GSPN-Based Reliability-Aware Performance Evaluation of IoT Services

Songyuan Li, Jiwei Huang
School of Computer Science
Beijing University of Posts and Telecommunications
Beijing, China
sylee1416@gmail.com

Outline

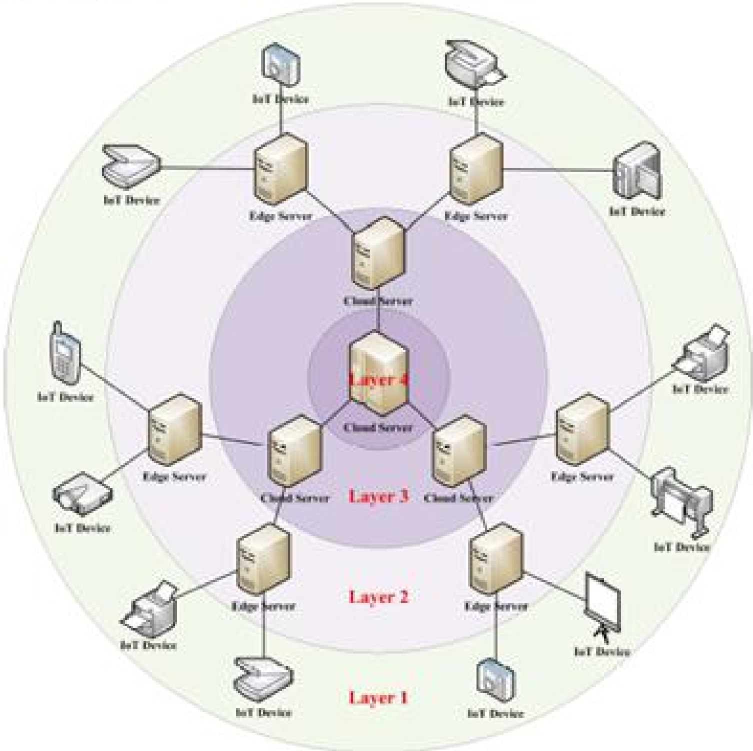
- Background
- GSPN Model for Performance Evaluation
- Real Data Based Evaluation
- Conclusion and Future Work

Background

❖ Internet of Things (IoT)
- Connect interrelated devices over the network
- Cooperate with each other to reach common goals
- Wide application in several aspects
  - ✓Industrial manufacturing
  - ✓Smart home
  - ✓Transportation
  - ✓...

Background

- Edge Computing
  - A novel computing paradigm for IoT service implementation



Background

- QoS of IoT Services
  - Service Level Agreement (SLA)
  - Reference for QoS Optimization
  - Measurement-Based Approach
  - Model-Based Approach

Directly applied in real-time systems

Conducted at the design phase before system implementation

Model-Based Approach

Much cheaper than the measurement-based approach

Background

❖ Reliability Issue in IoT Systems

- Services deployed in virtual machines (VMs) or virtualized cloud servers
- VM migration and recovery techniques allowing the fast restart of a virtual machine when failure occurs

❖ Reliability-Aware QoS Evaluation

Challenges which VM migration brings -The error probing, fault repairing and system restarting take time, affecting the overall end-to-end QoS evaluation of the services.

Background

❖ Our contribution

- A theoretical approach of reliability-aware performance evaluation of IoT services
  - ✓Formulating the dynamics of the IoT systems with GPSN models.
  - ✓Modeling both atomic services and systems
  - ✓Corresponding quantitative analyses
  - ✓Empirical experiments based on real-world data

Outline

- Background
- GSPN Model for Performance Evaluation
- Real Data Based Evaluation
- Conclusion and Future Work

GSPN Models

- Generalized Stochastic Petri Net (GSPN)

  a 7-tuple $\psi = \left(P, T, F, W, \Pi, M_0, \lambda\right)$

  - P: the finite set of places $\quad T = T_t \cup T_i$
  - T: the finite set of transitions
  - F: the finite set of arcs $\quad F \subseteq \left(S \times T\right) \cup \left(T \times S\right)$
  - W: the weight function of arcs $\quad F \rightarrow \square^{+}$
  - Π: the priority function $\quad \mathrm{T} \rightarrow [0,1]$
  - M_0: the initial state (marking) of the GSPN
  - λ: the set of firing rate of $T_t \quad T_t \rightarrow \square^{+}$

