Mobile & Cloud Computing: Research Challenges

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

Mobile Cloud Lab

Estonia pop: 1,300,000

TARTU

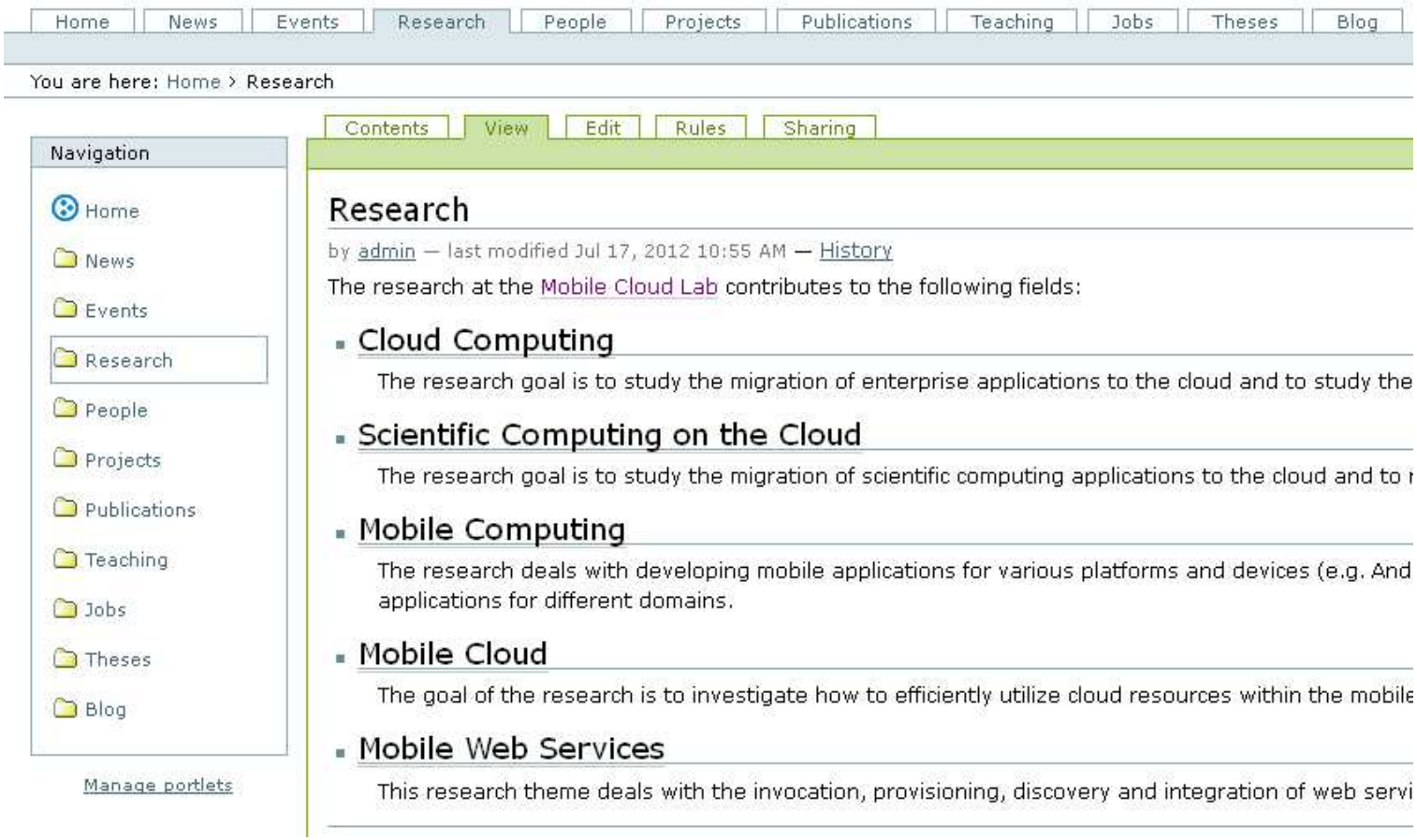
Pop: 100,000

Academic excellence since 1632

Main Research Activities

Mobile Cloud Lab

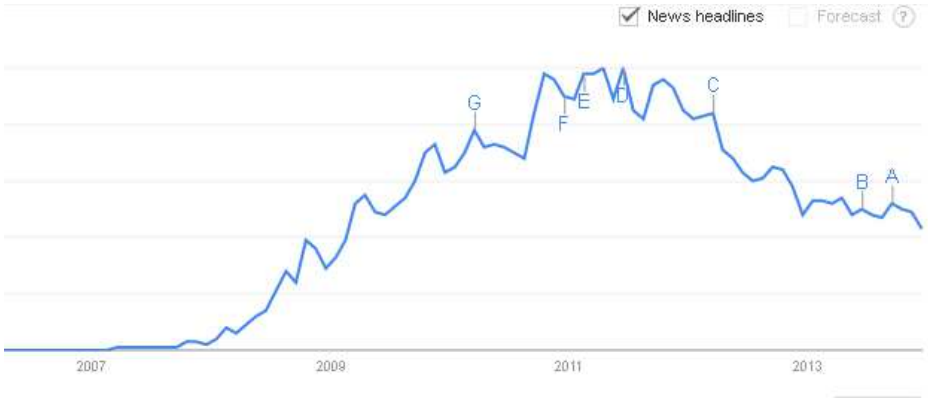
UNIVERSITY OF TARTU
Institute of Computer Science
1632



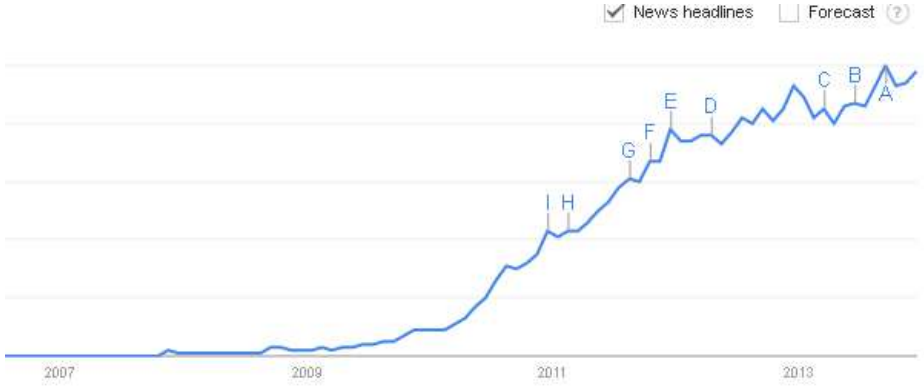
Outline

- Cloud computing
- Migrating enterprise/scientific applications to the cloud
- Adapting computing problems to the cloud
- Mobile Cloud

