Mobility-Aware Application Scheduling in Fog Computing

Luiz F. Bittencourt Javier Diaz-Montes Rajkumar Buyya
Omer F. Rana Manish Parashar

Presenter: Maicon Ança dos Santos

October 9, 2018

Laboratory of Ubiquitous and Parallel Systems

Introduction

- With new levels of computing capacity provides by Fog computing, new forms of resource allocation and management can be developed;
- The grow of the number of devices scattered and connected to Internet, producing and consuming data, requires a scalable resource management at unprecedented levels
  - *Focus on IoT*

Introduction

- Data also are produced at the edge. Data generation and consumption can occur at many differents places and times;
- Different applications can have different requirements, especially in terms of response time.

Introduction

The problem

Resource allocation considering the hierarchical infrastructure composed of edge capacity and cloud data centers, analyzing application classes along with different scheduling policies.

Fog Computing Model

- User applications that access the public cloud do so through an access point that allows data exchange through the core network to reach the cloud data center;
- *Cloudlet*: access point extended to also provide computing and storage services.

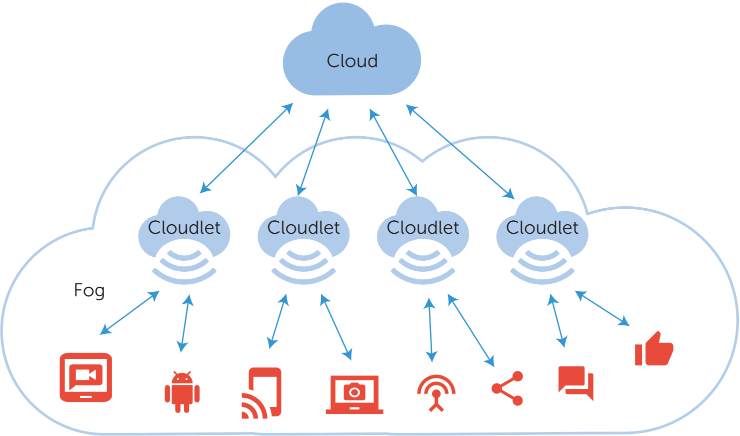


Figure 1: Fog computing: cloud, cloudlets and edge devices/applications ecosystem

Fog Computing Model

- Hierarchical, bi-directional computing infrastructure: edge devices communicate with cloudlets and cloudlets communicate with clouds.
  - *Cloudlets can also communicate with each other to perform data and process management.*

Fog Computing Model

- Processing and storage capacity in fog computing can benefit different types of applications
  - *Applications with low latency requirements;*
  - *Applications that currently rely on the cloud;*
  - *Cases in which raw data collected by many devices that generally do not need to be transferred to the cloud for long-term storage.*

Fog Computing Model

- Cloudlets can provide reduced latencies, however...
  - *New challenges:* ***what***, ***when*** *and* ***where*** *carry out processing to meet QoS;*
  - *Fog scheduling must bring users location to the resource allocation policies to uphold the benefits of proximity to the user.*

Related Work

- Fog computing as a platform to provide support for IoT;
- A programming model has been proposed to support fog computing applications, which includes event handlers and APIs;
- GigaSight is a virtual machine-based cloudlet infrastructure used to support video storage analytics at the edge.

Related Work

- Resource management objectives and challenges in cloud computing; resource management functions and network-aware resource allocation;
- Multi-clouds are platforms that aggregate computing resources from different cloud providers;
- The ETSI has launched the initiative to create standards for mobile edge computing platforms, having proposed a blueprint and also documents presenting technical requirements, terminology, and service scenarios.

Applications

- The fog architecture is hierarchical, where the decision is subject to application constraints and user geo-location
  - *Application constraints can be specified, for instance, as in the form of QoS constraints;*
  - *User geo-location depends on human behavior.*

"By acknowledging different application classes, one could employ different scheduling policies, algorithms, or mechanisms to deal with each class."

Applications

- Considering geo-location and different application classes, were identified two types of apps:
  - *Near real-time: EEGTBG game;*
  - *Delay-tolerant: VSOT application.*

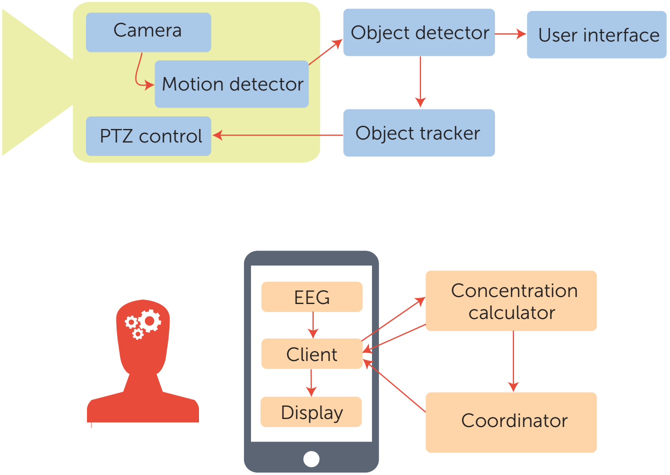


Figure 2: Example applications and their modules

Applications

- Electroencephalography tractor beam game (EEGTBG):
  - *Players try to gather items by concentrating on them. A player that has a better concentration on an item can attract it towards him/herself;*
  - *Fast processing and low response times achieved by edge computing devices can give players a true online, real-time experience.*

