International Journal of Interactive Multimedia and Artificial Intelligence

December 2012, Vol I, Number 7, ISSN: 1989-1660

One machine can do the work of fifty ordinary men.
No machine can do the work of one extraordinary man.
Elbert Green Hubbard

www.elasticbox.com
info@elasticbox.com
blog.elasticbox.com
Twitter: elasticbox

elasticbox

Deploy applications, not servers.

Challenges In Cloud Computing

Automation
- How long to deploy an application?
- What version do I use?
- How do I upgrade applications?

Portability
- How do I change providers?
- What is being used?
- How much does it cost?

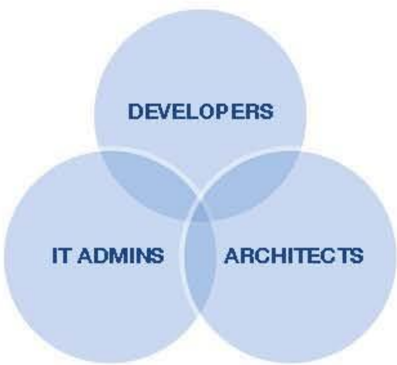
Auto-Scaling
- Can my application auto-scale?
- How do I configure auto-scaling?

Disaster Recovery Planning
- Can my application tolerare faults?
- How do I recover my system?

ElasticBox Solution

Automation
- ✔ Automatic Deployments
- ✔ Automatic Configuration
- ✔ Disaster Recovery

Runtime Environment
- ✔ Application Scaling
- ✔ Fault Tolerance
- ✔ Resource Clean-up
- ✔ Replication



Framework Design
- ✔ Architecture Policies
- ✔ Versioning
- ✔ Platform Management

Infrastructure Control
- ✔ Cost Analysis
- ✔ Policy Management
- ✔ Traceability

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE AND INTERACTIVE MULTIMEDIA

ISSN: 1989-1660 - VOL. I, NUMBER 7

Editor's Note

I. INTRODUCTION

The International Journal of Interactive Multimedia and Artificial Intelligence provides an interdisciplinary forum in which scientists and professionals can share their research results and report new advances on Artificial Intelligence and Interactive Multimedia techniques.

This regular issue includes extended versions of selected papers from The 6th International Conference on Multimedia and Ubiquitous Engineering, held in Leganes - Madrid, Spain, 2012. The issue includes, thus, nine selected papers, describing innovative research work, on several areas of Artificial Intelligence and Interactive Multimedia including, among others: Mobile and Cloud Computing, Semantic Brokering, Metaheuristic Optimization Algorithms, Video Services, Dynamic Conceptual Data, Accessing Wireless Sensor, Models of E-Commerce and Service Orchestration.

We would like to thank all the contributing authors, as well as all members of International Conference on Multimedia and Ubiquitous Engineering, for their hard and valuable work that assured the high scientific standard of the conference and enabled us to edit this issue. Finally, the Guest Editors would also like to thank the Editors-in-Chief of International Journal of Interactive Multimedia and Artificial Intelligence for the publication of this issue.

Jeús Carretero Pérez (Carlos III University), José Daniel García Sánchez (Carlos III University)

II. JOURNAL COVER

We want to thanks the cover of this issue to the professional photographer Nacho García Peña (www.nachogarciapena.es)

III. WELCOME TO NEW MEMBERS

Javier Bajo received a Ph. D. in Computer Science and Artificial Intelligence and a MSc in e-commerce from the University of Salamanca in 2007. He obtained his Information Technology degree from the University of Valladolid (Spain) in 2001 and Engineering in Computer Sciences degree from the Pontifical University of Salamanca in 2003. He hold the positions of Director of the Data Center and Associate Professor at the Pontifical University of Salamanca (Spain). At present he is Associate Professor at the Technical University of Madrid. He has been a member of the organizing and scientific committee of several international symposiums such as PAAMS, CAEPIA, IDEAL, HAIS, etc. and is co-author of more than 200 papers published in recognized journals, workshops and symposiums.

Prof. Dr. Ing. Armando Walter Colombo. Schneider Electric Group Senior Expert (Edison L2), Program Manager for Collaborative/Advanced Projects in the BU Industry. He became Full professor at the University of Applied Sciences Emden-leer from August 2010. He received the Doctor degree in Engineering from the University of Erlangen-Nuremberg, Germany, in 1998. From 1999 to 2000 was Adjunct Professor in the Group of Robotic Systems and CIM, Faculty of Technical Sciences, New University of Lisbon, Portugal. His research interests are in the fields of service-oriented architecture (SoA), collaborative automation, intelligent supervisory control, formal specification of flexible automation systems. He is a senior member of the IEEE and member of the Gesellschaft für Informatik e.V. Dr. Colombo served/serves as Associated Editor of the IEEE Transactions on Automation Systems Engineering (IEEE T-ASE) and Associated Editor of the IFAC Associated Journal ATP-International.

Dr. Héctor Fernández is a postdoctoral research scientist member of the ConPaaS team at VU University Amsterdam. He received his PhD in Computer Science from University of Rennes 1 in 2012 in the area of service-oriented computing, and in particular on the decentralized workflow coordination in distributed infrastructures. Hector's current research interests include service coordination, resource management and fault-tolerance in large scale infrastructures.

TABLE OF CONTENTS

OPEN ACCESS JOURNAL

ISSN: 1989-1660

Framework for Computation Offloading in Mobile Cloud Computing

Dejan Kovachev and Ralf Klamma
Department of Information Systems and Databases
RWTH Aachen University

***Abstract* — The inherently limited processing power and battery lifetime of mobile phones hinder the possible execution of computationally intensive applications like content-based video analysis or 3D modeling. Offloading of computationally intensive application parts from the mobile platform into a remote cloud infrastructure or nearby idle computers addresses this problem. This paper presents our Mobile Augmentation Cloud Services (MACS) middleware which enables adaptive extension of Android application execution from a mobile client into the cloud. Applications are developed by using the standard Android development pattern. The middleware does the heavy lifting of adaptive application partitioning, resource monitoring and computation offloading. These elastic mobile applications can run as usual mobile application, but they can also use remote computing resources transparently. Two prototype applications using the MACS middleware demonstrate the benefits of the approach. The evaluation shows that applications, which involve costly computations, can benefit from offloading with around 95% energy savings and significant performance gains compared to local execution only.**

***Keywords* — Application Virtualization, Middleware, Pervasive computing, Software Design, Android**

IV. INTRODUCTION

RESOURCE-DEMANDING multimedia applications such as 3D video games are being increasingly demanded on smart phones. Even if mobile hardware and mobile networks continue to evolve and to improve, mobile devices will always be resource-poor, less secure, with unstable connectivity, and with constrained energy. Resource poverty is major obstacle for many applications [14]. Therefore, computation on mobile devices will always involve a compromise. For example, on-the-fly editing of video clips on a mobile phone is prohibited by the energy and time consumption. Same performance and functionalities on mobile devices still cannot be obtained as on desktop PCs or even notebooks when dealing with high resource-demanding tasks.

Recently, the combination of cloud computing [11], wireless communication infrastructure, ubiquitous computing devices, location-based services, and mobile Web, has laid the foundation for a novel computing model, called mobile cloud computing [9]. It provides to users an online access to unlimited computing power and storage space. The cloud abstracts the complexities of provisioning computation and storage infrastructure. The end user uses them as utility and in reality they can be far-away data center or nearby idle hardware.

Offloading has gained big attention in mobile cloud computing research, because it has similar aims as the emerging cloud computing paradigm, i.e. to surmount mobile devices' shortcomings by augmenting their capabilities with external resources. Offloading or augmented execution refers to a technique used to overcome the limitations of mobile phones in terms of computation, memory and battery. Such applications, which can adaptively be split and parts offloaded [6, 18], are called elastic mobile applications. Basically, this model of elastic mobile applications gives the developers the illusion as if they are programming virtually much more powerful mobile devices than the actual capacities. Furthermore, elastic mobile applications can run as stand-alone mobile applications, but use also external resources adaptively. Which portions of the application are executed remotely is decided at runtime based on resource availability. In contrast, client/server applications have static partitioning of code, data and business logic between the server and client, which is decided in the development phase.

Our contributions include integration with the established Android application model for development of "offloadable" applications, a lightweight application partitioning and a mechanism for seamless adaptive computation offloading. We propose Mobile Augmentation Cloud Services (MACS), a services-based mobile cloud computing middleware. Android applications that use the MACS middleware benefit from seamless offloading of computation-intensive parts of the application into nearby or remote clouds. First, from the developer perspective, the application model stays the same as on the Android platform. The only requirement is that computation-intensive parts are developed as Android services, each of which encapsulates specific functionality. Second, according to different conditions/parameters, the modules of program are divided into two groups; one group runs locally, the other group is run on the cloud side. The decision for partitioning is an optimization problem according to the input conditions of the cloud and devices, such as CPU load, available memory, remaining battery power on devices, bandwidth between the cloud and devices. Third, based on the

DOI: 10.9781/ijimai.2012.171

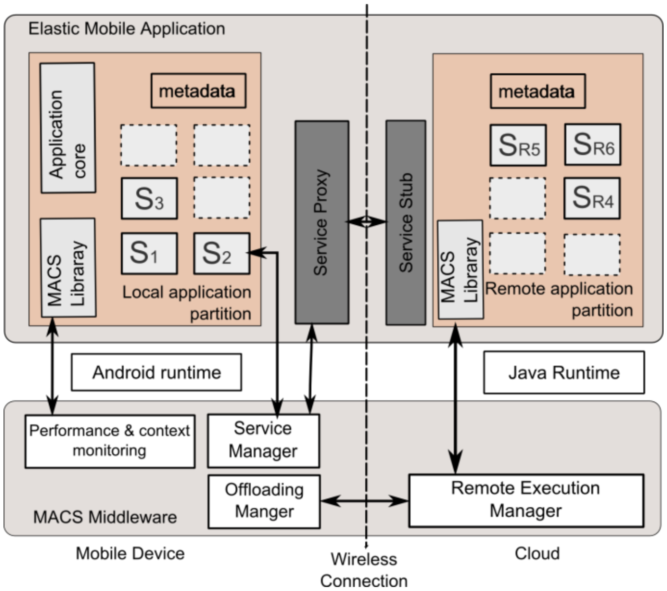


Fig. 1 MACS architecture. Application logic is structured from multiple Android services (S_i). Some of them can be offloaded into the cloud (S_{Ri}).

solution of the optimization problem, our middleware offloads parts to the remote clouds and returns the corresponding results back. Two Android applications on top of MACS demonstrate the potential of our approach.

In the rest of the paper, we first review related research in work mobile cloud computing in Section 2. Then we describe our MACS middleware with detailed descriptions of the implementation (Section 3). We explain the offloading model in our middleware in Section 4. In Section 5 we introduce two use case applications, the setup of the evaluation and the corresponding evaluation. After that, we discuss the results in Section 6. Finally, we draw conclusions and refer to the future work.

V. Related Work

Previous work has proposed many mechanisms that address the challenges of seamless offloaded execution from a device to a computational infrastructure (cloud). The encapsulation of the mobile device's software stack into a virtual machine image and executing it on a more powerful hardware can be considered as a "brute force" approach to offloading, such as proposed by Chun and Maniatis [1] or Satyanarayanan et al. [14]. More recently, Kosta et al. [8] further improved this idea. Although such virtualized offloading can be considered as simple and general solution, it lacks flexibility and control over offloadable components. Therefore, we consider that application developers can better organize their application logic using the established Android service design patterns and benefit from the MACS middleware.

Ou et al. [13] propose class instrumenting technique, i.e. a process to transform code classes into a form which is suitable for remote execution. Two new classes are generated from the original class, one is an instrumented class which has the real implementation and the same functionality as the original class, the other is a proxy class, whose responsibility is only to call the function written in the instrumented class. Then, the instrumented class can be offloaded to remote cloud, and the call will be invoked from the instrumented in the remote cloud. In MACS we adopt similar idea, but unlike Ou et al. [13] we use standardized language for the proxy interfaces which is already widespread in the Android platform. The Cuckoo framework [6] and MAUI system in [2] implement a similar idea. Our MACS middleware is inspired by these solutions. However, MACS middleware does extra profiling and resource monitoring of applications and adapts the partitioning decision at runtime.

An important challenge in partitioned elastic applications is how to determine which parts of code should be pushed to the remote clouds. The graph based approach to model the application has been used in several works. Giurgiu et al. [3] use "consumption" graphs and decide which part should be running locally or remotely. It finds a cut of the consumption graph with a goal function, which minimizes the total sum of communication cost, transmitting cost and the cost of building local proxies. The AIDE platform [4] uses a component-based offloading algorithm, which mainly focuses on minimum historical transmission between two partitions. The $(k+1)$ partitioning algorithm, introduced by Ou et al. [13], is applied to a multi-cost graph representing the class-based components. A similar approach is done by Gu et al. [4, 5]. Zhang et al. [19, 17] use a general Bayesian inference to make the partitioning decision. However, executing constantly executing graph or inference algorithms on the mobile device takes significant resources on the constrained device. We use an integer linear optimization model to describe the offloading so that it is not only easy to implement, but it can also be independently solved if the remote clouds are not available temporarily.

VI. Mobile Augmentation Cloud Services

The goal of our MACS middleware is to enable the execution of elastic mobile applications. Zhang et al. [17] consider elastic applications to have two distinct properties. First, an elastic application execution is divided partially on the device and partially on the cloud. Second, this division is not static, but is dynamically adjusted during application runtime. The benefits of having such an application model are that the mobile applications can still run independently on mobile platforms, but can also reach cloud resources on demand and availability. Thus, mobile applications are not limited by the constraints of the existing device capacities.

MACS architecture is depicted on Figure 1. In order to use MACS middleware, the application should be structured using established Android services pattern. Android is already established as the most prominent mobile phone platform. Additionally, its application architecture model allows decomposition of applications into service components which can be shared between applications. A MACS application consists of an application core (Android activities, GUI, access to devices sensors) which cannot be offloaded, and multiples services (S_i) that encapsulate separate application functionality (usually resource-demanding components) which can be offloaded (S_{Ri}). The services communicate with the application through an interface defined by the developer in the Android interface definition language (AIDL).

As service-based implementation is adopted, for each service we can profile following metadata:

- type: whether can be offloaded or not
- memory cost: the memory consumption of the service on the mobile device
- code size: size of compiled code of the service
- dependency information on other services, for each related module, we collect following:
  - transfer size: amount of data to be transferred
  - send size: amount of data to be sent
  - receive size: amount of data to be received

Metadata is obtained by monitoring the application execution and environment.

Android services are using Android inter-process communication (IPC) channels for RPC. The services are registered in the Service Manager, and a binder maintains a handle for each service. Then an application, that wants to use a service, can query the service in the Service Manager. Upon service discovery, the Android platform will create a service proxy for the client application. All the requests to access the service will be sent through the service proxy, and then forwarded to the service by the binder. After processing the requests, results are sent back to the service proxy on the client application through the binder. Finally, the client gets the result from the service proxy. From client's point of view, there is no difference between calling a remote service or calling a local function.

The offload manager determines the execution plan, and then the services to be offloaded are pushed to the cloud. The results are sent back to the application upon completion. Our approach is similar to the Cuckoo framework [6], however, MACS allows dynamic application partitioning at runtime, where Cuckoo only enables static partitioning at compile time. MACS monitors the execution of the services and the environment parameters. Whenever the situation changes, the middleware can adapt the offloading and partitioning.

The main goal of MACS is to enable transparent computation offloading for mobile applications. Therefore, our middleware tries to fit the usual Android development process and bring the developers an easier way to offload parts of their applications to remote clouds in a transparent way. MACS hooks into the Android compile system, makes modifications of generated Java files from AIDL in the pre-compile stage. Developers need to include MACS SDK libraries into their Android project.

Since our implementation wants to bring the developers an easier way to distribute their application to remote cloud, the low-level implementation should be transparent to them. The way to hide the low-level implementation is as follows. Recalling the Android compile system and combining with the idea of using the services on Android, the possible way to make modification of generated Java file from AIDL is in the pre-compile stage. Our code is embedded there to realize the transparency to the developers. Each time while the developers compile their projects by using Ant tool, our code will be embedded without notice. The way to add a customized process while building with Ant tool is to write a new target, which can be treated as a task.

At the cloud side, the MACS middleware handles the offload requests from the clients, installs of offloaded services, their initialization and method invokes (s. Figure 9 in the Appendix). The cloud-side MACS middleware is written with pure J2SE so that it can run on any machines with installed J2SE.

MACS middleware monitors the resources on the mobile execution environment and available clouds. It then forms an optimization problem whose solution is used to decide whether the service which contains the called function should be offloaded or not. When the service is determined to be offloaded to the remote cloud, our middleware tries first to execute the service remotely. If there is no such service on the remote clouds, our framework transmits the service code (jar file) to the cloud, and the corresponding results after the service execution are returned to the mobile device. The cloud caches the jar files for subsequent executions.

Except for the computation offloading, our framework also features simple data offloading. If files are needed to be accessed on the remote cloud, MACS file transmission (MACS-FTM) transfers automatically the non-existing files from the local device and vice versa. Basically, the middleware at pre-compile time inserts a line of code after a file object is created by using the File object in Java. This code snippet retrieves file information of the file at runtime. When no such file exists in the remote cloud, MACS-FTM throws an exception which is caught by the middleware which in turn obtains the file from the device.

VII. Adaptive Computation Offloading

The proposed model and corresponding algorithm are supposed to be applied for scenario which is computationally intensive [7]. The requirements for the developer are that the code should be structured in a model in advance.

The developer should also provide or use extra tools to extract meta-information from given modules and then tag each module with some parameters. The tagged parameters are used for deciding on code partition later.