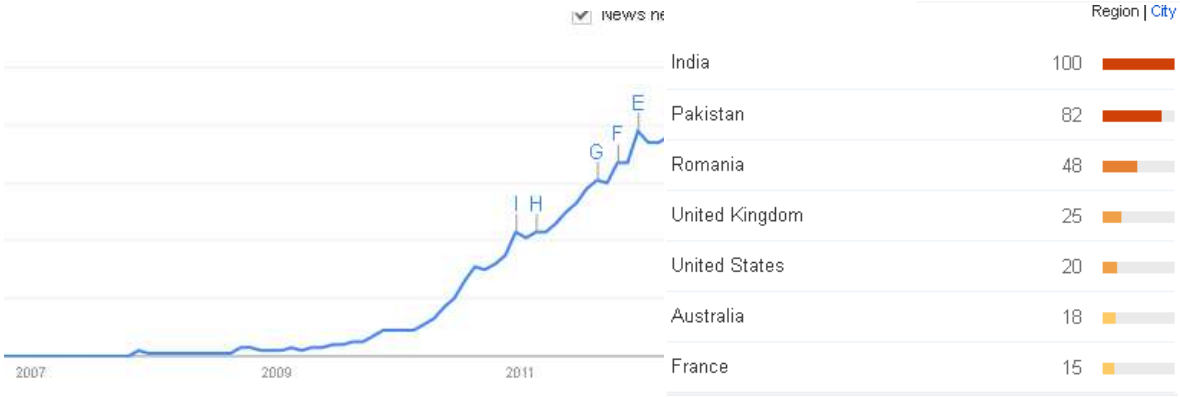
Some Recent Trends



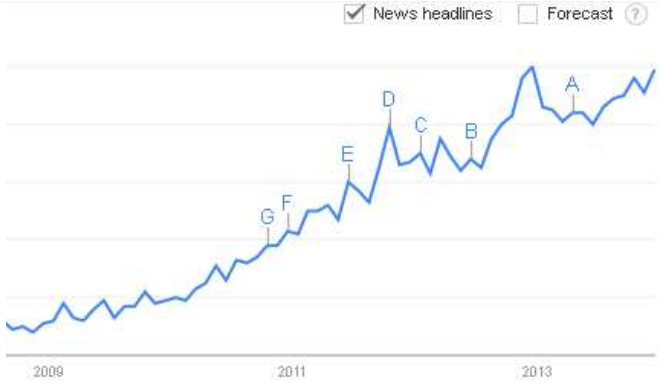
Cloud Computing



Smart Phone



Android



Mobile Cloud

What is Cloud Computing?

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

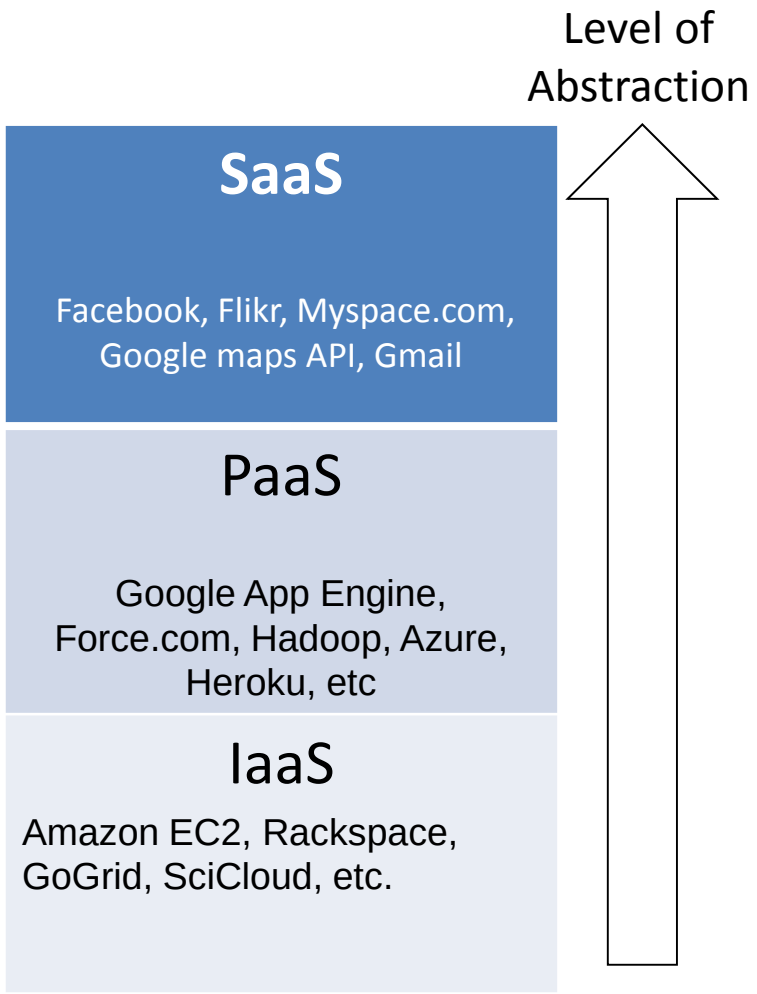
1969 – Leonard Kleinrock, ARPANET project

- "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

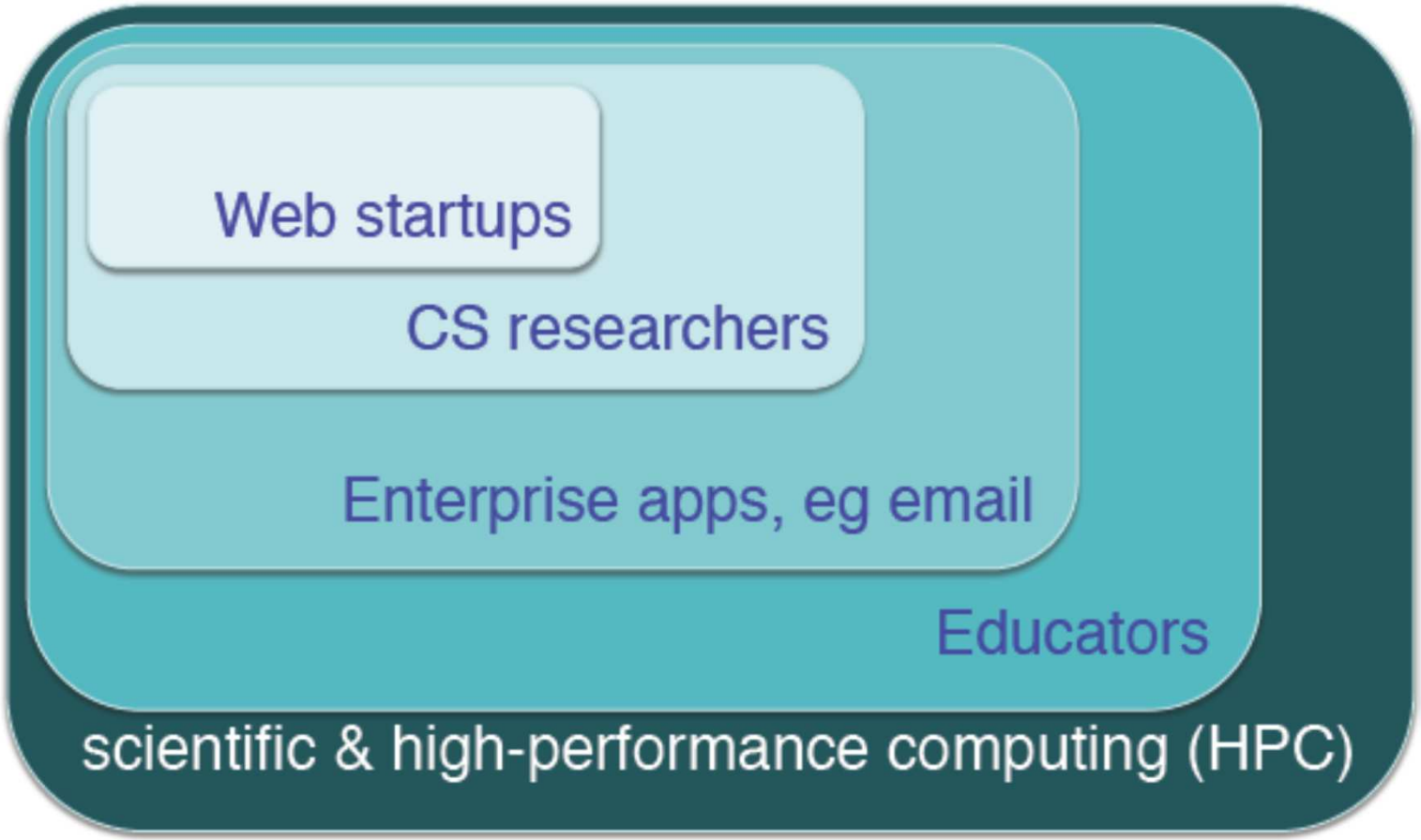


Cloud Computing - Themes

- 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 SCIENTIFIC/ENTERPRISE APPLICATION TO 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

