AWARDED STSM GRANTS IN THE SECOND GRANT PERIOD

CHIPSET COST

Short Term Scientific Mission

Data Fusion In Cloud Computing

Piotr Szuster

Cracow University of Technology | retsuz@gmail.com

Short Term Scientific Mission

COST ACTION: cHiPSet, High-Performance Modelling and Simulation for Big Data Application

TOPIC: Data Fusion In Cloud Computing

Working Groups related: WG1

Home Institution: Cracow University of Technology

Host Institution: Carlos III University of Madrid

Problem definition

- Timely acquisition and processing of data from different sources and extraction of accurate in-formation plays an important role in many realistic scenarios (such as emergency situations, evacuation systems, crowd management, remote health monitoring, etc.).

- The first work in this scenario is the data fusion of the information received from the multiples sources (sensors), this problem is known as sensor fusion. Sensor fusion problems have been widely analyzed in Air Traffic Control scenarios with multiple sensors and multiple targets.

Acheivements

- The development of the COST Action IC1406 WG1 research topics include: novel het-erogeneous models, algorithms and techniques for advanced Big Data exploitation, based on the current and emerging multicore system architectures, virtualised servers and data centres, mobile cloud and multi-cloud systems. The project subject is relevant to the WG1 research topics. The developed materials will be used in the work of Actions' group.

Achievements

- DFM as internal component of cloud system - the module that will support dis-tributed or hybrid model of fusion process, asynchronous communication model with event-driven flow control.
- DFM as external component of cloud system - the module that will support centralized model of data fusion process, that will be connected to the cloud system through the Internet.
- Theoretical architectures for whole cloud –DFM system
- Preparation and evaluation of achieved results in the form of the paper for chosen relat-ed International Conferences or Journals with impact factor.

Thank you!

Optimized haplotying based on whatshap

Beat Wolf
HES-SO Fribourg

The problem

- Through the analysis of next generation sequencing (NGS) data, differences between the a sample and the general population can be found.
- For heterozygous variants, it is unknown how they are related to each other, meaning, which variants belong on the same strand.
- To solve this problem, sequences overlapping multiple variant positions are used
- Due to the large data-sizes produced by NGS, the computational complexity of the problem is very large

Whatshap

- Whatshap is an algorithm which solves the haplotyping problem through the use of dynamic programming
- The algorithm has a complexity of O(2^n), where n is the maximum coverage in the sample, which heavily limits its usage
- Instead of using heuristics like other algorithms, whatshap finds an exact solution

pWhatshap

- The original whatshap algorithm is a sequential algorithm
- pWhatshap is parallel implementation of the whatshap algorithm
- The problem is parallized on the individual columns
- Due to the uneven repartion of the problem across the columns, the speedup using this approach is very limited

STSM work

- During the STSM, new ways to speed up the algorithms have been explored
- Three ways have been identified, with three of them having been implemented in a prototype
- The first one, which allowed to reduce the problem size by up to 70% on the sample datasets, reduced the dataset by identifying and removing redundant information
- The second one found new ways to split the dataset into independent parts, thus allowing for a higher parallelization
- The third one identified new ways to parallelize the calculations even more

Outcome

- The identification of multiple ways to speed up the whatshap haplotying algorithm
- The implementation of prototypes showing the potential of those approaches
- Scientific collaboration between the University of Stirling and the HES-SO Fribourg
- Plans to further collaborate on the subject through a joint research project

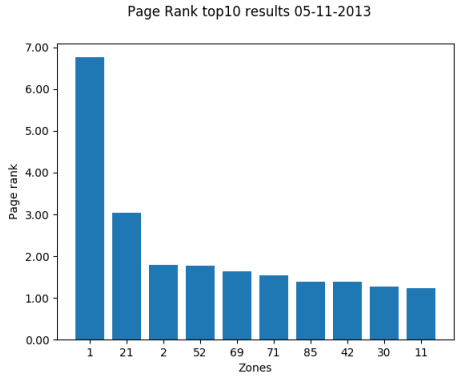
Big Data frameworks for processing mobile phone data

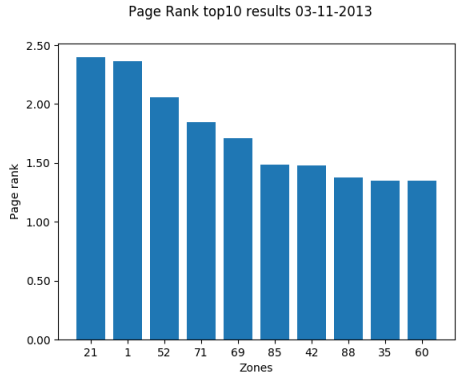
STSM Olivera Novović, Apostolos N. Papadopoulos

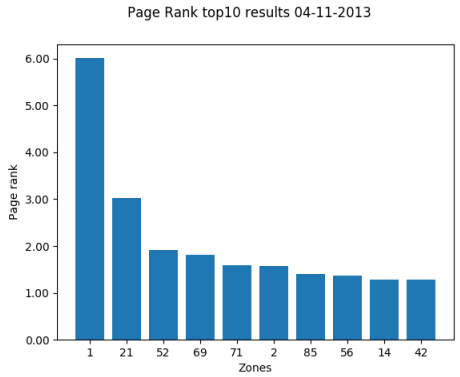
Summary of activities

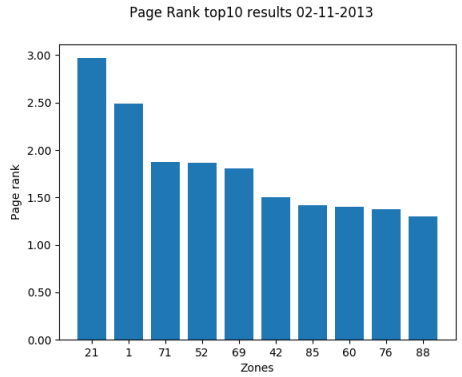
- Introduction to Apache Spark and programing language Scala
- Introduction to Spark's GraphX library for graph mining
- Discussion of possible algorithms that could be applied on mobile phone data graphs
- Installation and setting up IntelliJ IDEA environment for Scala
- Generating top10 page rank results using Scala and Spark
- Discussion about modularity based algorithms for community detection
- Introduction to Louvain modularity based algorithm
- Application of Louvain algorithm to mobile phone data graphs in Python and Spark
- Generating communities using Louvain algorithm

Top10 page rank results



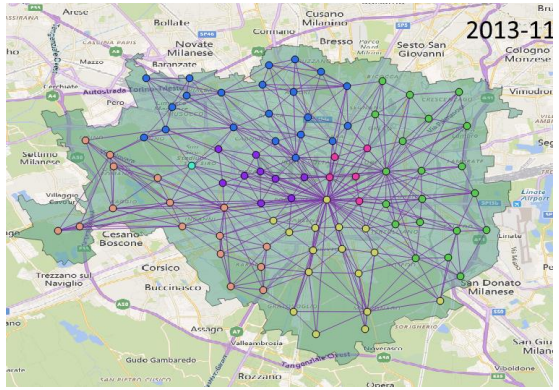


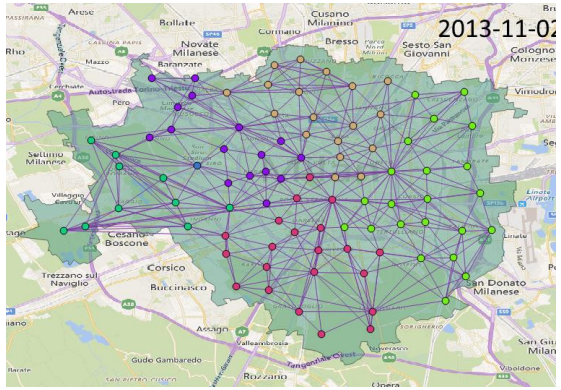


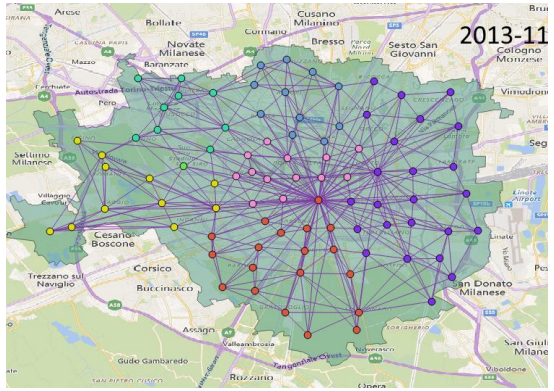


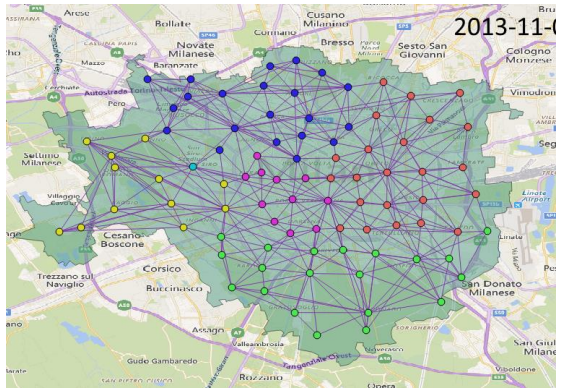
- Here are presented some of the top10 page rank results
- The results were generated for all graphs day by day
- The results were generated using Spark
- After this processing we were able to track evolution of page rank for specific nodes

