Mobile and Cloud Computing Laboratory

Major Research Interests

Satish Srirama

satish.srirama@ut.ee

Mobile Cloud Lab

iktp

European Union
Regional Development Fund

Investing in your future

Who am I

- Head of Mobile & Cloud Lab, Institute of Computer Science, University of Tartu, Estonia

http://mc.cs.ut.ee

Estonia pop: 1,300,000

TARTU

Pop: 100,000

Academic excellence since 1632

CS@UT – Key Numbers

600 students

60 staff

- 40 academic (professors, associate professors…)
- 20 full-time researchers

60 industry guest lecturers/year

60 PhD students

Highly international profile

- **60%** of academic staff have international experience
- **60%** international PhD and Masters students

6 Curricula

Bachelors (3 years)	Masters (2 years)		PhD (4y)
Computer Science	Computer Science	Software Engineering (joint TUT)	Computer Science

- **Masters of Security and Mobile Computing (with Aalto, KTH, DTU, NTNU)**
- **Masters of Cyber-Security (coordinated by TUT)**

6 Key Competence Areas

Software Engineering http://sep.cs.ut.ee	Distributed & Mobile and Cloud Computing http://mc.cs.ut.ee http://ds.cs.ut.ee	Data Mining, Bioinformatics & Neuroscience http://biit.cs.ut.ee
Programming Languages http://lambda.cs.ut.ee	Security, Crypto & Coding http://crypto.cs.ut.ee/	Estonian Language Technology

Software Technology Competence Centre (STACC.ee)

Mobile & Cloud Lab – Main Research Activities

Mobile Cloud Lab

UNIVERSITY OF TARTU
Institute of Computer Science
1632

Home | News | Events | Research | People | Projects | Publications | Teaching | Jobs | Theses | Blog

You are here: Home > Research

Contents | View | Edit | Rules | Sharing

Navigation
- Home
- News
- Events
- Research
  - Internet of Things
- People
- Projects
- Publications
- Teaching
- Jobs
- Theses
- Blog

Manage portlets

Research

by admin — last modified Oct 01, 2014 02:10 PM — History

The research at the Mobile & Cloud Lab contributes to the following fields:

- **Cloud Computing**
  The research goal is to study the migration of enterprise applications to the cloud and to study their performance on the cloud.
- **Scientific Computing on the Cloud**
  The research goal is to study the migration of scientific computing applications to the cloud and to reduce these applications and a
- **Mobile Computing**
  The research deals with developing mobile applications for various platforms and devices (e.g. Android, iOS, Windows Phone 7 etc
- **Mobile Cloud**
  The goal of the research is to investigate how to efficiently utilize cloud resources within the mobile applications (aka mobile cloud
- **Mobile Web Services**
  This research theme deals with the invocation, provisioning, discovery and integration of web services from smart phones, in develo
- **Internet of Things**
  The goal of this research is to overcome the challenges of cyber-physical systems in the Internet of Things. The challenges include: i efficiency, trustworthiness etc.

Outline

- Cloud computing
- Migrating enterprise applications to the cloud
- Scientific computing on the cloud
- Mobile Cloud & Internet of Things

What is Cloud Computing?

- Computing as a utility
  - Utility services e.g. water, electricity, gas etc
  - Consumers pay based on their usage

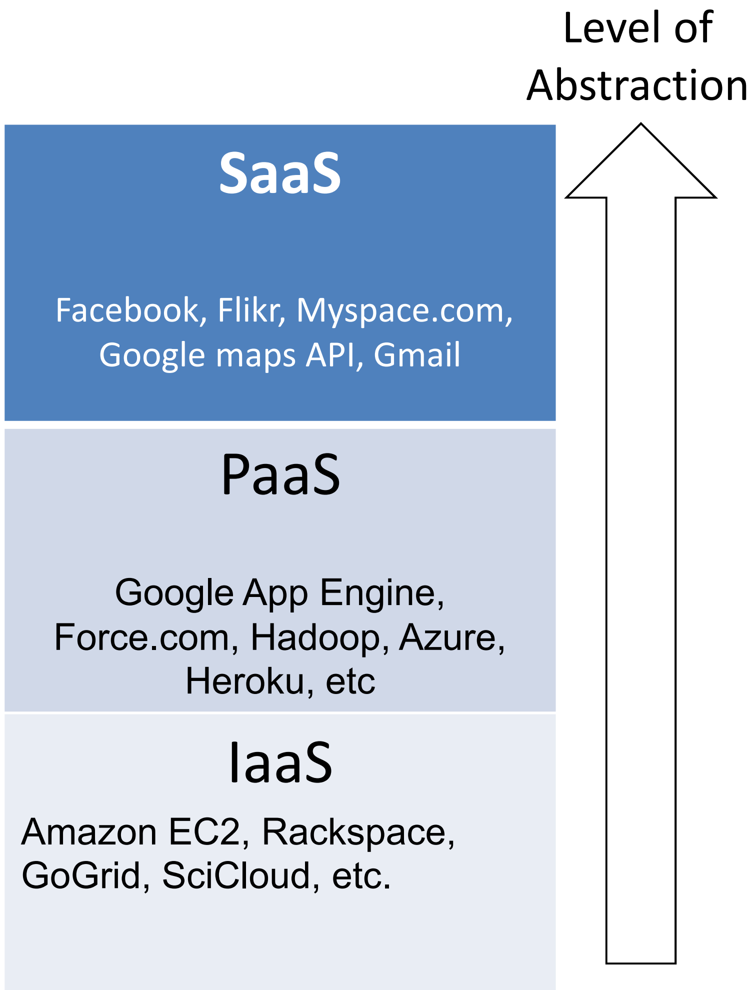
1969 – Leonard Kleinrock, ARPANET project

- "As of now, computer networks are still in their infancy, but as they grow up and become sophisticated, we will probably see the spread of 'computer utilities', which, like present electric and telephone utilities, will service individual homes and offices across the country"