Desktop to Cloud Migration

- Non computer-scientists do not have significant knowledge of computer science, clouds and migration procedures
  - They are only interested in submitting an experiment and collecting the results after some time.
- D2CM is an open source tool developed as part of EU FP7 projects REMICS & SITIO
  - REMICS - “Reuse and Migration of Legacy Applications to Interoperable Cloud Services”
  - SITIO – „Semantic Business Processes based on Software-as-a Service and Cloud Computing “
- Enables scientists to migrate their computational experiments running on their local desktops to the cloud

http://mc.cs.ut.ee/mcsite/projects/desktop-to-cloud-migration

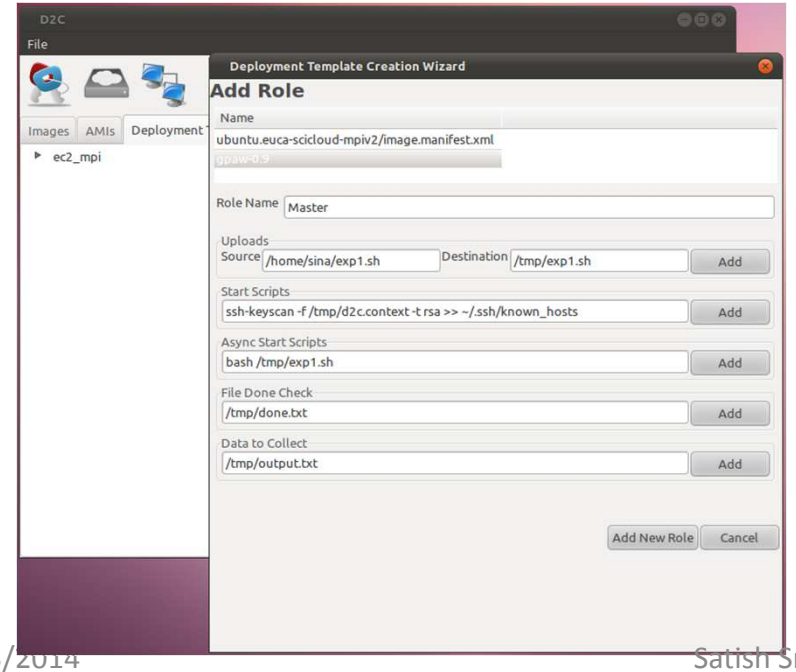
D2CM - continued

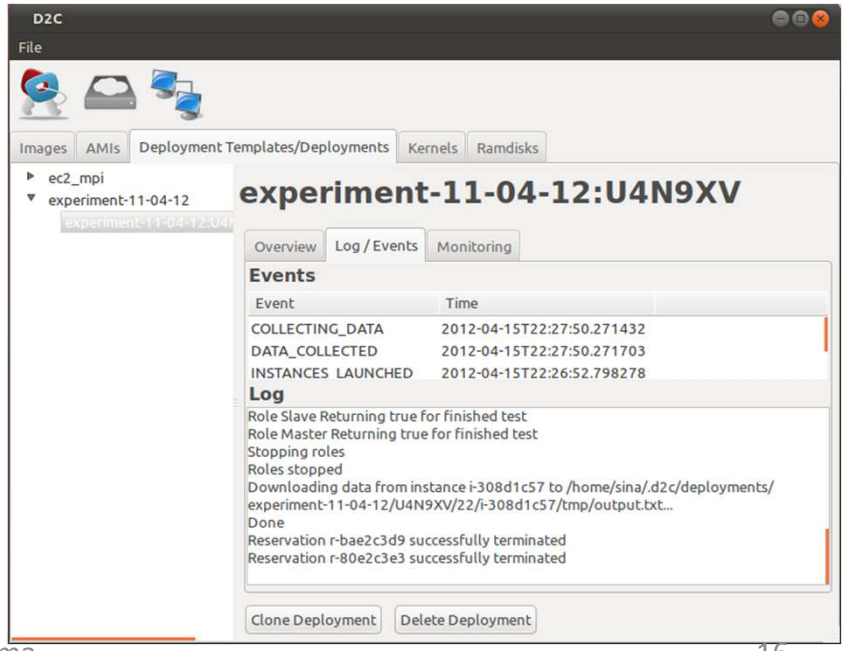
- Seamless migration of desktop virtual machine images to the Cloud
  - Transform local VM images into cloud compatible VM images
    - Extract the file system, package kernels, Install additional software etc.
  - Move it to the target cloud
- Create deployment template to describe the configuration
  - Define roles
    - Instance type, Number of instances
  - Define actions for each role
    - Uploads, Initialization commands, Run commands, Deployment ending conditions, Downloads
- Describe experiment once, run anywhere
  - Reuse the deployment template to generate new experiments and change the parameters as needed

[Srirama et al, HPCS 2013]

D2CM - continued

- Used extensively by the electrochemists for migrating their experiments
  - Running computational experiments to improve the design of supercapasitors
    - by simulating and studying room temperature Ionic liquids (RTIL)