Let us suppose that we have *n* number of modules which can be offloaded, S_1, S_2 ... S_n. Each of the modules has several properties described as metadata, i.e. for specific module *i*, its memory cost mem_i, code size $code_i$. Let us consider the k number of related module which can be offloaded. For each of them, we denote the transfer size tr_1, tr_2 ... tr_k, send size $send_1$, $send_2$...$send_k$, receive size rec_1, rec_2...rec_k, where $\{1..k\} \subseteq \{1..n\}$ and $send_k + rec_k = tr_k$. Meanwhile, we introduce x_i for module i, which indicates whether the module i is executed locally ($x_i = 0$) or remotely ($x_i = 1$). The solution x_1, x_2...x_n represents the required offloading partitioning of the application.

The cost function is represented as follows:

$$\min_{x \in 0,1}(c_{transfer} * w_{tr} + c_{memory} * w_{mem} + c_{CPU} * w_{CPU}) \tag{1}$$

where

$$c_{transfer} = \sum_{i=1}^{n} code_i * x_i + \sum_{i=1}^{n}\sum_{j=1}^{k} tr_j * (x_j \ XOR \ x_i) \tag{2}$$

$$c_{memory} = \sum_{i=1}^{n} mem_i * (1 - x_i) \tag{3}$$

$$c_{CPU} = \sum_{i=1}^{n} code_i * \alpha * (1 - x_i) \tag{4}$$

There are three parts in the cost function. The first part depicts the transfer costs for the remote execution of services, including the transfer cost of its related services which are not at the same execution location. The latter part of Eq.(2) implicitly includes the dependency relationship between modules, i.e. if the output of one module is an input of another. The c_{memory} contains the memory costs on the mobile device, and c_{CPU} the CPU costs on the mobile device is, where α is the convert factor mapping the relationship between code size and CPU instructions, which is taken to be 10 based on [12]. w_{tr}, w_{mem}, w_{CPU} are the weights of each costs, which can lead to different objectives, for example lowest memory costs, lowest CPU load or lowest interaction latency.

The three constraints are expressed as the following:

Minimized memory usage. First, the memory costs of resident service can not be more than available memory on the mobile device, i.e.

$$\sum_{i=1}^{n} mem_i * (1 - x_i) \le avail_{mem} * f_1 \tag{5}$$

where $avail_{mem}$ can be obtained from the mobile device, f_1 is the factor to determine the memory threshold to be used, because the application can not occupy the whole free memory on the mobile device.

Minimized energy usage. Second, for the offloaded services, the energy consumption of offloading should not be greater than not offloading [10], i.e.

$$E_{local} - E_{offload} > 0 \tag{6}$$

The local energy consumption can be expressed using the number of local instructions to be executed I_{local}, local execution speed S_{local} and the power used of local execution P_{local} [10]. At the first decision time, I_{local} is estimated according to the code size. After the first decision, this number is collected from our framework (by calling the statistic method provided from Android API). Obviously, there is relationship between the instruction number and the power used for that instruction while doing power profiling.

$$E_{local} = \frac{P_{local} * I_{local}}{S_{local}} \tag{7}$$

The energy costs of offloading some parts to remote cloud can be expressed as the sum of energy consumption during waiting for the results from the cloud E_{idle}, transferring (including sending E_s and receiving E_s) the services to be offloaded [10] and also the additional data which may be needed on the remote cloud E_{extra}.

$$\begin{aligned} E_{offload} &= E_s + E_{idle} + E_r + E_{extra} \\ &= P_s * (t_s + t_{extra}) + P_{idle} * t_{idle} + P_r * t_r \end{aligned}$$

The idle time of the mobile device waiting for the result from cloud can be treated as the execution time of remote cloud, so the formula becomes

$$\begin{aligned} E_{local} - E_{offload} &= \frac{P_{local} * I_{local}}{S_{local}} - \frac{P_{local} * I_{local}}{S_{cloud}} \\ &- \frac{P_s * (D_s + D_{extra})}{B_s} - \frac{P_r * D_r}{B_r} \end{aligned}$$

where D_s and D_r are the total data sizes to be sent and received, D_{extra} is the size of extra data needed because of offloading, which is determined at runtime, B_s and B_r are the bandwidths of sending and receiving data, and S_{cloud} is the remote execution speed. Additionally,

$$I_{local} = \sum_{i=1}^{n} code_i * type_i * x_i \tag{10}$$

$$D_s = \sum_{i=1}^{n} send_i * type_i * x_i \tag{11}$$

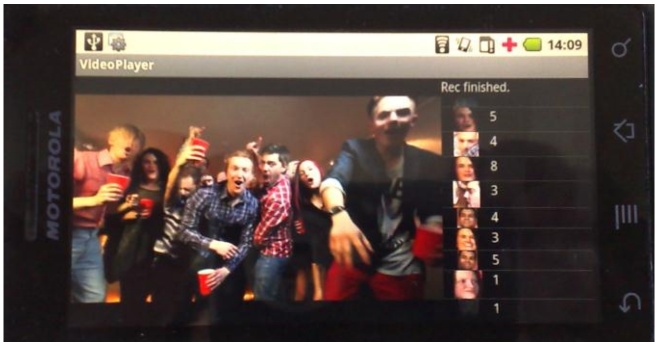


Fig. 2. Snapshot of prototype application of face detection and recognition using MACS middleware.

$$D_r = \sum_{i=1}^{n} rec_i * type_i * x_i \qquad (12)$$

where $type_i \in \{0,1\}$ represents whether a service is offloadable or not.

Minimized execution time. Third, the third constraint is enabled when the user prefers fast execution, i.e.

$$t_{local} - t_{offload} > 0 \qquad (13)$$

The local execution time can be expressed as the ratio of CPU instructions to local CPU frequency, meanwhile, the remote execution time consists of the time consumed by CPU, file transmission and the overhead of our middleware.

$$t_{local} = \frac{I_{local}}{S_{local}} * x_i \qquad (14)$$

$$t_{offload} = \left(\frac{I_{local}}{S_{cloud}} + \frac{D_{extra}}{B_s} + t_{overhead}\right) * x_i \qquad (15)$$

where $t_{overhead}$ is the overhead which our framework brings in. According to the constraints above, we now transform the partitioning problem to an optimization problem. The middleware determines the execution location by solving a linear integer optimization problem. The decision maker receives the input parameters and execution constraints. It then returns the corresponding running locations. The solution of $x_1, x_2 .. x_n$, is the optimized partitioning strategy. By using integer liner programming (ILP) on the mobile device, MACS gets a global optimization result. Whenever the the parameters in the model change, such as available memory or network bandwidth, the partitioning is adapted by solving the new optimization problem.

MACS middleware defines the abstraction of a decision maker so that we can apply different decision makers which determine the execution location of each "offloadable" service. In our experiments we used Cream[1], an open-source class library for constraint programming in Java. It provides enough features to run the decision solver on an Android platform, with acceptable solving speed in the order of tens of milliseconds. During the calculation, as an objective function is taken the sum of transmission cost, memory cost and CPU cost.

[1] http://bach.istc.kobe-u.ac.jp/cream/

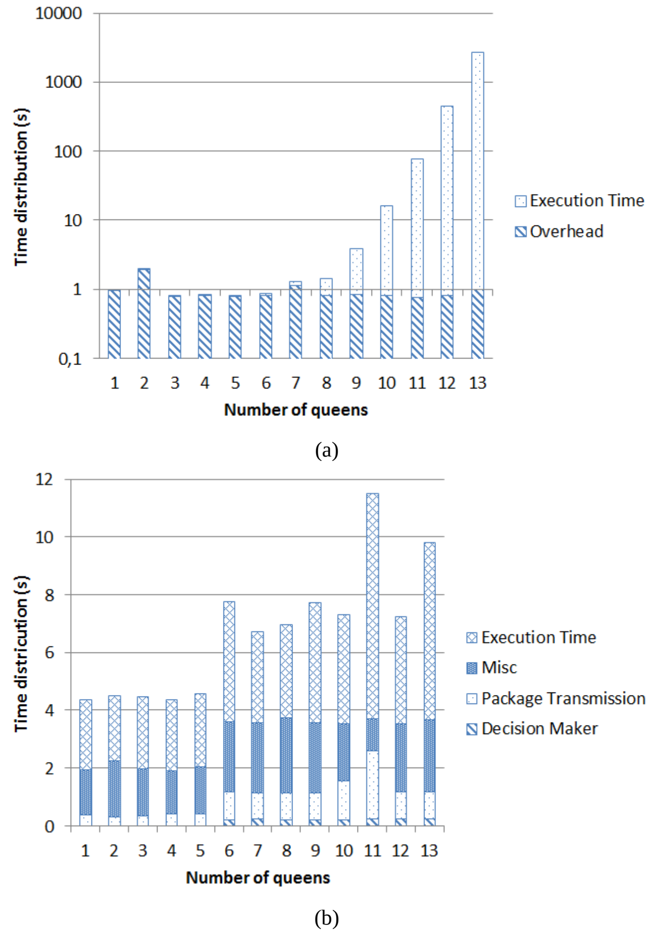


Fig. 4. Total time distribution of N-queens: (a) local execution and (b) remote execution.

Although MACS introduces the overhead because of using a proxy for communication between offloaded services and the mobile application, the overhead is relatively small, which is shown in the evaluation. MACS also profiles each offloaded module/service to dynamically change its execution plan and adjust the partitioning.

VIII. Experimental Evaluation of Offloading

We evaluate our MACS middleware with two use case smartphone applications. The first application implements the well-known N-Queens problem. It is chosen because the performance bottleneck represents a pure computation problem. This use case can easily show the overhead introduced by MACS middleware. The second application involves face detection and recognition in video files. This use case involves lots of computation, but also requires much more memory resources to process and obtain results. Table 8 shows the problem space in terms of N.

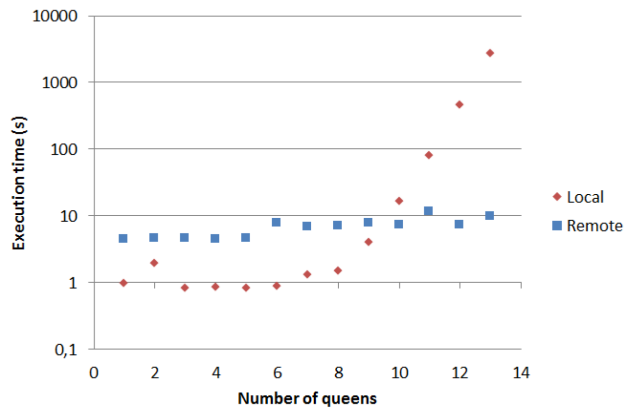


Fig. 3. Snapshot of prototype application of face detection and recognition using MACS middleware.

TABLE I
HARDWARE COMPONENTS OF MOBILE DEVICES AND DESKTOP COMPUTER

Hardware Component	Android Milestone	PC
Processor	ARM A8 600MHz	QuadCore 2.83GHz
Memory	256MB	8GB
WLAN	Wi-Fi 802.11 b/g	N/A
OS	Android 2.2	Windows XP x64

The second case can process a video file, and detect faces from the video file, cluster them and provide the time point cues for video navigation. The results can then be used for faster video navigation on small screen devices (Figure 2). The video file is processed with OpenCV[2] and FFmpeg[3] libraries. We use FFmpeg to open video files, and scan it frame by frame. Face detection in video files is done by detecting faces in video frames. In the processing, faces in the video file are detected by the existed implementation in OpenCV, and then the detected faces are recognized by the method proposed by Turk and Pentland [15], and after that, the faces are clustered.

In the implementation we used JavaCV[4] for video processing. When the application gets the results from the processing, it shows all detected faces as a clustered view. The user can select a cluster, and then navigate to the time points where that face occurs in the video. Thus, the application can accelerate navigation in a video based on persons that occur within.

A. Setup of the Evaluation

Hardware. The hardware we are using in the evaluation is as follows. A Motorola Milestone mobile phone based on Android platform 2.2 is used in the evaluation. A desktop computer which includes quad-core CPU acts as a cloud provider that can host the offloaded computation. The details about the hardware components for the mobile device and desktop computer are shown in Table 1.

TABLE II
ESTIMATED ENERGY CONSUMPTION OF A MOBILE DEVICE

Hardware Component	Estimated Energy Consumption (W)
Processor	0.4 (Idle: 0.05)
WLAN	0.75 (Low: 0.03)
LCD	0.9

Network Topology. While offloading services to the remote cloud, the mobile phone connects to a nearby access point. Since the wireless local area network is encrypted with Wi-Fi Protected Access 2 (WPA2) security protocol, the data speed is not as fast as non-encrypted considering of the overhead introduced by the security protocol. The desktop computer is connected to the Internet directly by network cable, whose bandwidth is 100 Mbps.

Energy Estimation Model. We adopt a method as the one proposed by Zhang et al. [16], a power model for an Android phone and a measurement application for the energy consumption on the Android-based mobile device on the fly. Using their software, the energy consumption of each hardware component of the Motorola Milestone such as LCD, CPU and Wi-Fi can be measured separately (see Table 2).

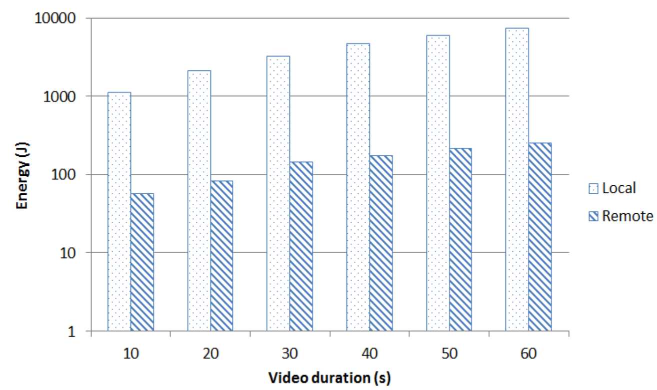


Fig. 7 Energy consumption of face detection.

B. Results of the Use Case 1

We use the algorithm by Sedgewick and Wayne[5]. The basic idea is to use recursion and back-tracking to enumerate all possible solutions. Although it is not the best algorithm, it is often used for solving the N-queens puzzle. It is clear that with the increase of N, much more steps are spent to find solutions, which is extremely time-consuming for the mobile device.

We run the N-Queens on the local device and offload to the remote cloud separately, for $N=1$ to $N=13$. For $N=14$, it will take hours to finish on the local device, it is not realistic not to be offloaded while doing computation after $N=14$. Table 8 in the Appendix provides overview of the problem space in terms of N.

[2] http://opencv.willowgarage.com
[3] http://ffmpeg.org
[4] http://code.google.com/p/javacv
[5] http://introcs.cs.princeton.edu/java/23recursion/Queens.java.html

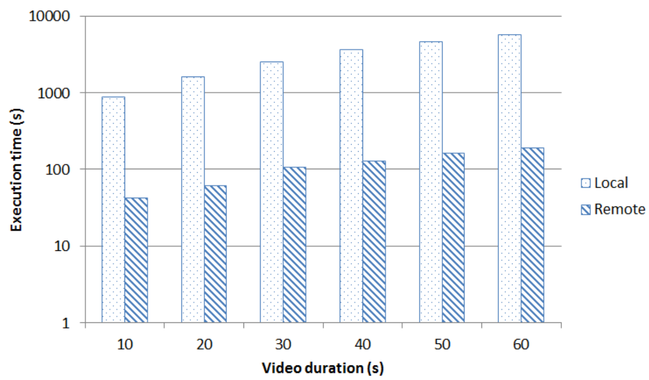


Fig. 6. Execution time of face detection

Figure 3 shows the time duration of execution of the specific calculation service. From $N = 1$ to $N = 9$, the execution speed on the local device is acceptable compared with the remote speed and to run the method locally is better, but after $N = 10$, the remote speed dominates to be the better option as the computation time dominates the total time in the rest cases, and the remote execution speed is also relative stable, there is no huge variation for remote speed.

Figure 4 shows the different times, which are made up of the total spent time. With the increase of the queens number, the local execution time increases outstandingly, especially from $N = 9$, the execution time of calculating solution occupies more than half of the total time. Meanwhile, the overhead, our framework brings in, remains constant. As for the remote execution, the overhead is broken down into three parts, one is the package offloading time, one is the decision making time, the rest is the residual overhead. It shows that our decision model costs only little time to finish the determination. The transmission time of remote package occupies also few periods of total time, since the remote package is small. The execution time of solving the N-queens is relative stable, except for $N = 11$, which is a deviation during the execution and measurement.

The last Figure 5 shows the results of consumed energy with and without offloading. As for the local execution, most of the time is spent on computation, since our energy model involves CPU and LCD, and the LCD is always on while computation, so that the energy consumption of CPU and LCD dominates the total energy consumption of local execution. The execution time is significantly increased from $N = 9$ compared to the remote execution, which leads to the highest energy value. In contrast to those, the remote execution time is nearly stable, so that the consumed energy is almost at the same level.

C. Results of the Use Case 2

Six video files are used in the evaluation. All of them belong to a same original video file with different length of time, 10, 20, 30, 40, 50 and 60 seconds (see Table 3). The video resolution is 720 pixels × 480 pixels, the fps is set to 30, the overall bit rate is 1500 Kbps and the video is compressed with MPEG-4 format, 3GPP Media Release 5 profile. The audio codec is AAC, and the bit rate for audio is

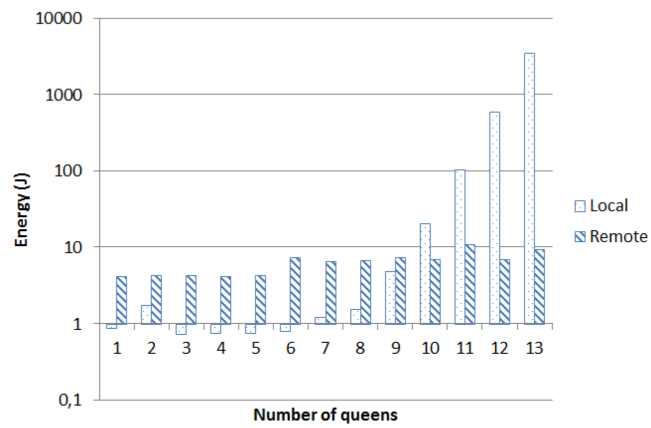


Fig. 5. Energy consumption of the N-queens app.