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

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

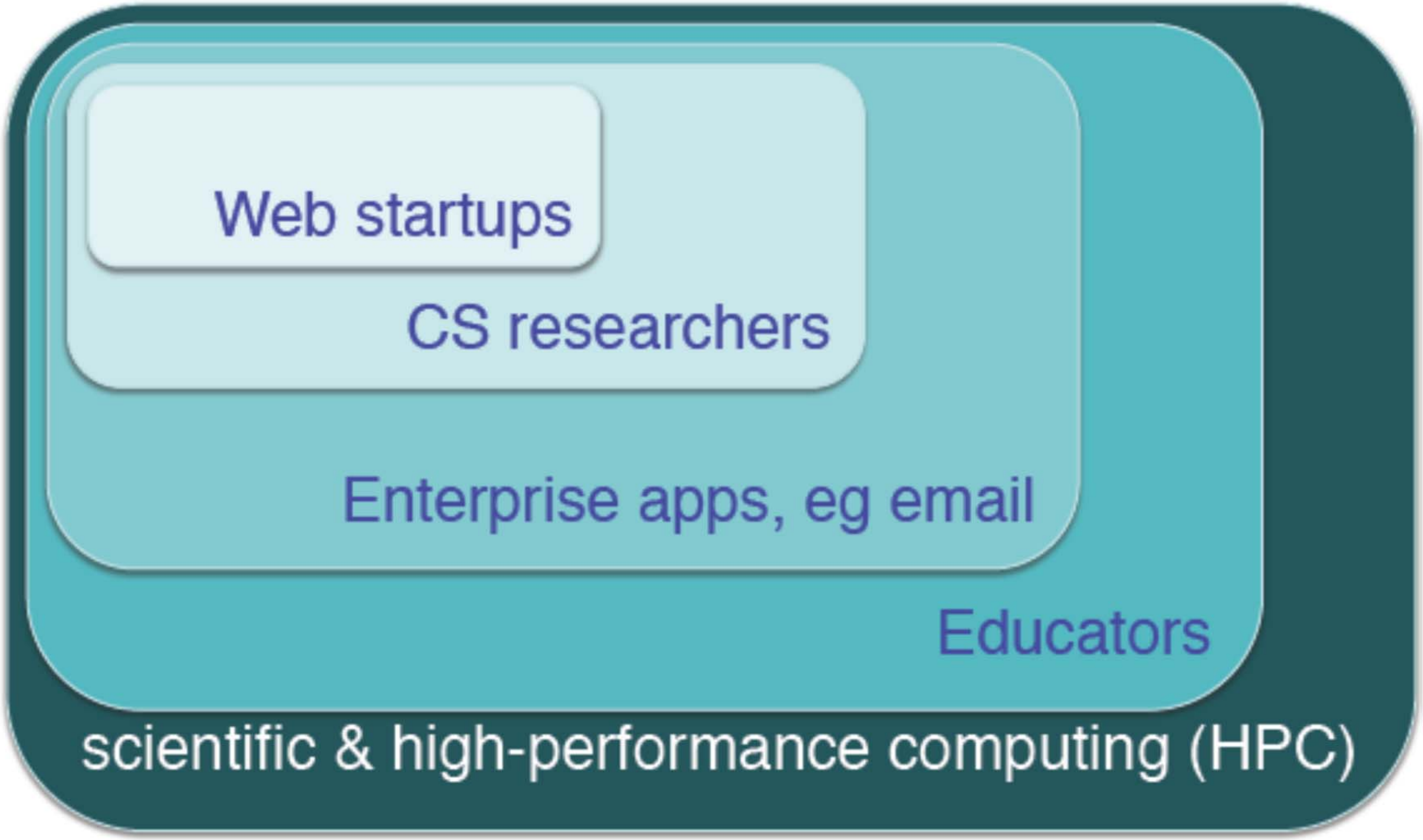


Cloud Computing - Themes

- 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 SLAs

SLA – Service Level Agreement

Cloud Computing Progress



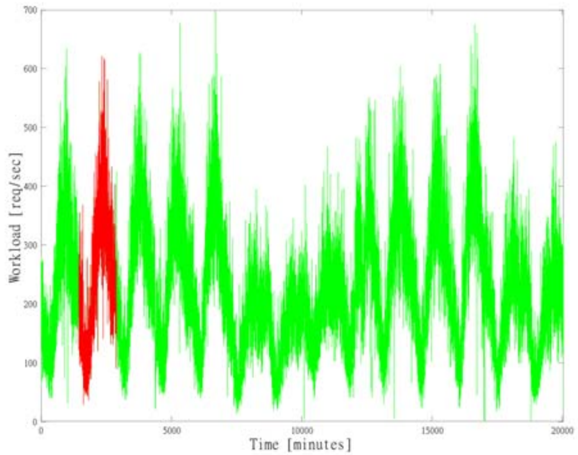
[Armando Fox, 2010]

Research Challenges

MIGRATING ENTERPRISE APPLICATIONS TO THE CLOUD

Enterprise applications on the cloud

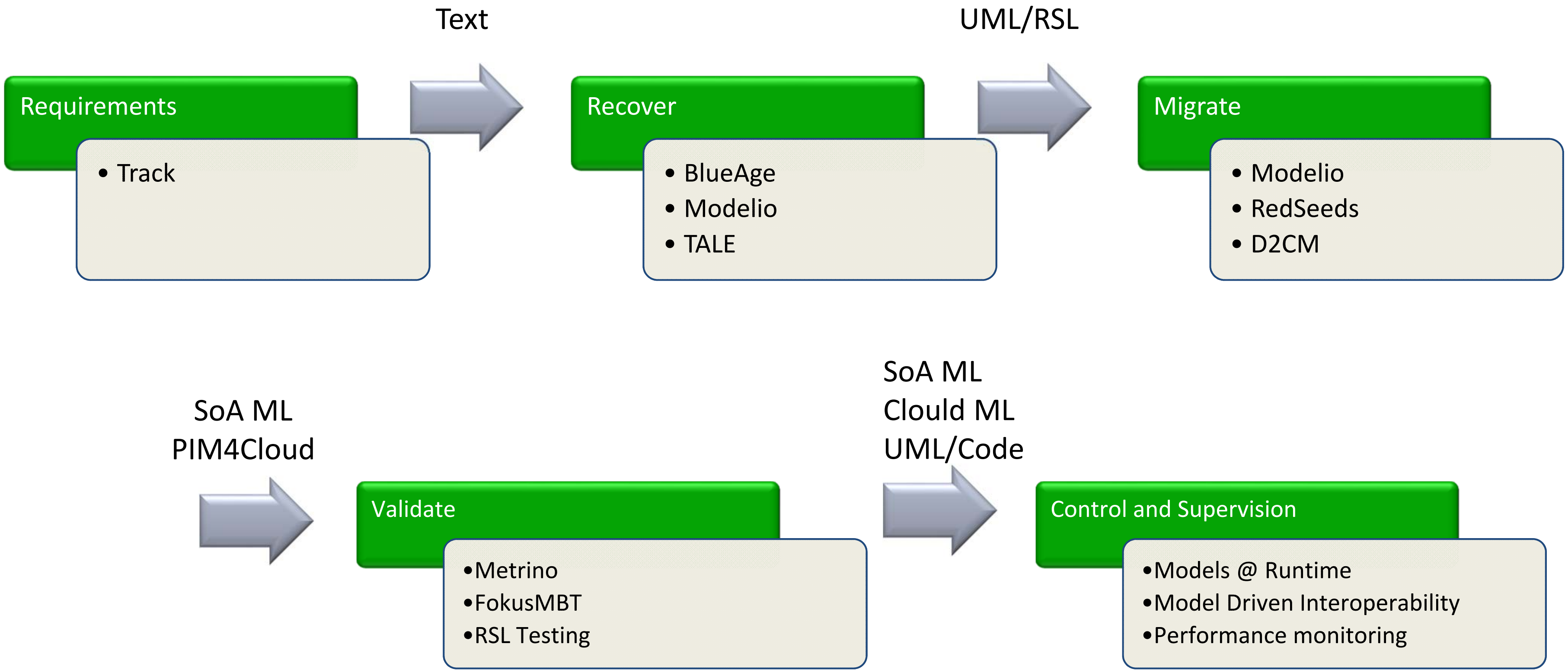
- Enterprise applications are mostly based on SOA and componentized models
- Fault tolerance, high availability & scalability
  - Essential prerequisites for any information system
- Cloud with its promise of virtually unlimited resources can offer the above prerequisites
  - Availability zones
  - Elasticity and horizontal scaling
  - Utility computing