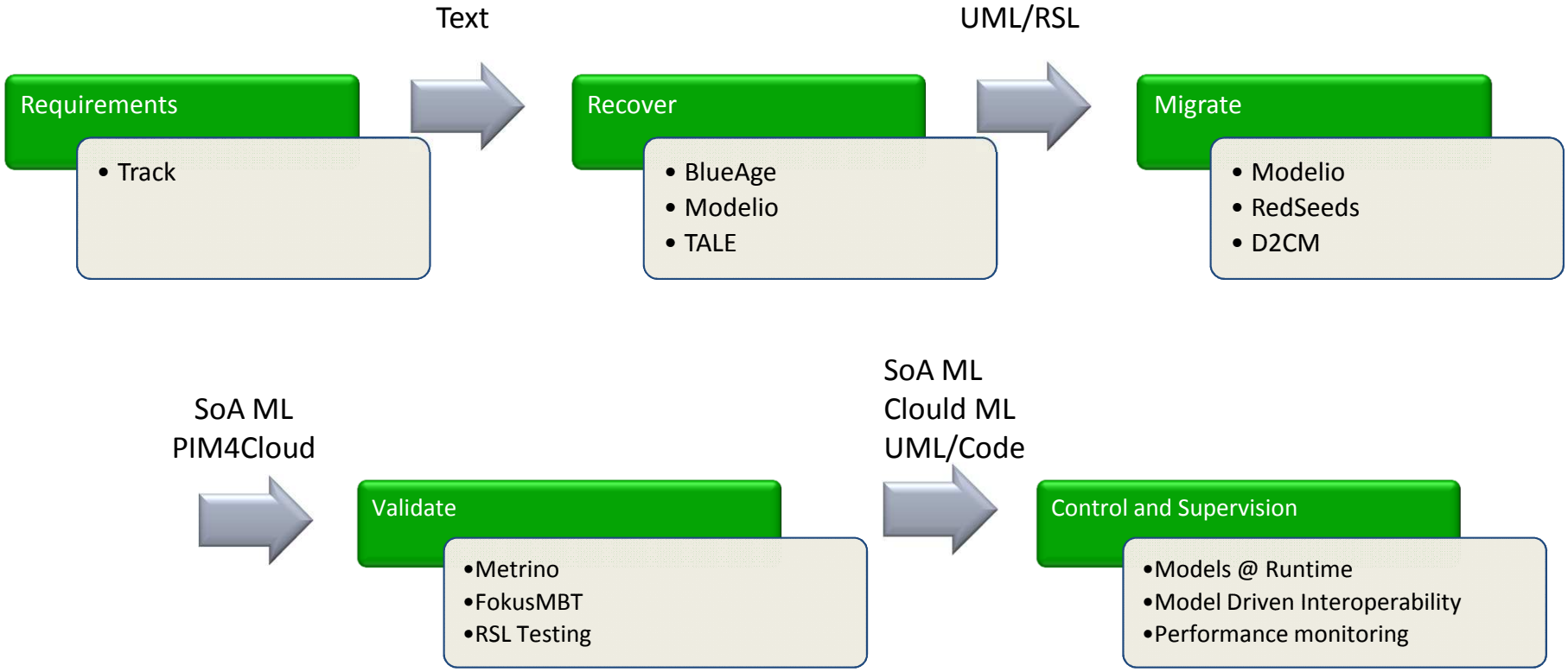


[Srirama et al, HPCS 2013]