GSPN Models

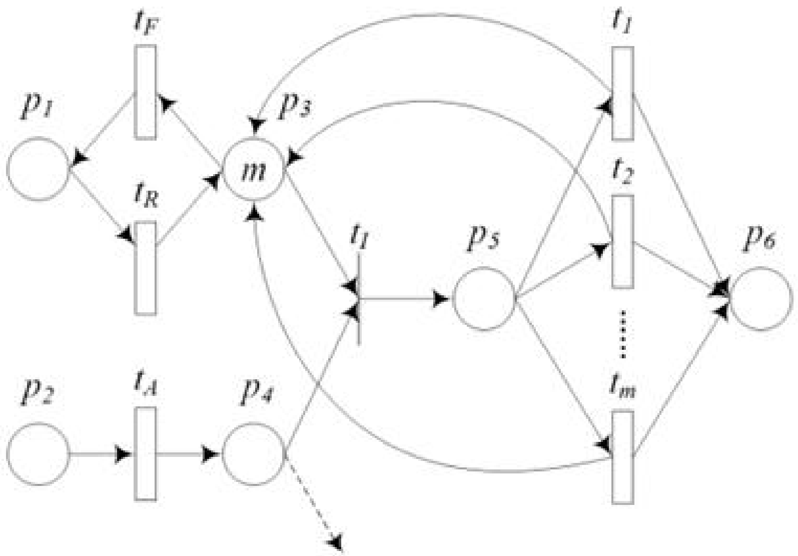
❖ Cluster Model

- Maximum of Performed Machines: m
- Failure Rate: λ_F
- Repair Rate: λ_R
- Task Arrival Rate: λ_A
- Transition Probability:

$$1 - \pi_I$$

- Average Response Time:

$$RS = \frac{q_4 \cdot \pi_I + q_5}{\lambda_A \cdot \pi_I}$$



With *Little's Law*

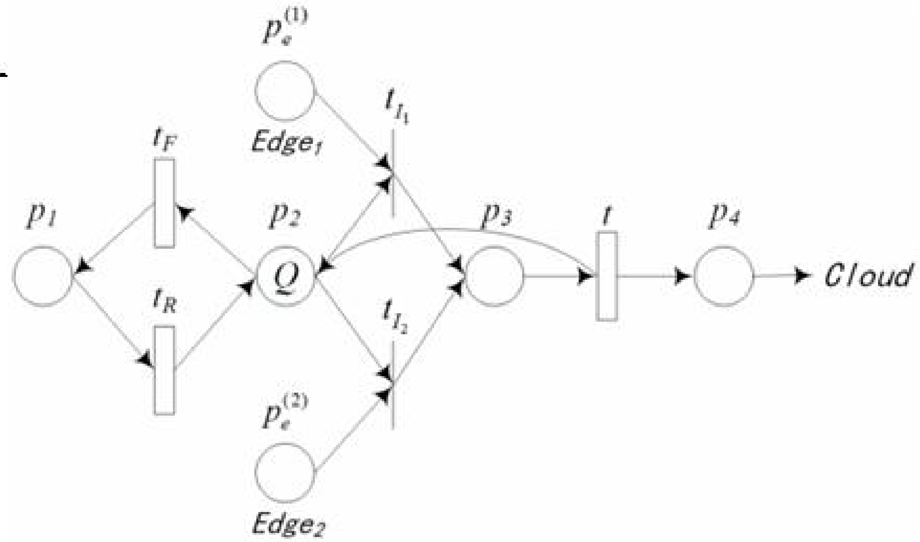
q: the average amount of tokens

GSPN Models

❖Communication Model

- Failure Rate: λ_F　　　　Repair Rate: λ_A
- Average Communication Delay:

$$TR = \frac{q_e^{(i)} \cdot \left(1 - \pi_I^{(i)}\right) + q_3}{\lambda_A^{(i)} \cdot \left(1 - \pi_I^{(i)}\right)}$$