30 Kbps.

TABLE III
VIDEO DURATION AND FILE SIZE

Duration (seconds)	File size (Bytes)
10	1864984
20	3864612
30	5827420
40	7754219
50	9633240
60	11584020

In order to get a more accurate estimation of execution time which is used in the model, we first run the face detection services locally, and keep track of the spent time (second) and the file size (bytes), and then a linear regression is used to reflect the relationship between the spent time and the file size. Considering the number of CPU instructions provided by Android API, it can only be used to make estimation on the execution which involves no native calls, we don’t directly use that count, but focus on the execution time. The regression shows that,

$$Time = 0.0005 * FileSize - 246.09 \tag{16}$$

We use this heuristic equation in our model to determine the execution time.

On Figure 6 can be seen clearly that the execution time is reduced hugely while offloading compared with the local execution. Even dealing with the 10 seconds video file, the local device spends more than 15 minutes on processing and detecting, but the corresponding remote offloading takes only less than 1 minute. Each time the computation is offloaded to the remote cloud, the execution speed can be reduced by more than 20 times, see Table 4. It is absolutely not acceptable to let the CPU of local mobile device 100% load for such long time, and it confirms that the video processing task is still a huge burden for the mobile device.

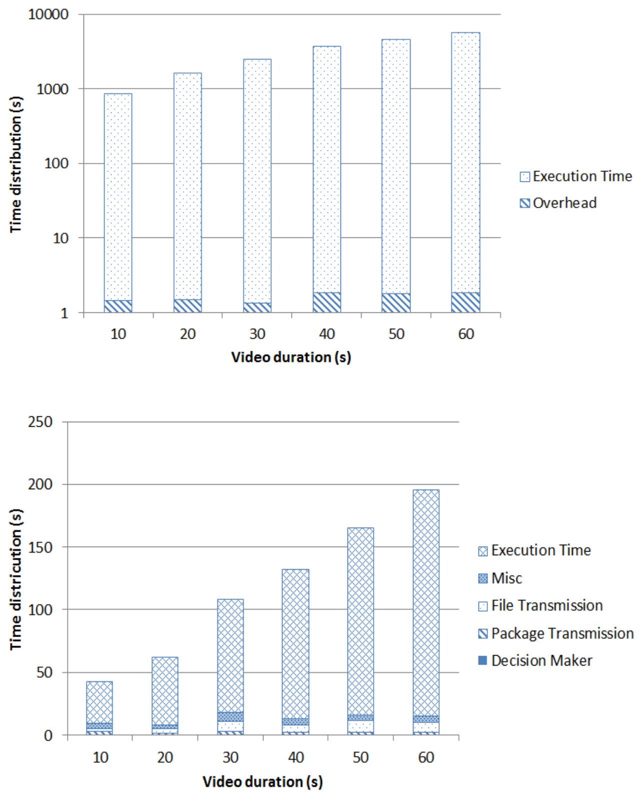


Fig. 8. Total time distribution of face detection: (a) local execution and (b) remote execution.

TABLE IV
VIDEO DURATION AND SPEED UP OF FACE DETECTION IN VIDEO FILES

Duration (seconds)	10	20	30	40	50	60
Speed up	x20	X26	X23	X28	X28	X29

With the huge difference between local and remote execution time, we can conclude that the local energy consumption is worse than the remote ones, because most of the time is spent on CPU and LCD, which are the top two energy consuming components (s. Figure 7). Table 5 also depicts the energy saving situation while offloading. Energy can be saved by more than 94% thanks to the offloading.

Figure 8 describes the composition of the local and remote total spent time in details. As the execution time increases with the bigger video file size, the overhead our framework brings occupies about only 0.1%, which can be nearly omitted. Regarding to the remote execution, the total time spent consists of execution time, transmission time of needed files, package transmission (service offloading) time, and decision making time. With the increase of the video file size, the file transmission time also rises, but compared to the total time, it is not significant. The decision making does its determination in less than 1 second, which is only 1% of the total time spent. The total overhead our framework brings is about 5% of the total time, which is acceptable considering about the speed up and energy save above.

The face clustering can only be done on the remote cloud because of the software limitations on the local mobile device. Detailed results can be seen in Table 6 in the Appendix. Most of the execution time is spent on building/rebuilding the training set. If the training set is already available before the remote execution, then the estimated execution time can be significantly reduced.

TABLE V
VIDEO DURATION AND SPEED UP OF SAVED ENERGY

Duration (seconds)	10	20	30	40	50	60
Saved Energy (%)	94.98	96.07	95.59	96.37	96.33	96.55

IX. DISCUSSION

Offloading perhaps is not suited for every mobile application, but from the results of the two use cases, we see that when an application uses complex or time consuming algorithms such as recursion, by offloading those parts into the cloud, time and energy consumption are reduced, so that the local execution time is reduced to an acceptable level. Offloading can lower the CPU load on a mobile device significantly. It can also save lots of energy, which indicates that the battery lifetime can be increased compared to the local execution, as shown in the second use case, where more than 90% of energy is saved and the calculation speed is up to 20 times over local execution.

The results also prove that the overhead of our framework is small and acceptable with the increase of needed computation, it is better to push those computations which cost considerable resources to the remote cloud. But for the small N in the N-Queens problem, the overhead occupies almost half of the total execution time because of the needed computation is small so that it takes only little time to obtain the results. This shows a clear advantage of local execution over remote offloading when the needed computation is not much. In a word, the more computation is needed, offloading has more advantages. Since we use Wi-Fi in the evaluation, the time of sending files and receiving results has small proportion, but if 3G or GPRS are used, the offloading time will surely increase.

X. CONCLUSIONS AND FUTURE WORK

The results show that the local execution time can be reduced a lot through offloading, which is sometimes not acceptable for users to wait for, and by pushing the computation to the remote cloud can lower the CPU load on mobile devices significantly thanks to the remote cloud, since most of the computations are offloaded to the remote cloud. Meanwhile, lots of energy can be saved which indicates users can have more battery time compared to the local execution. The results also prove that the overhead of our framework is small.

Our framework supports offloading of multiple Android services. If there are multiple services in one application and all of those services can be offloaded to the remote clouds, our resource monitor natively supports this situation and can make the corresponding allocation determination, so that some of the services should be offloaded and the rest of the services should be run locally.

The next steps are to enable parallelization of the offloaded services. Additionally, we can extend the current middleware so that it supports automatic partitioning arbitrary mobile applications. A great challenge is how to estimate the characteristics of an application depending on different input parameters, which is precisely the relationship between the input of the invoked method and the execution time. We could characterize the relationship between execution time and input parameters by running the target application several times and adapt the offloading algorithm accordingly.

APPENDIX

Figure 9 presents the registration flow of “offloadable” services (a) and the optimization process (b).

In Table 6, the data on the left side of the slash sign is for local execution, whereas on the right side is for remote execution.

TABLE VI
EVALUATION RESULTS OF USE CASE 1 IN DETAILS

Num-ber	Execution Time (s)	Overhead (s)	CPU Instructions	Energy (W)
1	0.973 / 4.383	0.97 / 1.97	155986 / 847523	0.877 / 4.188
2	1.953 / 4.534	1.95 / 2.269	162317 / 847189	1.759 / 4.329
3	0.82 / 4.486	0.817 / 2.012	155925 / 847421	0.739 / 4.285
4	0.847 / 4.378	0.839 / 1.943	164070 / 847368	0.766 / 4.183
5	0.827 / 4.602	0.811 / 2.062	164021 / 847941	0.751 / 4.397
6	0.883 / 7.786	0.84 / 3.629	201058 / 848234	0.812 / 7.513
7	1.318 / 6.752	1.178 / 3.592	383161 / 848363	1.242 / 6.531
8	1.473 / 7.006	0.841 / 3.779	1008483 / 848173	1.579 / 6.770
9	3.994 / 7.746	0.868 / 3.611	4475993 / 848379	4.845 / 7.475
10	16.431 / 7.337	0.836 / 3.549	22916437 / 849381	21.026 / 7.094
11	80.055 / 11.527	0.794 / 3.72	128838160 / 850019	103.754 / 11.112
12	459.055 / 7.256	0.832 / 3.571	780925799 / 849810	596.439 / 7.013
13	2766.508 / 9.839	1.015 / 3.685	6937052630 / 849590	3596.054 / 9.468

Detailed results of use case 2. The data of Table 7 in the left side of slash is local execution, the middle one is remote execution of face detection, the last one is remote execution of face detection with face recognition.

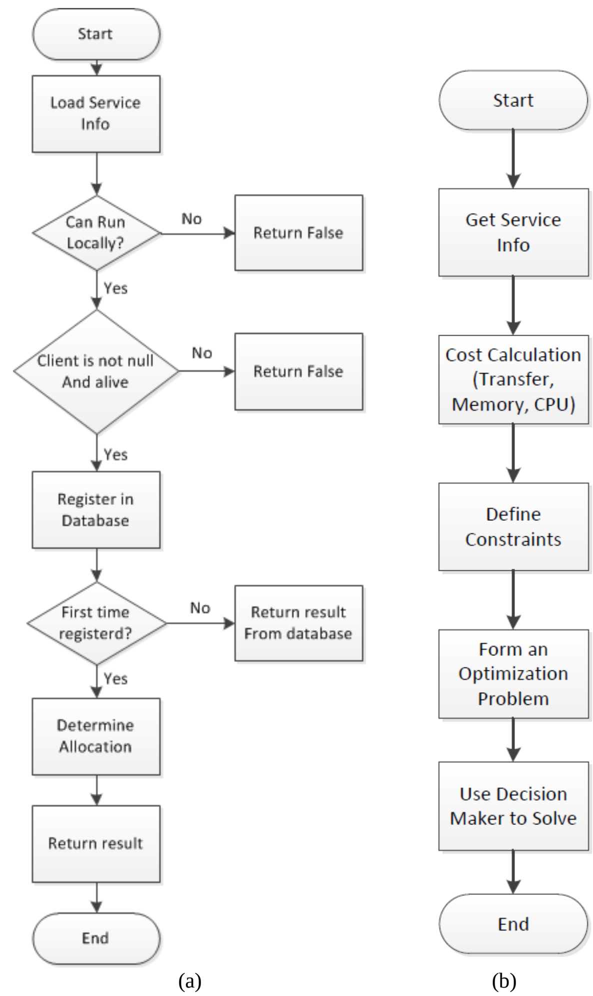


Fig. 9. Flow charts: (a) process registration and (b) allocation determination.

TABLE VII
EVALUATION RESULTS OF USE CASE 2 IN DETAILS

Dura-tion (s)	Execution Time (min)	Overhead (s)	CPU Instructions (Million)	Energy (kW)
10	14.8 / 0.7 / 1.9	1.5 / 5.0 / 4.6	7.4 / 1.8 / 2.1	0.9 / 0.1 / 0.2
20	27.6 / 1.0 / 12.3	1.5 / 2.9 / 5.9	7.2 / 2.6 / 5.5	1.7 / 0.1 / 1.0
30	42.6 / 1.8 / 25.0	1.4 / 7.3 / 5.9	8.3 / 4.1 / 7.6	2.6 / 0.1 / 2.0
40	62.5 / 2.2 / 70.0	1.9 / 5.5 / 5.8	10.1 / 5.7 / 8.4	3.7 / 0.2 / 5.5
50	77.7 / 2.8 / 122.5	1.8 / 5.1 / 7.2	11.8 / 7.8 / 9.6	4.7 / 0.2 / 9.6
60	96.7 / 3.3 / 152.5	1.9 / 5.8 / 6.5	13.3 / 9.0 / 11.1	5.8 / 0.3 / 11.9

Table 8 shows the number of possible solutions of the N-Queens problem in terms of N.

TABLE VIII
N-QUEENS PROBLEM SPACE

N	4	6	8	10	12	14
No. Distinct Solutions	2	4	92	724	14200	365596

ACKNOWLEDGMENT

This work is supported by the Excellence Initiative of German National Science Foundation (DFG) within the research cluster Ultra High-Speed Mobile Information and Communication (UMIC) and in part by NRW State within the B-IT Research School. We thank Tian Yu for the prototype implementation.

REFERENCES

[1] B.-G. Chun and P. Maniatis, "Augmented Smartphone Applications Through Clone Cloud Execution," in *Proceedings of the 12th Workshop on Hot Topics in Operating Systems (HotOS XII)*, pp. 1-8, Monte Verita, Switzerland, 2009. USENIX.

[2] E. Cuervo, A. Balasubramanian, D. Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl, "MAUI: Making Smartphones Last Longer with Code Offload," In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services (ACM MobiSys '10)*, pp. 49–62, San Francisco, CA, USA, 2010. ACM.

[3] I. Giurgiu, O. Riva, D. Juric, I. Krivulev, and G. Alonso, "Calling the Cloud: Enabling Mobile Phones as Interfaces to Cloud Applications," In *Proceedings of the 10th ACM/IFIP/USENIX International Conference on Middleware (Middleware '09)*, pp. 1–20, Urbana Champaign, IL, USA, November 2009. Springer.

[4] X. Gu, A. Messer, I. Greenberg, D. Milojicic, and K. Nahrstedt, „Adaptive Offloading for Pervasive Computing," *IEEE Pervasive Computing*, vol.3, pp.66–73, Jul. 2004.

[5] X. Gu, K. Nahrstedt, A. Messer, I. Greenberg, and D. Milojicic, "Adaptive Offloading Inference for Delivering Applications in Pervasive Computing Environments," in *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications (PerCom 2003)*, pp. 107–114, Dallas-Fort Worth, TX, USA, 2003. IEEE.

[6] R. Kemp, N. Palmer, T. Kielmann, and H. Bal, "Cuckoo: a Computation Offloading Framework for Smartphones," in *Proceedings of the 2nd International ICST Conference on Mobile Computing, Applications, and Services (MobiCASE 2010)*, pp. 59-79 Santa Clara, CA, USA, October 2010.

[7] R. Kemp, N. Palmer, T. Kielmann, F. Seinstra, N. Drost, J. Maassen, and H. Bal, „eyeDentify: Multimedia Cyber Foraging from a Smartphone," in *Proceedings of the 11th IEEE International Symposium on Multimedia (ISM 2009)*, pp. 392–399, 2009.

[8] S. Kosta, A. Aucinas, P. Hui, R. Mortier and X. Zhang, "Unleashing the Power of Mobile Cloud Computing using ThinkAir," *CoRR, abs/1105.3232, 2011*, informal publication.

[9] D. Kovachev, Y. Cao, and R. Klamma, "Mobile Cloud Computing: A Comparison of Application Models," *CoRR, abs/1107.4940*, 2011.

[10] K. Kumar and Y.-H. Lu, "Cloud Computing for Mobile Users: Can Offloading Computation Save Energy?," *Computer*, vol. 43, no. (4), pp. 51–56, April 2010.

[11] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," 2009.

[12] J. K. Ousterhout, "Scripting: Higher Level Programming for the 21st Century," in *IEEE Computer*, vol. 31, pp. 23–30, 1997.

[13] S. Ou, K. Yang, and J. Zhang, "An Effective Offloading Middleware for Pervasive Services on Mobile Devices," *Pervasive Mob. Comput.*,vol.3, pp.362–385, August 2007.

[14] M. Satyanarayanan, P. Bahl, R. Cáceres, and N. Davies, "The Case for VM-Based Cloudlets in Mobile Computing," *IEEE Pervasive Computing*, vol. 8, no. 4, pp. 14–23, October 2009.

[15] M. Turk and A. Pentland. "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, MIT Press, vol. 3, pp. 71–86, January 1991.

[16] L. Zhang, B. Tiwana, Z. Qian, Z. Wang, R. P. Dick, Z. Morley Mao, and L. Yang, "Accurate Online Power Estimation and Automatic Battery Behavior Based Power Model Generation for Smartphones," in *Proceedings of the Eighth IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis*, pp. 105–114. ACM, 2010.

[17] X. Zhang, S. Jeong, A. Kunjithapatham, and S. Gibbs, "Towards an Elastic Application Model for Augmenting Computing Capabilities of Mobile Platforms," in *Third International ICST Conference on Mobile Wireless Middleware, Operating Systems, and Applications*, pp.161-174, Chicago, IL, USA, 2010.

[18] X. Zhang, A. Kunjithapatham, S. Jeong, and S. Gibbs, "Towards an Elastic Application Model for Augmenting the Computing Capabilities of Mobile Devices with Cloud Computing," Mobile Networks and Applications, vol. 16, pp. 270–284, 2011.

[19] X. Zhang, J. Schiffman, S. Gibbs, A. Kunjithapatham, and S. Jeong, "Securing Elastic Applications on Mobile Devices for Cloud Computing," in *CCSW '09: Proceedings of the 2009 ACM Workshop on Cloud Computing Security*, pp. 127–134, Chicago, IL, USA, November 2009. ACM.

Dejan Kovachev has a diploma in computer science and engineering (MSc.) from "Ss. Cyril and Methodius" University in Skopje, Macedonia. Currently, he pursues PhD title at the information systems chair, RWTH Aachen University. He is supported by the B-IT Research School and the German excellence cluster for Ultra-high Mobile Information and Communication (UMIC). His research interests are mobile multimedia and cloud computing, metadata and context-awareness.