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

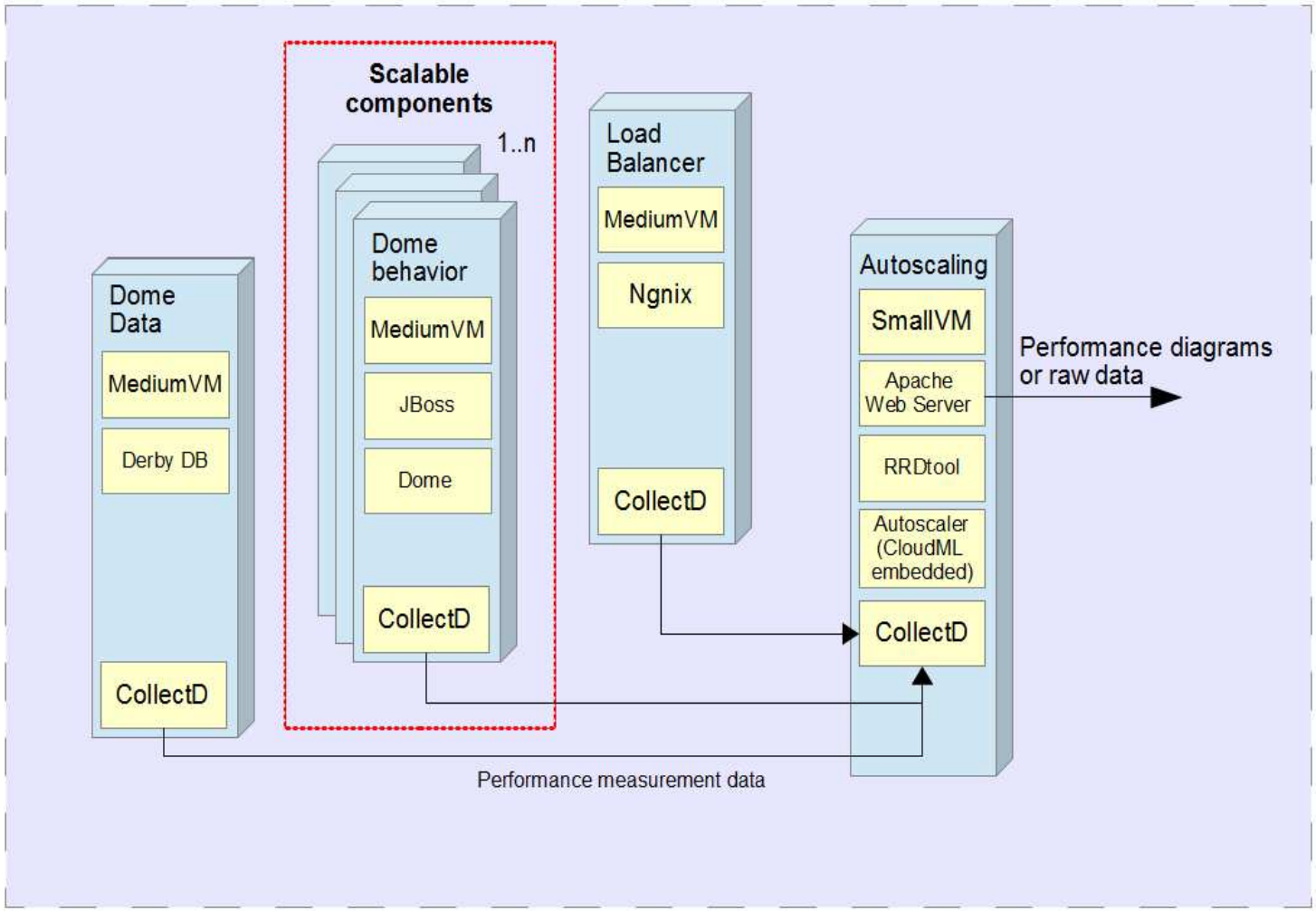
```
"nodeTypes": [
    {
        "id": "SmallGNULinux",
        "os": "GNULinux",
        "compute": [ 2, 4 ],
        "memory": [ 2048, 4096 ],
        "storage": [ 10240 ],
        "location": "eu",
        "provides": [
            { "id": "SSHCapability" }
        ]
    }
]

"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]

Control and Supervision

- Enables REMICS migrated applications to exploit cloud advantages, such as:
  - Elasticity
  - On-demand and in real-time resource provisioning
- Introduced automatic load-based scaling to the migrated application deployment model
  - Avoid vendor lock-in to any single cloud provider

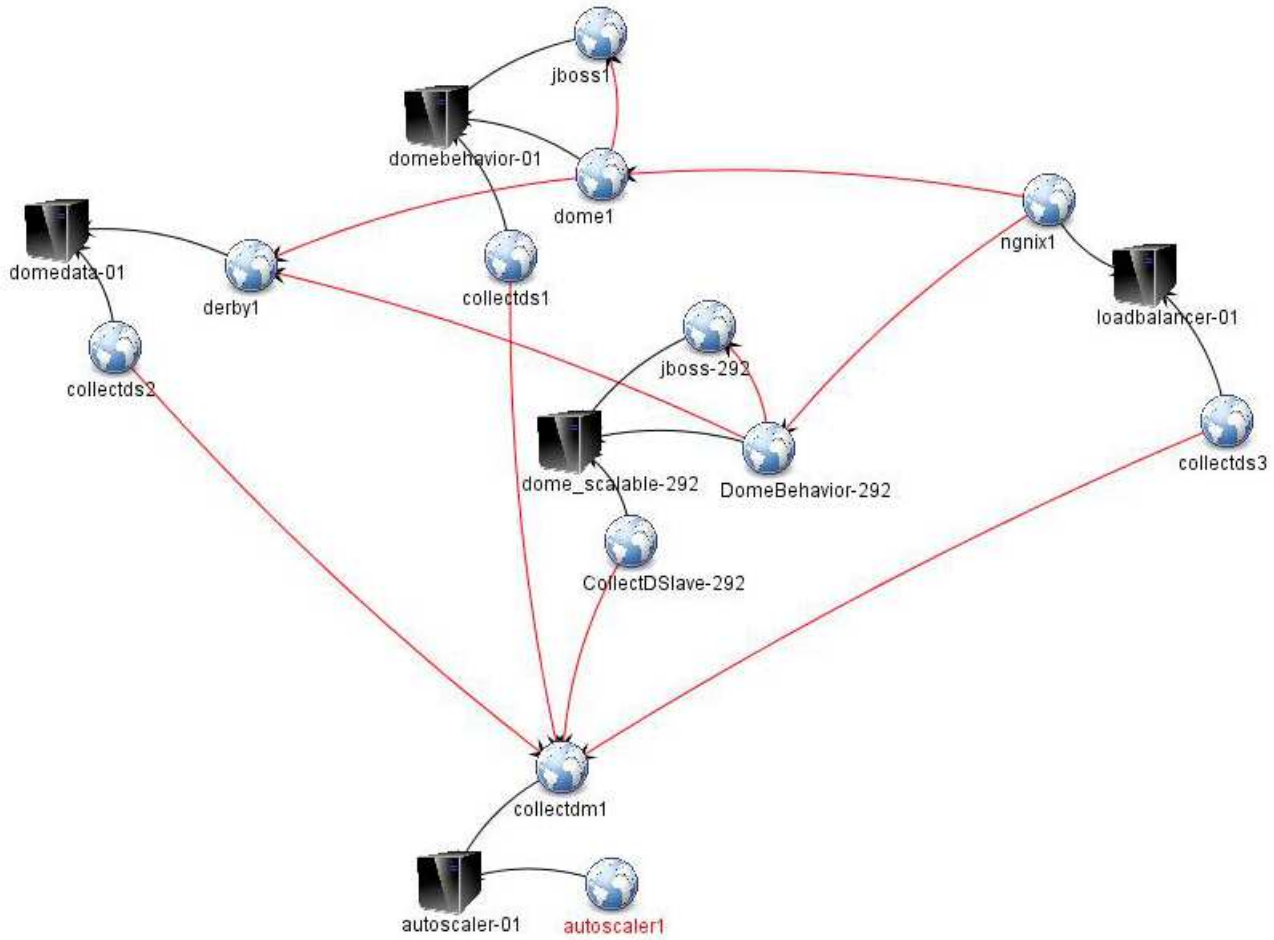
Performance Monitoring + Autoscaling



Control and Supervision - continued

- Autoscaler Java daemon (UT)
  - CollectD – collecting distributed performance metrics
- CloudML engine and API (SINTEF)
  - Modifying CloudML models
  - Re-deploying the modified model
- CloudML Bridge (SINTEF & UT)
  - Generating autoscaling configuration from the CloudML model
- RRDtool - generating visual performance graphs
- Apache web server – displaying performance results

Visualizing CloudML Model



Research Challenges

ADAPTING COMPUTING PROBLEMS TO THE CLOUD

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