REMICS

http://www.remics.eu/

- Reuse and migration of legacy applications to the cloud



CloudML

- Developed to tame cloud heterogeneity
- Domain-specific language (DSL) for modelling the provisioning and deployment at design-time
  - Nodes, artefacts and bindings can be defined
- Different means to manipulate CloudML models
  - Programmatically via Java API
  - Declaratively, via serialized model (JSON)
- Models@Runtime
  - Dynamic deployment of CloudML based models

```
"nodeTypes": [
    {
        "id": "SmallGNULinux",
        "os": "GNULinux",
        "compute": [ 2, 4 ],
        "memory": [ 2048, 4096 ],
        "storage": [ 10240 ],
        "location": "eu",
        "provides": [
            { "id": "SSHCapability" }
        ]
    }
]
"artefactsTypes": [
    {
        "id": "Docs",
        "retrieval": "wget http://cloudml.org/a
http://cloudml.org/apps/docs_configure; wget
http://cloudml.org/apps/docs_deploy",
        "configuration": "sudo docs_configure",
        "deployment": "sudo docs_deploy",
        "requires": [
            { "id": "JettyCapability" },
            { "id": "MongoDBCapability" }
        ]
    }
    …
]
```

[Ferry et al, Cloud 2013]

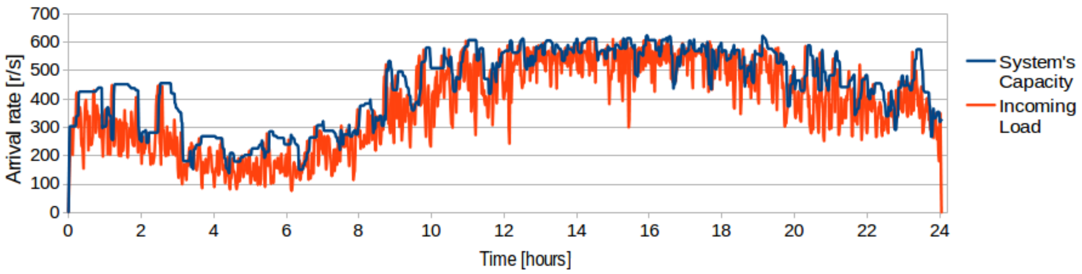
Auto-Scaling enterprise applications on the cloud

- Auto-Scaling
  - Scaling policy **->** When to Scale
  - Resource provisioning policy **->** How to scale
- Threshold-based scaling policies are very popular due to their simplicity
  - Observe metrics such as CPU usage, disk I/O, network traffic etc.
  - E.g. Amazon AutoScale, RightScale etc.
  - However, configuring them optimally is not easy

Optimal Resource Provisioning for Auto-Scaling Enterprise Applications

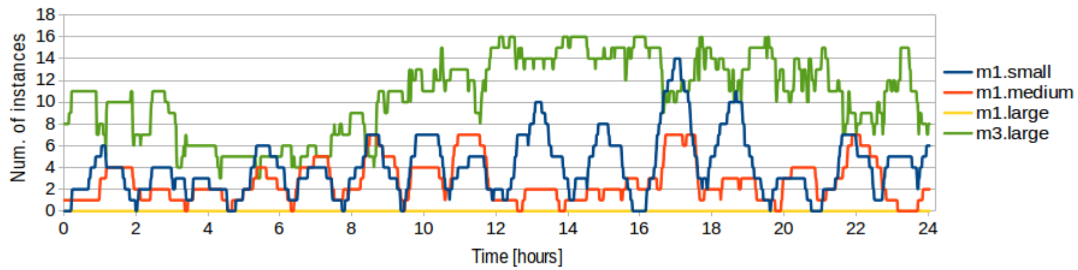
- Cloud providers offer various instance types with different processing power and price
  - Can it be exploited in deciding the resource provisioning policy?
  - Makes the policy to be aware of current deployment configuration
- Another challenge: Cloud providers charge the resource usage for fixed time periods
  - E.g. Hourly prices of Amazon cloud
- Developed an LP based optimization model which considers both the issues [Srirama and Ostovar, CloudCom 2014]

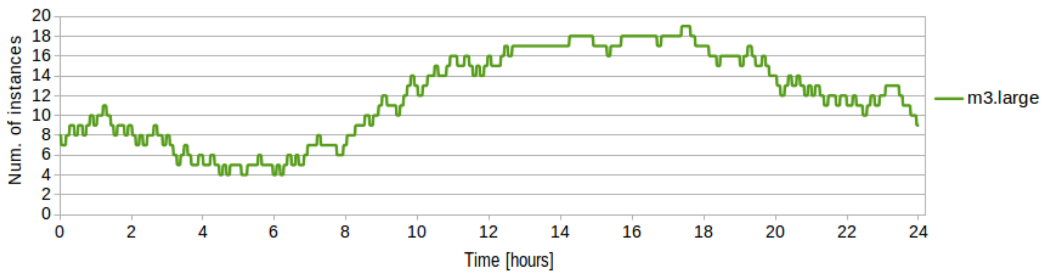
Scaling enterprise application with the optimization model



Incoming load and scaling curves of Optimization model

Instance type usage curves of Optimization model





Scaling with Amazon AutoScale

[Srirama and Ostovar, CloudCom 2014]

Optimization Model

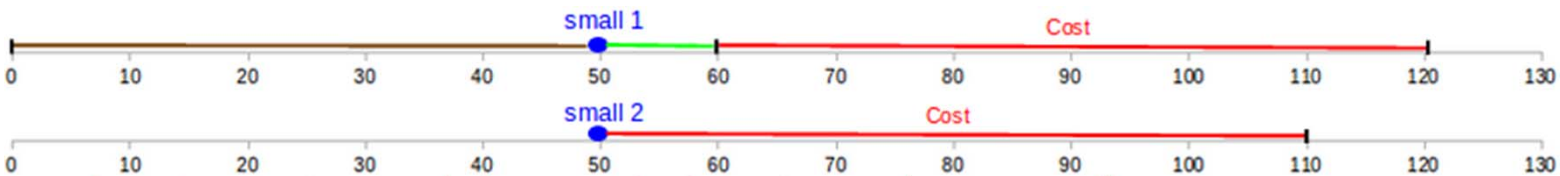
Intuition behind instance lifetime consideration

- Consider 2 instance types
  - Small instance(PW = 6r/s, Price = $0.25/h),
  - Medium instance(PW = 12r/s, Price = $0.4/h)

Load is 6r/s

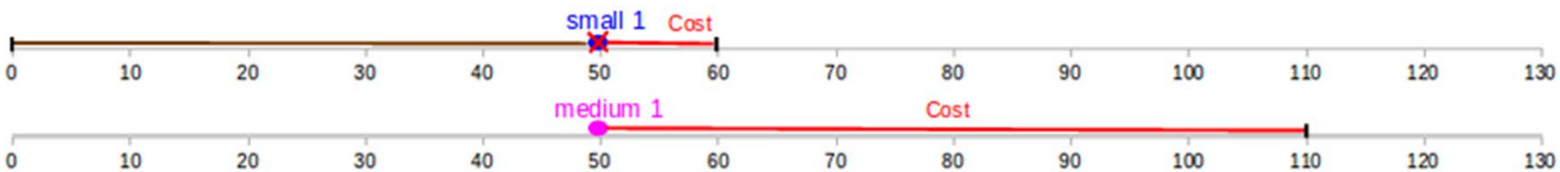
Load increases to 12r/s => ?

