Hong-wu Lv, Jun-yu Lin, Hui-qiang Wang, Guang-sheng Feng, Mo Zhou, 2015. Analyzing the service availability of mobile cloud computing systems by fluid-flow approximation. *Frontiers of Information Technology & Electronic Engineering*, **15**(7):553-567. [doi:10.1631/FITEE.1400410]

Analyzing the service availability of mobile cloud computing systems by fluid-flow approximation

Key words: Service availability, Mobile cloud computing, Fluid-flow approximation, Ordinary differential equations

Contact: Hong-wu Lv
E-mail: lvhongwu@hrbeu.edu.cn
ORCID: http://orcid.org/0000-0002-1917-3978

Introduction

- Cloud computing is changing the way that network services are provided.
- Mobile Internet users have accounted for 70% of the entire Internet users.
- Mobile cloud computing (MCC) has become a promising technique to deal with computation- or data-intensive tasks.
- MCC argues that the service must be available for users at anytime and anywhere.
- For large-scale and complex types of services in MCC systems, simulation-based methods may be costly and the traditional state-based methods always suffer from the problem of state-space explosion.

MCC model based on PEPA

Master=(broadcast_requesting, r_1).Waiting;
Waiting=(return, r_2).Splitting;
Splitting=(split, r_3).Mapping;
Mapping=(map, r_4).Shuffling;
Shuffling=(shuffle, r_5).Reducing;
Reducing=(reduce, r_6).Master;

The PEPA model of the cloud part

Device=(broadcast_requesting, w_1).Awake;
Awake=(sensoring, w_2).Sensing;
Sensing=(return, w_3).Device;

The PEPA model of the device part

The MCC model can be described as follows:

$$\mathrm{Master}[M] \underset{L}{\bowtie} (\mathrm{Device}_1[N_1] \,\|\, \mathrm{Device}_2[N_2] \,\|\, \ldots \,\|\, \mathrm{Device}_k[N_k]),$$

where N_i represents the number of the ith kind of device, and the parallel computing capability of the cloud is M.

Critical metrics

- Response time of service (RTS)

For a certain type of service, RTS refers to the shortest possible time from receiving the service request to returning the final results for the first time.

- Minimum sensing time of devices

Minimum sensing time of the device is the least time used by devices from receiving a command to returning a final result

- Minimum number of nodes chosen

To deal with unexpected faults and random high loads, each service is always allocated with redundancy processing capabilities. However, the fact that fewer devices are needed means less energy consumption and fewer links unexpectedly interrupted. Thus, it is important to estimate the minimum number of nodes needed and determine the range of devices chosen.

- Action throughput

Action throughput mainly represents the frequency of a type of action implemented, assessing the transaction processing capability.

Experimental results

Table 4 The comparison of computing time

Parameter	Number of components		Computing time (ms)			
	Markov chains before simplification	Markov chains after simplification	Markov chains before simplification	Markov chains after simplification	Monte-Carlo simulation	Fluid-flow approximation
M=N=3	6104	546	10 438	328	3562	31.4
M=N=10	3.65×10^{15}	940 576	–	–	8688	31.4
M=N=100	4.27×10^{155}	2.43×10^{18}	–	–	51 219	31.4
M=N=200	1.82×10^{211}	1.44×10^{22}	–	–	100 938	31.4
M=N=1000	2.01×10^{1556}	1.34×10^{31}	–	–	506 281	31.4

'–' means it is beyond the machine's computing power

Conclusions

- The proposed method is based on a formal language, instead of SLA or measuring a real-world system. Thus, we can analyze the service availability of MCC even before a target system is deployed. Meanwhile, it avoids a variety of random interferences from different real systems under testing.
- The fluid-flow approximation approach is used to convert the PEPA model into ODEs, avoiding the problem of state-space explosion which is widely existing in current models, such as the Markov chains or Petri net.
- Our method costs less time in analyzing the service availability of MCC than the Markov- or simulation-based methods.