Mobile Cloud, Internet of Things and Cloud Computing Summary

Satish Srirama
Outline

• Continue with Research at Mobile & Cloud Lab
• Quick recap of what we have learnt as part of this course
• How to prepare for the examination
The Seven Mass Media

The proliferation of smartphones, tablets and other mobile devices is contributing to change in the private sector, as businesses continue to leverage these gadgets in an attempt to enhance efficiency and potentially gain a competitive advantage. According to a new report by Global Industry Analysts, the evolution of mobility is also changing the cloud computing landscape, pushing the mobile cloud market to generate more than $11 billion in revenue by 2018.

Verizon's Stratton: The Future Of IT Is Mobile And Cloud
Mobile Applications

• One can do interesting things on mobiles directly
  – Today’s mobiles are far more capable
  – Location-based services (LBSs), mobile social networking, mobile commerce, context-aware services etc.

• It is also possible to make the mobile a service provider
  – Mobile web service provisioning [Srirama et al, ICIW 2006; Srirama and Paniagua, MS 2013]
  – Challenges in security, scalability, discovery and middleware are studied [Srirama, PhD 2008]
  – Mobile Social Network in Proximity [Chang et al, PMC 2014]
However, we still have not achieved

• Longer battery life
  – Battery lasts only for 1-2 hours for continuous computing

• Same quality of experience as on desktops
  – Weaker CPU and memory
  – Storage capacity

• Still it is a good idea to take the support of external resources for building resource intensive mobile applications
Mobile Cloud Applications

• Bring the cloud infrastructure to the proximity of the mobile user

• Mobile has significant advantage by going cloud-aware
  – Increased data storage capacity
  – Availability of unlimited processing power
  – PC-like functionality for mobile applications
  – Extended battery life (energy efficiency)
Mobile Cloud – Our Interpretation

• We should **not** see Mobile Cloud to be just a scenario where mobile is taking the help of a much powerful machine!!!
• We should **not** see cloud as just a pool of virtual machines

• Mobile Cloud based system should take advantage of some of the key intrinsic characteristics of cloud efficiently
  – Elasticity & AutoScaling
  – Utility computing models
  – Parallelization (e.g., using MapReduce)
Mobile Cloud - Binding models

Task Delegation

[Flores and Srirama, JSS 2014]

Code Offloading

[Flores et al, IEEE Communications Mag 2015; Zhou et al, Cloud 2015; TSC 2017]
Task Delegation

• Follows traditional SOA model to invoke services
  – Similar to mobile Web service client

• Typical scenarios
  – Face recognition, sensor mining etc.

• Critical challenges were (~2010)
  – Hybrid clouds and cloud interoperability
  – Unavailability of standards and mobile platform specific API
  – Fat applications and significant network latencies

SOA - service-oriented architecture
Mobile Cloud Middleware (MCM)

[Warren et al, IEEE PC 2014]
Mobile Cloud Middleware – enables

- Interoperability between different Cloud Services (IaaS, SaaS, PaaS) and Providers (Amazon, Eucalyptus, etc.)
- Provides an abstraction layer on top of API
- Composition of different Cloud Services
- Asynchronous communication between the device and MCM [Warren et al, IEEE PC 2014]
- Means to parallelize the tasks and take advantage of Cloud’s intrinsic characteristics [Flores and Srirama, JSS 2014]
MCM applications

• CroudSTag [Srirama et al, MobiWIS 2011]
  – Social group formation with people identified in Pictures/Videos
  – Cloud services used
    • Media storage on Amazon S3
    • Processing videos on Elastic MapReduce
    • face.com to recognize people on facebook
    • Starting social group on facebook
Code Offloading - Major Components

• Major research challenges
  – What, when, where and how to offload?