Community detection









- Here are presented results for community detection for some days, different graphs
- Communities are detected using Louvain modularity based algorithm in Spark
- Change detection of communities on daily bases

Next steps

- For future community detection we should consider dataset were the links are not aggregated
- With higher resolution of data we could detect communities with better accuracy
- Using Spark on server with 40 cores, or on a cluster, we could process original dataset in more efficient way
- In our future analyses we aim to include element of spatial distribution of nodes in graph mining algorithm

Enjoying my time in Thessaloniki

From the top of White Tower

Presenting my work at DeLab

Sunny day in Thessaloniki, Nea Paralia

Agia Sofia in Thessaloniki

Designing a blockchain query system

STSM Livio Pompianu

Home institution: Università di Cagliari
Host institution: University of Stirling

Blockchains

A blockchain is a distributed database in which it is not possible neither modify nor remove data. The blockchain size continuously grows.

The first system that uses a blockchain is the Bitcoin cryptocurrency (2008). The Bitcoin blockchain is a list of blocks: each one contains a list of currency transactions moving funds between Bitcoin addresses.

Today blockchains are also used for recording events, tracking physical and digital assets, storing and executing decentralized programs.

Scalability issues

Blockchains are rich sources of data for researchers, that usually develop an own tool for extracting and analysing data.

However, the increasing size of the blockchain, causes scalability issues performing blockchain analysis.

The Bitcoin blockchain size increased from 50GB to 100GB the last year.

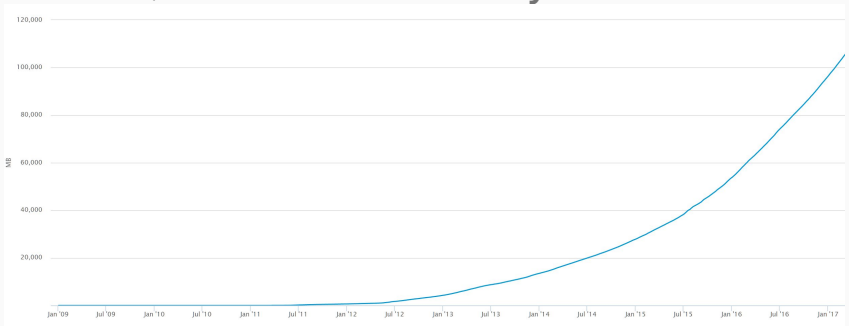


Figure: Bitcoin blockchain size

A blockchain query system

My STMS lays the basis for the development of a blockchain query system, a valuable resource for researchers that wish to analyse blockchains.

This work will help researchers in several ways.

1. It will avoid to build a new ad-hoc tool for each new analysis.
2. It will provide a general blockchain model abstracting from any blockchain.
3. It will offer an efficient way for querying blockchains.

A blockchain query system

We discussed a blockchain model, defining the entities that researchers can manipulate with their queries.

We discussed a minimum set of queries that must be supported by our tool.

We discussed a tool for compiling and executing queries efficiently.

Main results

I gave a seminar at the University of Stirling, presenting blockchain research topics and my STMS. Since the topics discussed are interdisciplinary, my seminar opened new collaboration perspectives.

This STSM contributes to the development of a longer-term research collaboration on the topic between my research group in the University of Cagliari (led by Dr. Massimo Bartoletti) and the research group of the University of Stirling (led by Dr. Andrea Bracciali).

We are extending the work and writing a paper describing the results.

SHORT TERM SCIENTIFIC MISSION: DEPLOYMENT AND PROVISION OF CLOUD-BASED SERVICES DEPENDING ON INTERNET OF THINGS (IOT) DEVICES

ELLI RAPTI*

*Applied Informatics Laboratory, Department of Furniture Design and Technology, TEI of Thessaly, Karditsa, Greece, elli.rapti@gmail.com

SHORT TERM SCIENTIFIC MISSION

- **COST Action:** cHiPSet, *High-Performance Modelling and Simulation for Big Data Applications*
- **Topic:** Deployment and provision of cloud-based services depending on Internet of Things (IoT) devices
- **Working Groups related:** WG1
- **Home Institution:** TEI of Thessaly, Karditsa, Greece
- **Host Institution:** Cracow University of Technology, Krakow, Poland

Cracow University of technology

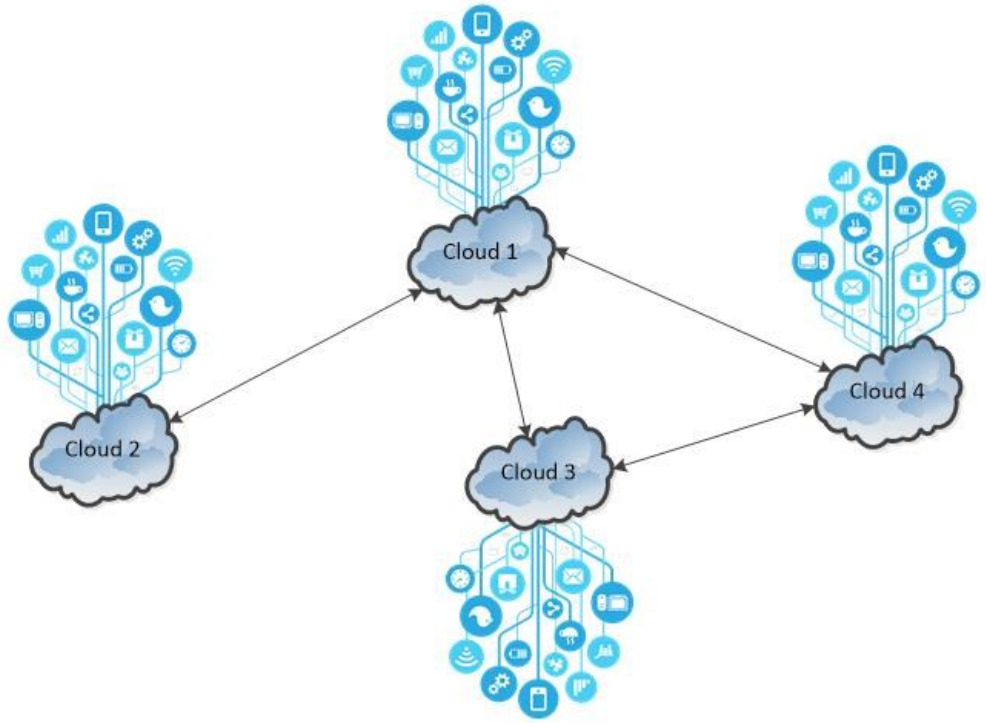

Krakow Old Town Square

STSM: WORKPLAN

- The aim of STSM was to deepen the knowledge on the deployment and provision of cloud-based services for Internet of Things (IoT) applications.
- The workplan was realised as follows:
  - Formal definition of the problem
  - Literature review of bio-inspired approaches
  - Establishment of subsequent collaboration
- Deliverables:
  - Report of the formal definition of the problem
  - Literature review report
  - Paper outline

STSM: PROBLEM DEFINITION

- IoT devices (mobile phones, sensors) provide their functionality as a service s on a cloud service base c
- Clouds are interconnected (preestablished agreements) creating a distributed network $G = (C, E)$
  - C → set of all cloud service bases
  - E → set of links among cloud service bases that can communicate directly
- Each cloud service base has partial view of the network (neighborhood) N^c
- IoT services can be discovered and composed
- There is no central authority:
  - Decentralized approach for discovery of service compositions



STSM: NEXT STEPS

- Finalization of the problem formulation
- Detailed description of the solution approach:
  - System design model
  - Description of the honey bee inspired algorithm for discovery of cloud-based service compositions
  - Service selection
- Development of a custom built simulator
- Experimentation
- Final corrections
- Paper deadline: June 2017

THANK YOU!