Solution 1



Cost = (cost of two small instances) - (10-min profit of a small instance) = 0.5 - 0.04 = $0.46

Solution 2

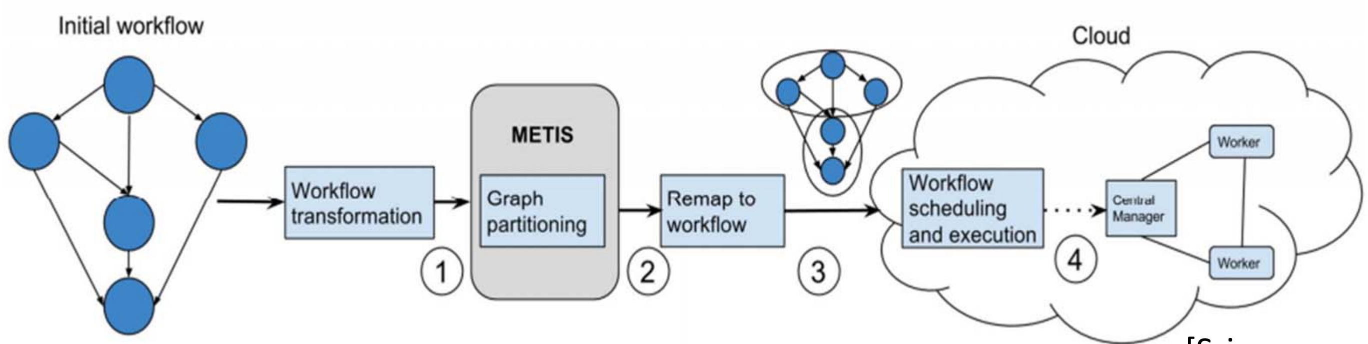


Cost = (cost of a medium instance) + (10-min cost of a small instance) = 0.4 + 0.04 = $0.44

- Saved cost with solution 2 : 0.46 – 0.44 = 0.02$
- The model can find this automatically

Current Interests

- Remodeling enterprise applications for the cloud migration
  - Cloud has huge troubles with communication/transmission latencies [Srirama et al, SPJ 2011]
  - Intuition: Reduce inter-node communication and to increase the intra-node communication
- Auto-scale them based on optimization model and CloudML



[Srirama and Viil, HPCC 2014]

Research Challenges

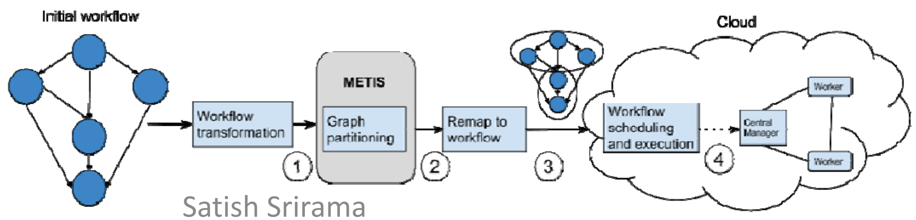
SCIENTIFIC COMPUTING ON THE CLOUD

Scientific Computing on the Cloud

- Public clouds provide very convenient access to computing resources
  - On-demand and in real-time
  - As long as you can afford them
- High performance computing (HPC) on cloud
  - Virtualization and communication latencies are major hindrances [Srirama et al, SPJ 2011; Batrashev et al, HPCS 2011]
    - Things have improved significantly over the years
  - Research at scale
    - Cost-to-value of experiments
- Desktop to Cloud Migration (D2CM) tool for domain scientists [Srirama et al, HPCS 2013]

Migrating Scientific Workflows to the Cloud

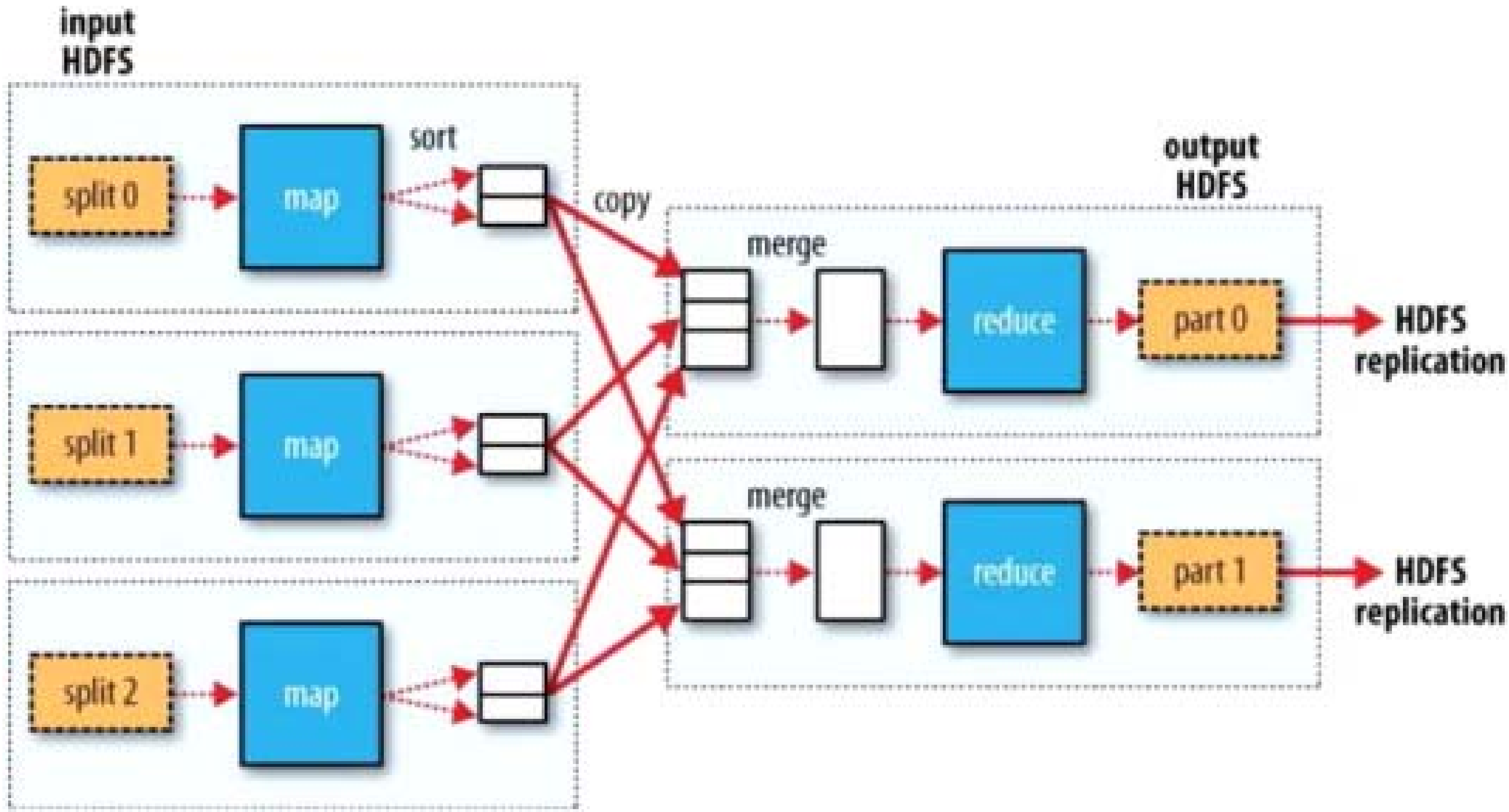
- Workflow can be represented as weighted directed acyclic graph (DAG)
- Partitioning the workflow into groups with graph partitioning techniques [Srirama and Viil, HPCC 2014]
  - Such that the sum of the weights of the edges connecting to vertices in different groups is minimized
  - Utilized Metis' multilevel k-way partitioning
- Scheduling the workflows with tools like Pegasus
  - Considered peer-to-peer file manager (Mule) for Pegasus