Ralf Klamma has diploma, doctoral and habilitation degrees in computer science from RWTH Aachen University. He leads the research group "advanced community information systems" (ACIS) at the information systems chair, RWTH Aachen University. He is coordinating and working in four major EU projects for Technology Enhanced Learning (ROLE, TELMAP, GALA and TELLNET), He is member of the research excellence cluster "Ultra High Speed Mobile Information and Communication" (UMIC) specialized in mobile multimedia. Ralf organized doctoral summer schools and conferences in Technology Enhanced Learning, and Social Network Analysis. He is on the editorial board of IEEE Transactions on Technology Enhanced Learning and Social Network Analysis and Mining (SNAM). His research interests are community information systems, multimedia metadata, social network analysis and technology enhanced learning.

Semantic Brokering of Multimedia Contents for Smart Delivery of Ubiquitous Services in Pervasive Environments

Alba Amato, Beniamino Di Martino, Salvatore Venticinque
Department of Information Engineering,
Second University of Naples, Aversa, Italy

***Abstract* — With the proliferation of modern mobile devices having the capability to interact each other and with the environment in a transparent manner, there is an increase in the development of those applications that are specifically designed for pervasive and ubiquitous environments. Those applications are able to provide a service of interest for the user that depends on context information, such as the user's position, his preferences, the capability of the device and its available resources. Services have to respond in a rational way in many different situations choosing the actions with the best expected result by the user, so making environment not only more connected and efficient, but smarter. Here we present a semantic framework that provides the technology for the development of intelligent, context aware services and their delivery in pervasive and ubiquitous environments.**

***Keywords* — Intelligent Agents, Context Awareness, Mobile Augmented Reality.**

I. Introduction

With the proliferation of modern mobile devices having the capability to interact each other and with the environment in a transparent manner, there is an increase in the development of applications specifically designed for pervasive and ubiquitous environments[14]. These environments are characterized by the presence of devices such as sensors, actuators, RFID tags, etc... which make them vastly interconnected, with distributed intelligence. Thus, context-aware applications are able to provide a service of interest for the user that depends on context information, such as the user's position, his preferences, the ability of the device and its available resources (memory, cpu, peripherals). Moreover these services have to be able to respond in a rational way in many different situations choosing the actions with the best expected result, so making environment not just more connected and efficient, but smarter. This leads, as a result, the improvement of the quality of life.

Location awareness is relevant in various fields, in order to provide more efficient services to the user. Popular examples provide support to the users on a visit of a museum or of an archaeological site, for shopping at the supermarket, on the go within a city, to a nurse for the administration of drugs, etc...

Here we present a semantic framework that provides the technology for the development of intelligent, context aware services and their delivery in pervasive and ubiquitous environments. Besides semantic-based techniques have been used for dynamically and efficiently data processing, retrieval and delivery. In order to provide a pervasive environment with advanced smart capabilities services, we integrated those technologies that enable objects of the real world to become part of the *Internet of Things*.

The paper is organized as follows: Section II introduces related work. The problem statement is described in Section III. In Section IV the framework is described and in Section V two case studies are presented. In the end conclusion and future work are discussed.

II. Related Work

Many research contributions focus on the development of software/hardware architecture and frameworks for mobile context-aware applications, also based on agent technology and semantic techniques.

Context Broker Architecture (CoBrA) project from Chen et al. [1] is an agent based architecture for supporting context-aware systems in smart spaces (e.g., intelligent meeting rooms, smart homes, and smart vehicles). Central to this architecture is an intelligent agent called context broker, which maintains a shared model of context on the behalf of a community of agents, services, and devices in the space, and provides privacy protections for the users in the space, by enforcing the policy rules that they define. However their approaches did not address various characteristics of context information such as classification and dependency. The Service-Oriented Context-Aware Middleware (SOCAM) project introduces by Gu et al. [2] is an architecture for the building and rapid prototyping of context-aware services. It provides efficient support for acquiring, discovering, interpreting and accessing various contexts to build context-aware services. The main contribution of this work is the use of an ontology that allows for the description of the context in a semantic way that is independent of programming language. However the

DOI: 10.9781/ijimai.2012.172

implementation can provide and process only contextual information that is associated the meeting domain.

In particular, several frameworks for development and delivery mobile context-aware applications are used in the domain of cultural heritage and recommendation systems. CIMAD [3] (Common Infrastructure/Context Influenced Mobile Acquisition and Delivery of cultural heritage data) is a framework supporting the development of cultural heritage “services” designed and implemented within the EU Network of Excellence EPOCH (Excellence in Processing Open Cultural Heritage). CIMAD aims to address a wide range of context-aware and multi-channel services in the cultural heritage domain [4], i.e. data acquisition, content delivery, monitoring, and management. The innovative aspect of CIMAD is that it helps the development team in mastering the complexity of building, customizing and integrating site specific cultural heritage services, thus making the different implementation activities more productive and cost-effective. This framework has many similarities with our work, but it does not focus on the backend support for experts in the domain of the Cultural Heritage in order to augment the archaeological site as an editor of that site or a guide for the tourist (i.e. map editing, semantic annotation and localization of multimedia contents).

The EasiShop framework [5] assists consumers with their shopping experience. The Shopping Agent is a mobile agent that is responsible for managing the shopping list and finding the best suited products that match consumers preferences, interacting with the retailers Retailer Agents that is hosted by the wireless device in the shops. Shopping Agent is able to cross-compare the products and their prices so making informed decisions and deriving recommendation that could be passed to the consumer. In our approach semantic is used for knowledge representation and management, but the ontology is designed by experts of the application domain. It is used both as a common vocabulary to support interoperability among heterogeneous remote services, local applications and users, and for intelligent discovery of media contents. Besides we also provide tools for experts in the domain of the application in order to augment the market (i.e. map editor, semantic annotator).

III. Problem Model

The formal objective of the framework is to define a model for delivering contents and services that provide the greatest utility for the user according to the knowledge that can be inferred using pervasive devices. The autonomic decision process is driven both by an updating process of the knowledge itself and by a set of rules that are defined by the expert of the application domain. The problem is modeled as an event driven system described by the following set of parameters:

-- A set of Events $E = \{(e_1, t_1), \cdots, (e_{ne}, t_{ne})\}$ represent perceptions about the environment, detected by the device at a certain time;

-- A set of Believes $B = \{b_1, \cdots, b_{nb}\}$, to be used for describing the knowledge in the application context is described by a domain ontology. The ontology has to be defined for the specific use case, by an expert of the application domain;

-- The user knowledge at the time t, about the environment and about his own preferences, is described by a set of concepts and individuals belonging to the ontology B.

For each user i we have a profile $P_i \subset B$;

-- A function $F : (E,P) \rightarrow P'$ updates the user’s profile when new events have been notified to the system;

-- A set of goals: $G = \{g_1, \cdots, g_{ng}\}$, whose achievement improves the user’s utility;

-- A set of actions: $A = \{a_1, \cdots, a_{na}\}$ can be performed by the system to achieve the goals, so increasing the user’s utility;

-- A set of triggers, one for each goal:
$\{\overrightarrow{w_1}(P), \overrightarrow{w_2}(P), \ldots, \overrightarrow{w_{nw}}(P)\}, \overrightarrow{w_i}: P \rightarrow B^{na.}$

W is a $[nw, na]$ matrix of boolean values, that must be defined to start a number of actions, when some believes of the user’s profile have been updated. A trigger is computed if the correspondent goal needs to be achieved. $A'^T = A^T W^T W I$, where I represents a $[na, nw]$ diagonal matrix of boolean values and A'^T is the resulting subset of action to be started.

For a chosen case study we have to define the set of perceptions, the ontology, the available actions, and the set of triggers. Let us define a multi-objective metric for the user utility as $U(G)$, that is the measure of the achievement of each goal. Objective of the framework is the maximization of such utility executing the minimum subset of actions that help to achieve the active goals.

The set of triggers must satisfy the following condition: $U(G(A)) > U(G(A \cup a_j)) \; \forall j: a_j \subset A - A'$.

In fact if an action cannot improve the user’s utility, it can bother the user and at least it affects the performance of the system. In a general scenario the function F can be implemented by an expert system, but if $E \in B$ it simply updates the user profile by adding, removing or replacing perception in the user’s profile P.

In the following we introduce a technological framework that allows for implementing the proposed model and two case studies.

IV. The Framework

In Fig. 1 the architectural solution of the framework is shown. The framework is composed of different tools and applications.

On the left side of Fig. 1, the user is using his device that hosts a light agent that is able to perceive information from the field by pervasive sensors. The local agent executes

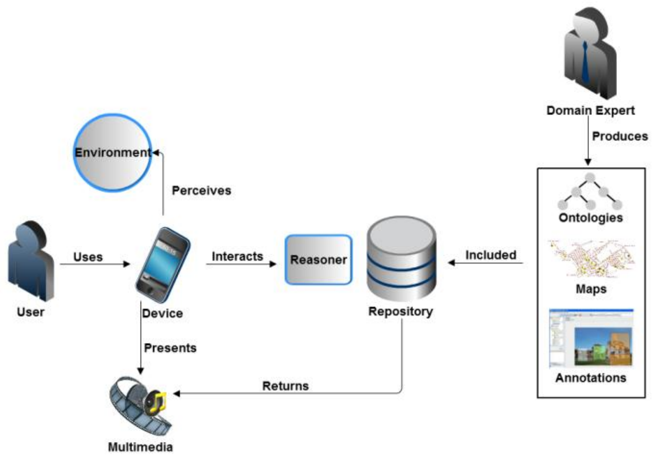


Fig. 1: Architecture and roles

autonomously and proactively in order to support the user's activity within the environment where he is moving. It discovers surrounding objects, it uses them to update the representation of the user's knowledge, reacts using the local knowledge to organize and propose the available contents and facilities by an interactive interface. If the connection works the device can access remote services, which can access a wider knowledge and have greater reasoning capabilities to look for additional contents and applications.

Remote applications implement context aware services. They use personal devices to collect perceptions and for content delivery. An ontology implements the representation of the global knowledge that is necessary to share a common dictionary and to describe the relationships among the entities/objects, which are part of the model. In our model a common ontology include all the general concepts, which are useful for a description of a pervasive environment where mobile users are moving, using their devices and interacting with available facilities and other users. The general ontology is complemented with a domain ontology that is designed by an expert of the specific application field. Experts of the application field define the domain ontology for the specific case study. They use or design a map to represent the environment. They add POIs to the map to geo-refer multimedia contents and can link them to concepts of the ontology. Furthermore they select relevant contents and annotate them using concepts and individuals of the ontology.

Concepts of the ontology are used at client side to describe a representation of the reality as it is perceived by the user. On the back-end the ontology is used to annotate digital resources like point of interests, contents, applications. It is also used to support reasoning. User's behaviors, information from pervasive devices or from other users, device properties, external events are heterogeneous data that are perceived by the device and that are used to build a dynamic changing representation of the user knowledge about the reality, within which he is moving. The applications are knowledge driven. The user's knowledge can be used by the application that is running on the device to adapt its logic locally, and it is updated remotely to improve the awareness of services at server side. Application are events based. Events can be updates of the user's knowledge or can be explicit service requests raised by the user. At each invocations semantic queries, that depend on the user's knowledge, are built and processed to get the action to be performed and the contents to be delivered. Results of the query are individuals of the ontology that are described by semantic annotation. The user's knowledge is composed of many semantic concepts with *static*, *dynamic* or *locked* properties. Components of the knowledge are:

-- *Device technology and capability*. Among the static properties here we means hardware resources, on board peripherals, display size, total memory and storage.
Dynamic ones can be power level, available memory and storage and bandwidth ;

-- *User's position*. It is a dynamic property that can change over the time and can be evaluated using different techniques and technologies, depending on the devices, on the available infrastructures and the kind of environment (indoor or outdoor);

-- *Pervasive objects*. They are dynamically discovery by the device. They can be sensors which provide information about the environment or can be used by the services itself by the device or eventually by thorough network if they are connected;

-- *Time information*. We intend the current time at user side and the time that the user is spending, or has spent within the environment.

-- *User's interests*. This part of the knowledge could be dynamically changed by the application according to the user's behaviors and to his feedbacks. The user could choose to start from an empty or a standard profile, to change it or to lock some properties interactively.

Semantic techniques are used for intelligent content and application discovery and delivery. Knowledge representation, ontology and annotations of digital resources are used to filter, organize and deliver contents and software to the device. Different techniques for reasoning can be experimented such as graph matching, description logics, neural networks or more simple ones like SPARQL queries. Here we integrate semantic techniques with discrete optimization methods to take into account constraints such as user's position and available time for exploitation.

V. Case Studies

Here we present two different case studies, which have been chosen to test the proposed model and the developed framework.

A. Cultural Heritage

Exploitation of archaeological sites can be very difficult because of a lack of supporting infrastructures and because of the complex recognition and comprehension of the relevant ruins, artworks and artifacts[10]. The availability of personal devices can be used to plan and support the tourist by suggesting him the itineraries, the points of interest and by providing multimedia contents in the form of digital objects which can semantically augment the perceived reality. In this context relevant issues are the profiling of the user, the discovery and the delivery of the contents that can improve the user's satisfaction, new models of interactions with reality. Three case studies have been chosen to test the approach and the framework [12]. The S. Angelo in Formis Basilica, in Campania, near S. Maria Capua Vetere, the ancient town of Norba and the amphitheater of Capua. In these sites we cannot install complex infrastructures and they are difficult to be understood without a tourist guide. For the presented case studies we need to provide a technological solution that does not need infrastructures for letting the software know the user location and his feeling about the environment. It means that Bluetooth, RFID, GPS, electronic compass, camera, network connection and others are the technologies which have to be used together or independently to get information about the user perceptions and to augment his exploitation of the archaeological site. The first thing to be considered when modeling this kind of problem is the environment. In fact we have to model the environment where the user is moving and to reconstruct the perceptions of the user himself in order to get his particular vision about what is surrounding him. A real representation of the environment is necessary to identify landmarks and possibilities of intervention using pervasive actuators and sensors whose input will be stored in the memory of the agent and will be updated as the environment changes. The environment will be modeled as a geo-referred map with itineraries, landmarks and points of interest. The user will be able to download the map of the area to be visited at home, before to leave, or on site, if the network will be available. The map will include all the points of interest that identify the relevant objects of that area and different cultural itineraries which could be exploited on site. Also contents can be discovered and downloaded in advance according to the device capabilities. On board software and remote services will assist the cultural visit by augmenting the reality by the user's personal device. A software agent executes on the user's device to support services exploitation. It percepts the surrounding environment using the on-board peripherals and executes plans which are chosen by an ad-hoc reasoning to optimize the user's satisfaction. The knowledge of the environment acquired by the agent represents part of its own beliefs. Another set of believes describes user's knowledge about the surrounding environment that can be acquired by using peripherals of his personal mobile device, by recording and evaluating user actions or explicitly asking for user feedbacks. Some examples are user's position, interest, nearby objects, landscape, etc. Interaction with the environment and presentation of contents can be done again using user's device and his peripherals. Of course the way of localization of users and objects depends on the device technology, the available infrastructures and the kind of environment.

Indoor or outdoor localization can be implemented using heterogeneous technologies, and often absolute localization can not be performed, but it is only possible to detect nearby landmarks or objects. Besides as the device is a PDA, it has limited battery, storage and computation capacities so also the agent running on it has limited capabilities.

In order to extend the agent's capability another remote intelligent agent interacts with the local one to achieve individual or shared goals. This two-agent system is part of a loosely coupled network of problem-solver entities that work together to find solution to problems that are beyond the individual capabilities or knowledge of each entity [6]. In fact the agent running on the local device is able to collect perceptions, to communicate believes and to execute actions locally. The remote one can access a wide knowledge base and can perform a more complex reasoning. Of course different components of the user's knowledge could have different weights according to their relevance or the time of perception. We could consider the knowledge at the current time, in a time windows, or its complete history. In order to augment user's knowledge and capability to interact with the environment, services have to choose, according to their context awareness[11]; i) what content and application it has to deliver; ii) when it needs to present the content; iii) how this should be done. Finally another set of believes is the set of contents in the remote knowledge base which are available for delivery, to be used for augmenting the reality perceived by the user. The maximization of user satisfaction represents the goal to achieve. With this aim agents have to make a decision about the optimal set of contents and their organization according to user's profile, device and time availability.

B.Smart Market

The second proposed case study is an e-commerce service that profiles the customers and supports their purchases recommending products and shops [13]. On the other side the service is used by vendors, who upload information about their offers, shops and sales. The choice of this application field is due to the fact that the rising usage of mobiles as the primary route for consumer interaction with brands, and its importance as a catalyst for content sharing and recommendation, are still at odds with the level of focus and budget currently afforded to mobile in the market field. Additional advantages in this field brought by the technology are the traceability, which will help the consumers to verify the origins of the products. Besides, in case of a food related disease, the traceability of the purchased food will enable faster detection of the origin of the infection and thus contain its impact better and faster. A QRCode or an electronic tag can be read by the device and may carry out information related to drugs, making it easier for the customer to be acquainted with adverse effects and optimal dosage.

Besides, in case of an accident because of drug abuse or misuse it will be possible to quickly identify the taken drug by asking the smart device, which may also inform about the right antidote and dosage to enable the emergency treatment to react faster and better, and thereby saving lives.

Let us imagine users of different ages and habits, and with different purposes, in a market. Each of them has a smartphone with a smart application that perceives the environment, provides simple facilities, and accesses the e-service. All the customers continuously update, by the smart application, their profile in the system. They can insert information about themselves such as name, sex, age, home country, e-mail address, languages spoken, level of education, lists of interests, etc. They also set a few preferences about what information the device can autonomously communicate, such as position, nearby objects, user's behaviour. Customers use a grocery list by their smartphone updating their shopping chart. Vendors add or remove their marketplaces and products, and update their description in the knowledge base, also by semantic annotations. Just to simplify our requirements, we suppose that customers cannot add, describe and recommend products on their own. Using the autonomic behavior of the service, the objective of the system is to propose the best set of recommendations, which can help the user to improve his own utility, exploiting the available information about products, about the environment within which customers are moving and about the customers themselves.