Typical Large-Data problem

Map

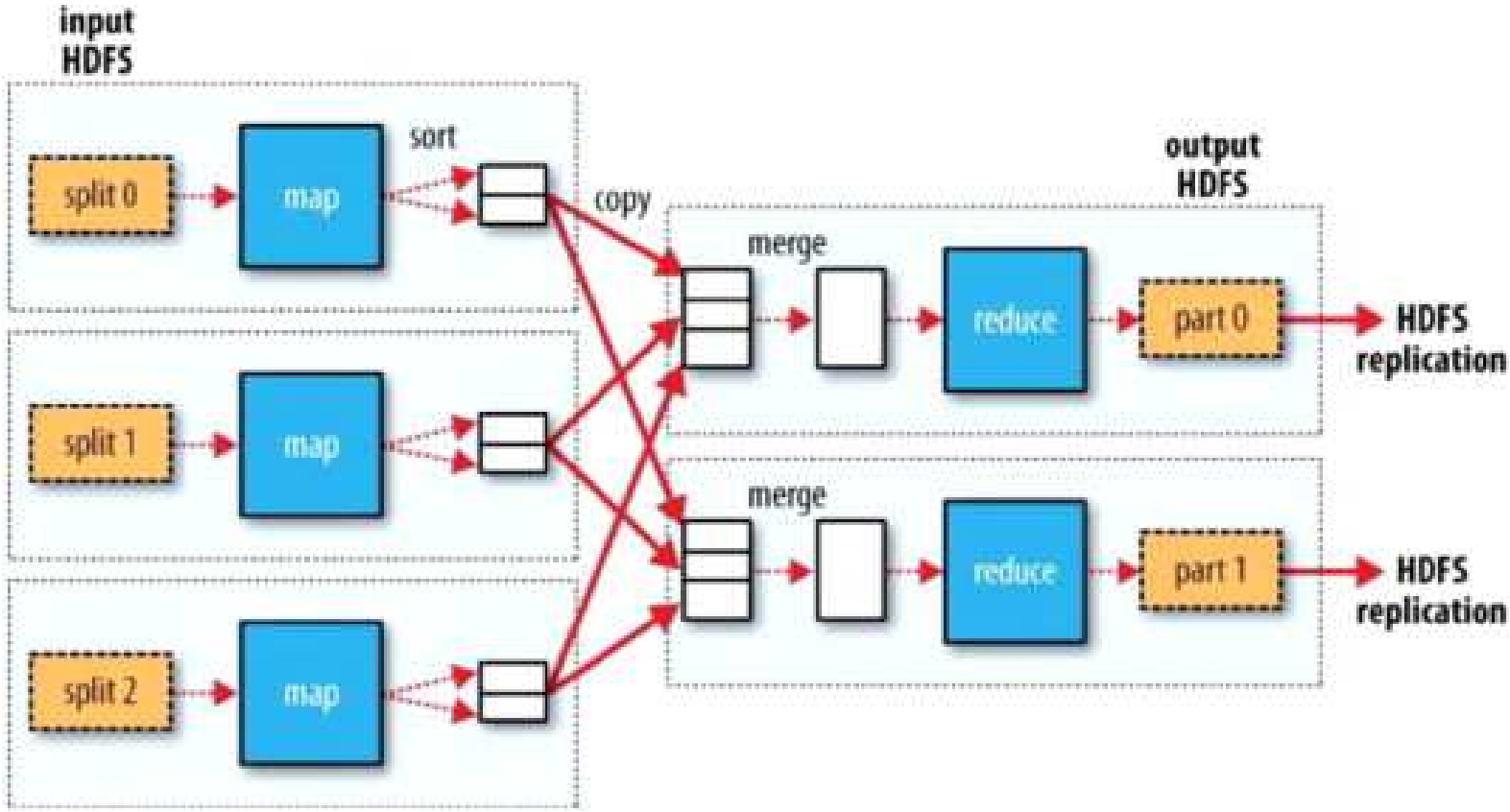
- Iterate over a large number of records
- Extract something of interest from each
- Shuffle and sort intermediate results

Reduce

- Aggregate intermediate results
- Generate final output

Some material adapted from slides by Jimmy Lin, Christophe Bisciglia, Aaron Kimball, & Sierra Michels-Slettvet, Google Distributed Computing Seminar, 2007 (licensed under Creation Commons Attribution 3.0 License)

MapReduce model



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

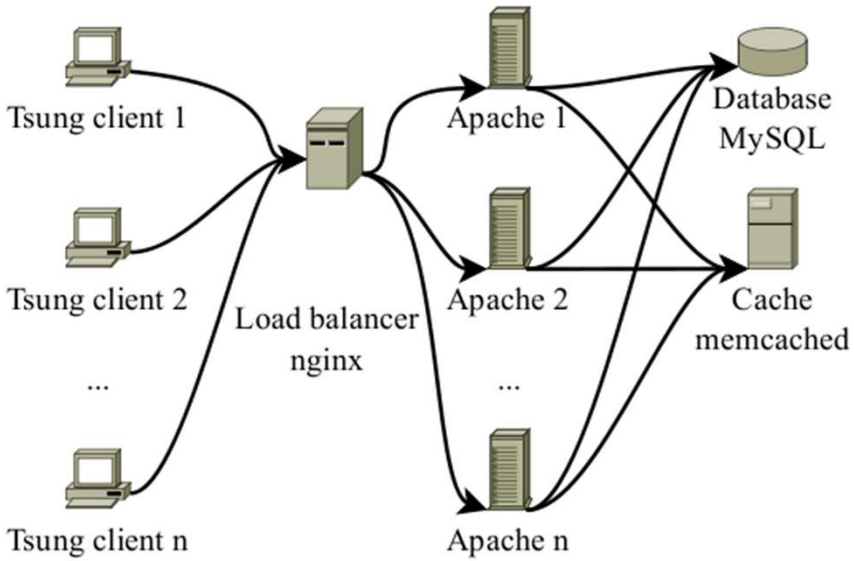
- Restructuring algorithms into non-iterative versions
  - CLARA instead of PAM [Jakovits & Srirama, Nordicloud 2013]
- Alternative MapReduce implementations that are designed to handle iterative algorithms
  - E.g. Twister [Jakovits et al, ParCo 2011], HaLoop, Spark
- Alternative distributed computing models
  - Bulk Synchronous Parallel model [Valiant, 1990] [Jakovits et al, HPCS 2013]
  - Building a fault-tolerant BSP framework (NEWT) [Kromonov et al, UCC 2013]

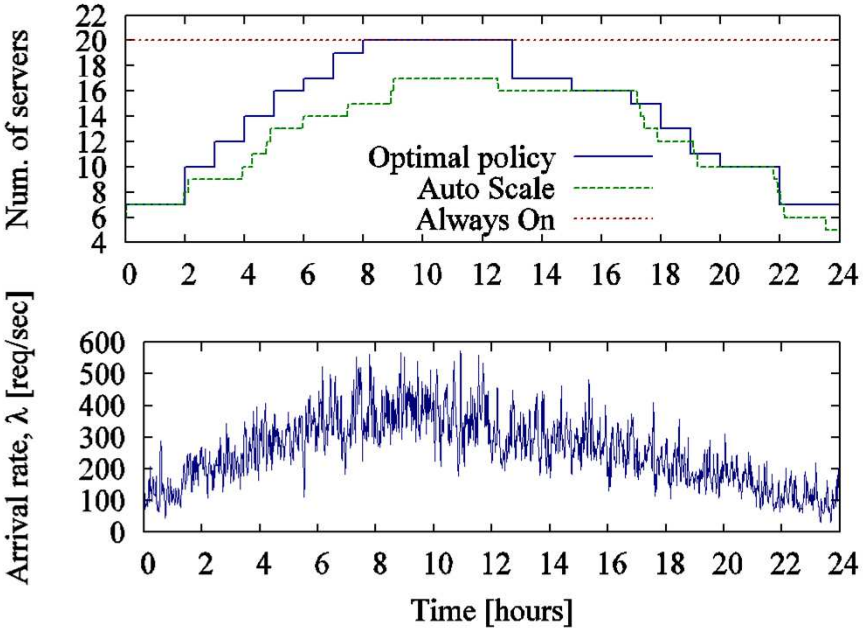
Remodeling Enterprise Applications for the Cloud

- Remodeling workflow based applications for the cloud
  - To reduce communication latencies among the components
  - Intuition: Reduce inter-node communication and to increase the intra-node communication