STSM Grant Report

COST STSM Ref. Number: COST-STSM-ECOST-STSM-IC1406-050217-081940
COST Action: IC1406
STSM type: Regular
Grant Period (exact dates): 2017-02-05 to 2017-02-17

Implementation of the H-matrices in order to reduce a numerical complexity of the boundary integral approaches applied for the high-resolution Earth's gravity field modelling

***Applicant Name*: Róbert Čunderlík**

***Host Name*: Prof. Francesca Vipiana**

Home Institution: ***Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Slovak University of Technology, Bratislava, Slovakia (SK)***

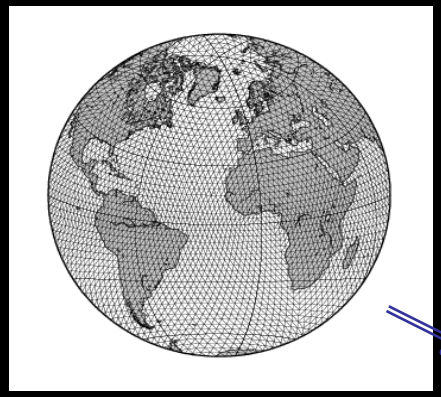
Host Institution: ***Department of Electronics and Telecommunications (DET), Politecnico di Torino, Turin, Italy (IT)***

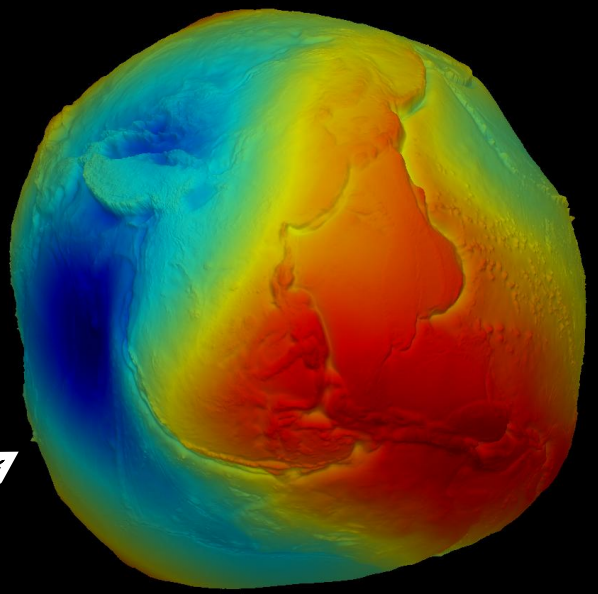
***cHiPSet Working Group related*:**
WG4: HPC-enabled Modelling for Socio-Economical and Physical Sciences

Boundary integral approaches for the Earth's gravity field modelling

BEM for numerical solutions of the geodetic boundary-value problems

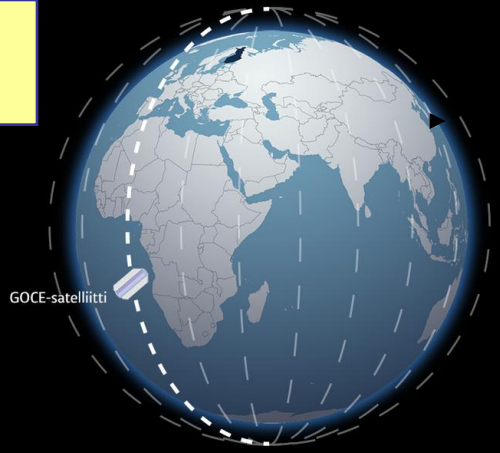
- processing terrestrial gravimetric data





MFS or SBM for processing the GOCE satellite data

- the GOCE satellite mission (ESA)



!!! Enormous memory requirements !!! ⇒ ***reduction using the H-matrix compression techniques***

STSM workplan and pre-phase

Proposed workplan

- An introduction of the BEM, MFS or SBM approaches for gravity field modelling to Experts from the HI including a detailed study of the algorithms
- An advanced introduction of the H-matrices for the elliptic partial differential equation by Experts from the HI to Applicant
- A proposal of an optimal strategy to implement the H-matrices taking into account properties of the system matrices
- An implementation of the H-matrices into the existing algorithms and programming
- Debugging programs
- Testing of efficiency and accuracy on artificial experiments

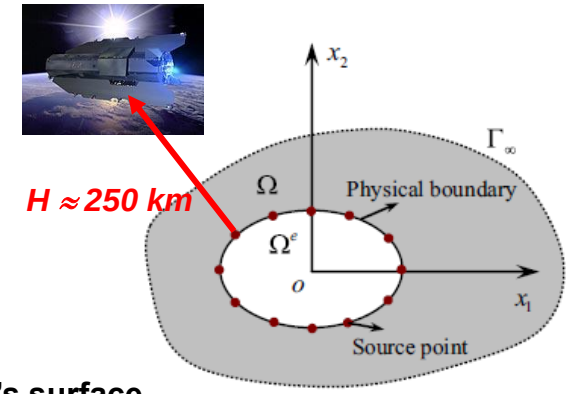
Steps before the STSM visit

- Host's revision of the boundary integral approaches used by the Applicant's research group for the high-resolution Earth's gravity field modelling
- Applicant's study of a theoretical background of the H-matrix and H^2-matrix compression techniques, namely of the Adaptive Cross Approximation (ACA) and Nested Skeletonization Scheme (NSS)

Implementation of the ACA algorithm into the MFS approach

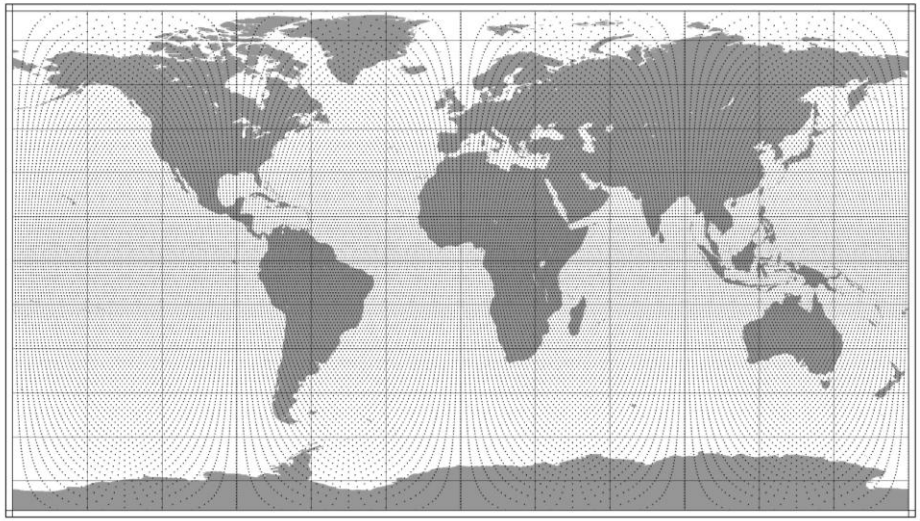
MFS for simulated example

- reconstruction of the gravitational potential generated by a spherical approximation of the Earth defined by its radius R and geocentric gravitational constant GM.



Developing program in MATLAB

- Source points regularly distributed on a spherical approximation of the Earth's surface
- Observation points with input data generated at altitude 250 km above the sphere
- Geometrical partitioning of the source points as well as observation points into clustered groups (not yet optimal for spherical domains, just very simple for this simulated experiment)
- Treating relations between the created groups by defining the far field and near field 'neighbours' for each group
- Near field: matrix coefficients computed using the standard MFS
- Far field: an implementation of the ACA algorithm
- Debugging program
- Testing efficiency and accuracy



Implementation of the ACA algorithm into the MFS approach

ACA algorithm

- **numerically rank-deficient sub-blocks, which correspond to interactions of well-separated groups, efficiently compressed through an approach very similar to the column-pivoted LU decomposition**
- **an approximation of every original sub-block by a product of two dense rectangular matrices whose dimensions derived from an effective rank of the original sub-block and for a given tolerance**
  - ***approximating the sub-block*** $\mathbf{Z}^{m\times n}$ by $\tilde{\mathbf{Z}}^{m\times n}$

$$\tilde{\mathbf{Z}}^{m\times n} = \mathbf{U}^{m\times r}\mathbf{V}^{r\times n} = \sum_{i=1}^{r} \mathbf{u}_i^{m\times 1}\mathbf{v}_i^{1\times n}$$

- ***memory compression from*** $O(m \times n)$ ***to*** $O(r(m+n))$

while $r << m$ or n

Initialization:

1) Initialize the 1st row index $I_1 = 1$ and set $\tilde{\mathbf{Z}} = 0$.
2) Initialize the 1st row of the approximate error matrix: $\tilde{\mathbf{R}}(I_1,:) = \mathbf{Z}(I_1,:)$.
3) Find the 1st column index $J_1 : |\tilde{\mathbf{R}}(I_1,J_1)| = \max_j(|\tilde{\mathbf{R}}(I_1,j)|)$.
4) $\mathbf{v}_1 = \tilde{\mathbf{R}}(I_1,:)/\tilde{\mathbf{R}}(I_1,J_1)$.
5) Initialize the 1st column of the approximate error matrix: $\tilde{\mathbf{R}}(:,J_1) = \mathbf{Z}(:,J_1)$.
6) $\mathbf{u}_1 = \tilde{\mathbf{R}}(:,J_1)$.
7) $\|\tilde{\mathbf{Z}}^{(1)}\|^2 = \|\tilde{\mathbf{Z}}^{(0)}\|^2 + \|\mathbf{u}_1\|^2\|\mathbf{v}_1\|^2$.
8) Find 2nd row index $I_2 : |\tilde{\mathbf{R}}(I_2,J_1)| = \max_i(|\tilde{\mathbf{R}}(i,J_1)|), i \neq I_1$.

kth Iteration:

1) Update (I_k)th row of the approximate error matrix: $\tilde{\mathbf{R}}(I_k,:) = \mathbf{Z}(I_k,:) - \sum_{l=1}^{k-1}(\mathbf{u}_l)_{I_k}\mathbf{v}_l$.
2) Find kth column index $J_k : |\tilde{\mathbf{R}}(I_k,J_k)| = \max_j(|\tilde{\mathbf{R}}(I_k,j)|), j \neq J_1,\ldots,J_{k-1}$.
3) $\mathbf{v}_k = \tilde{\mathbf{R}}(I_k,:)/\tilde{\mathbf{R}}(I_k,J_k)$.
4) Update (J_k)th column of the approximate error matrix: $\tilde{\mathbf{R}}(:,J_k) = \mathbf{Z}(:,J_k) - \sum_{l=1}^{k-1}(\mathbf{v}_l)_{J_k}\mathbf{u}_l$.
5) $\mathbf{u}_k = \tilde{\mathbf{R}}(:,J_k)$.
6) $\|\tilde{\mathbf{Z}}^{(k)}\|^2 = \|\tilde{\mathbf{Z}}^{(k-1)}\|^2 + 2\sum_{j=1}^{k-1}|\mathbf{u}_j^T\mathbf{u}_k|\cdot|\mathbf{v}_j^T\mathbf{v}_k| + \|\mathbf{u}_k\|^2\|\mathbf{v}_k\|^2$.
7) Check convergence: If $\|\mathbf{u}_k\|\|\mathbf{v}_k\| \leq \varepsilon\|\tilde{\mathbf{Z}}^{(k)}\|$, end iteration.
8) Find next row index $I_{k+1} : |\tilde{\mathbf{R}}(I_{k+1},J_k)| = \max_i(|\tilde{\mathbf{R}}(i,J_k)|), i \neq I_1,\ldots,I_k$.

The ACA algorithm published in Zhao et al. 2005

Results/Achievements

Obtained efficiency of the implemented ACA compression technique

N	Tol.	Memory requirements		Ratio Full/ACA	Speed-up (CPU time)	Accuracy			
		Full	ACA			Min	Max	Mean	STD
32 402	10^{-3}	1001 MB	125 MB	7.99	~2.5	-0.00567	0.02806	0.00714	0.00798
129 602	10^{-3}	15.64 GB	1.44 GB	10.84	~4.7	-0.11607	0.15290	0.03898	0.03427
129 602	10^{-7}	15.64 GB	1.74 GB	8.99	~4.5	-0.00001	0.00002	1.5E-07	1.3E-06

Remarks:

- **the partitioning is, so far, not optimal and the number of clustered groups (144 groups) is the same in all three cases**
- **we expect that more appropriate partitioning will lead to more efficient compression**

For the tolerance 10^{-7}, the ACA algorithm gives practically the same results as the standard MFS approach with full matrix while the memory requirements are 9 times smaller!

Further steps discussed at the final STSM meeting

In order to improve efficiency of the compression techniques

- **How to propose optimal partitioning of the source and testing points in case of the spherical computational domains taking into account a row-wise parallelization of the system matrix using MPI.**

- **How to implement the ACA algorithms into the existing C codes and how to treat the matrix-vector products within the BiCGStab linear solver.**

- **How to apply a multilevel version of the ACA in case of the spherical computational domains.**

- **How to implement the ACA algorithms into the existing application of the direct BEM and collocation.**

COST-STSM-ECOST-STSM-IC1406-130217-081941

STSM Topic:

Application of stream clustering techniques to the analysis of e-commerce data

Grant Period: 13-24.02.2017

Applicant Name: **Grażyna Suchacka**

Supervisor: prof. Stefano Rovetta

Working Groups related: WG4

Home Institution: Opole University,
Institute of Mathematics and Informatics (IMiI)
Opole, Poland

Host Institution: University of Genoa,
Department of Informatics, Bioengineering,
Robotics, and Systems Engineering (DIBRIS)
Genoa, Italy

Grażyna Suchacka's seminar at DIBRIS

On 15.02.2017 Grażyna Suchacka presented the seminar at DIBRIS (Department of Informatics, Bioengineering, Robotics, and Systems Engineering) at the University of Genoa in Italy

Seminar title: Web usage mining in e-commerce environment

- The participants of the seminar were students and PhD students in Computer Science, as well as the research staff of DIBRIS
- The topic of the seminar was related to the STSM topic
- The goal was to introduce the participants into the subject of application of data mining methods into e-commerce data

Scientific cooperation of G. Suchacka and S. Rovetta

- **The following research questions were formulated:**

  Q1: Is it possible to learn to classify bots from non-bots solely from a reduced set of web log entries?

  Q2: Is it possible to identify different categories of bots?

  Q3: How many requests do we need to observe at the Web server before being able to decide if a session is performed by a bot or by a human user?

  A few data mining methods were chosen to be applied

- **3 methods were chosen to be applied first for Q1:** *k*-means clustering, Support Vector Machine (SVM) and neural networks

Scientific cooperation of G. Suchacka and S. Rovetta

- **A novel approach** based on unsupervised machine learning was proposed and its practical aspects, design, imlementation and verification were discussed
- **22 features describing user sessions in a Web store** were selected as predictor variables. Data format for the use in the analytical sofware was prepared
- **Preliminary research** using the three data mining methods were performed
- Comparing **the results** of classification in the fully supervised case (SVM, neural network) with those in the unsupervised + labelling case (*k*-means clustering) showed that the level performance is surprisingly similar

Results of STSM

STSM allowed to strengthen the collaboration between the Institute of Mathematics and Informatics, Opole University, Poland and Department of Informatics, Bioengineering, Robotics, and Systems Engineering, University of Genoa, Italy

Tangible results of the STSM include:

- Proposal of a new approach for bot recognition at the e-commerce Web server
- Conducting preliminary research
- Preparing a draft paper to be submitted to the journal
- Proposing a research plan for future common work

COST STSM Reference Number: : COST-STSM-ECOST-STSM-IC1406-040317-080946

Grant Period 04/03/2017 to 13/03/2017

Applicant Name: Dr Agnieszka Jakóbik (Krok)

Home Institution:
Cracow University of Technology, ul. Warszawska 24, 31-155 Kraków, tel. 12 628 20 00, fax 12 628 20 71

Host Institution:
Department of Computer Science, University of Salerno, Via Giovanni Paolo II, 132 - 84084 - Fisciano (SA)

TOPIC: SECURITY IN CLOUD ENVIRONMENT

The aim of the visit…

- …was to deepen knowledge in the field of Cloud Computing Security.
- The main topic was automatic security assessment techniques for Cloud Environments.
- The research included selection of the most effective security assessment techniques for Clouds: data encryption, key management and secure communication methods and others.
- The STSM activities included discussing the broad number of methods for Secure Cloud establishment . The meetings during STSM resulted in formulating the state-of-the-art survey on cloud computing security threats. The game theory solutions, in the form of Stackelberg games, to the presented problems were discussed

Stackelberg games for modeling defending scenarios against security threats in the Cloud systems

- a game-theoretic model to deal with the security attack on the cloud by multiple adversaries one defender
- a utility functions for modeling the Cloud provider gain and the attacker payoff
- the set of possible actions of the defender and the attacker
- a reasonable assumption for modeling the attacker actions
- the game as a non zero-sum game between the attacker and the defender that minimize the reward for the attacker and maximizes the gain of the defender
- a game equilibrium and fond a best response defense strategy for the defender against attacks
- the transition into next round in the game played
- Based on our proposed game model, the defender can rationally choose the defence to cope with the attack launched by security adversaries.

