ICT, FET Open

LIFT ICT-FP7-255951

Using Local Inference in Massively Distributed Systems

Collaborative Project

D 5.1

Evaluation Criteria

Contractual Date of Delivery:	30.03.2012
Actual Date of Delivery:	30.09.2012
Author(s):	C. Kopp, A. Deligiannakis, M. Gabel, M. Garofalakis, M. Kamp, D. Keren, M. May, A. Monreale, M. Nanni
Institution:	HU
Workpackage:	WP 5
Security:	PU
Nature:	R
Total number of pages:	9

Project coordinator name: Michael May

Project coordinator organisation name:
Fraunhofer Institute for Intelligent Analysis
and Information Systems (IAIS)
Schloss Birlinghoven, 53754 Sankt Augustin, Germany
URL: http://www.iais.fraunhofer.de

Revision: 10

Abstract:

This document is the LIFT deliverable of WP5 for the second review period (01.10.2011 – 30.09.2012). The document contains the description of the evaluation criteria applied to the application scenarios specified in deliverable 4.1.

Revision history

Administration Status			
Project acronym: LIFT			**ID:** ICT-FP7-255951
Document identifier:			D 5.1 Evaluation Criteria (01.10.2011 – 30.09.2012)
Leading Partner:			HU
Report version:			10
Report preparation date:			30.03.2012
Classification:			PU
Nature:			REPORT
Author(s) and contributors:			C. Kopp (FHG), A. Deligiannakis (TUC), M. Gabel (Technion), M. Garofalakis (TUC), M. Kamp (FHG), D. Keren (HU), M. May (FHG), A. Monreale (CNR), M. Nanni (CNR)
Status:	-	Plan	
	-	Draft	
	-	Working	
	-	Final	
	x	Submitted	

Copyright

Project funded by the European Community under the Information and Communication Technologies Programme
Contract ICT-FP7-255951

LIFT Deliverable 5.1
Evaluation Criteria

C. Kopp, A. Deligiannakis, M. Gabel, M. Garofalakis, M. Kamp, D. Keren,
M. May, A. Monreale, M. Nanni

Contents

Chapter 1

Introduction

The application scenarios in our project cover a variety of requirements and practical applications and are thus evaluated in a different manner for each scenario. In the following, we introduce the evaluation criteria and list the application scenarios they are applied to. We then elaborate the evaluation criteria for each application scenario in detail.

All application scenarios are evaluated in terms of scalability and communication reduction. Communication reduction is measured in the number of messages and their size, i.e., the overall communication throughput. Scalability is measured application-dependently. For sensor applications, such as *Sensor Networks* and *Monitoring Movement Behavior Using Stationary Sensors*, scalability is measured in terms of sensors that can be employed and entities that can be monitored. For those scenarios, also battery savings is an important evaluation criterion. Other applications use other notions of scalability which are described in detail in chapter 2. For the majority of our application scenarios, privacy is a major concern. Thus, they are evaluated in terms of privacy guarantees they provide. Some of the applications provide only approximate solutions, because of privacy concerns or data size. In these cases, the approximation quality is an important evaluation criterion. For many threshold monitoring tasks, the amount of false alarms is important and will be evaluated. Furthermore, for some applications the correctness of the final results has to be guaranteed.

In table 1, we list our application scenarios and the evaluation criteria that are applied. Additionally, for the scenarios *Cloud Data Centers* and *Financial Monitoring* the prediction models are evaluated in terms of accuracy.

Scenario	Scalability	Approx. Quality	Comm. Reduction	Privacy	False Alarms	Correctness	Battery Saving
Distributed Density Map Computation	✓		✓	✓			
Distributed Estimation of Traffic Flows		✓	✓	✓			
Monitoring Movement Behavior Using Stationary Sensors	✓	✓	✓	✓	✓		✓
Quantitative Monitoring of Visit Patterns	✓	✓	✓	✓	✓	✓	✓
Internet Scale Query Processing	✓		✓	✓		✓	
Sensor Networks	✓		✓		✓		✓
Cloud Data Centers	✓	✓	✓		✓	✓	
Financial Monitoring	✓		✓			✓	

Table 1.1: Application scenarios and the evaluation criteria that are applied to them. A detailed explanation of each evaluation criteria in the context of the specific application scenarios is given in chapter 2

Chapter 2

Evaluation Criteria

In the following, we elaborate how our application scenarios will be evaluated in detail, e.g. quality of safe zones, quality of approximated values, false alarm rate, privacy guarantees.