- LP based mathematical models to find ideal deployment configuration [Paniagua et al, iiWAS 2011]
  - Based on the loads and regions

Monitoring and Testing Web Application Scalability on the Cloud

- WAACS framework
  - Supports to run and setup experiments in the cloud





[Vasar et al, Nordicloud 2012]

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

1/ + Comment Now + Follow Comments

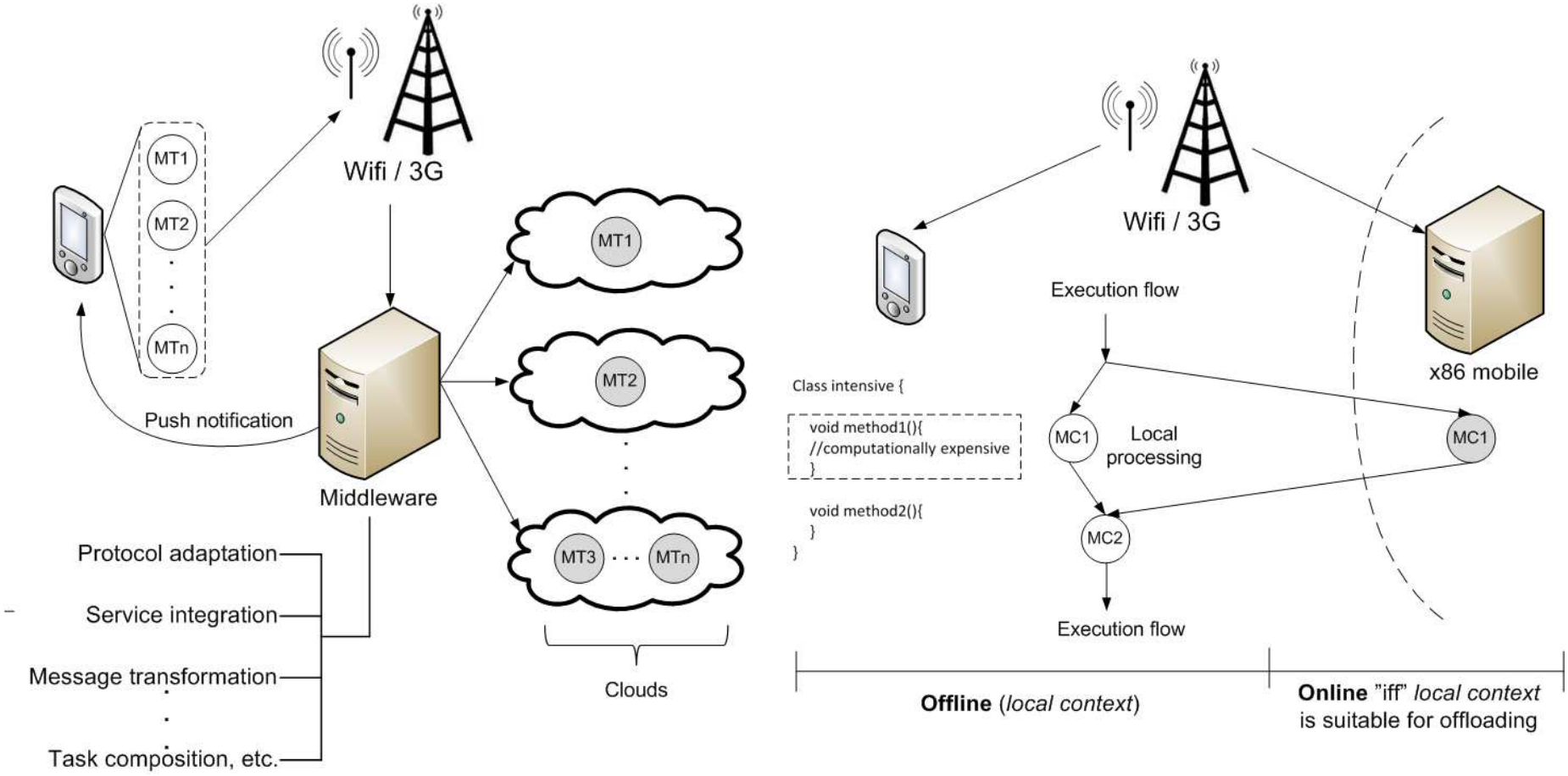
Mobile Applications

- 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 2013]

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

Mobile Cloud Binding Models

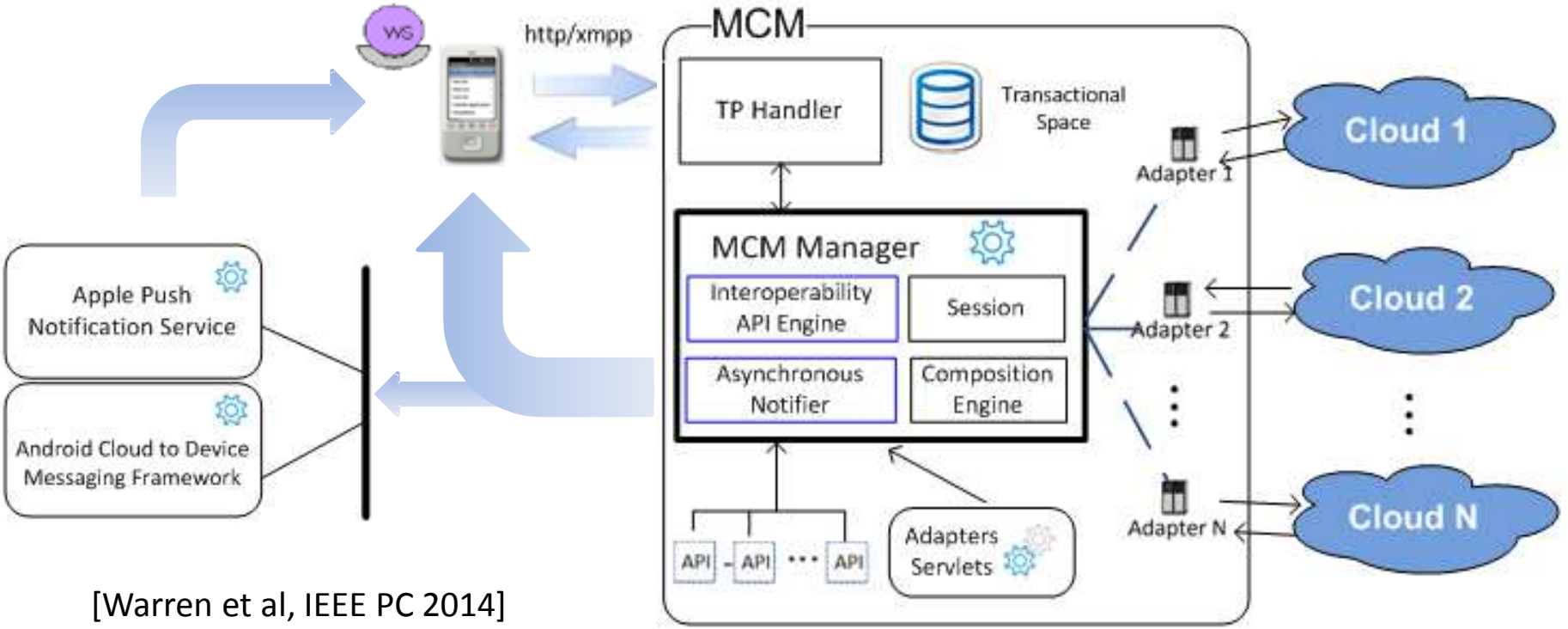


Task Delegation **Code Offloading**

[Flores & Srirama, JSS 2013]

Mobile Cloud Middleware

[Srirama and Paniagua, MS 2013]



[Warren et al, IEEE PC 2014]

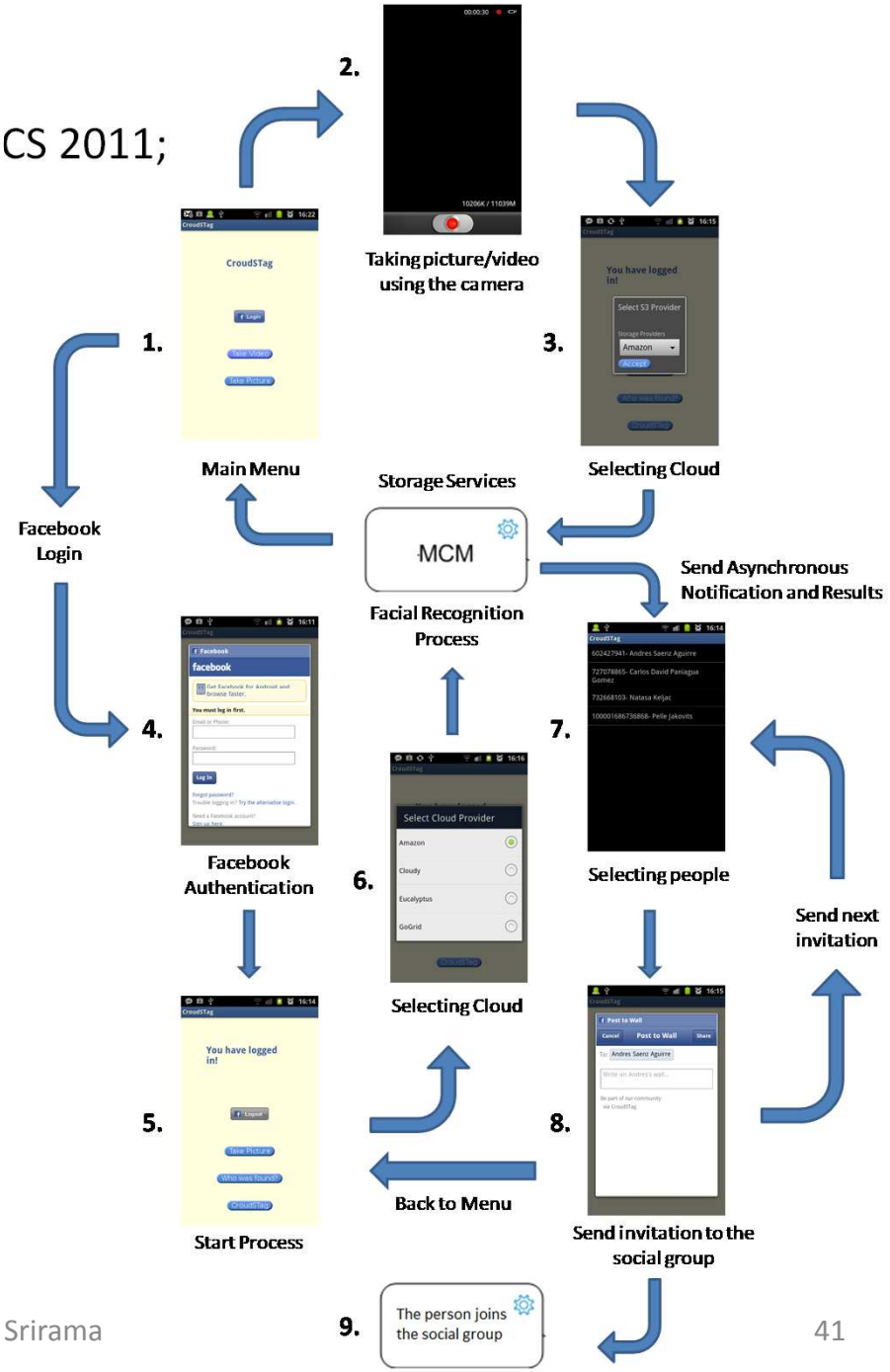
[Flores et al, MoMM 2011; Flores and Srirama, JSS 2013]

CroudSTag – Scenario

- CroudSTag takes the pictures/videos from the cloud and tries to recognize people
  - Pictures/Videos are actually taken by the phone
  - Processes the videos
  - Recognizes people using facial recognition technologies
- Reports the user a list of people recognized in the pictures
- The user decides whether to add them or not to the social group
- The people selected by the user receive a message in facebook inviting them to join the social group