VI. Context Aware Semantic Discovery

For sake of simplicity, in both the proposed case studies we suppose that the perceptions notified by the device to the remote services are the user's location, his/her available time, an concepts of the ontology. A common goal is the maximization of the user's satisfaction by proposing the best set of contents, which are relevant to the user's interest. Each time a new perception has been received, the user's profile is evaluated in order to eventually trigger the discovery service that updates the optimal set of contents to recommend.

The semantic discovery service provided by our framework returns a set of digital objects related to POIs in the pervasive environments. Each content is annotated by concepts from the ontology and can be discovered by a SPARQL query to the content repository. The result of the query is a set of *N* instances of digital objects whose relevance to the user context has not been considered yet. The context awareness of the services is exploited by computing the relevance of each content to the profile *p*, which is the semantic representation of the context, including interest, position, etc.. The annotation and the user's profile are represented using the Vector Space Model (VSM). VSM is a model for semantic representations of contents as vectors of items created by G. Salton [9]. In our case $\overrightarrow{a_j}$ is the vector of concepts $c_{i,j}$ of the domain ontology.

-- $\overrightarrow{a_j} = < \{c_{1,j}, o_{1,j}\}, \{c_{2,j}, o_{2,j}\}, ..., \{c_{l,j}, o_{l,j}\} > \quad \forall j = 1...N$

-- $\vec{p} = < c_1, c_2, ..., c_m >$

Sizes *l* and *m* are the number of different concepts that appear in the annotation and in the profile . If a term occurs in the annotation, $o_{k,j}$ is the number of occurrences of that term in the annotation.

Sizes *l* and *m* are the number of different concepts that appear in the annotation $\overrightarrow{a_j}$ and in the profile $\vec{p}$. If a term occurs in the annotation, $o_{k,j}$ is the number of occurrences of that term in the annotation.

We defined a score $A(\overrightarrow{a_j}, \vec{p})$ to measure the relevance of an annotation $\overrightarrow{a_j}$ to the profile $\vec{p}$ by the following formula (1)

$$w_j = w(\vec{a}_j, \vec{p}) = \sum_{k=1}^{l} r_k \qquad (1)$$

where

$$r_k = o_{k,j} * \sum_{i=1}^{m} \frac{1}{d_{k,i}+1}$$

where $d_{k,i}$ is the minimum number of edges that connect the node representing the concept $c_{k,j}$ of the annotation to c_i of the profile. In (1), for each item $c_{k,j}$ of the vector $\overrightarrow{a_j}$ the relevance to the profile $\vec{p}$ is computed by adding the relevance of that concept to each concept of the profile, and by multiplying each contribution for the number of occurrences $o_{k,j}$. The relevance between two concepts is calculated by dividing 1 by the number of edges of the ontology that connect the node representing the concept c_k of the profile to $c_{k,j}$ plus 1. As a result we have a score for each annotated item that is associated to a POIs so that it is possible to order the items and the POIs according to user's preferences.

However a user has some additional constraints.

In the case of the archaeological guide the contents to be delivered can be limited by the device technology or the available time for the experience.

In the case of smart market, constraints can be the price, the distance he wants to go away, the purchase frequency, etc. For this reason it is necessary, to select the valid recommendations excluding, those that cannot be enjoyed according the current constraints. For example it is necessary that the the budget does not exceed the amounts that is available to the user. Here we aim at recommending the best products that: i) maximize the score, ii) are compliant with the user's budget, iii)can be purchased without exceeding time and space limits.

This problem can be reduced to a *discrete optimization problem* that consists in searching the optimal value (maximum or minimum) of a function $f : \vec{x} \in Z^n \rightarrow R$, and the solution $\vec{x} = \{x_1, . . . , x_n\}$ in which the function's value is optimal. $f(\vec{x})$ is said *cost function*, and its domain is generally defined by means of a set of *m* constraints on the points of the definition space. Constraints are generally expressed by a set of inequalities:

$$\sum_{i=1}^{n} b_{i,j} x_i \leq a_j \quad \forall j \in \{1...m\} \qquad (2)$$

and they define the set of feasible values for the x_i variables (the *solutions space* of the problem). In our case:

$w_i \geq 0 \; \forall j = 1...N$, $W = \{\overrightarrow{b_1}, .., \overrightarrow{b_m}\}$

where:

-- w_i represents a score and

-- *B* a set of constraints $\overrightarrow{b_i}$, each one composed of *N* +1 integer.

We have to compute:

$$max \sum_{k=1}^{N} w_i x_i$$

so that (3)

$$\sum_{i=1}^{N} b_{i,j} x_i \leq b_{j,N+1}$$

with $x_i \in \{0, 1\}$ $\forall i = 1...N$. The goal is to maximize the value delivered. The vector $\vec{x}$ represents a possible solution: its components x_i are 1 or 0 depending on whether the object is included or not in the best set. To set the constraints we create a table with a column for each content and a row for each requirement. The rows of the matrix *B* are contain in each cell $b_{i,j} = 1$ if the requirement is necessary for the content in that column, $b_{i,j} = 0$ otherwise.

In Table II we specify price and shopping time for each product.

TABLE II
RAPPRESENTATION OF PRODUCT PROPERTIES AND USER'S CONSTRAINTS

	C_1	C_2	...	C_n	User
Price	10	20		15	35
Time	10	5		5	15

Furthermore in the last column of the table we have $b_{i,N+1} = 0$the budget and the time availability of the user. For example, the first row tells us that the user can spend not more than 30 , so it is necessary to exclude one content. The same consideration must be done for the second row. Finally we have to exclude that product that affects the affinity with a minimum penalty. Further details about the algorithm and its performance evaluation are out of the scope here.

In the following section we detail the technological choices which have been taken to implement each component of the framework.

VII. Technological Implementation

In the following sections, we present the techniques and technology choices, explaining their role in the overall picture. We refer to the introduced case studies to explain how the framework can be used in different fields.

A.Environment Map

To provide a description of the environment within which the user is moving we need a geo-referenced map that describes buildings, roads, bans, itineraries in the case of cultural heritage or departments, sectors, shelves, routes in the case of smart marketa and, in general, Points of Interest (POIs). We used the OpenStreetMap format to design open maps. In Fig. 2 the way we used to build the map of a supermarket is shown.

In the picture there are two different views of the JSOM tool, the image level and the geo-referenced map under construction. The user can choose to build the map following the edge of a reference image or can import other vectorial and geo-referenced formats like GPX. Paths, walls and POIs are outlined. Each point of interest represents departments, sectors, shelves, or any other entities of relevance and can be described using a list of keyvalue pairs. The tool allows to export the map in an open format that can be read and used by the client application.

A. Ontology

An ontology has been designed to describe the sites of interest and to annotate the related media. A general part includes the concepts which are common to all the class of applications that can be modeled according to the proposed approach.

Among the others the *Time* class and its properties (*CurrentTime*, *ElapsedTime)*, allow to organize and assist the user taking into account time information and handling time constraints. *Position* class and its properties allow to localize the user and objects around him. An application specific part of the ontology includes the concepts that belong to the domain in question.

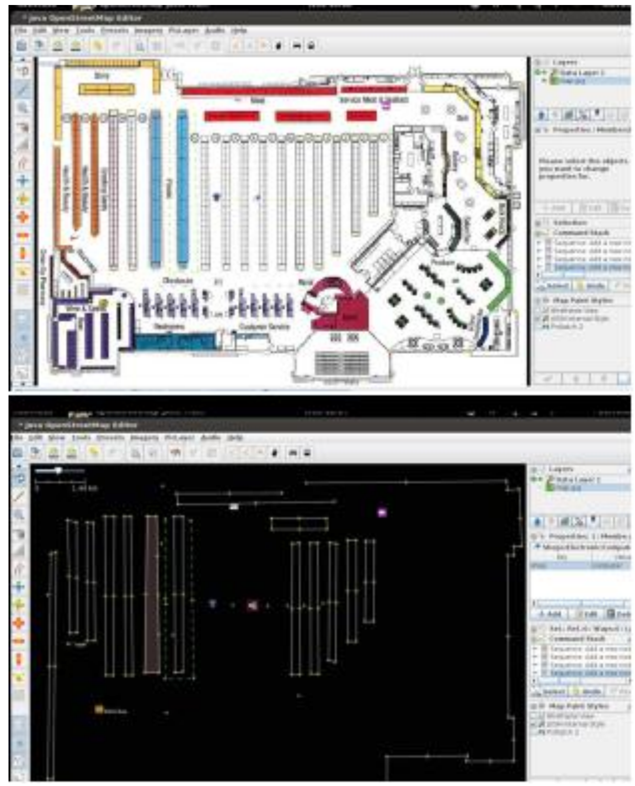


Fig. 2: Marketplace map

In the construction of ontology for the case of the archaeological guide it has been developed by experts of the cultural heritage, and we avoid here to go in deep on details which are out of our knowledge. In the case of the smart market case study, we used the following classification based on the buying habits of consumers [7]. In fact an appropriate classification allows to make better recommendations, since it is possible to identify the type of purchase made by the customer, the frequency he buys a certain product, and on the other hand it is possible to make more accurate cross-sell suggestions.

-- *Convenience products* are purchased frequently by the consumer, they do not require any major effort in the process of choice and any risk of making the wrong choice. Often, their purchase is systematic, and affordability. This also facilitates impulse buying by the individual. They are commonly illustrated by commodities, *unsought* (emergency) items, and impulse products. Examples of consumer goods that fall into the convenience category include fresh produce and grocery staples, umbrellas, gum, and batteries.

-- *Preference products* are distinguished from the previous item primarily on the basis of a perceived greater risk in their purchase. The individual, mainly because of marketing efforts, however, tends to address the problem of the choice of the brand, trying to find one that can better meet his needs. The most prominent examples of preference products are in the consumer package goods industry.

-- *Shopping products* are goods that the individual acquires after careful considerations, and after a prolonged comparison among different brands. These assets may be slightly different from the others, making the price the main driver of choice, or differentiated, making their attributes to play a key role in purchasing decisions. Examples of shopping goods are clothes, and furniture for end consumers, and equipment and components parts for industrial users.

-- *Speciality products* are goods with special features, often characterized by prestigious brands and very high prices. Examples of speciality goods include vintage imported wines, expensive sports cars, and paintings
by well-known artists.

-- *Unknown products* are not known by the consumer , or the consumer would not show any interest if it were not informed about them through advertising.

In [8], three product types are identified: *convenience*, *shopping*, and *speciality* and three steps of purchase/consumption are considered: *pre*, *during*, and *post* transaction. In Table 1 each stage is further divided into areas, in which consumers may seek benefits. Information discovery is more important in the case of shopping, rather than convenience products. Each cell in the upper half of the table depicts whether that particular benefit is important in the context of the given product type.

TABLE I
PRODUCT CLASSIFICATION: THE MATCH BETWEEN PRODUCT TYPES AND MEDIA ATTRIBUTES

	Pre-Transaction		During Transaction	
	Information Search	**Product Trial**	**Payment Options**	**Immediate Delivery**
Convenience products	*Not Important*	*Not Important*	*Important*	*Important*
Shopping Products	*Important**	*Important*	*Important*	*Moderately Important*
Speciality Products	*Important***	*Important*	*Important*	*Moderately Important*

	Post-Transaction			
	Installation Service	**Consumption**	**Return**	**After Sales Enhance-ment**
Convenience Products	*Not Important*	*Important*	*Not Import ant*	*Not Important*
Shopping Products	*May be Important*	*Important*	*Import ant*	*May be Important*
Speciality Products	*May be Important*	*Important*	*Import ant*	*May be Important*

Since the goal here is not to design a complete profile, for the creation of the ontology we focused on general concepts that make it possible to achieve the successful personalization we aim for. We identified two main disjoint classes: *Product* and *Customer*. The *Customer* class contains personal information and has two disjoint subclasses: *Member* and *Non member* to distinguish the users that are regular customers enrolled from the others. Personal information, inserted as datatype properties, consists of information common for all users that is useful to identify and partially segment users. Datatype properties included are *hasName* (type String), *hasAge* (type int), *hasEmail* (type String) and *hasGender* (type Man or Woman). In Fig. 3 a picture the relevant concepts of the ontology for a simple use case of Smart Market are shown.

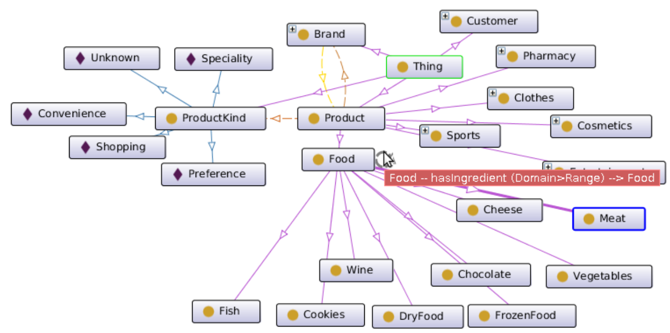


Fig. 3: Smart Market Ontology

The *Product* class has six subclasses: *Food*, *Cosmetics*, *Sports*, *Clothes*, *Entertainment*, *Pharmacy*.

The *isProductType* object property supports a classification based on the purchasing habits of consumers according to the model explained before. Besides *Food* class has the data properties: *isGlutenFree (type boolean) indicates the presence of gluten*, *hasExpirationDate* (type Date Time) registers the sell-by date expiration and *isFrom* (type string) provides information about the production chain. Another part of the ontology describes the type of purchase: *Convenience* , *Preference*, *Shopping* or *Speciality*. In particular Convenience products are characterized by the *hasPurchaseFrequency* that can be instantiated for each couple *(product, user profile)*.

B. Annotation

The ontology is used also for annotating the multimedia contents. To annotate texts, images and any kind of contents we chose the AktiveMedia tool.

In Fig. 4 a picture of the Amphitheater of S. Maria Capua Vetere is annotate with the *Column* and the *Arc* classes which are part of this kind of building. The output produced by the annotator is an RDF file with concepts and properties of the AktiveMedia ontology and of the domain ontology. The same has been done for the smart-market, where images show those products to be recommended.

C. Digital Repository

The Fedora repository is used to store digital objects and supports their retrieval. Into the Fedora repository a digital object is composed of a set of files which are:

-- *object metadata*: used by the client application to understand how to deliver the content;

-- *binary streams*: which are images, video, text ... any kind of raw information to be delivered;

-- *RDF annotation*: that describe the semantic of the object according to the ontology;

-- *disseminations*: filters to be eventually used for adapting the object according to the target client.

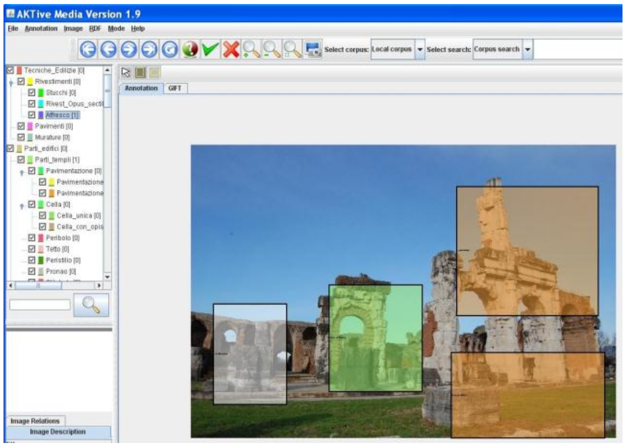


Fig. 4: The annotator

We loaded the Aktive-Media ontology and the domain ontology into the Fedora repository in order to exploit its embedded SPARQL engine that is used to select the optimal set of individuals (i.e. contents). Multimedia contents are automatically stored into the repository after the annotation phase. The RDF output is automatically processed using an XSL transformation to make it compliant with the model used by the Fedora repository.

D. Content's Delivery for the Mobile Archaeological Guide

The diversity of media content to deliver and their quality, is especially relevant in the case of the archaeological guide. In this case we have defined the following content models.

-- Multiple images whose transparency can be graduated by the user to compare changes in different periods. In the same way real pictures can be compared with paintings. Old picture can be compared with what is seen by the camera.

-- Parts of the image acquired by the camera are recognized and linked to related multimedia contents.

-- Virtual reconstructions which are synchronized with the camera output or the detected RFIDs.

-- Text, audio, video and composite media. A content descriptor is attached to every digital object. It is used by the device when the content is being delivered. The descriptor defines the right player for that media, configuration parameter and necessary input. In Fig. 5 an example of delivered content is shown. The user focuses the camera on a particular view of the S. Angelo in Formis Basilica and sees the original temple of *Diana Tifatina*. In particular the perspective viewed by the camera is the same shown by the video.

E. The Client for Mobile Archaeological Guide

At client side we extended the android version of an open source software navigator called Navit. The navigator provides basic facilities for map visualization and to guide the user along some itineraries by using the on-board GPS receiver, or other technologies. By new extension the guide is able to sense the environment by the available peripherals, to understand the situations according to which the visit will be adapted, and to enhance user's experience.