Applications

- Video surveillance/object tracking application (VSOT):
  - *Set of distributed intelligent cameras that are able to track movement, having 6 modules: camera, motion detector, object detector, object tracker, user interface, and pan, tilt, and zoom (PTZ) control.*

Applications

- EEGTBG (delay-sensitive) and VSOT (delay-tolerant) can be set up in a fog to take advantage of low latency due to the use of cloudlets
  - *VSOT is able to work under data center-distance latencies >100 miliseconds;*
  - *In EEGTBG, higher delays can impact the players real-time perception, making the game unreal and impairing its playability.*

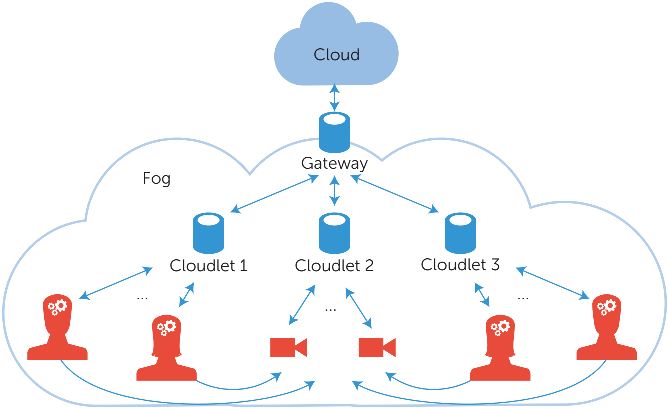
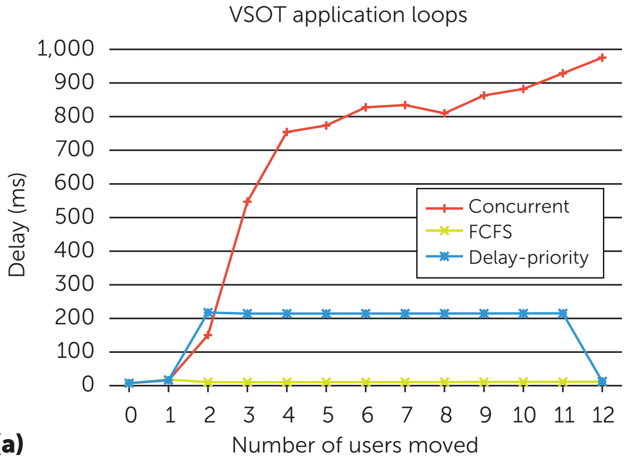


Figure 3: Mobility scenario: mobile concentration game users electroencephalography tractor beam game (EEGTBG) move and compete for the same cloudlet resources with existing surveillance (VSOT) application

Allocation Policies

Scheduling Strategies

- ***Concurrent***: application requests that arrive at a cloudlet are simply allocated to such cloudlet, regardless of capacity or current usage;
- ***First Come-First Served (FCFS)***: requests are served in the order of their arrival, until there are no more computing resources available;
- ***Delay-priority***: applications requiring lower-delay are scheduled first.



(a)