CroudSTag [Srirama et al, PCS 2011; SOCA 2012]

- 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

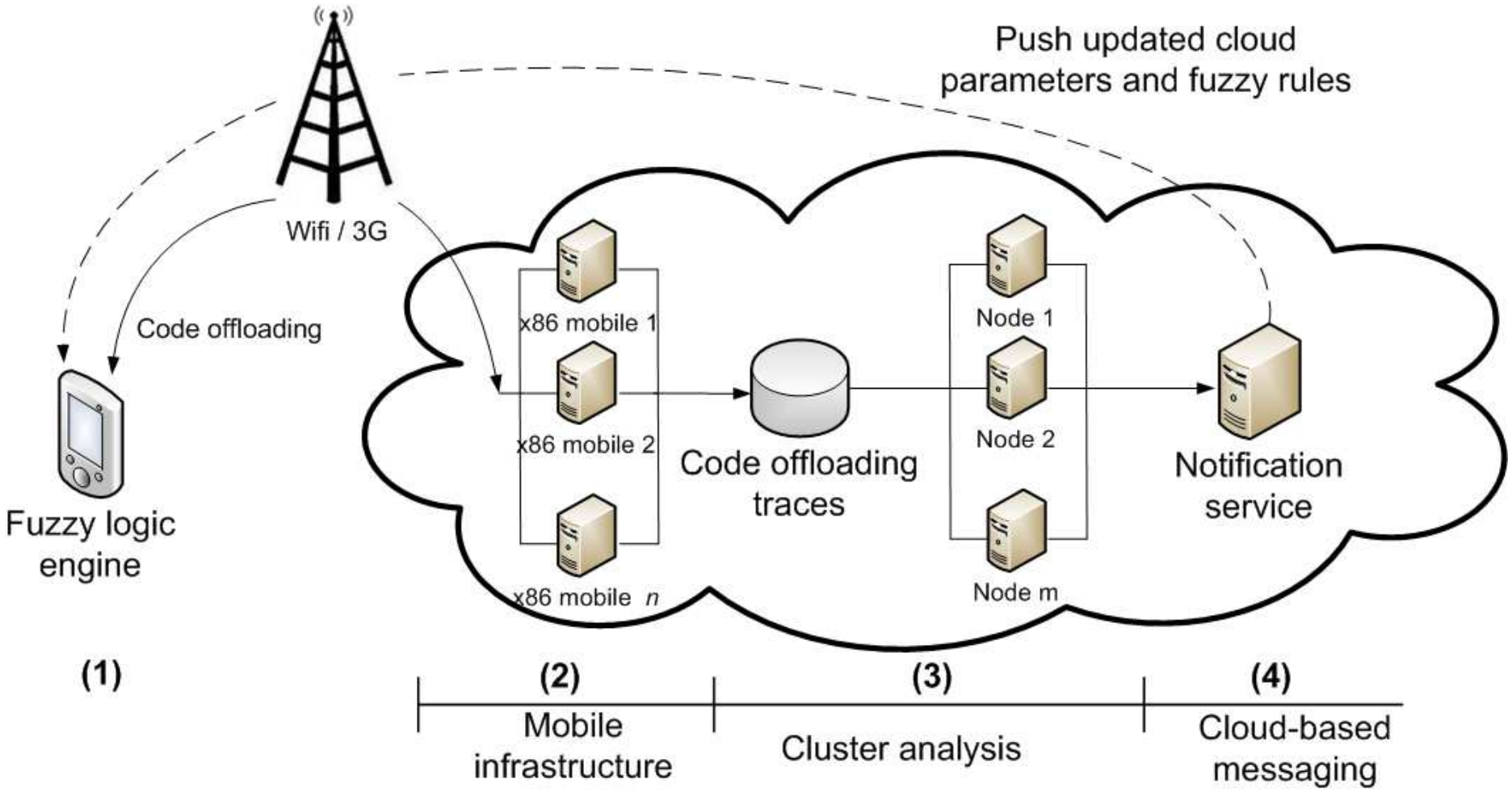


Code Offloading

- Studied extensively by community [MAUI, Cloudlets etc.]
- Is *Mobile Cloud* taking full advantage of *Cloud Computing*?
  - Parallelization and elasticity are not exploited
- Offloading from a different perspective
  - *"Offloading is a global learning process rather than local decision process"* [Flores and Srirama, MCS 2013]
- How it can learn?
  - Analysis of code offloading traces which are generated by the massive amount of devices that connect to cloud

"EMCO: Evidence-based mobile code offloading"

Evidence-based Mobile Code Offloading



[Flores and Srirama, MCS 2013]

Process-intensive Tasks on Cloud

- Media processing
  - CroudSTag demonstrates image and video processing
- Sensor data analysis
  - Human activity recognition [Srirama et al, NGMAST 2011]
  - Context aware gaming
  - MapReduce based sensor data analysis [Paniagua et al, MobiWIS 2012]

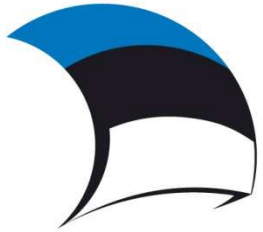
Research Results

- 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, ZeroTurnaround etc.

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

e-ESTONIA

European Union
Regional Development Fund



Investing in your future

Eesti Teadusagentuur
Estonian Research Council

srirama@ut.ee

THANK YOU FOR YOUR ATTENTION