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 model



https://tedwon.atlassian.net/wiki/display/SE/Apache+Hadoop

Apache Hadoop MapReduce

- Most prominent Open Source solution
- The user only has to write Map and Reduce functions
- Framework handles everything else for the user
  - Scheduling, data distribution, synchronization, errors and faults
- Parallelism is achieved by executing Map and Reduce tasks concurrently

Adapting Computing Problems to Cloud

- Reducing the algorithms to cloud computing frameworks like MapReduce [Srirama et al, FGCS 2012]
- Designed a classification on how the algorithms can be adapted to MR
  - Algorithm → single MapReduce job
    - Monte Carlo, RSA breaking
  - Algorithm → *n* MapReduce jobs
    - CLARA (Clustering), Matrix Multiplication
  - Each iteration in algorithm → single MapReduce job
    - PAM (Clustering)
  - Each iteration in algorithm → *n* MapReduce jobs
    - Conjugate Gradient
- Applicable especially for Hadoop MapReduce

Issues with Hadoop MapReduce

- It is designed and suitable for:
  - Data processing tasks
  - Embarrassingly parallel tasks
- Has serious issues with iterative algorithms
  - Long „*start up*“ and „*clean up*“ times **~17** seconds
  - No way to keep important data in memory between MapReduce job executions
  - At each iteration, all data is read again from HDFS and written back there at the end
  - Results in a significant overhead in every iteration

Alternative Approaches

- Restructuring algorithms into non-iterative versions
  - CLARA instead of PAM [Jakovits & Srirama, Nordicloud 2013]
- Alternative MapReduce implementations that are designed to handle iterative algorithms [Jakovits and Srirama, HPCS 2014]
  - E.g. Twister, HaLoop, Spark
- Alternative distributed computing models
  - Bulk Synchronous Parallel model [Valiant, 1990] [Jakovits et al, HPCS 2013]
  - Built a fault-tolerant BSP framework (NEWT) [Kromonov et al, HPCS 2014]

Research Challenges

MOBILE CLOUD

The Seven Mass Media

First Mass Media Channel - ***Print*** from the 1500s

Second Mass Media Channel - ***Recordings*** from 1900s

Third Mass Media Channel - ***Cinema*** from 1910s

Fourth Mass Media Channel - ***Radio*** from 1920s

Fifth Mass Media Channel - ***TV*** from 1950s

Sixth Mass Media Channel - ***Internet*** from 1990s

Seventh Mass Media Channel - ***Mobile*** from 2000s

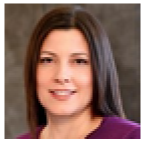
[Tomi T Ahonen]

Rank	Country or region	Number of mobile phones	Population	Phones per 100 citizens	Data evaluaton date
-	World	6,800,000,000+	7,012,000,000[1]	87	2013[2][3]
01	China	1,206,553,000[4]	1,349,585,838[5]	89.2	September 2013[4]
02	India	867,800,000	1,220,800,359[6]	70.72	30 April 2013[7]
03	United States	327,577,529	310,866,000[8]	103.9	June 2013[9]
04	Brazil	268,440,423	192,379,287[10]	135.4	August 2013[11]
05	Russia	256,116,000	142,905,200[10]	155.5	July 2013[12]
06	Indonesia	236,800,000	237,556,363	99.68	September 2013[10]
07	Pakistan	129,583,076	178,854,781[13]	72.45	September 2013[14]
08	Japan	121,246,700	127,628,095	95.1	June 2013[15]
09	Nigeria	114,000,000	165,200,000	69	May 2013[16]
10	Bangladesh	110,675,000	165,039,000	73.8	September 2013[17]

Report: Mobile cloud to grow beyond $11 billion in 2018

Written by CopperEgg // July 12, 2012 // No Comment // Cloud Performance

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.



Maribel Lopez,

I track how mobile

+ Follow (87)

TECH | 4/18/2012 @ 7:43AM | 18,825 views

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

+ Comment Now + Follow Comments

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, ICSOC 2012; 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)

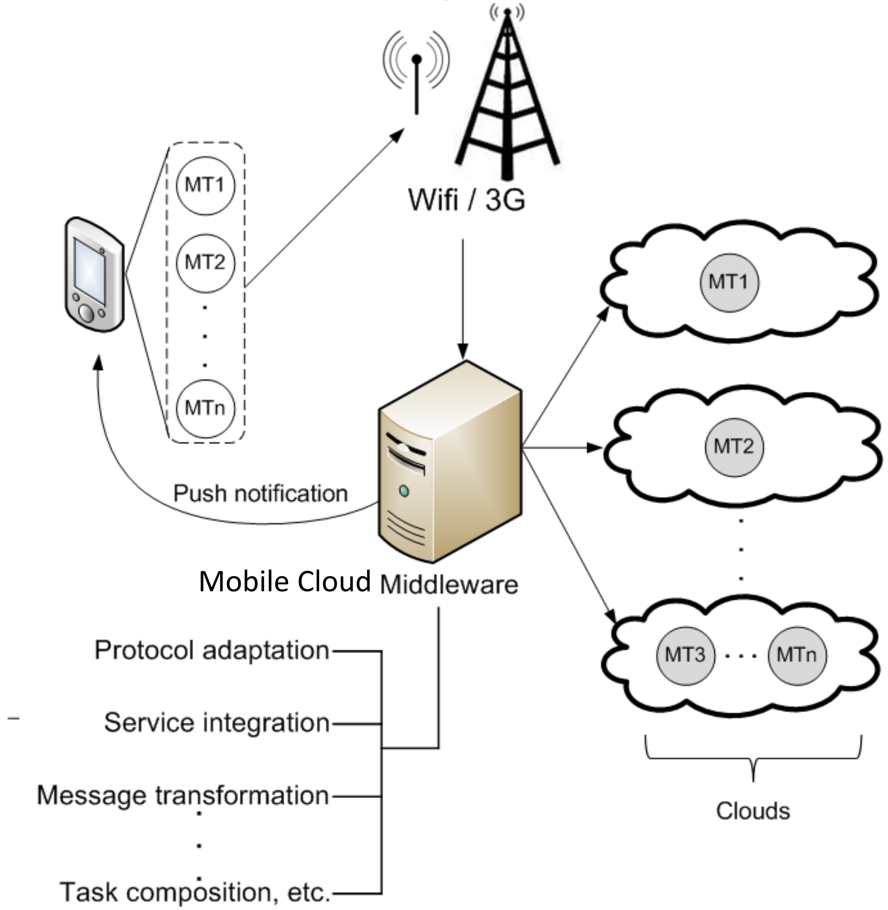
Mobile Cloud – Our interpretation

- We do not see Mobile Cloud to be just a scenario where mobile is taking the help of a much powerful machine!!!
- We do 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)