• Mobile
  – Code profiler
  – System profilers
  – Decision engine

• Cloud based surrogate platform

[Flores and Srirama, MCS 2013]
Challenges and technical problems

• Inaccurate code profiling
  – Code has non-deterministic behaviour during runtime
    • Based on factors such as input, type of device, execution environment, CPU, memory etc.
  – Some code cannot be profiled (e.g. REST)

• Integration complexity
  – Dynamic behaviour vs Static annotations
    • E.g. Static annotations cause unnecessary offloading

• Dynamic configuration of the system

• Offloading scalability and offloading as a service
  – Surrogate should have similar execution environment
  – Should also consider about resource availability of Cloud

[Flores et al, IEEE Communications Mag 2015]
Applications that can benefit became limited with increase in device capacities

[Srirama, CSIICT 2017]
Internet of Things (IoT)

• “The Internet of Things allows people and things to be connected Anytime, Anyplace, with Anything and Anyone, ideally using Any path/network and Any service.” [European Research Cluster on IoT]

• More connected devices than people

• Cisco believes the market size will be $19 trillion by 2025

Source: Cisco IBSG, April 2011
IoT - Scenarios

• Environment Protection
• Smart Home

[Kip Compton]
[Perera et al, TETT 2014]
Internet of Things – Challenges

- How to provide energy efficient services?
- How do we communicate automatically?
- How to interact with ‘things’ directly?

Sensors

Tags

Mobile Things

Appliances & Facilities

[Chang et al, ICWS 2015]

[Chang et al, SCC 2015; Liyanage et al, MS 2015]
Cloud-centric IoT

- Remote Cloud-based processing
- Connectivity nodes & Embedded processing
- Sensing and smart devices

[Srirama, CSIICT 2017]
IoT Data Processing on Cloud

• Enormous amounts of unstructured data
  – In Zetabytes ($10^{21}$ bytes) by 2020 [TelecomEngine]
  – Has to be properly stored, analysed and interpreted and presented

• Big data acquisition and analytics
  – Is MapReduce sufficient?
    • MapReduce is not good for iterative algorithms [Srirama et al, FGCS 2012]
  – IoT mostly deals with streaming data
    • Message queues such as Apache Kafka can be used to buffer and feed the data into stream processing systems such as Apache Storm
    • Apache Spark streaming

[Distributed Data Processing on the Cloud - LTAT.06.005 (Fall 2018)]
Issues with Cloud-centric IoT

- Latency issues for applications with sub-second response requirements
- Certain scenarios do not let the data move to cloud
- Fog computing [Chang et al, AINA 2017]
  - Processing across all the layers, including network switches/routers
- Challenges in Fog computing
  - Mobility, task migration, discovery, scalability and containerization [Soo et al, IJMCMC 2017; Chang et al, IEEE Computer 2017]
  - QoE-aware application placement across Fog topology [Mahmud et al, JPDC 2018]
Research Roadmap – IoT & Fog Computing

- Distributed data processing on the Cloud
  E.g. MapReduce, Spark

- Distributed data processing across the Cloud and Fog layers
  E.g. Personalized data, privacy etc.

- Fog topology management and scheduling the tasks
  E.g. tasks run across the fog topology such as stream data processing, smart streetlights etc.

- Edge analytics
  E.g. filter, error detection, consolidation etc.

- Intelligent sensors
  E.g. vehicular networks

Cloud

Core Network

Edge Nodes

Fog

Gateways

End points

03/04/2018 Satish Srirama 22/40
WE ALWAYS WELCOME NEW IDEAS!

03/04/2018

Satish Srirama
WHAT WE LEARNT IN THE COURSE!
What is Cloud Computing?

• Computing as a utility
  – Consumers pay based on their usage

• Cloud Computing characteristics
  – Illusion of infinite resources
  – No up-front cost
  – Fine-grained billing (e.g. hourly)

• Gartner: “Cloud computing is a style of computing where massively scalable IT-related capabilities are provided ‘as a service’ across the Internet to multiple external customers”
Cloud Computing - Services

- **Software as a Service – SaaS**
  - A way to access applications hosted on the web through your web browser

- **Platform as a Service – PaaS**
  - Provides a computing platform and a solution stack (e.g. LAMP) as a service

- **Infrastructure as a Service – IaaS**
  - Use of commodity computers, distributed across Internet, to perform parallel processing, distributed storage, indexing and mining of data
  - Virtualization

- **SaaS**
  - Facebook, Flikr, Myspace.com, Google maps API, Gmail

- **PaaS**
  - Google App Engine, Force.com, Hadoop, Azure, Heroku, etc.