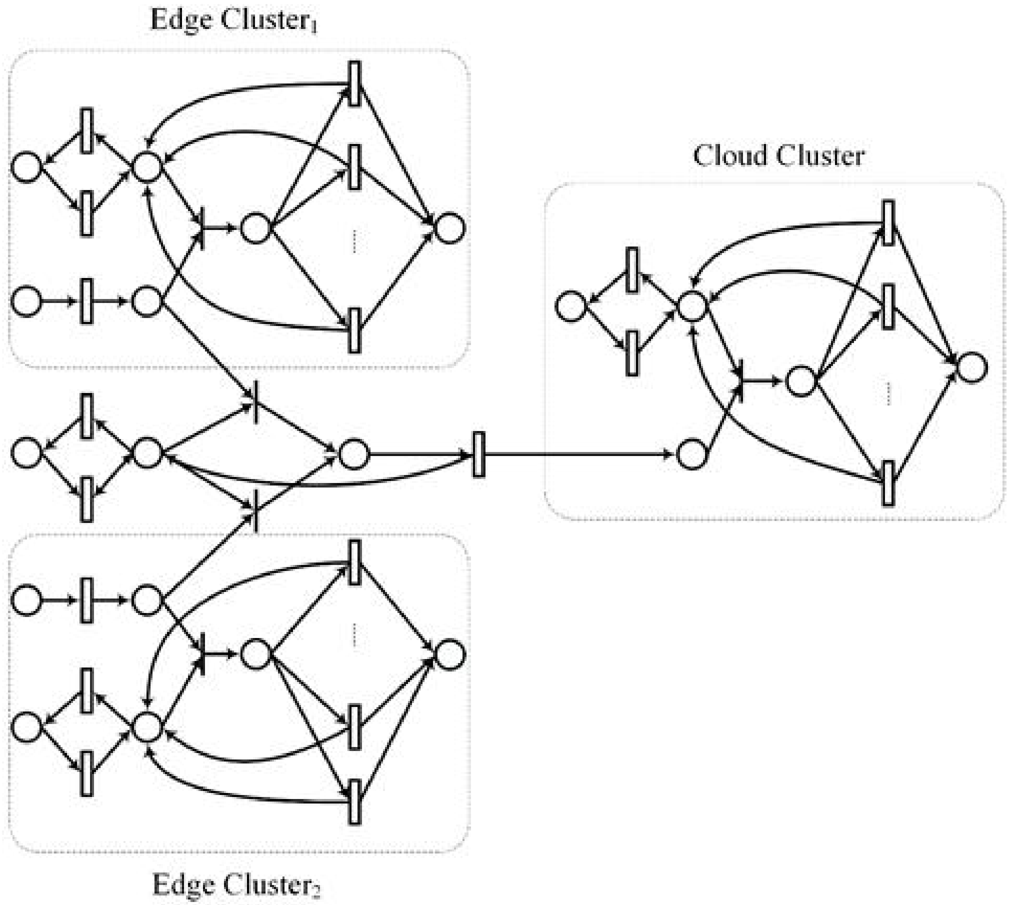


With *Little's Law*

q: the average amount of tokens

GSPN Models

❖ System Model



GSPN Models

❖ System Model

- Average response time of tasks at the cloud

$$RS^{(c)} = ST^{(c)} + TR$$

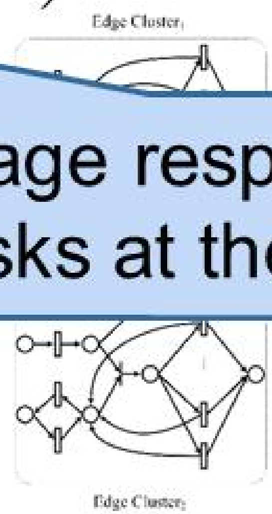
- Total average response time for the system

$$RS^{(sys)} = \pi \cdot RS^{(e)} + (1 - \pi) \cdot RS^{(c)}$$

Cluster Model:

$$RS = \frac{q_4 \cdot \pi_I + q_5}{\lambda_A \cdot \pi_I}$$

Average response time of tasks at the edge



Outline

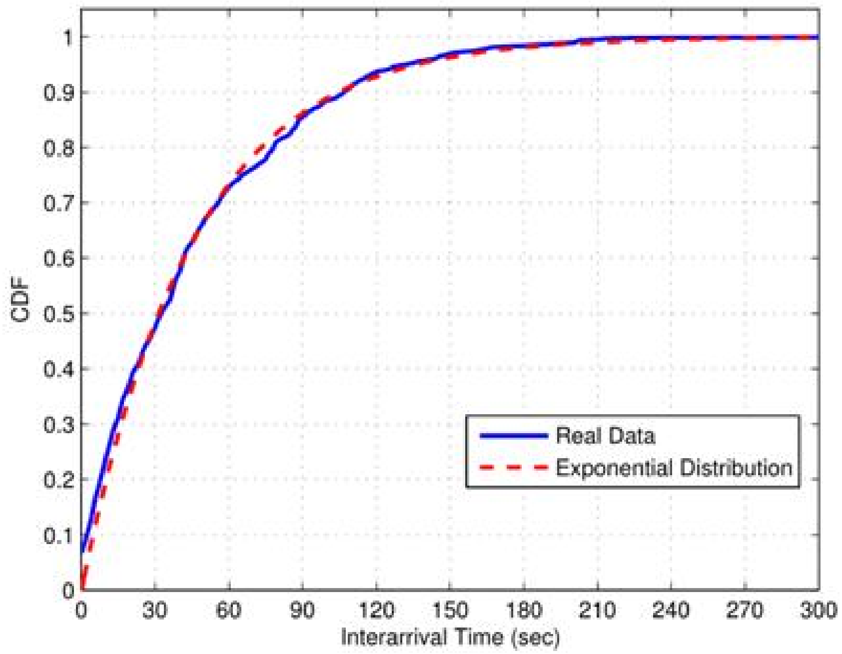
- Background
- GSPN Model for Performance Evaluation
- Real Data Based Evaluation
- Conclusion and Future Work

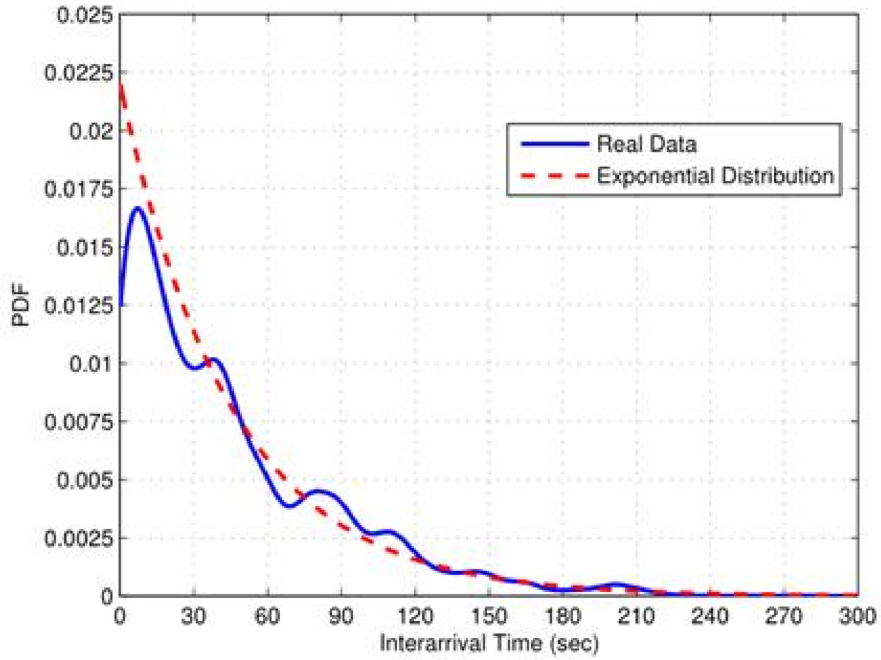
Evaluation

- Datasets
  - T-Drive
    - ✓ the GPS trajectories of 10,357 taxis within the city of Beijing during a period of one week in 2008
    - ✓ 4 fields for each piece of data, including taxi id, timestamp, longitude and latitude
  - LA-UR Failure Data
    - ✓ 23,739 failure situations of 23 High Performance Computing (HPC) systems
    - ✓ the failure logs on software, hardware and network failures, and etc.

