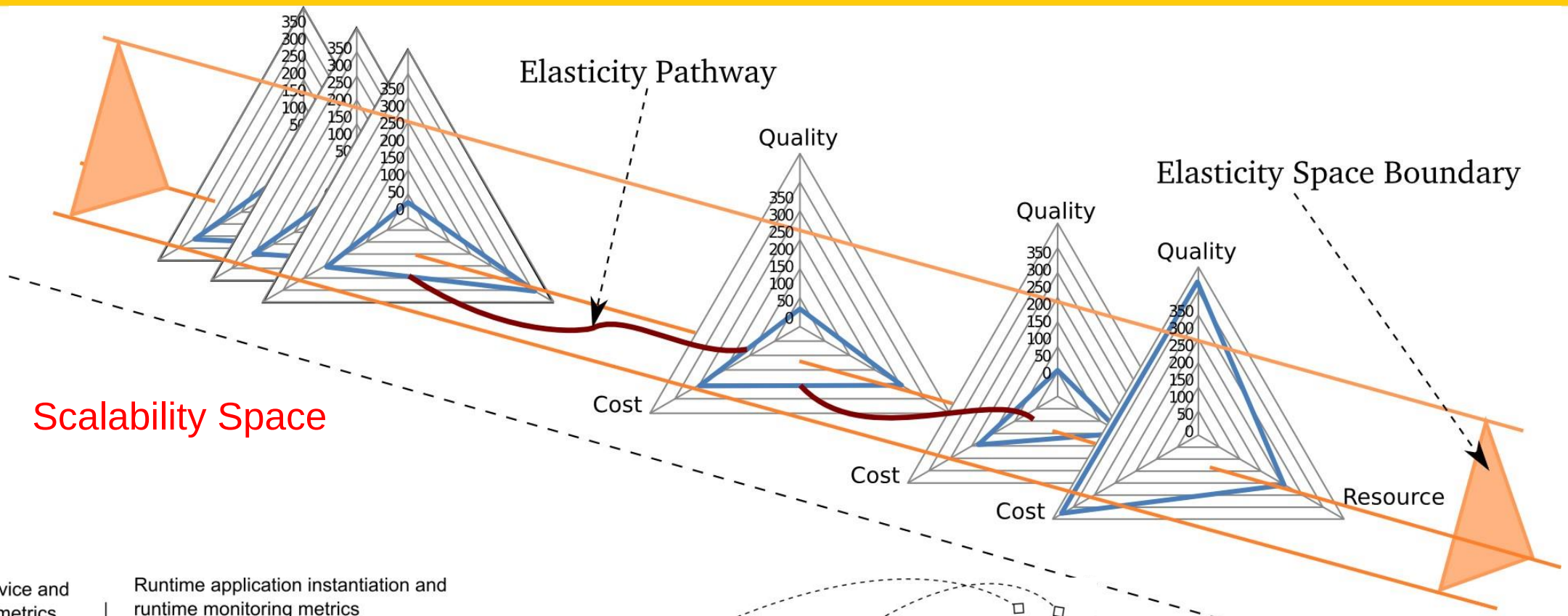
➔Requirement: An abstraction of multi-systems to control, manage