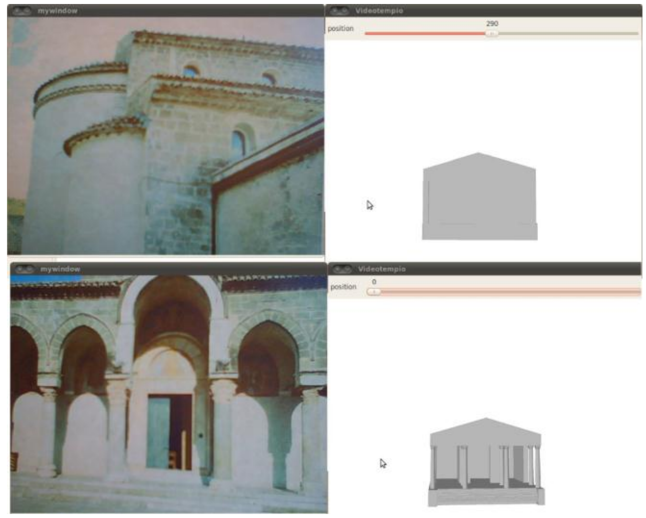


Fig.5: Synchronized video

Even if many experiences on augmented reality are proposed nowadays, usually the exploitation of vision is used only to overlap real and virtual images. Our extensions provide the following functionalities:

-- access to device peripherals to sense the environment and to update the representation of the user's knowledge about the reality around him;

-- access to remote services to update the user knowledge and to ask for available contents

-- a local cache of objects for exploiting the visit without connection;

-- a limited reasoner that is able to organize the content by itself when the connection does not work.

Perceptors are implemented by:

-- GPS positioning to localize the user in open spaces and to guide him on cultural itineraries;

-- RFID for positioning and detection of nearby POIs. This technology can be used to alert the user but also for his positioning in indoor environments;

-- CODBAR recognition to get information about artifact, monuments, ... when a RFID reader is not available;

-- image recognition by search by sample techniques which are speed-up and improved using a position based filtering;

-- monitoring of device resources and configuration;

-- collection of user interest by feedback and by an analysis of his behavior;

-- time monitoring.

On the other hand a list of functionalities are provided to deliver contents and to guide the visit. Content management (discovery, organization and fruition) is supported both at client side and at server side. Knowledge visualization and management are provided to allow the explicit specification of user's own interest. The output of the camera is used as a component of the user's knowledge as well as a map on which semantic additions are anchored, not simply superposed. The user will be able to ask for and exploit available multimedia contents, which are related to points of interest, or to personal interests expressed by semantic concepts. Multimedia contents will be adapted at server side according to the device/user/session profile to provide to the user the best quality of service. Fig. 6 shows the result of the content discovery service. The service invocation starts the camera that is used to take a photo of the landscape or of a subject of interest. In background, the client uploads the image and waits for the content retrieval. The discovery of relevant contents and the download of retrieved information runs in background meanwhile the user is interacting with other facilities provided by the client. At server side the picture is received together with the user's position. The received sample image is processed to extract a list of properties, which are invariant with the color, the zoom and little rotations by the OpenCV library. The server owns a list of images of landscapes and relevant points of interest the user can shoot. These images have already been processed by the same OpenCV library. The user's position is used to perform a filtering of the points of

Fig. 6: The Client

interest close to him, in order to speed-up the comparison of sample features with the ones of candidate images. The best matching images are returned. If they have been annotated, annotations are used to suggest other contents that are semantically related. Furthermore frames of the images, which have been semantically annotated, can be discovered and pointed in the picture that has been just taken. Other events, which open dialog for suggesting actions, itineraries, POIs or media can be related to new perceptions.

F. *The Client for Smart Market*

To access the smart market the user will exploit, by his smartphone, an Android application using an interface shown in Fig. 7 that is able to get and communicate his position and to localize POIs or entities close to him. Furthermore the user is able, by the application, to search for products in a grocery list and to add the products to be bought in his shopping cart. Let us suppose that the user wants to buy cookies. Its device updates periodically a remote agent about the user's position. Each time this component of the user's profile changes, the agent checks for a POIs that is linked to a product semantically related to the cookies concept.

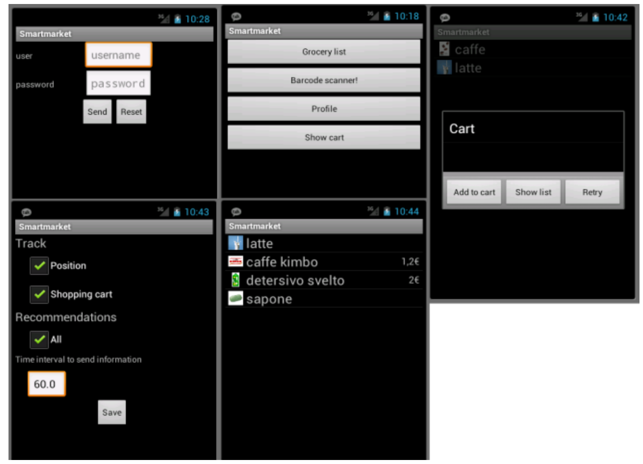


Fig. 7: The Android Client

If the POIs is found the user is notified. Starting from the POI the user can browse the list of products, which are sold at that position. In the same scenario we can imagine that the user has bought the product. Another client application of our smart service is able to recognize the user, by a fidelity card, or by any kind of personal identification mechanism, without getting any additional private information. Scanning the barcode of the purchased product the smart market service can be informed about new convenience goods in the user's habit. A new timer will be created to alert the user at the sellby expiration date of the cookies. The purchase frequency is not known and can be put equals to the expiration date. When the same, or an equivalent product, will be bought, the frequency timer will be updated and set again. Finally a list of concepts have been added to the user's profile for personalizing the discovery of similar products to recommend. In the provided example we will have *glutenFree*, *chocolate*, the *cookies brand*. At the expiration of the timer the user will be alerted about the eventual opportunity to buy the same cookies, but alternative products will be recommended according to the context aware semantic discovery strategy described in the following section. In particular among the available products the system recommends the one annotated with the *cookies* concept, the *hasBrand* property, the *glutenFree* property, and a *hasIngredient* property. Annotated texts and images are currently presented by the web interface shown in Fig. 8. In particular you have on the left the list of relevant concepts, and the picture of the product on the right.

At this point the user can move the mouse over the concepts and see the annotation on the picture. He can also move the mouse over the picture and see a tool-tip that explains the annotation.

Concepts from the list, or from the tool-tip, can be easily added to the user profile in order to let the system learning about the user's preferences for a personalized recommendation. Using the mouse button or the form is also possible to search for other products, which are semantically related to the selected concepts, in order to choose an alternative brand or similar kinds of cookies.

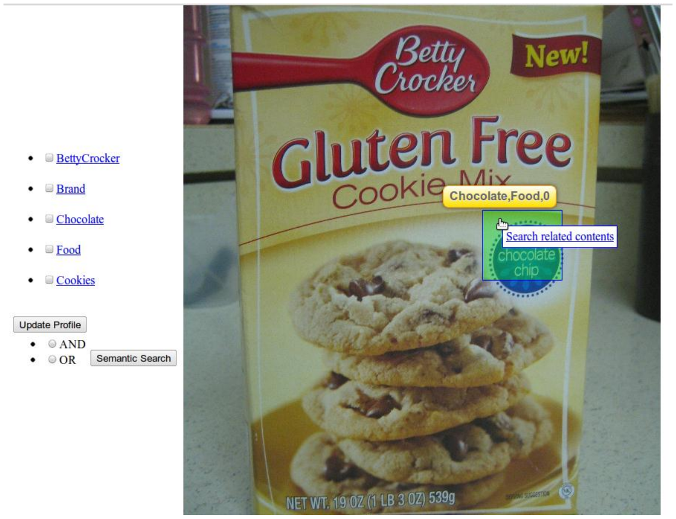


Fig. 8: Semantic browsing of a recommended product

VIII. Conclusion

In this paper, we proposed a proactive semantic framework for the intelligent delivery of services in pervasive environments. In such environments, users need to have access to the most relevant and interested services within a rich and dynamic context. The main goal is the exploitation of pervasive devices to maximize the expected satisfaction of the user by adapting applications and contents to be delivered and by augmenting and improving his utility in several contexts. Personal devices are used to interact with the environment, to run interactive applications and to present contents to the user. A smart application runs on the user device, uses the peripherals to perceive the environment, provides context awareness to remote discovery services. We described a framework that implements the proposed approach and described its application for cultural heritage and shopping recommendation. Future works will address quantitative analysis of the approach and the support of collaboration among users by personal semantic based recommendation.

Acknowledgment

Author thank Massimiliano Diodati and Vincenzo Reccia for their technical support.

References

[1] H. Chen, “An Intelligent Broker Architecture for Pervasive Context-Aware Systems”. *PhD thesis, University of Maryland*, Baltimore County, 2004.

[2] T. Gu, H. K. Pung, D. Q. Zhang, “A service-oriented middleware for building context-aware services”, Journal of Network and Computer Applications, Vol . 28, pp. 1-18, Jan. 2005.

[3] Manzaroli D., “Architectures for Context Aware Services in Smart Environments”. *PhD thesis,* May 2010

[4] G. Raffa, P. H. Mohr, N. Ryan, D. Manzaroli, M. Pettinari, L. Roffia, S. Bartolini, L. Sklenar, F. Garzotto, P. Paolini, T. Salmon Cinotti, “CIMAD - A Framework for the Development of Context-aware and Multi-channel Cultural Heritage Services”, in ICHIM07 September 30, 2007 Toronto

[5] S. Kurkovsky and K. Harihar, “Using ubiquitous computing in interactive mobile marketing,” *Personal Ubiquitous Computing*, vol. 10, no. 4, pp. 227–240, Mar. 2006.

[6] V.R., Durfee, E.H., Lesser and Corkill, D.D., *Trends in Cooperative Distributed Problem Solving*, IEEE Transactions on Knowledge and Data Engineering, Vol. March 1989, KDE-1(1), pages 63-83.

[7] E. B. Murphy, P., “Classifying products strategically,” *Journal of Marketing*, vol. 50, no. 3, pp. 24–42.

[8] B. Sindhav and A. L. Balazs, “A model of factors affecting the growth of retailing on the internet,” *Journal of Market-Focused Management*, vol. 4, pp. 319–339, 1999.

[9] G. Salton and M. E. Lesk, “Computer evaluation of indexing and text processing,” *J. ACM*, vol. 15, no. 1, pp. 8–36, Jan.1968.

[10] A. Amato, B. Di Martino, S. Venticinque “BDI Intelligent Agents for Augmented Exploitation of Pervasive Environments”. In: Proceedings of the 12th Workshop on Objects and Agents. Università della Calabria, 2011, vol. 741, p. 81-88, ISBN/ISSN: 1613-0073

[11] S. Venticinque, A.Amato, B. Di Martino Ontology driven Augmented Exploitation of Pervasive Environments. In: AFIN2011: The Third International Conference on Advances in Future Internet. Nice/Saint Laurent du Var, France, August 21-27IARIA, p. 41-48, ISBN/ISSN: ISBN: 978-1-61208-148-9

[12] A. Amato, B. Di Martino, S. Venticinque “Semantically Augmented Exploitation of Pervasive Environments by IntelligentAgents”. In: Mue 2012: The Sixth International Conference on Multimedia and Ubiquitous Engineering. Madrid, Spain, July 10 – 13, ISBN/ISSN: ISBN: 978-0-7695-4701-5/12

[13] A. Amato, B. Di Martino, S. Venticinque (2012) A Semantic Framework for Delivery of Context-Aware Ubiquitous Services in Pervasive Environments. In INCoS 2012: 4-th International Conference on Intelligent Networking and Collaborative Systems. Bucharest, Romania, September 19 – 21, 2012, ISBN/ISSN: ISBN: 978-0-7695-4808-1/12

[14] Aversa, R., Di Martino, B., Venticinque, S. Distributed agents network for ubiquitous monitoring and services exploitation (2009) Proceedings - 12th IEEE International Conference on Computational Science and Engineering, CSE 2009, 2, art. no. 5284221, pp. 197-204.

Alba Amato is a PhD student in Computer Engineering at the Information Engineering Department of the Second University of Naples, Italy. She graduated with a Masters degree in Computer Science from the University of Napoli Federico II. Her research interests focus on intelligent systems, agent architectures, multi-agent systems and their applications in cloud computing.

Salvatore Venticinque is an Assistant Professor at Department of Information Engineering of the Second University of Naples. He received his PhD in Electronic Engineering in 2003. He has been an Adjunct Professor in Computer Programming and Computer Architecture in regular courses. He is involved in research activities dealing with parallel and grid computing and mobile agents programming for distributed systems.

Beniamino Di Martino is Full Professor of Information Systems at the Second University of Naples (Italy) since 2005. He participated to various research projects supported by national and international organizations. He is vice Chair of the Executive Board of the IEEE CS Technical Committee on Scalable Computing. His research interests include: Knowledge Discovery and Management, Semantic Web and Semantic Web Services, Semantic based Information Retrieval, Cloud Computing, High Performance Computing and Architectures, Mobile and Intelligent Agents and Mobile Computing, Reverse Engineering, Automated Program Analysis and Transformation, Algorithmic Patterns Recognition and Program Comprehension.

A fuzzy c-means bi-sonar-based Metaheuristic Optimization Algorithm

[1]Koffka Khan, [2]Ashok Sahai,
[1]*Department of Computing and Information Technology,* The University of the West Indies, St. Augustine Campus @ TRINIDAD,
[2]*Department of Mathematics and Statistics,* The University of the West Indies, St. Augustine Campus @ TRINIDAD

***Abstract* — Fuzzy clustering is an important problem which is the subject of active research in several real world applications. Fuzzy c-means (FCM) algorithm is one of the most popular fuzzy clustering techniques because it is efficient, straightforward, and easy to implement. Fuzzy clustering methods allow the objects to belong to several clusters simultaneously, with different degrees of membership. Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership. However FCM is sensitive to initialization and is easily trapped in local optima. Bi-sonar optimization (BSO) is a stochastic global Metaheuristic optimization tool and is a relatively new algorithm. In this paper a hybrid fuzzy clustering method FCB based on FCM and BSO is proposed which makes use of the merits of both algorithms. Experimental results show that this proposed method is efficient and reveals encouraging results.**

***Keywords* — Fuzzy, Clustering, Bi-sonar, Metaheuristic, Optimization.**

I. INTRODUCTION

Clustering is the process of assigning data objects into a set of disjoint groups called clusters so that objects in each cluster are more similar to each other than objects from different clusters. Let {x(q): q = 1,…,Q} be a set of Q feature vectors. Each feature vector $x(q) = (x_1(q), \dots, x_N(q))$ has N components with weights $w(q) = (w_1(q), \dots, w_N(q))$ and distances metrics $D(q) = (d_1(q), \dots, d_N(q))$. The process of clustering is to assign the Q feature vectors into K clusters {c(k): k = 1, …, K} usually by the minimum distance assignment principle. Choosing the representation of cluster centers (or prototypes) is crucial to the clustering. Feature vectors that are farther away from the cluster center should not have as much weight as those that are close. These more distant feature vectors are outliers usually caused by errors in one or more measurements or a deviation in the processes that formed the object.

The simplest weighting method is arithmetic averaging. It adds all feature vectors in a cluster and takes the average as prototype. Because of its simplicity, it is still widely used in the clustering initialization. The arithmetic averaging gives the central located feature vectors the same weights as outliers. To lower the influence of the outliers, median vectors are used in some proposed algorithms. To be more immune to outliers and more representatives, the fuzzy weighted average is introduced to represent prototypes:

$$Z_n^{(k)} = \Sigma_{\{q:\, q\in k\}} w_{qk} x^{(q)}_n \qquad (1)$$

Rather than a Boolean value 1 (true, which means it belongs to the cluster) or 0 (false, does not belong), the weight wqk in equation (1) represent partial membership to a cluster. It is called a fuzzy weight. There are different means to generate fuzzy weights. One way of generating fuzzy weights is the reciprocal of distance.

$$w_{qk} = 1/ D_{qk}\,,\ \ w_{qk}= 1 \text{ if } D_{qk} \qquad (2)$$

When the distance between the feature vector and the prototype is large, the weight is small. On the other hand, it is large when the distance is small. Using Gaussian functions to generate fuzzy weights is the most natural way for clustering. It is not only immune to outliers but also provides appropriate weighting for more centrally and densely located vectors. It is used in the fuzzy c-means (FCM) algorithm.

Clustering techniques are applied in many application areas such as pattern recognition [13], data mining [12], and machine learning [1]. Clustering algorithms can be broadly classified as Hard, Fuzzy, Possibilistic, and Probabilistic [6]. K-means [15] is one of the most popular hard clustering algorithms which partitions data objects into k clusters where the number of clusters, k, is decided in advance according to application purposes. This model is inappropriate for real data sets in which there are no definite boundaries between the clusters. After the fuzzy theory introduced by Lotfi Zadeh, the researchers put the fuzzy theory into clustering. Fuzzy algorithms can assign data object partially to multiple clusters. The degree of membership in the fuzzy clusters depends on the closeness of the data object to the cluster centers. The most popular fuzzy clustering algorithm is fuzzy c-means (FCM) which was introduced by Bezdek [8] in 1974 and now it is widely used.

Fuzzy clustering [9] is an important problem which is the subject of active research in several real world applications. Fuzzy c-means (FCM) algorithm is one of the most popular fuzzy clustering techniques because it is efficient,

DOI: 10.9781/ijimai.2012.173

straightforward, and easy to implement. However FCM is sensitive to initialization and is easily trapped in local optima because of the random selection in center points. It generalizes c-means (also known by k-means). While c-means builds a crisp partition with c clusters, fuzzy c-means builds a fuzzy one (also with c clusters). u_{ik} is used to formalize the membership of element x_k to the i-cluster. The crisp case corresponds to have u_{ik} as either 0 or 1 (boolean membership) while the fuzzy case corresponds to have u_{ik} in [0; 1]. In this latter case, $u_{ik} = 0$ corresponds to non-membership and $u_{ik} = 1$ corresponds to full membership to cluster i. Values in-between correspond to partial membership (the largest the value, the greatest the membership). Due to this fuzzy nature, in this latter case elements are allowed to belong to more than one cluster.