- **IaaS**
  - Amazon EC2, Rackspace, GoGrid, SciCloud, etc.
Cloud Computing - Themes

- Massively scalable
- On-demand & dynamic
- Only use what you need - Elastic
  - No upfront commitments, use on short term basis
- Accessible via Internet, location independent
- Transparent
  - Complexity concealed from users, virtualized, abstracted
- Service oriented
  - Easy to use Service Level Agreements
Cloud Computing Progress

• Short term and long term implications of cloud
• Economics of cloud users and cloud providers
• Challenges and opportunities offered by cloud computing
Cloud Providers

• Amazon Web Services
  – EC2, S3, EBS, Elastic Load Balancing, Amazon Auto Scale, Amazon CloudWatch, IAM, CloudFormation, Elastic MapReduce
• Private Cloud enabling technologies
  – Eucalyptus
  – OpenStack
    • Worked with SciCloud
• PaaS
  – Google AppEngine
  – Windows Azure
• Serverless computing
  – Apache OpenWhisk
    • Triggers, Actions, Rules
Scaling Applications on the Cloud

• Two basic models of scaling
  – Vertical scaling, aka Scale-up
  – Horizontal scaling, aka Scale-out

• Scaling Enterprise Applications in the Cloud

• Load balancing
  – Types and algorithms

• Autoscaling
Economics of Cloud Providers

• Cloud Computing providers bring a shift from high reliability/availability servers to commodity servers
  – At least one failure per day in large datacenter

• Why?
  – Significant economic incentives
    • much lower per-server cost

• Caveat: User software has to adapt to failures
  – Very hard problem!

• Solution: Replicate data and computation
  – MapReduce & Distributed File System
MapReduce

- Programmers specify two functions:
  - `map (k, v) → <k', v'>`
  - `reduce (k', v') → <k', v'>`
    - All values with the same key are reduced together
- The execution framework handles everything else...
- Not quite...usually, programmers also specify:
  - `partition (k', number of partitions) → partition for k'`
    - Often a simple hash of the key, e.g., `hash(k') mod n`
    - Divides up key space for parallel reduce operations
  - `combine (k', v') → <k', v'>`
    - Mini-reducers that run in memory after the map phase
    - Used as an optimization to reduce network traffic
Hadoop Processing Model

• Create or allocate a cluster
• Put data onto the file system (HDFS)
  – Data is split into blocks
  – Replicated and stored in the cluster
• Run your job
  – Copy Map code to the allocated nodes
    • Move computation to data, not data to computation
  – Gather output of Map, sort and partition on key
  – Run Reduce tasks
• Results are stored in the HDFS
MapReduce Examples

- Distributed Grep
- Count of URL Access Frequency
- Reverse Web-Link Graph
- Inverted Index
- Distributed Sort
Synchronization in Hadoop

• Approach 1: turn synchronization into an ordering problem
  – Sort keys into correct order of computation
  – Partition key space so that each reducer gets the appropriate set of partial results
  – Hold state in reducer across multiple key-value pairs to perform computation
  – Illustrated by the “pairs” approach in calculating conditional probability of words

• Approach 2: construct data structures that “bring the pieces together”
  – Each reducer receives all the data it needs to complete the computation
  – Illustrated by the “stripes” approach
Research @ Mobile & Cloud Lab

- Scientific computing on the cloud
- Classification on how the algorithms can be adapted to MR
- Limitations of Hadoop MapReduce
  - Alternatives
- Mobile Cloud
  - Mobile Cloud Binding Models
- Cloud-centric Internet of Things
- Fog computing
This week in lab

• Work with cloud functions
  – Function as a service based on OpenWhisk
  – Severless computing
Examination

• Exam I : Monday, 09.04.2018, 10:00 – 12:00 Ülikooli 17 /Paabel/ room 220

• Exam II : Tuesday, 10.04.2018, 10:00 – 12:00 in room J. Liivi 2 - 402
Preparation for Examination

- One of the earlier exam papers is kept online
- Slides are mostly self sufficient
- References are mentioned for further reading
- Mainly focus at keywords
References