Cloud Controls and Security Risks

- **Cloud Controls Matrix proposed by** Cloud Security Alliance that fulfils the ISO 27002/27017/27018 Security Controls norm was chosen as the most brad list of the security in the Cloud assessment methods. The most up to date list of Cloud systems threads was chosen as OWASP **Cloud Top 10 Security Risks**.
- Additional methods was chosen in the form **NIST Cloud Computing Standards Roadmap** published by Working Group NIST Cloud Computing Program Information Technology Laboratory.
- We prepared the model based on Stackelberg games dedicated to automatization of the security assessment in the Cloud systems.
-

The modelling includes

- defining the leader of the game (assumed as Cloud provider) and the follower of the game (assumed as attacker into Cloud system)
- selecting the actions for the leader of the game (actions from the Cloud Controls Matrix was proposed) and the follower of the game (actions from the Cloud Top 10 Security Risks)
- defining utility functions for the leader and the follower
- defining the aim of the game
- introducing the mathematical model of the game assuming different cases for example simple model for the follower , adversarial behavior uncertainty, Bayesian Stalkerberg model for the follower, model for unknown adversary strategy, random walk adversary strategy
- selecting the tools for solving the game (numerical methods in the form of Simplex method was chosen) and selecting the tools and environments for implementing the model and preliminary tests: Java and Matlab Environment .

The result of the STSM….

- is the development of the COST Action IC1406 WG1 research topic:
- novel heterogeneous models, algorithms and techniques for advanced Big Data exploitation, based on the current and emerging multicore system architectures, virtualised servers and data centres, mobile cloud and multi-cloud systems.

- The main result of the visit in Salerno was the formulating the model of Stackelberg game for automatization the security assessment in the Cloud systems.
-
- The next result will be preparing the paper for Journal of Network and Computer Applications, Special Issue on "Security in Cloud Computing".

e-COST-STSM Exchange on High-Performance Computing for Many-Core Architectures

Nicolas Melot

[1] Linköping University
Dept. of Computer and Inf. Science
Linköping, Sweden

December 9, 2016

Outline

Introduction

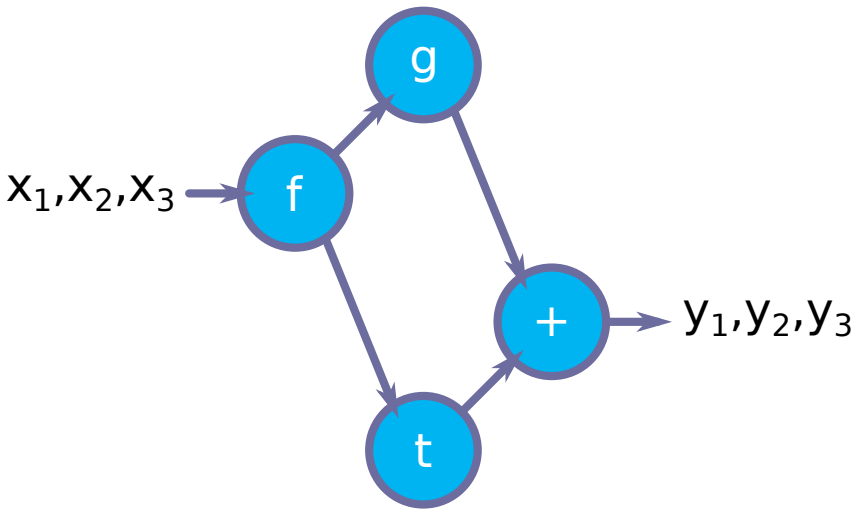
- Guest: Nicolas Melot (Linköping University, Sweden)
- Host: Marco Aldinucci (Turin University, Italy)
- High performance on Many-Core architectures
- Programming frameworks
  - Drake (Linköping)
  - FastFlow (Turin)

Themes explored

- Frameworks semantic
- Communication queues
- Projection for joined work

Actor and Tasks

Steaming application: $y = (g \circ f)(x) + (t \circ f)(x)$

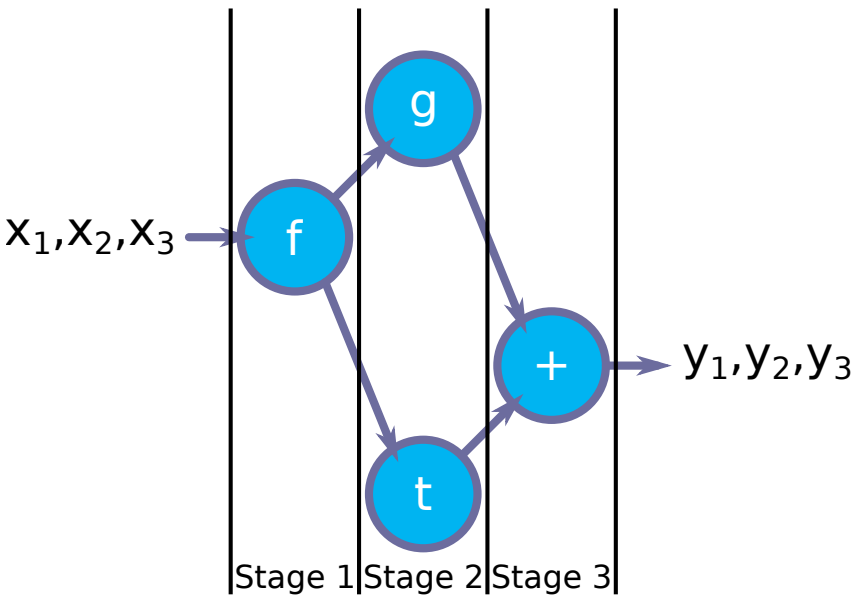


Semantic matrix

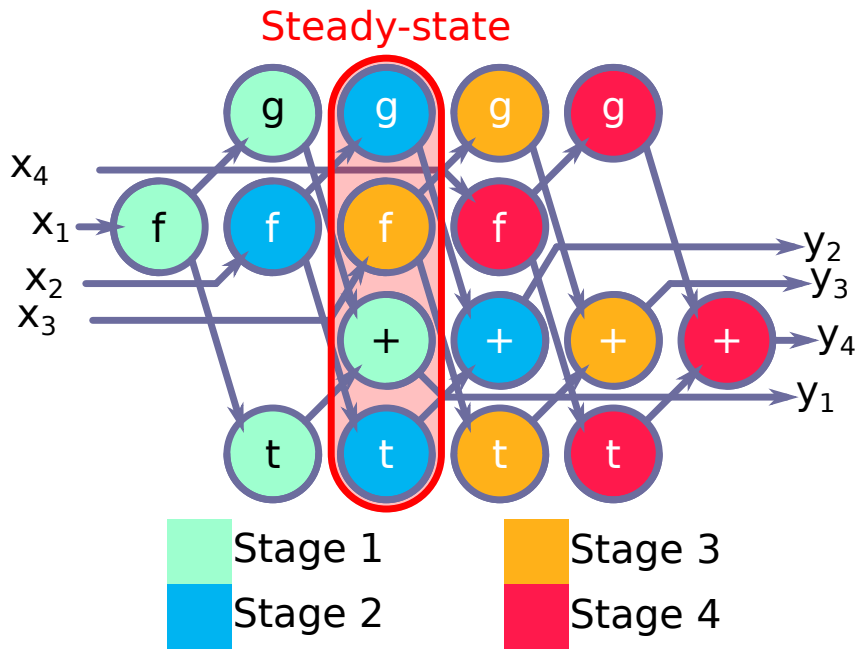
Symbol	FastFlow	Drake	Streaming
$f, g, t, +$	Actor	Task	Process
x_1,x_2,x_3, y_1,y_2,y_3	Task	(input/output) Data	Token