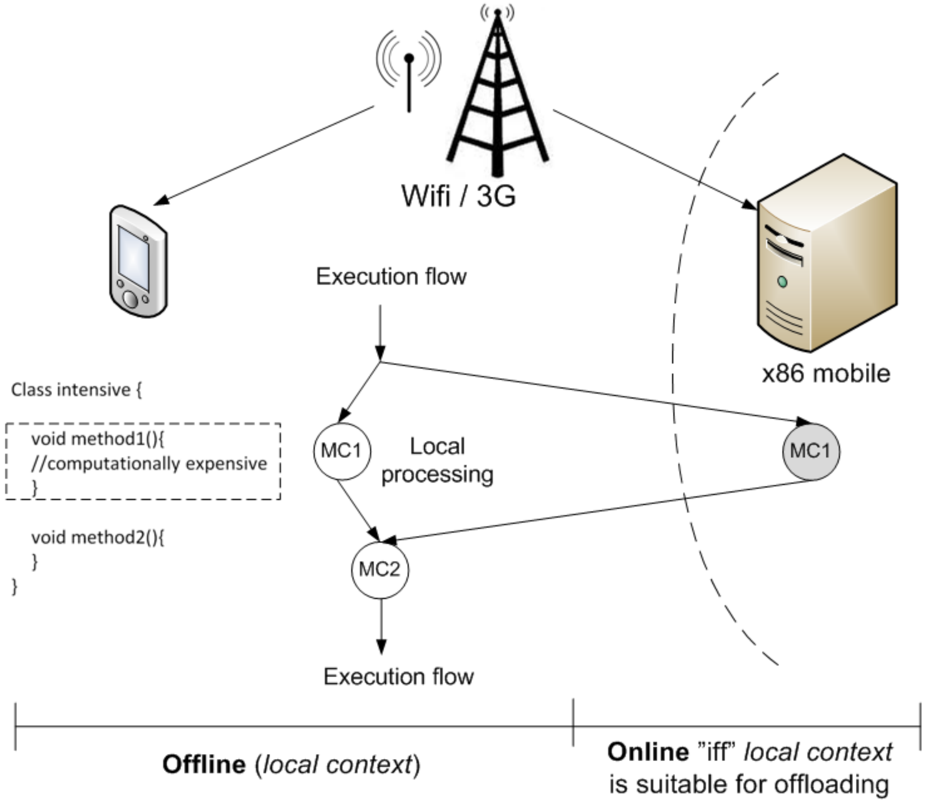
Mobile Cloud Binding Models

[Flores et al, MoMM 2011]



Task Delegation

[Flores and Srirama, MCS 2013]



Code Offloading

[Flores & Srirama, JSS 2014]

[Flores et al, MoMM 2011] MCM – 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

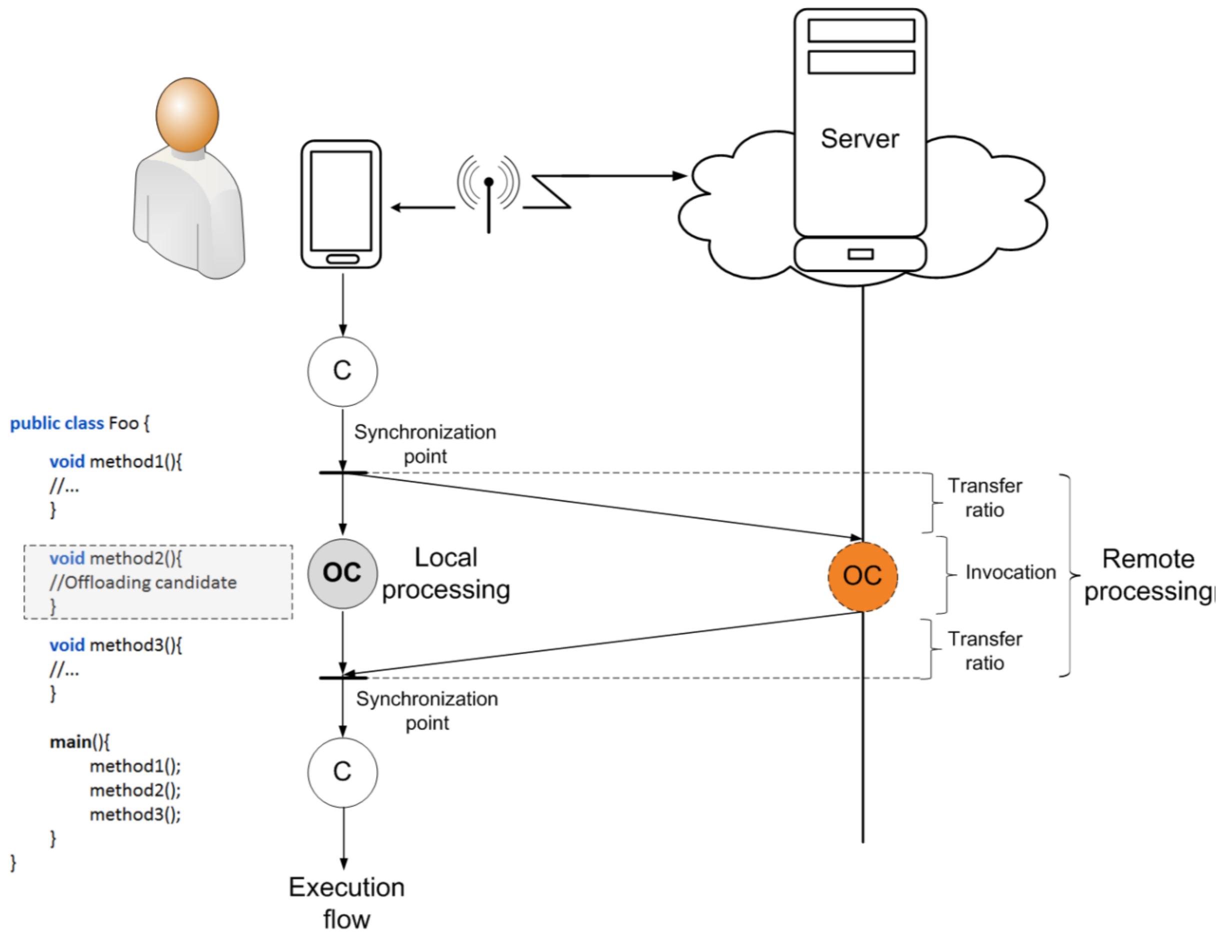
MCM applications

- CroudSTag [Srirama et al, MobiWIS 2011]
  - Social group formation with people identified in Pictures/Videos
- Zompopo [Srirama et al, NGMAST 2011]
  - Intelligent calendar, by mining accelerometer sensor data
- Bakabs [Paniagua et al, iiWAS-2011]
  - Managing the Cloud resources from mobile
- Sensor data analysis
  - Human activity recognition
  - Context aware gaming
  - MapReduce based sensor data analysis [Paniagua et al, MobiWIS 2012]