In the 1970ies, a new kind of approximate algorithm has emerged which tries to combine basic heuristic methods in higher level frameworks aimed at efficiently and effectively exploring a search space. It is defined in St¨utzle, T. Local Search Algorithms for Combinatorial Problems – Analysis, Algorithms and New Applications. DISKI – Dissertationen zur K¨unstliken Intelligenz. infix, Sankt Augustin, Germany, 1999.St¨utzle, T. Local Search Algorithms for Combinatorial Problems – Analysis, Algorithms and New Applications. DISKI – Dissertationen zur K¨unstliken Intelligenz. infix, Sankt Augustin, Germany, 1999. as “Metaheuristics are typically high-level strategies which guide an underlying, more problem specific heuristic, to increase their performance. The main goal is to avoid the disadvantages of iterative improvement and, in particular, multiple descent by allowing the local search to escape from local minima. This is achieved by either allowing worsening moves or generating new starting solutions for the local search in a more “intelligent” way than just providing random initial solutions. Many of the methods can be interpreted as introducing a bias such that high quality solutions are produced quickly. This bias can be of various forms and can be cast as descent bias (based on the objective function), memory bias (based on previously made decisions) or experience bias (based on prior performance). Many of the metaheuristic approaches rely on probabilistic decisions made during the search. But, the main difference to pure random search is that in metaheuristic algorithms randomness is not used blindly but in an intelligent, biased form.”

The performance of simple iterative improvement local search procedures is in general unsatisfactory, for example in Figure 1 the final solution, Trial, is still not the optimal or best for this arbitrary objective function. The quality of the obtained local minimum heavily depends on the starting point for the local search process. As the basin of attraction of a global minimum is generally not known, iterative improvement local search might end up in a poor quality local minimum.

There are different ways to classify and describe metaheuristic algorithms, each of them being the result of a specific viewpoint. For example, we might classify metaheuristics as nature-inspired metaheuristics vs. non-nature inspired metaheuristics. This classification is based on the origins of the different algorithms. There are nature-inspired algorithms, such as evolutionary computation and ant colony optimization, and non nature-inspired ones such as tabu search and iterated local search. We might also classify metaheuristics as memory-based vs. memory-less methods. This classification scheme refers to the use metaheuristics make of the search history, that is, whether they use memory or not. Memory-less algorithms, for example, perform a Markov process, as the information they exclusively use to determine the next action is the current state of the search process. The use of memory is nowadays recognized as one of the fundamental elements of a powerful metaheuristic.

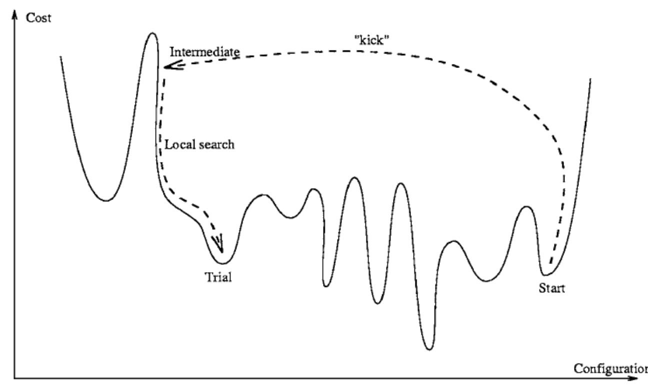


Fig. 1. Schematic representation of the objective function used in a hypothetical metaheuristic algorithm.

Finally, metaheuristics may also be classified into methods that perform a single point vs. population-based search. This classification refers to the number of solutions used by a metaheuristic at any time. Generally, algorithms that work on a single solution at any time are referred to as trajectory methods. They comprise all metaheuristics that are based on local search, such as tabu search, iterated local search and variable neighborhood search. They all share the property that the search process describes a trajectory in the search space.

Population-based metaheuristics, on the contrary, either perform search processes which can be described as the evolution of a set of points in the search space (as for example in evolutionary computation), or they perform search processes which can be described as the evolution of a probability distribution over the search space (as for example in ant colony optimization).

For solving this problem, recently evolutionary metaheuristic algorithms such as genetic algorithm (GA) Vas, P. Artificial-intelligence-based Electrical Machines And Drives: Application Of Fuzzy, Neural, Fuzzy-neural, And Genetic-algorithm-based Techniques (monographs In Electrical And Electronic Engineering). Oxford University Press, 1999., simulated annealing (SA) Wang, J.X., Garibaldi, J. Simulated Annealing Fuzzy Clustering in Cancer Diagnosis. Informatica, 29:61-70, 2005., ant colony optimization (ACO) Ganji, M.F. Using fuzzy ant colony optimization for diagnosis of diabetes disease, IEEE, 18th Iranian Electrical Engineering

(ICEE) Conference, pp. 501-505, 2010., particle swarm optimization (PSO) [14], [16] and Bi-sonar optimization (BSO) [11] have been successfully applied. BSO is a population based optimization tool, which could be implemented and applied easily to solve various function optimization problems, or the problems that can be transformed to functions where fitness can be used in optimization problems [8]. In this paper, a hybrid fuzzy clustering algorithm based on FCM and BSO called FCB is proposed. The experimental results over three real-life data sets indicate the FCB algorithm is superior to the FCM algorithm and BSO algorithm.

The rest of the paper is organized in the following manner. Section 2 introduces FCM, BSO and FCB. In Section 3 parameter settings for FCB algorithm for clustering is presented with experimental results. Finally section 4 concludes this work.

II. METHODS

Different algorithms have been developed using different approaches and considering different underlying assumptions on the data and on the final set of clusters. c-means, fuzzy c-means, self-organizing maps are some of the well known clustering algorithms. Existing algorithms can be classified according to several dimensions. Some of them are described below. One of such dimensions is the direction of the clustering process. In this case, methods are divided into agglomerative ones and partitive ones. Agglomerative algorithms build clusters gathering together those records that are similar. This situation corresponds to a bottom-up strategy (or a bottom-up direction) i.e. from individual records to the set that contains all records. Partitive algorithms, instead, follow a top-down strategy. This is, clusters are defined by partitioning larger sets of records.

Another dimension corresponds to the membership of records to clusters. In this case, we can distinguish among crisp, fuzzy and probabilistic clusters. In crisp clusters, membership of a record into a cluster is boolean. This is, the record either belongs or not to the cluster. Instead, in the case of fuzzy clusters, membership is a matter of degree (in [0; 1]). At the same time, individual records can belong to several clusters. In the case of probabilistic clusters, membership is boolean but there is a distribution of probability of belonging to clusters.

A third dimension is the structure of the clusters. In short, this is whether the clusters themselves define a structure and, if so, which is the structure they define. The simplest case is when no structure is defined. Each cluster is understood as an independent object. Alternatively, clusters can define hierarchies or other complex structures. Such dimensions can be used to classify clustering methods. For example, agglomerative clustering methods are bottom-up (agglomerative) crisp methods that naturally lead to hierarchical cluster structures. c-means is a top-down (partitive) crisp method where clusters do not have any particular relation. Fuzzy c-means is also a top-down (partitive) algorithm that leads to fuzzy clusters (fuzzy memberships of elements to clusters). Self-Organizing Maps (SOM) is also a partitive crisp algorithm but in this case, a grid structure is established among clusters.

A. Fuzzy c-means (FCM)

The fuzzy c-means (FCM) clustering algorithm [3] generates fuzzy partitions for any set of numerical data, allowing one piece of data to belong to two or more clusters. FCM partitions a set of patterns $X_i = \{x_1, x_2,..., x_n\}$ with n features [2] into c (1<c<n) fuzzy clusters with a set of cluster centers $Z_j = \{z_1, z_2, ... , z_c\}$ each being initialized.

$$z_j = \frac{\sum_{i=1}^{n} \mu_{ij}^{m} p_i}{\sum_{i=1}^{n} \mu_{ij}^{m}} \tag{3}$$

Here, the membership degree $\mu_{ij} \in [0, 1]$ quantifies the grade of membership of the ith pattern to jth cluster. The aim of FCM is to minimize the objective function J_{fcm} with d_{ij} being the Euclidean distance [5], [4] measure taken from pattern feature data point x_i to the cluster center z_j. m (m>1) is a scalar which controls the fuzziness of the resulting clusters.

In this formulation, x_i corresponds to the centroid (cluster center/cluster representative) of the i-th cluster and m is a parameter ($m \geq 1$) that plays a central role. With values of m near to 1, solutions tends to be crisp (with the particular case that m = 1 corresponds to the crisp c-means). Instead, larger values of m yield to clusters with increasing fuzziness in their boundaries. To solve this problem, an iterative process is applied. The method interleaves two steps. One that estimates the optimal membership functions of elements to clusters (when centroids are fixed) and another that estimates the centroids for each cluster (when membership functions are fixed).

$$J_{fcm} = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^{m} d_{ij} \tag{4}$$

$$d_{ij} = \|x_i - z_j\| \tag{5}$$

The membership degree is μ. This method does not assure to find the optimal solution of the minimization problem given above but a local optimum. Different starting points can lead to different solutions.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij}}{d_{ik}})^{\frac{2}{m-1}}} \tag{6}$$

1. Select m(m>1); initialize the membership function values μ_{ij}, i = 1, 2,..., n; j = 1,2,...,c. (6)
2. Compute the cluster centers z_j, j = 1,2,...,c. (3)
3. Compute the Euclidian distance d_{ij}, i = 1,2,...,n; j = 1,2,...,c. (5)
4. Update the membership function μ_{ij}, i = 1,2,...,n; j=1,2,...c.

(6)
5. Calculate the objective function J_{fcm}. (4)
6. If not converged go to step 2.

Fig. 2. FCM procedure.

B. Bi-sonar optimization (BSO)

Global optimization algorithms are often classified as either deterministic or stochastic. A stochastic method usually refers to an algorithm that uses some kind of randomness (typically a pseudo-random number generator), and may be called a Monte Carlo method. Examples include pure random search, simulated annealing, and genetic algorithms. Random search methods have been shown to have a potential to solve large problems efficiently in a way that is not possible for deterministic algorithms. An advantage to stochastic methods is that they are relatively easy to implement on complex problems.

A common experience is that the stochastic algorithms perform well and are "robust" in the sense that they give useful information quickly for ill-structured global optimization problems. Bat-inspired algorithm is a metaheuristic optimization algorithm developed by Xin-She Yang [14]. This bat algorithm is based on the bi-sonar/echolocation behaviour of microbats with varying pulse rates of emission and loudness. The idealization of the echolocation of microbats can be summarized as follows: Each virtual bat flies randomly with a velocity v_i at position (solution) xi with a varying frequency or wavelength and loudness A_i. As it searches and finds its prey, it changes frequency, loudness and pulse emission rate r.

Search is intensified by a local random walk. Selection of the best continues until certain stop criteria are met. This essentially uses a frequency-tuning technique to control the dynamic behaviour of a swarm of bats, and the balance between exploration and exploitation can be controlled by tuning algorithm-dependent parameters in bat algorithm. We have to define the rules how bats frequencies f_i, positions x_i and velocities v_i in a d-dimensional search space are updated. The new solutions $x_i(t)$ and velocities $v_i(t)$ at time step t are given by:

$$f = \frac{v}{\lambda}$$

(7)

$$f_i = f_{min} + (f_{max} - f_{min}) \times \delta$$

(8)

$$v_i^t = v_i^{t-1} + \left(x_i^t - x_{gbest}^t\right)$$

(9)

$$x_i^t = x_i^{t-1} + v_i^t$$

(10)

where δ [0, 1] is a random vector drawn from a uniform distribution. Here $x(t_{gbest})$ is the current global best location or hunting space or solution which is located after comparing all the solutions among all the n bats. As the product $\lambda_i f_i$ is the velocity increment, we can use either f_i (or λ_i) to adjust the velocity change while fixing the other factor λ_i (or f_i), depending on the type of the problem of interest. The domain size of the problem in context determines the values of fmin and f_{max}. Initially, each bat is randomly assigned a frequency which is drawn uniformly from $[f_{min}, f_{max}]$.

Bat algorithm has been used for engineering Yang, X. S. and Gandomi, A. H., Bat algorithm: a novel approach for global engineering optimization, Engineering Computations, Vol. 29, No. 5, pp. 464-483, 2012. classifications S. Mishra, K. Shaw, D. Mishra, A new metaheuristic classification approach for microarray data,Procedia Technology, Vol. 4, pp. 802-806, 2012. A fuzzy bat clustering method has been developed to solve ergonomic workplace problems Khan, K., Nikov, A., Sahai A., A Fuzzy Bat Clustering Method for Ergonomic Screening of Office Workplaces,S3T 2011, Advances in Intelligent and Soft Computing, 2011, Volume 101/2011, 59-66, 2011.. An interesting approach using fuzzy systems and bat algorithm has shown a reliable match between prediction and actual data for energy modeling Lemma, T. A. Use of fuzzy systems and bat algorithm for exergy modelling in a gas turbine generator, IEEE Colloquium on Humanities, Science and Engineering (CHUSER'2011), pp. 305-310, 2011. A detailed comparison of bat algorithm (BA) with genetic algorithm (GA), PSO and other methods for training feed forward neural networks concluded clearly that BA has advantages over other algorithmsKhan, K. and Sahai, A. A comparison of BA, GA, PSO, BP and LM for training feed forward neural networks in e-learning context, Int. J. Intelligent Systems and Applications (IJISA), Vol. 4, No. 7, pp. 23-29, 2012..

C. Fuzzy c-means bi-sonar (FCB) optimization for clustering

Stochastic methods, such as simulated annealing and genetic algorithms, are gaining in popularity among practitioners and engineers because they are relatively easy to program on a computer and may be applied to a broad class of global optimization problems. However, the theoretical performance of these stochastic methods is not well understood. The stochastic and fuzzy set theories cannot be considered to be an omnipotent mean which will solve all the problems automatically. They have to be understood as an appropriate instrument for modeling the indeterminateness. As the main objective of fuzzy sets is the modeling of the semantics of a natural language there exist numerous specializations in which the fuzzy sets can be applied.

Besides the most often used probabilistic models and the stochastic analysis techniques newer uncertainty models have been developed that offer the chance to take account of non-stochastic uncertainty that frequently appears in real world problems. The quantified uncertain parameters are introduced in the respective analysis algorithm: Fuzzy c-means and bi-sonar optimization algorithm. A modified bat algorithm for cluster analysis is proposed. The velocities (cf. equation (7)) of bats are redefined to update the fuzzy relation between variables.

$$v_i^t = v_i^{t-1} + \left(x_i^t - x_{gbest}^t\right) + \left(x_i^t - x_{pbest}^t\right) \quad (11)$$

The variable $x(t_{pbest})$ is the personal best hunting space for a bat. The inclusion of this in the algorithm should enhance clustering by increasing exploitation of the algorithm towards favorable cluster centers. For evaluating the generalized solutions of the FBC algorithm's fitness function f(x) the objective function J_{fcm} of the FCM algorithm is used:

$$f(X) = \frac{K}{J_{fcm}} \; f(X) = \frac{K}{J_{fcm}} \quad (12)$$

where K is a constant. The smaller is J_{fcm} the better is the clustering effect and the higher is the individual fitness.

1. Initialize the parameters including population size, frequencies and the maximum iterative count. (7) & (8)
2. Create a swarm with P bats.
3. Calculate the cluster centers and distance matrix for bi-sonar. (3) & (5)
4. Calculate the fitness value of each bat. (4), (6) & (12)
5. Update the velocity matrix for each bat. (11)
6. Update the location vector for each bat. (10)
7. Calculate the (global) best and (local) personal best location for all the bats and each bat.
8. If terminating condition is not met, go to step 3.

Fig. 3. FCB procedure.

III. EXPERIMENTAL RESULTS

A. Parameter settings

Optimization techniques traditionally depend on the setting of one or more parameters. Depending on the problem and the techniques the number of parameters can be one, two or even dozens of them. One of the main difficulties of applying an evolutionary algorithm (or, as a matter of fact, any heuristic method) to a given problem is to decide on an appropriate set of parameter values. The tuning process, when dealing with several parameters, is a time consuming and critical step. Typically these are specified before the algorithm is run and include population size, selection rate, operator probabilities, not to mention the representation and the operators themselves.

In order to optimize the performance of the FCB, fine tuning has been performed and best values for their parameters are selected. The parameters were tuned (meta-optimized) to perform well on the problem sets. Based on experimental results these algorithms perform best under the following settings: $\alpha=\gamma=0.9$, initial loudness $A_i=1.35$ and initial emission rate $r_i=0.001$. The FCB terminating condition is the maximum number of iterations 3000 or no changes in g_{best} in 400 iterations. In all of algorithms m, the weighting exponent is set to 2. Parameter settings for FCM and BSO are shown in [7].

B. Findings

For evaluating FCB, three well-known real-world data sets UCI Machine Learning Repository, Center for Machine Learning and Intelligent Systems, 2012. Available online: http://archive.ics.uci.edu/ml/ have been considered:

1. Glass, which consists of 214 objects and 6 different types of glasses. Each type has 9 features,
2. Vowel data set, which consists of 871 Indian Telugu vowel sounds, the data set has three features and six overlapping clusters,
3. Contraceptive Method Choice (CMC), which consists of 1473 objects and 3 different types characterized by 9 features.

FCB obtained superior results than others in all three data sets and it can escape from local optima (cf. Table 1). These data sets cover examples of data of low, medium and high dimensions. These algorithms are implemented using Matlab. The experimental results of over 50 independent runs for FCM and 20 independent runs for BSO and FCB are summarized in Table 1. There are differences in the run times as the BSO Khan, K., Nikov, A., Sahai A., A Fuzzy Bat Clustering Method for Ergonomic Screening of Office Workplaces,S3T 2011, Advances in Intelligent and Soft Computing, 2011, Volume 101/2011, 59-66, 2011. and FCB Khan, K., Nikov, A., Sahai A., A Fuzzy Bat Clustering Method for Ergonomic Screening of Office Workplaces,S3T 2011, Advances in Intelligent and Soft Computing, 2011, Volume 101/2011, 59-66, 2011. was validated from a previous study. The figures in this table are the objective function values (equation (2)).