Evaluation

- Arrival Distributions
  - Exponential -- Poisson

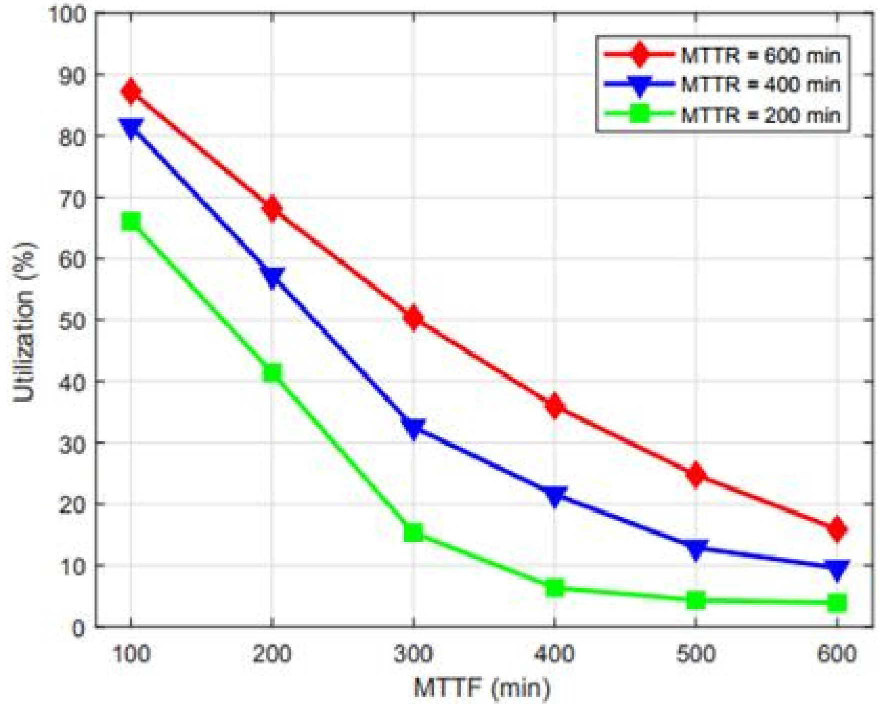




Evaluation

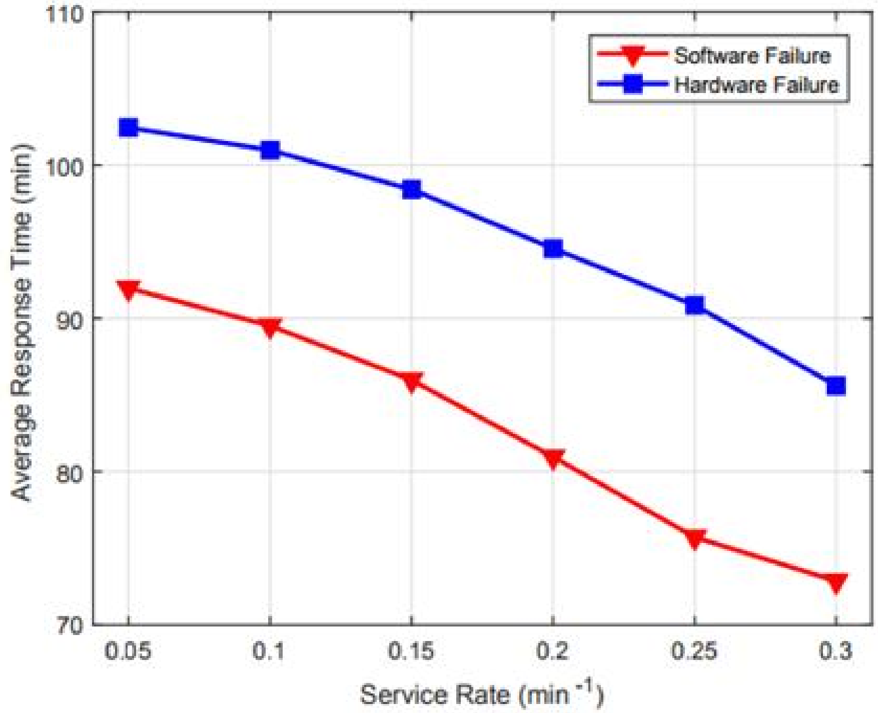
❖ Experimental Results

- Cluster Utilization



Evaluation

- Experimental Results
  - Average Response Time



Outline

- Background
- GSPN Model for Performance Evaluation
- Real Data Based Evaluation
- Conclusion and Future Work

Conclusion

- A modeling approach of performance evaluation in edge computing paradigm
- GSPN models formulating the dynamics of IoT services with reliability issues fully considered
- Quantitative analyses on performance metrics for IoT services
- Simulation experiments based on real-world data from IoT and cloud systems

Future Work

- Detailed specifications on models
  - More variety of failures in real-life IoT systems
  - More precise descriptions on the statistical distributions of the failures, repairs, arrivals and service processes
- Elaborate experimental design
  - Further validation of our approach
  - Further exploration of parameter settings

Thank you!

Q&A

Songyuan Li
Beijing University of Posts and Telecommunications
Beijing, China
sylee1416@gmail.com