Figure 4: Application loop delays according to the scheduling strategy

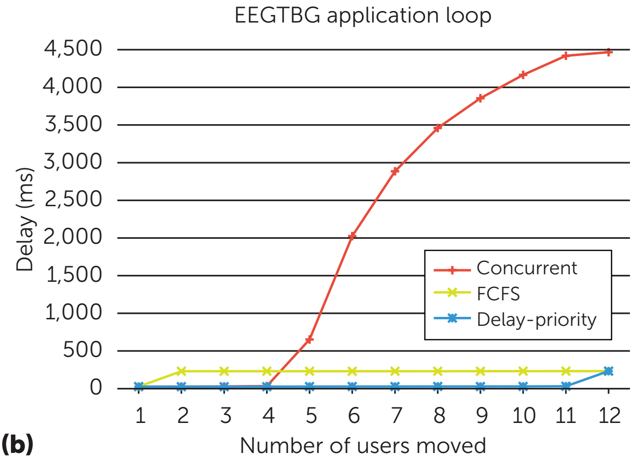


Figure 5: Application loop delays according to the scheduling strategy

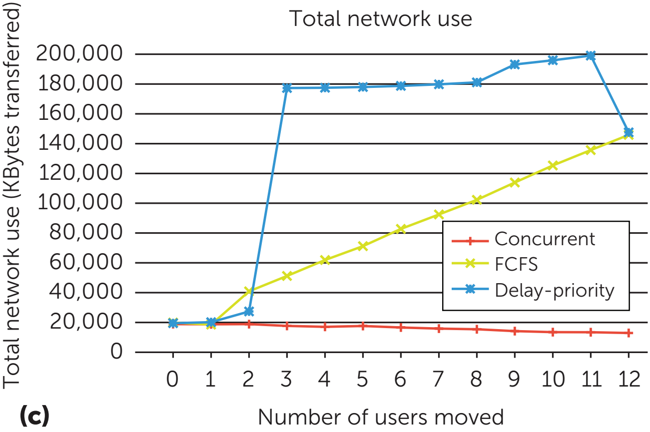


Figure 6: Network usage to the scheduling strategy

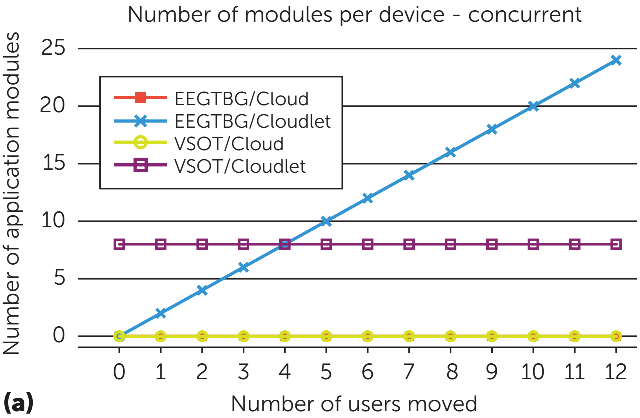


Figure 7: Number of modules per device - concurrent

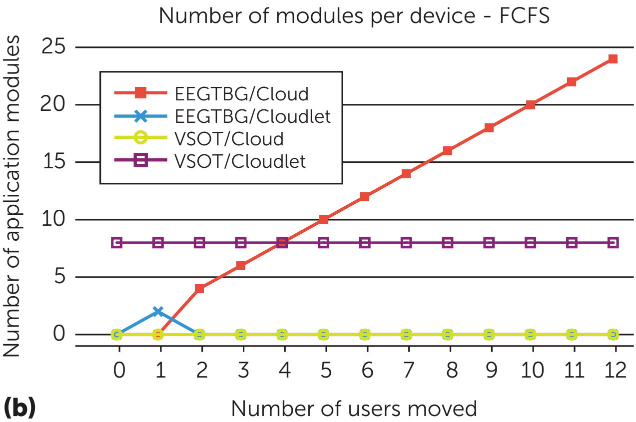


Figure 8: Number of modules per device - FCFS

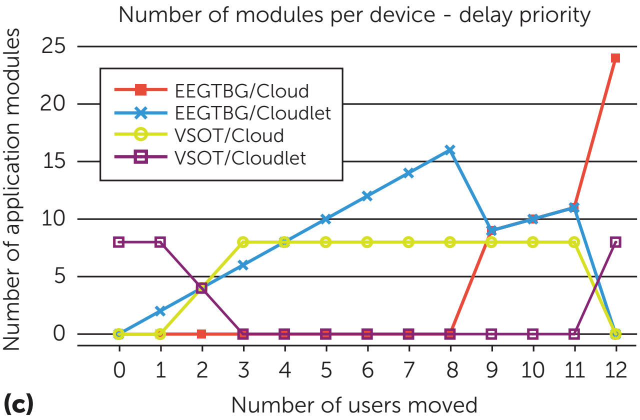


Figure 9: Number of modules per device - delay priority

Challenges and Future Directions

- **Application classification** and **user mobility** are key aspects to be associated with scheduling;
  - *The former, provide the scheduler with information about application requirements, which will allow the scheduler to prioritize the cloudlet use and other optimizations;*
  - *The latter, improve resource management by better planning the applications scheduling beforehand. This planning is crucial to avoid application delays during user movement.*

Challenges and Future Directions

Some interesting areas to explore

- Strategies to deal with mobility prediction failure;
- Maintaining connectivity without service disruption while migration occurs (resource virtualization, SDN);
- Application execution costs in a fog utility model.

Conclusions

- Fog computing provides lower communication latencies and computing capacity closer to the final user;
- For this infrastructure to become efficient, proper resource management mechanisms must be deployed;
- Scheduling strategies can be designed to cope with different application classes according to the demand coming from mobile users.

Mobility-Aware Application Scheduling in Fog Computing

Luiz F. Bittencourt Javier Diaz-Montes Rajkumar Buyya
Omer F. Rana Manish Parashar

Presenter: Maicon Ança dos Santos

October 9, 2018