- SPiCa: A Social Private Cloud Computing Application Framework [Chang et al, MUM 2014]

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



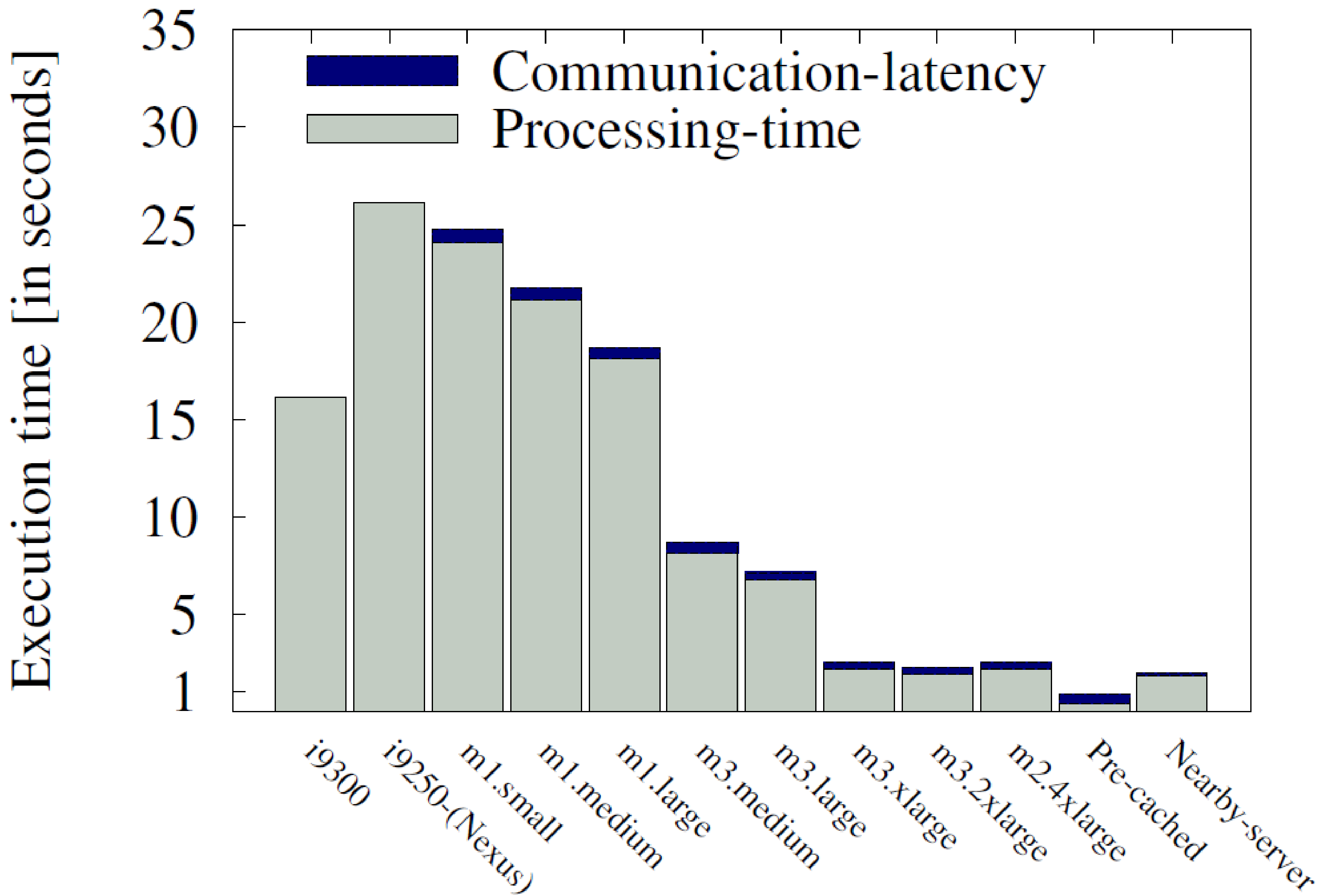
[Flores and Srirama, MCS 2013]

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]

Practical adaptability of offloading



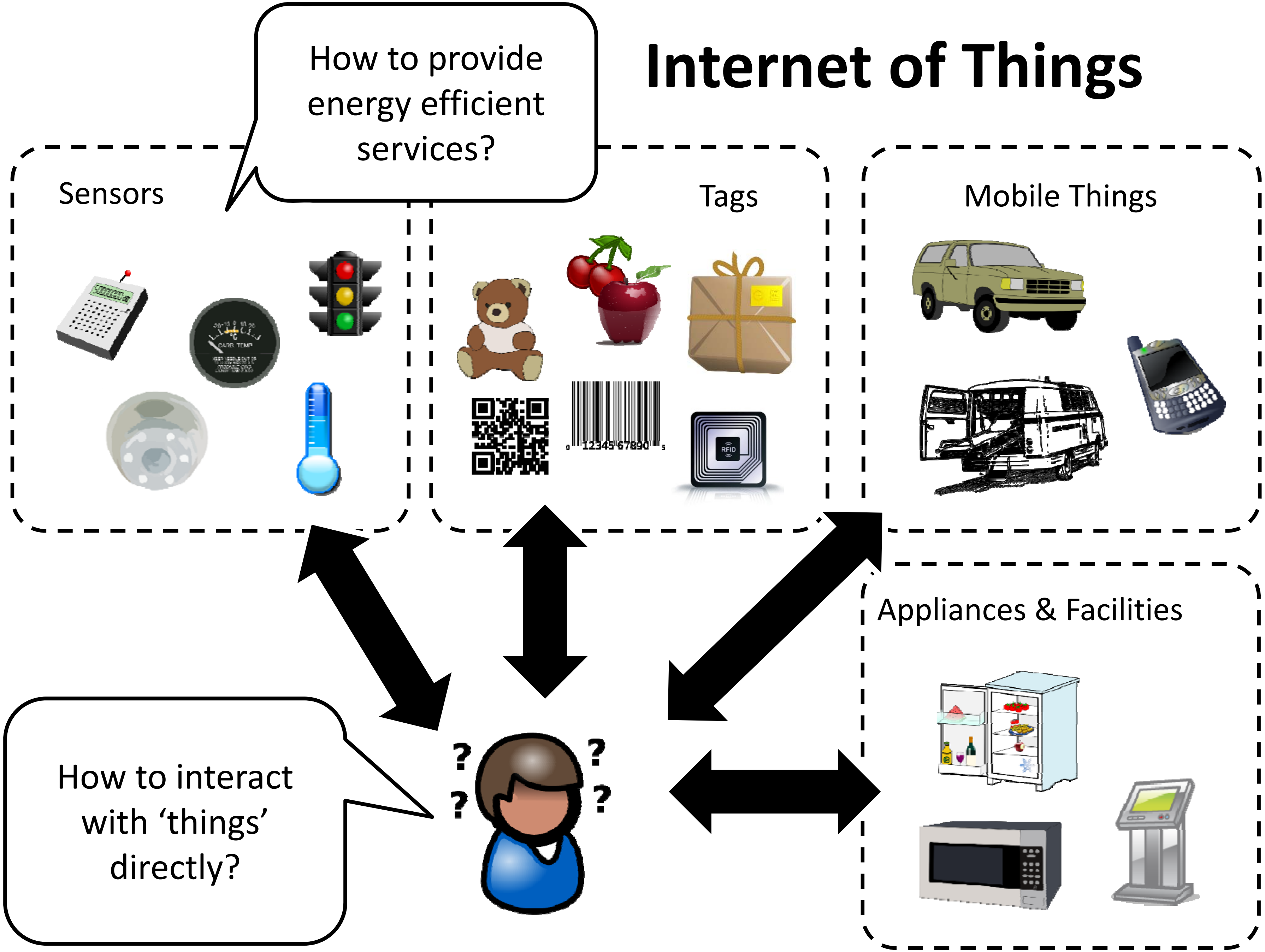
Applications that can benefit became limited with increase in device capacities