Pipeline stage

Pipeline stage



Set of tasks running in parallel

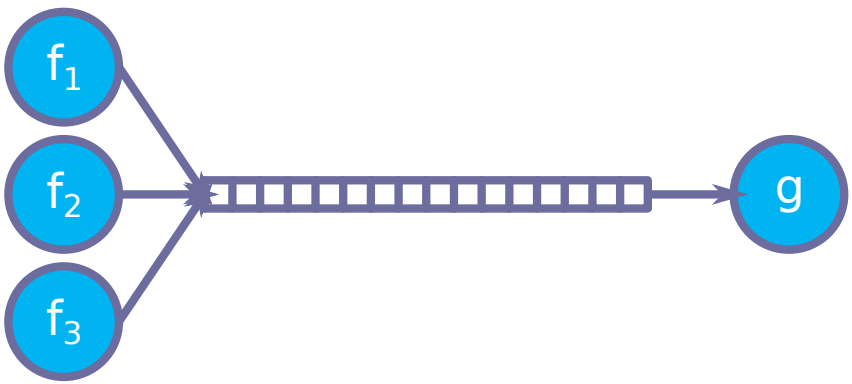


Set of tasks run for each token

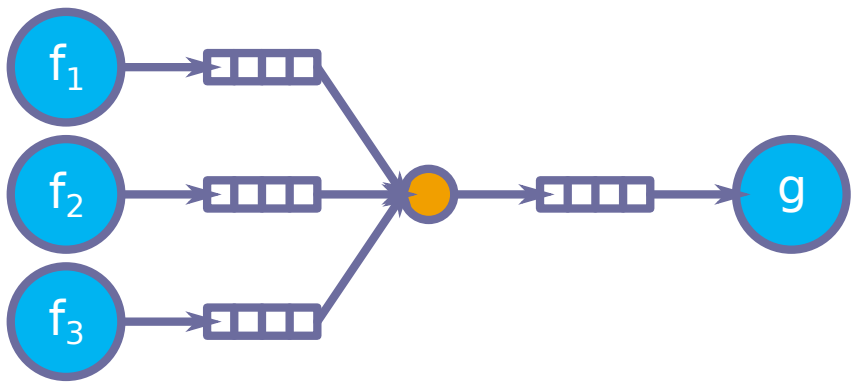
Communication queues

Communication queues

- Fundamentally: one-to-one (Drake & FastFlow)
- FastFlow: One-to-Many, Many-to-One, Many-to-Many
- Finite memory buffer written and read by producer and consumer tasks
- Low-level: dispatcher task (orange)
  - Arbitrary token forwarding policy
  - Low latency overhead (Aldinucci et al. [2012])



High-level view

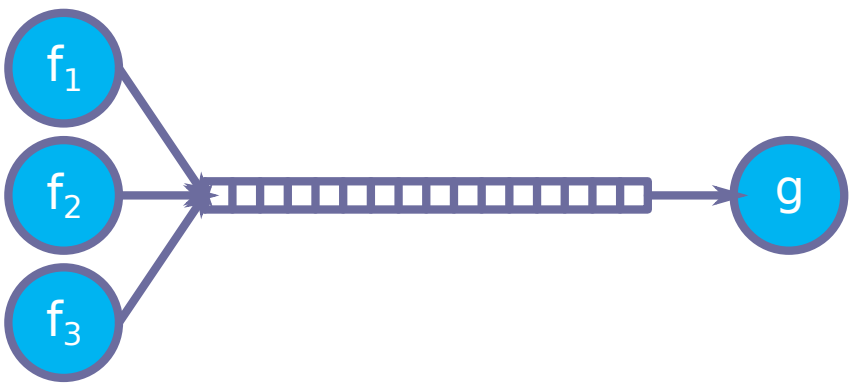


Low-level FastFlow design

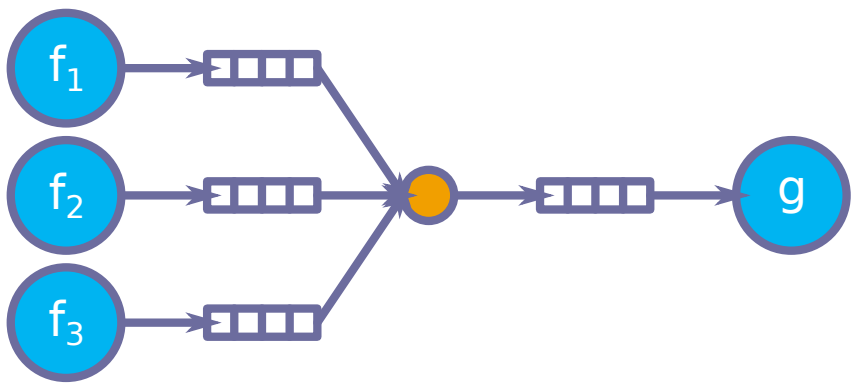
Communication queues

Communication queues

- Fundamentally: one-to-one (Drake & FastFlow)
- FastFlow: One-to-Many, Many-to-One, Many-to-Many
- Finite memory buffer written and read by producer and consumer tasks
- Low-level: dispatcher task (orange)
  - Arbitrary token forwarding policy
  - Low latency overhead (Aldinucci et al. [2012])



High-level view

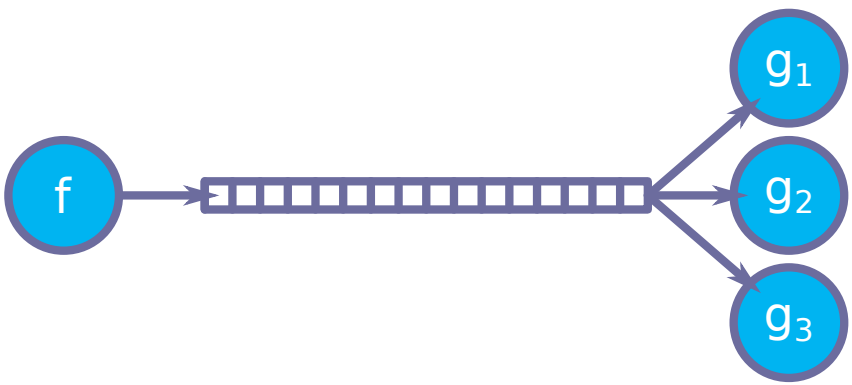


Low-level FastFlow design

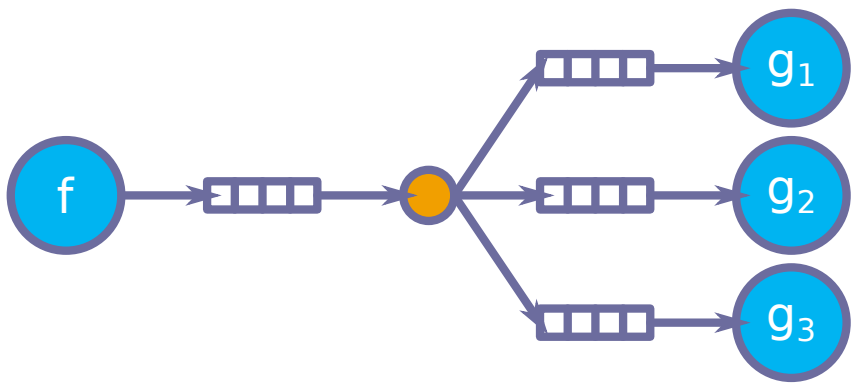
Communication queues

Communication queues

- Fundamentally: one-to-one (Drake & FastFlow)
- FastFlow: One-to-Many, Many-to-One, Many-to-Many
- Finite memory buffer written and read by producer and consumer tasks
- Low-level: dispatcher task (orange)
  - Arbitrary token forwarding policy
  - Low latency overhead (Aldinucci et al. [2012])



High-level view

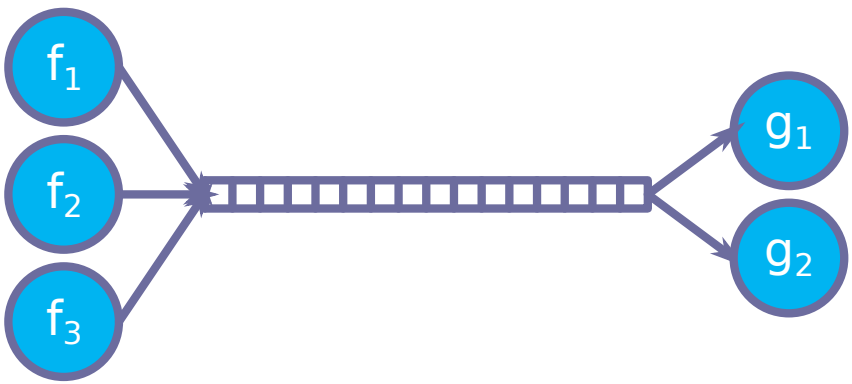


Low-level FastFlow design

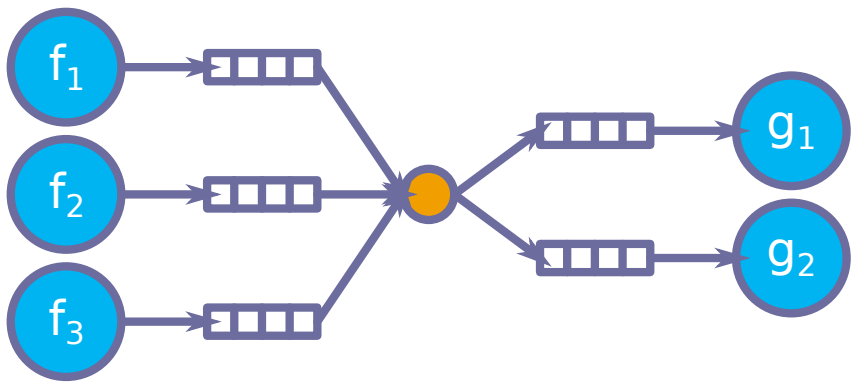
Communication queues

Communication queues

- Fundamentally: one-to-one (Drake & FastFlow)
- FastFlow: One-to-Many, Many-to-One, Many-to-Many
- Finite memory buffer written and read by producer and consumer tasks
- Low-level: dispatcher task (orange)
  - Arbitrary token forwarding policy
  - Low latency overhead (Aldinucci et al. [2012])