2.1 Mobility Scenarios

2.1.1 Distributed Density Map Computation

The proposed solutions are mainly evaluated in terms of percentage of saved communications, also in relation to the parameters of the application (e.g. the size of the Gaussian kernel adopted in the density estimation), the parameters for the construction of the predictive models adopted, and the data sampling rate. In addition, the scalability on the number of nodes is analyzed, and privacy protection is evaluated.

2.1.2 Distributed Estimation of Traffic Flows

The privacy requirements will be evaluated by considering various attack models based on different assumptions about the knowledge used by the adversary during the attack. We will define specific measures able to evaluate how much private information an attacker can possibly derive from the information transmitted by the node. We will evaluate the accuracy of the application measuring the approximation and so the error introduced by the sketching transformation and the data transformation due to the privacy requirements comparing the results obtained after the data transformations with those obtained without application of any transformation. Lastly, we also measure the amount of information in the local vector before and after the sketching transformation in order to evaluate the performance of the system in terms of amount of transmitted information.

2.1.3 Monitoring Movement Behavior Using Stationary Scanners

Our proposed method is evaluated both in terms of privacy and accuracy. Furthermore, the scalability is analyzed. The Safe Zone approach applied to this scenario is evaluated in terms of communication reduction. Additionally, battery power saving through application of the Safe Zone approach is analyzed on real Bluetooth scanners.

2.1.4 Quantitative Monitoring of Visit Patterns

The privacy requirements will be evaluated based on different attacking scenarios assessing how much information an attacker can possibly derive w.r.t. current and historic location information of a user.

For the performance evaluation of the local nodes we will test how many POI and visit patterns can be concurrently monitored for a given GPS update frequency. In addition we will measure memory requirements.

The performance evaluation of the coordinator will test how many updates the coordinator can handle and how much memory will be required to store the frequency distributions.

Finally, we will evaluate the accuracy of the application by comparing the different visit quantities when derived from the ground truth and the results obtained by the global coordinator.

2.2 Internet Scale Query Processing

The key performance criteria for the ISQP application include: (1) *Communication efficiency* compared to not only simplistic "push-all-updates" solutions but other existing solutions (if any); (2) *Scalability* in terms of data and network size, as well as query response time; and, (3) *Approximation quality* (including the number of false negatives and false alarms), especially when summarization tools (such as sketches) are employed.

2.3 Sensor Networks

The performance criteria for the Sensor Networks application are as follows:

- Reduction in communication cost, expressed as the number of messages required by our techniques, when compared to naive techniques that centrally collect the sensor data.
- When our techniques include further specialized optimizations for safe zones, an additional criterion also involves the reduction in communication cost compared to using geometric monitoring.
- For the case of centrally computing the safe zones for all central nodes, an additional criterion involves the running time efficiency of the proposed algorithms, when compared to a flat (non-hierarchical) organization of the sensor nodes.

2.4 Cloud Data Centers

Performance must be evaluated on two criteria: accuracy of the latent fault detection compared against real machine health logs, and the reduction in communications. In other words, there is the LIFT-based criteria of the score error introduced by LIFT techniques, and the communication costs. For the application criteria, there is the effectiveness and precision of the latent fault detection.

For the LIFT criteria, we can compare the resulting scores from the original tests to the sketch-based tests, as the general approach and the original tests have already been evaluated. Communication cost will be compared to a centralized solution.

Health logs for machines found are examined for up to 15 days after their counters are compared. Using standard measures found in literature, we report on false positive rate, precision, and recall, and investigate accuracy trade-offs using standard ROC and P-R curves. In the LIFT scenario with the sketch-based tests we plan to use similar methodology, if comprehensive health logs are available.

2.5 Financial Monitoring

Our method has to be evaluated on two criteria. First, the error of the predictions of the deviations of asset prices from their fundamental value has to be evaluated. Because we model our problem as a regression task, the root mean square error is a natural means of evaluating this error. We use standard evaluation techniques, i.e., 10-fold cross-validation, to estimate the true error of our method.

Second, the communication reduction has to be evaluated. Therefore, we will simulate distributed systems of different sizes and measure the amount of communication induced by our method using LIFT techniques. We then compare this to the amount of communication necessary when not using LIFT techniques. The communication reduction rate obtained by these experiments is then compared to standard network capacities of distributed systems to show the applicability of our method and its limits.