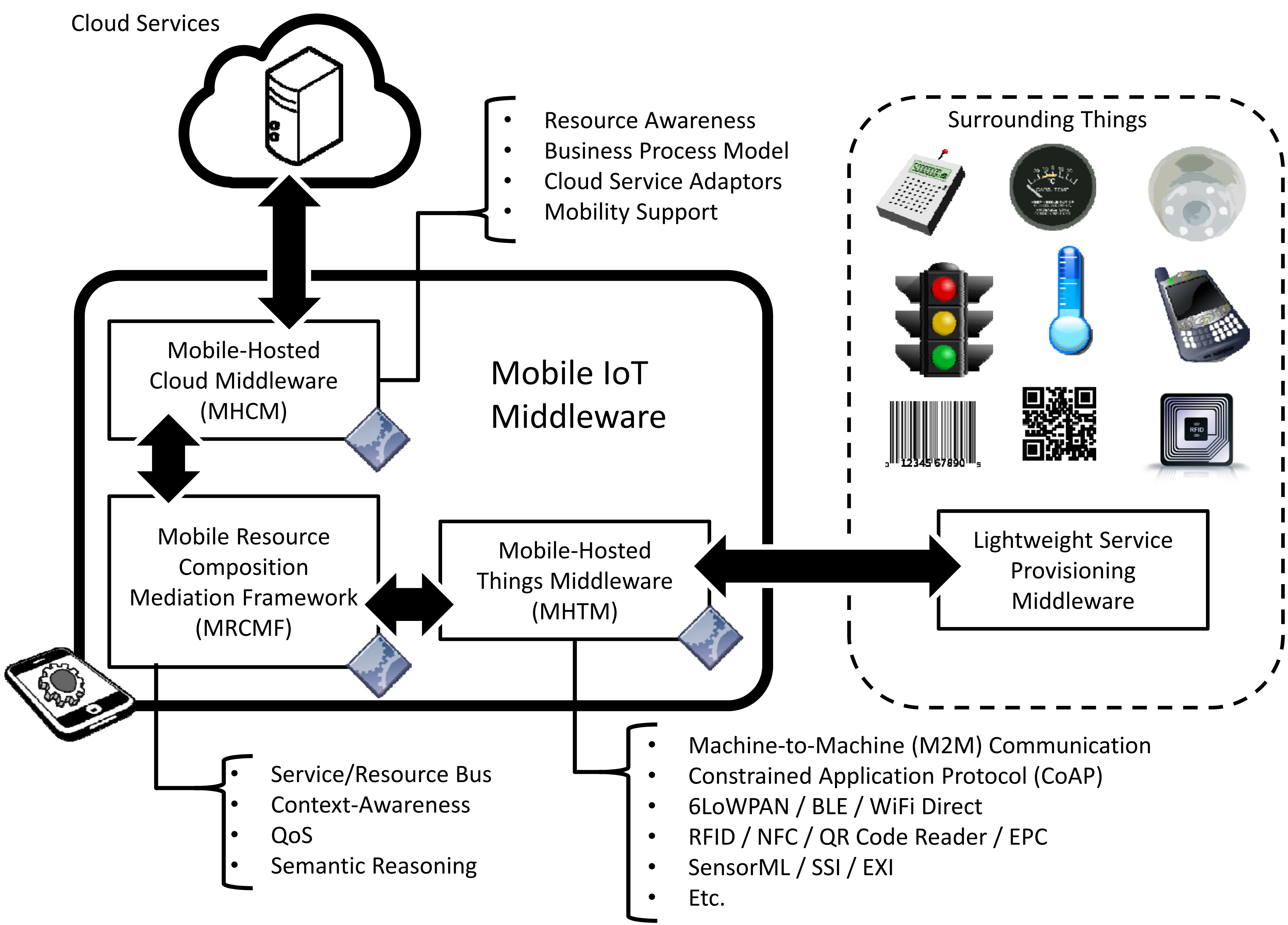
Way to proceed?

- Code offloading is not yet a reality!!!
- Take advantage of crowdsourcing
  - Computational offloading customized by data analytics
  - By analysing how a particular app behaves in a community of devices
  - E.g. Carat detects energy anomalies [Oliner et al, 2013]
    - By studying over ~328,000 apps gets an idea on what is resource – intensive app
    - Determines energy drain distribution of an app
- Decision models can also benefit from crowdsourcing
  - Analysis of code offloading traces [Flores and Srirama, MCS 2013]

[Flores et al, IEEE Communications Mag 2015]

Internet of Things




Cloud Services
Resource Awareness
Business Process Model
Cloud Service Adaptors
Mobility Support
Surrounding Things
Mobile-Hosted
Cloud Middleware
(MHCM)
Mobile IoT
Middleware
Mobile Resource
Composition
Mediation Framework
(MRCMF)
Mobile-Hosted
Things Middleware
(MHTM)
Lightweight Service
Provisioning
Middleware
Service/Resource Bus
Context-Awareness
QoS
Semantic Reasoning
Machine-to-Machine (M2M) Communication
Constrained Application Protocol (CoAP)
6LoWPAN / BLE / WiFi Direct
RFID / NFC / QR Code Reader / EPC
SensorML / SSI / EXI
Etc.

Research Results

- Participated in a number of EU-funded projects
- Partner in the Estonian Center of Excellence in Computer Science
- Partner in Software Technology and Applications Competence Centre (STACC)
  - An R&D center that conducts industry-driven research projects in the fields of software engineering and data mining
- Output resulted in several SMEs
  - Plumbr [Sor and Srirama, JSS 2014; Sor et al, SPE 2015], ZeroTurnaround etc.

Garage48, Startups, SME-s, .. #estonianmafia

European Union
Regional Development Fund

Investing in your future

Eesti Teadusagentuur
Estonian Research Council

srirama@ut.ee

THANK YOU FOR YOUR ATTENTION