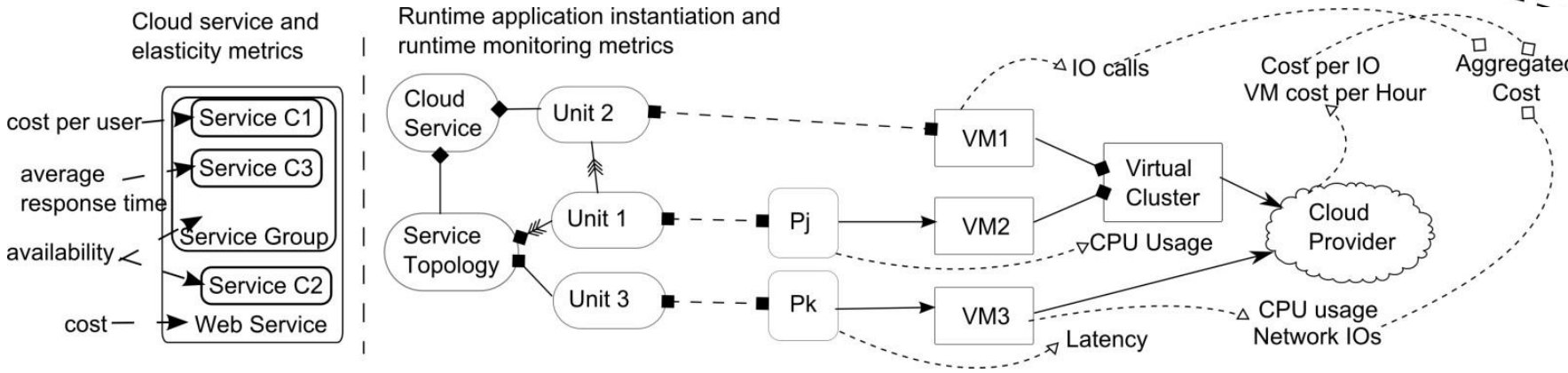
  - Modeling data
  - Semantic control

➔ Optimizing scaling mechanisms on the systems



Proactive Scalability Model for Distributed Systems





Blockchain R&D direction

- Consensus protocols:
  - Decreasing cost (computation, communication)
  - Increasing fairness among participant nodes
    - Nguyen, Binh Minh, et al. **"MPoC-A Metaheuristic Proof of Criteria Consensus Protocol for Blockchain Network**." 2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC). IEEE, 2021.
    - **Event-link consensus: US patent application**
- Optimizing stored data on-chain
    - Tran, Canh-Tuan, et al. "**A Novel Approach for Developing Decentralized Storage and Sharing Systems**." Proceedings of the 3rd ACM International Symposium on Blockchain and Secure Critical Infrastructure. 2021.
    - Pham, Van-Duy, et al. "**B-Box-A Decentralized Storage System Using IPFS, Attributed-based Encryption, and Blockchain**." 2020 RIVF International Conference on Computing and Communication Technologies (RIVF). IEEE, 2020.
- Indexing on-chain data for analysis (credit score, market trend prediction, risk estimation, abnormal detection)
    - On-chain Data
- Cross-chain (interoperability among blockchains)

Blockchain-based Dapp/Fintech solutions

- Nguyen, Binh Minh, Thanh-Chung Dao, and Ba-Lam Do. "**Towards a blockchain-based certificate authentication system in Vietnam**." PeerJ Computer Science 6 (2020): e266.

- Dao, Thanh Chung, Binh Minh Nguyen, and Ba Lam Do. "**Challenges and strategies for developing decentralized applications based on blockchain technology**." *International Conference on Advanced Information Networking and Applications*. Springer, Cham, 2019.

- Nguyen, Ngoc-Hoa, et al. "**Towards Blockchainizing Land Valuation Certificate Management Procedures in Vietnam**." *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*. IEEE, 2020.

HOW WE DID LAST YEAR

V-CHAIN PLATFORM

The **V-Chain** platform delivers its capabilities through pre-written Core and Utility Services.

https://v-chain.vn

CV GENERATION BASED-ON IPFS TECHNOLOGIES

CIVIZEN

BLOCKCHAIN FOR EDUCATION

B4E

GIẢI THƯỞNG NHÂN TÀI ĐẤT VIỆT 2019

VINGROUP

KHOA HỌC VÀ CÔNG NGHỆ

TÀI NGUYÊN VÀ MÔI TRƯỜNG

IBM

Demo with IBM Vietnam for Vietnam Airline, Sơn Kim Group

We can only see a short distance ahead, but we can see plenty there that needs to be done

- Alan Turing -

Thank you for your attention!

minh.nguyenbinh@hust.edu.vn