High-level view

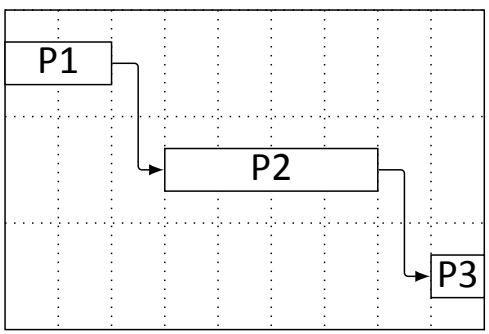


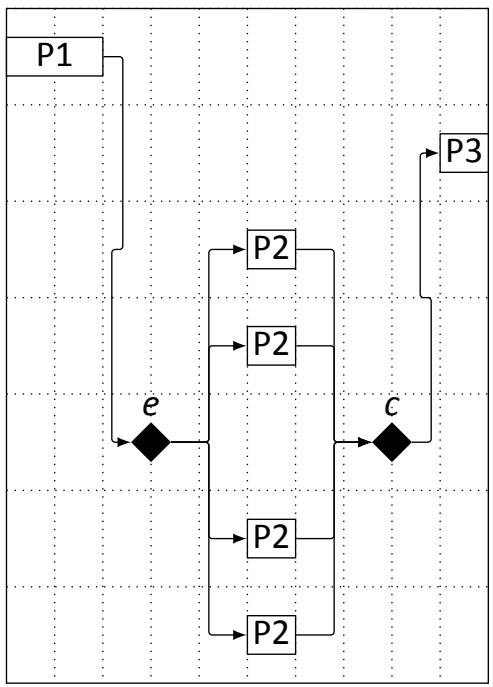
Low-level FastFlow design

Scheduling

Performance dependent on static or dynamic scheduling

- How many core per process?
- What core(s) run what process?
- At what frequency runs a process?

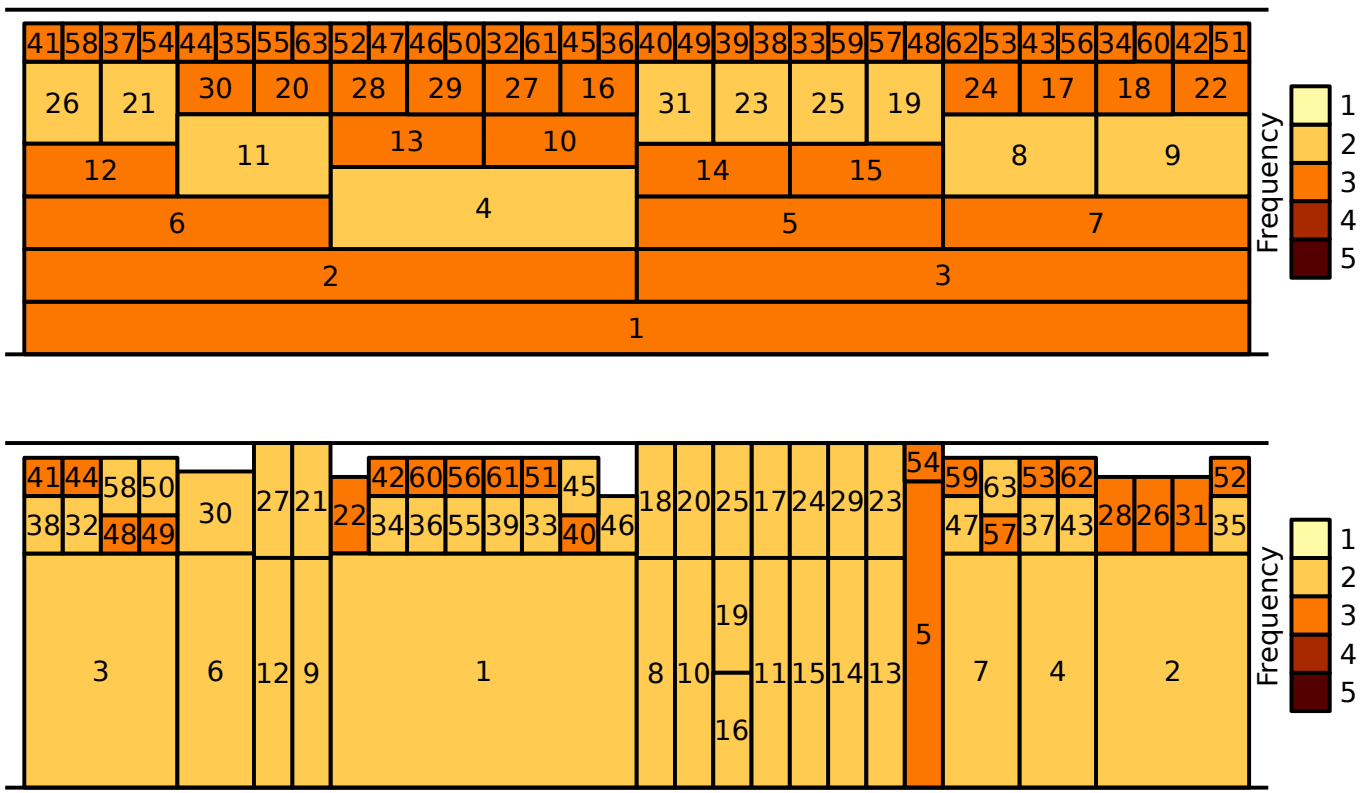




Scheduling (2)

Crown Scheduling (Melot et al. [2015])

- Schedule parallel streaming tasks for energy efficiency
- Below: FFT for 32 cores with two crown schedulers



Darker tasks: higher frequency, more energy consumption

Scheduling (3)

Modeling FastFlow application for Crown Scheduling

- Process j's workload τ_j
- Duplicate process into FastFlow *farms*
- Extra emitter and collector tasks
- For d duplications of process j, use $m = d + 1$ cores
- Parallel efficiency $e_j(m)$ for m cores:

$$e_j(m) = \begin{cases} 1 & \text{if } m = 1 \text{ (no emitter or collector extra process)} \\ \frac{m-1}{m} & \text{Otherwise (1 extra core for emitter and collector).} \end{cases}$$

Use Crown Scheduling to compute a good static schedule for FastFlow

Conclusion

- 2 weeks STSM at Turin University
  - Guest: Nicolas Melot (Linköping University)
  - Host: Marco Aldinucci (Turin University)
- Exchange on Drake and FastFlow details
  - Semantic
  - Communication queues
- Opening toward collaboration on Scheduling for FastFlow

www.liu.se

Bibliography

Marco Aldinucci, Marco Danelutto, Peter Kilpatrick, Massimiliano Meneghin, and Massimo Torquati. An efficient unbounded lock-free queue for multi-core systems. In *Proc. of 18th Intl. Euro-Par 2012 Parallel Processing*, volume 7484 of *LNCS*, pages 662–673, Rhodes Island, Greece, August 2012. Springer. doi: 10.1007/978-3-642-32820-6_65. URL http://calvados.di.unipi.it/storage/paper_files/2012_spsc_europar.pdf.

Nicolas Melot, Christoph Kessler, Jörg Keller, and Patrick Eitschberger. Fast Crown Scheduling Heuristics for Energy-Efficient Mapping and Scaling of Moldable Streaming Tasks on Manycore Systems. *ACM Trans. Archit. Code Optim.*, 11(4):62:1–62:24, January 2015. ISSN 1544-3566. doi: 10.1145/2687653. URL http://doi.acm.org/10.1145/2687653.

STSM Grant

Name: Fátima Manuela da Silva Leal

Affiliation: University of Vigo

Grant Research Topic: Big Data Techniques for Trust and Reputation Modelling using Crowdsourcing Information on Tourism Domain

Hosting Institution: National College of Ireland

Period: 01 of November 2016 to 30 of January 2017

The overarching aim of this STSM was to carry out research Machine Learning (ML) and Data Mining (DM) techniques as well as Big Data processing technologies in the tourism domain.

Specific objectives

- Create user profiling methods using crowdsourcing information;
- Develop personalised trust and reputation mechanisms for tourism recommendation systems;
- Deploy a cloud architecture to process large repositories of crowd-sourced data.

Work Accomplished

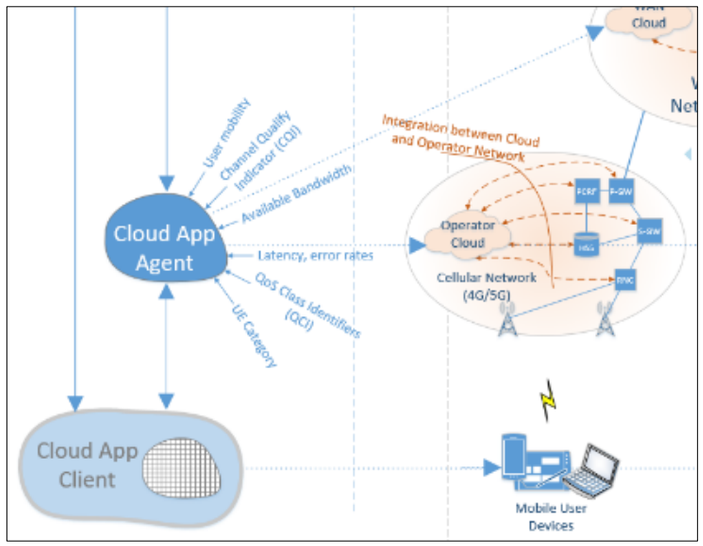
- Review the state of the art in ML and DM;
- Develop models for multiple criterion information i.e., textual reviews and ratings;
- Design an experimental cloud setup in OpenStack integrating Big Data processing software;
- Extract, transform, and load a crowd-sourced tourism dataset;
- Create distinct recommendation filters using the different tourism crowdsourced information;
- Include trust & reputation to recommendation filters;
- Test, compare and discuss the results;
- Write scientific papers reporting the outcomes.

STSM Grant

- COST Action: IC1406
- Grant Period: 19.12.2016 – 30.12.2016
- Applicant Name: Dzmitry KLIAZOVICH
- Topic: Communication Network Integrated Cloud Computing and Big Data Systems
- Home Institution: University of Luxembourg, Luxembourg
- Host Institution: Belarussian State University, Minsk, Belarus

Accomplishments

- Developed a framework of novel techniques for
  a) new communication network integrated cloud computing and big data systems
  b) efficient, communication-aware dynamic partitioning of applications between cloud computing environments located at different levels of a hierarchy, and featuring different levels of integration with network infrastructures



- Meeting with the team, presentations of research areas and exchange of the core expertise

- Working on idea for a possible joint Horizon 2020 proposal

- Planning student and staff exchange between our groups in Luxembourg and Belarus for the upcoming year

cHiPSet

High-Performance Modelling and Simulation

for Big Data Applications

www.chipset-cost.eu

cost
EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

Novel optimization methods for HPDA on emerging computing infrastructures

Sabri Pllana and Joanna Kolodziej

COST STSM Ref.: COST-STSM-IC1406-36581

COST Action: IC1406

Working Groups related: WG1, WG2

Overview

- STSM institutions
  - home: Linnaeus University, Växjö (SE)
  - host: Cracow University of Technology, Cracow (PL)
- STSM topic
  - High Performance Data Analytics (HPDA)
  - address computational and data-intensive problems
  - consider emerging computing infrastructures

Major tasks

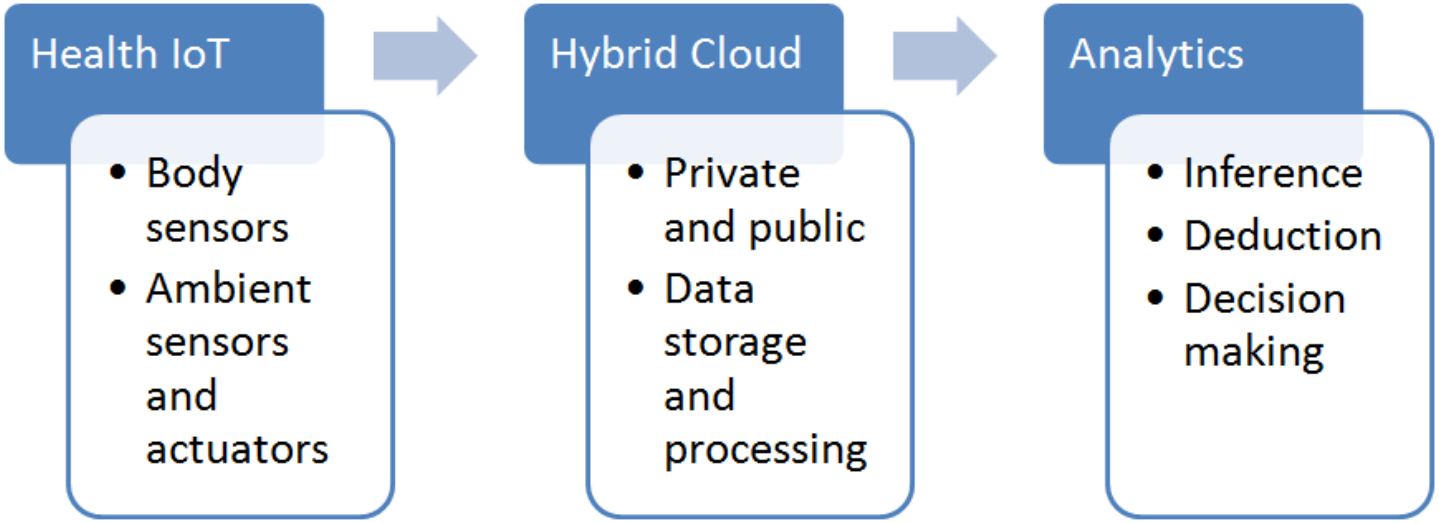
- Study the state of the art solutions for remote health monitoring
- Devise a hybrid cloud architecture for health monitoring and assistance
- Exemplify our approach with a case study

Context

- Chronic diseases (such as, heart diseases)
  - cause 60% of all deaths in the world
  - 50% of caused deaths are premature deaths
- It is possible to avoid fatal consequences
  - timely detection of critical health conditions
  - proper decision of medical personnel
- Use modern communication and computation infrastructures
  - Internet of Things (IoT)
  - Cloud computing

Our approach

- Health IoT generates big data that is stored and analyzed in the hybrid cloud



Case study

- Focus on heart-related chronic diseases

- Target group: professional drivers (such as, taxi drivers)
  - long hours in a single body posture
  - exposure to vibration, exhaust, and noise
  - demands constant vigilance
  - risk for ischemic heart disease

Summary and future work

- A productive visit that led to elaboration of a CHIPSET use case, and development of plans for joint papers and future project proposals

High Performance Modelling and Simulation for Big Data Applications

STSM Summary – 1st Grant Period

SPEAKER: Mauro Iacono -Seconda Università degli Studi di Napoli, Italy

Models for advanced Big Data platforms

- COST STSM Reference Number:
- COST-STSM-ECOST-STSM-IC1406-150117-080952
- COST Action: IC1406
- STSM type: Regular
- Grant Period (exact dates): 15/1/2017 to 16/2/2017
- Applicant Name: Dr Mauro Iacono
- Topic: Models for advanced Big Data platforms
- Working Groups related: WG1
- Home Institution: Università degli Studi della Campania “Luigi Vanvitelli”, viale Lincoln 5, 81100 Caserta, Italy

Host Institution: University Politehnica of Bucharest, Romania, Bucharest, Romania

Models for advanced Big Data platforms

Host Institution: University Politehnica of Bucharest, Romania, Bucharest, Romania

- Aim: exchange expertise on performance evaluation methods and techniques and comparison
- Goal: establish a stable collaboration to develop a unified comprehensive approach between system modeling view (UniCampania) and application modeling (UPB)
- Activities: knowledge transfer about modeling techniques for the evaluation of Big Data architecture and Large Scale computing systems
- Results: a research agreement has been established for long term collaboration and a research plan about modeling and evaluation of Microservice Based Architectures

Output: joint work is ongoing, one journal paper (Submission ID: IJPP-D-17-00027, "Adapting MCP and HLFET algorithms to multiple simultaneous scheduling", Emilia Ioana Popa; Mauro Iacono, Ph.D.; Florin Pop, Ph.D.) is under revision for publication in International Journal of Parallel Programming, two journal papers are planned on Microservice Based Architectures; other results are the proposal of 2 special issues (one under evaluation on Journal of Parallel and Distributed Computing) and an Erasmus agreement in phase of definition.