The experimental results show that when the size of data set (number of objects or clusters) is small (glass and vowel), the FCB surpasses FCM and with increasing the size of data set (CMC), FCB still obtains better results than FCM. It also performs better than fuzzy BSO (Fuzzy Bat Swarm Optimization) [10] in all test cases. The computation time for FCB algorithm is only about 51 seconds per instance on average with a maximum of 192 seconds for some of the largest instances. Here running times was used as a metric for the performance analysis of the clustering algorithms Zhao, B. An Ant Colony Clustering Algorithm. Proc. Conference on Machine Learning and Cybernetics, 3933-3938, 2007..

TABLE 1. RESULTS OF FCM

	FCM		
	Worst	Average	Best
Glass (214, 6, 9)	73.37	72.87	72.26
Vowel (871, 6, 3)	73390.8	71504.7	69069.1
CMC (1473, 3, 9)	3548.5	3334.7	3423.2

TABLE 2. RESULTS OF BSO

	BSO		
	Worst	Average	Best
Glass (214, 6, 9)	87.37	86.97	86.26
Vowel (871, 6,3)	100021.5	99394.0	98834.2

CMC (1473, 3,9)	4190.1	4095.6	4025.2

TABLE 3. RESULTS OF FCB

	FCB		
	Worst	Average	Best
Glass (214, 6, 9)	72.91	72.42	71.34
Vowel (871, 6, 3)	67342.2	66972.1	65682.3
CMC (1473, 3, 9)	3533.8	3329.8	3413.7

IV. CONCLUSION

This paper presented a derivation of a swarm family of stochastic algorithms. The fuzzy c-means algorithm is sensitive to initialization and is easily trapped in local optima. On the other hand the bi-sonar optimization algorithm is a stochastic tool which could be implemented and applied easily to solve various function optimizations. In this paper in order to overcome the shortcomings of the fuzzy c-means we integrate it with bi-sonar optimization algorithm to produce the FCB algorithm. Experimental results over three well known data sets, Glass, Vowel and CMC, show that the proposed hybrid method is efficient and reveals very encouraging results in term of quality of solution found. Interpretation of this reformulated functional underlying the FCM model as a generalized mean of order might lead to new results for other families of metaheuristic swarm-based fuzzy models, for example, using cuckoo searchYang, X.-S.; Deb, S. "Cuckoo search via Lévy flights". World Congress on Nature & Biologically Inspired Computing (NaBIC 2009). IEEE Publications. pp. 210–214, 2009., fireflyYang, X. S. Nature-Inspired Metaheuristic Algorithms. Frome: Luniver Press, 2008. or Krishnanand, K.N. and Ghose, D. Detection of multiple source locations using a glowworm metaphor with applications to collective robotics. IEEE Swarm Intelligence Symposium, Pasadena, California, USA, pp. 84–91, 2005. swarm optimization algorithms.

Apart of the disambiguation of assignment of objects in clusters this approach is more robust in terms of finding the local minima of the given objective function. The conjecture that this method is more robust than deterministic (crisp) clustering is supported by the experimental results. The FCM is a global stochastic tool which could be implemented and applied easily to solve various function optimization problems, or the problems that can be transformed to other function-based optimization problems.

The following properties are important research areas that can be taken in order to increase the efficiency and effectiveness of the FCB algorithm. The FCB algorithm should be able to generate arbitrary shapes of clusters rather than be confined to some particular shape, handle large volume of data as well as high-dimensional features with acceptable time and storage complexities, detect and remove possible outliers and noise, decrease the reliance of algorithms on users-dependent parameters, have the capability of dealing with newly occurring data without relearning from the scratch, be immune to the effects of order of input patterns; provide some insight for the number of potential clusters without prior knowledge, show good data visualization and provide users with results that can simplify further analysis and be capable of handling both numerical and nominal data or be easily adaptable to some other data type.

However, it is important to emphasis that ultimately, the tradeoff among different criteria and methods is still dependent on the applications themselves. Further work can be done on using multi-criteria analysis of the algorithm's performance, for example, space and data size. The advantages shown in using this approach can be applied in many areas including medical image segmentation, classification and soil-landform interrelationships, estimation and segmentation of magnetic resonance imaging (MRI) data, clustering of microarray data, image segmentation, color image segmentation, application to non-linear mapping to geochemical datasets, analysis of metabolomics, web document and snippet clustering, classification of remotely sensed images, eigenspace projections and pixel classification.

REFERENCES

[1] Konar, A. Computational Intelligence, principles, techniques and applications, 1, 2005.

[2] Bezdek, C. Numerical taxonomy with fuzzy sets. Journal of Mathematical Biology, 1:57-71, 1984.

[3] Bezdek, J, Fuzzy mathematics in pattern classification. Ph.D. thesis, Ithaca, NY: Cornell University, 1974.

[4] Gan, G., Wu, J., Yang, z. A genetic fuzzy k-Modes algorithm for clustering categorical data. Expert Systems with Applications (36): 1615-1620, 2009.

[5] Ganji, M.F. Using fuzzy ant colony optimization for diagnosis of diabetes disease, IEEE, 18th Iranian Electrical Engineering (ICEE) Conference, pp. 501-505, 2010.

[6] Gower J.C. Properties of Euclidean and non-Euclidean distance matrices. Linear Algebra and its Applications, 67:81-97, 1985.

[7] Hathway, J.R., and Bczdek, C.J. Optimization of clustering criteria by reformulation. IEEE transactions on fuzzy systems, 241-245, 1995.

[8] Izakian, H., Abraham, A. Fuzzy C-means and fuzzy swarm for fuzzy clustering problem. Expert Systems with Applications, 38(3): 1835-1838, 2011.

[9] Izakian, H., Abraham, A., Snasel, V. Fuzzy Clustering Using Hybrid Fuzzy c-means and Fuzzy Particle Swarm Optimization. IEEE: World Congress on Nature & Biologically Inspired Computing, pp. 1690-1694, 2009.

[10] Izakian. H, Abraham., A, Snášel. V. Fuzzy clustering using hybrid fuzzy c-means and fuzzy particle swarm optimization. In: IEEE world congress on nature & biologically inspired computing, pp 1690–1694, 2009.

[11] Khan, K. and Sahai, A. A comparison of BA, GA, PSO, BP and LM for training feed forward neural networks in e-learning context, Int. J. Intelligent Systems and Applications (IJISA), Vol. 4, No. 7, pp. 23-29, 2012.

[12] Khan, K., Nikov, A., Sahai A., A Fuzzy Bat Clustering Method for Ergonomic Screening of Office Workplaces,S3T 2011, Advances in Intelligent and Soft Computing, 2011, Volume 101/2011, 59-66, 2011.

[13] Krishnanand, K.N. and Ghose, D. Detection of multiple source locations using a glowworm metaphor with applications to collective robotics. IEEE Swarm Intelligence Symposium, Pasadena, California, USA, pp. 84–91, 2005.

[14] Lemma, T. A. Use of fuzzy systems and bat algorithm for exergy modelling in a gas turbine generator, IEEE Colloquium on Humanities, Science and Engineering (CHUSER'2011), pp. 305-310, 2011.
[15] Li, L., Liu, X., Xu, M. A Novel Fuzzy Clustering Based on Particle Swarm Optimization. Conf. Information Technologies and Applications in Education, 88-90, 2007.
[16] Meyerson, A., Callaghan, L.O, and Plotkin, S. A k-median algorithm with running time independent of data size. Machine Learning, 56(1–3):61–87, 2004.
[17] Pang, W., Wang, K., Zhou, C., & Dong, L. Fuzzy discrete particle swarm optimization for solving traveling salesman problem. Proc. Computer and Information Technology, 796–800, 2004.
[18] Runkler, A. T., Katz, C. Fuzzy Clustering by Particle Swarm Optimization. Conf. Fuzzy Systems, 601-608, 2006.
[19] S. Mishra, K. Shaw, D. Mishra, A new metaheuristic classification approach for microarray data,Procedia Technology, Vol. 4, pp. 802-806, 2012.
[20] St¨utzle, T. Local Search Algorithms for Combinatorial Problems – Analysis, Algorithms and New Applications. DISKI – Dissertationen zur K¨unstliken Intelligenz. infix, Sankt Augustin, Germany, 1999.
[21] Tan, N.P., Steinbach, M., Kumar, V. Introduction to data mining. Boston, Addison-Wesley, 2005.
[22] UCI Machine Learning Repository, Center for Machine Learning and Intelligent Systems, 2012. Available online: http://archive.ics.uci.edu/ml/
[23] Vas, P. Artificial-intelligence-based Electrical Machines And Drives: Application Of Fuzzy, Neural, Fuzzy-neural, And Genetic-algorithm-based Techniques (monographs In Electrical And Electronic Engineering). Oxford University Press, 1999.
[24] Wang, J.X., Garibaldi, J. Simulated Annealing Fuzzy Clustering in Cancer Diagnosis. Informatica, 29:61-70, 2005.
[25] Webb, A. Statistical pattern recognition. John Wiley & Sons, 2002.
[26] Xin-She Yang. A New Metaheuristic Bat-Inspired Algorithm. Nature Inspired Cooperative Strategies for Optimization (NICSO). Studies in Computational Intelligence, 284:65-74, 2010.
[27] Yang, F., Sun, T., Zhang, C. An efficient hybrid data clustering method based on K-harmonic means, and Particle Swarm Optimization. Expert Systems with Applications, 36: 9847–9852, 2009.
[28] Yang, W., Rueda, L., Acun, D., A Simulated Annealing Approach to Find the Optimal Parameters for Fuzzy Clustering Microarray Data. Conf. Chilean Computer Science Society, 2005.
[29] Yang, X. S. and Gandomi, A. H., Bat algorithm: a novel approach for global engineering optimization, Engineering Computations, Vol. 29, No. 5, pp. 464-483, 2012.
[30] Yang, X. S. Nature-Inspired Metaheuristic Algorithms. Frome: Luniver Press, 2008.
[31] Yang, X.-S.; Deb, S. "Cuckoo search via Lévy flights". World Congress on Nature & Biologically Inspired Computing (NaBIC 2009). IEEE Publications. pp. 210–214, 2009.
[32] Zhao, B. An Ant Colony Clustering Algorithm. Proc. Conference on Machine Learning and Cybernetics, 3933-3938, 2007.

Koffka Khan was born in San Fernando, Trinidad and Tobago in 1978. He received the B.Sc. and M.Sc. degrees from University of the West Indies, in 2002 and 2008, respectively. He was awarded by the University of the West Indies for his contributions made in postgraduate work in 2009 as a research assistant.

He is working presently at The University of The West Indies; St. Augustine Campus (TRINIDAD & TOBAGO) as a Tutor in Computer Science in the Department of Computing and Information Technology (Faculty of Science & Agriculture) since September 2006. Mr. Khan started his teaching-n-research career as a Demonstrator in Computer Science at the University of The West Indies at the Department of Mathematics and Computer Science. He has up-to-date, published ten research and co-authored four papers in journals of international repute & in the proceedings of international conferences.

Ashok Sahai is working presently at The University of The West Indies; St. Augustine Campus (TRINIDAD & TOBAGO) as a Professor of Statistics in the Department of Mathematics & Statistics (Faculty of Science & Agriculture) since February 2006. Dr. Sahai started his teaching-n-research career as a Lecturer in Statistics Department at Lucknow University (INDIA) in July 1966, and continued thereat till April 1980. He has, up-to-date, published more than one hundred research papers in peer-reviewed journals of international repute & in the peer-reviewed proceedings of international conferences.

He worked as Reader in Statistics and as Professor of Statistics in the Department of Mathematics at University of Roorkee (Now IIT Roorkee) during the period: April 1980- July 1995. Prof. Sahai had also worked as an Assoc. Professor of Statistics at University of Dar-Es-Salaam; TANZANIA (East Africa) during the period: July 1982- June 1984, and as a Professor of Statistics at University of Swaziland (Southern Africa) during the period: July 1993- June 2003. He worked as a Guest Scholar @ PharmacoEconomic Research Centre; University of Arizona, TUCSON during the period from July 2003 to October 2003 and as Visiting Professor @ Hyderabad; INDIA during December 2003 to January 2006 in ICFAI Tech. University, Medchel Rd.; in Aurora School of Management at Chikkadpally; and in St. Ann’s P.G. (Management) College For Women at Mallapur.

A Client-Server System for Ubiquitous Video Service

[1]Ronit Nossenson, [2]Orit Yudilevich, [2]Omer Marlowitz
[1]*Faculty of Computer Science*, Jerusalem College of Technology, Jerusalem, Israel
[2]*School of Computer Science,* The Interdisciplinary Center, Jerusalem, Israel

***Abstract* — In this work we introduce a simple client-server system architecture and algorithms for ubiquitous live video and VOD service support. The main features of the system are: efficient usage of network resources, emphasis on user personalization, and ease of implementation. The system supports many continuous service requirements such as QoS provision, user mobility between networks and between different communication devices, and simultaneous usage of a device by a number of users.**

***Keywords* — Seamless Content Delivery, Ubiquitous Multimedia Service, Personal Multimedia Delivery, Live Video Transmission, VOD Transmission.**

I. Introduction

A new generation of distributed services ranging from entertainment services such as live video streaming, video on demand and on-line gaming, to life- saving applications such as medical services and monitoring, are being deployed in heterogeneous and ubiquitous environments. To be accepted by both users and network operators, these ubiquitous services must deliver continuous service, as well as adaptive and satisfactory Quality-of-Service with a minimum overhead of network resources. Providing ubiquitous services entails a number of complex issues, such as supporting the required QoS during a session, seamless handovers between different radio access technologies (RATs), supporting user mobility, etc.

In this article we introduce a simple client-server architecture and algorithms for live video and Video On Demand (VOD) ubiquitous services. To achieve satisfactory continuous service with minimum overhead, collaboration and coordination between small number of agents uses several communication methods including wireless or cellular connections. This article is an extended version of our previous results [13].

The main features of the architecture are as follows:

1. Efficient usage of network resources complying with the required/preferred QoS.
2. User- driven architecture which enables easy personalization.
3. Ease of implementation.

Previous work on ubiquitous multimedia services has focused on middleware solutions (see, for example, [1-7, 9, 10]). These ideas are good and effective but they require the cooperation of network operators. Since the conventional business model is defined only between the content provider and its content consumers, the readiness of operators to deploy such solutions is limited.

As illustrated in Fig. 1, our novel architecture relies on both the user's communication devices (for example, smart phone, PDA or laptop) and the continuous service/content provider system; no changes or extensions are needed in the operator network. The system can completely handle a range of continuous service requirements: QoS provision, user mobility between RATs and between different communication devices, simple user interface and personalization, simultaneous usage of the same device by a number of users (while protecting privacy) and so on. The description focuses on ubiquitous live video and VOD services, but the system can be easily extended to other services as well.

The architecture is based on the "best k" algorithm to ensure efficient use of network resources [8]. This algorithm provides high quality live-video transmission by using few agents and proposes ways to minimize the usage of network resources. Experimental results show that by using the best-k algorithm, high quality video can be delivered with an overhead factor of 1.65%.

This paper is organized as follows. In the next section the

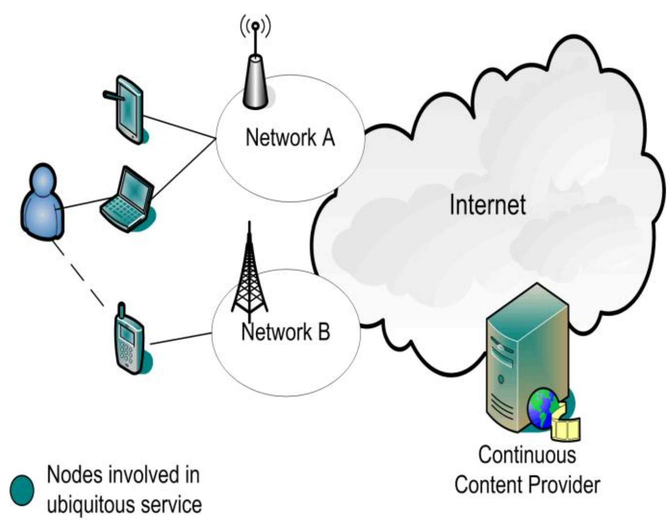


Fig. 1. High level description of a client-server system for ubiquitous service. The subscriber can receive a service via laptop, smart phone, etc.

DOI: 10.9781/ijimai.2012.174

ubiquitous system requirements are outlined. In section III, we describe the specifications of the new system by topic. These include the system building blocks, state machines and essential procedures. Next, some experimental results are described in Section IV. Finally, we suggest future research directions.

II. System Requirements

Providing ubiquitous service raises a number of issues that must be addressed. In this section we specify the technical requirements for a ubiquitous service system. First we define the terms used in this work. Next, we list the requirements according to their functionality: general, usability, QoS, multiple users' and privacy.

A. Definitions

Content provider functions as the server side. Its responsibilities include user management, QoS management and content provision. *Service* is a video service such as live video or VOD. *User* is a person who subscribes to the service. *Client* is the software which provides a special service on a user device. A user has one client on each device, per service. For example, a user who subscribes to 3 services via 4 devices has 12 clients. The client is responsible for communication between the user and the system, measuring and reporting QoS, agent management and combining the received data when necessary.

Agent is the software that is responsible for receiving and transmitting data for the service via a specific communication interface/technology. For example, a device with cellular, WiFi and BT connections has one client and up to three agents per service and user (see Fig. 2). Two or more users can use the same service via one device simultaneously (watching a movie together, for example). In this case the first user who activated the service is the *primary* user and the other users are referred to as *secondary* users. The primary user and secondary users together are referred to as the *service users' group on a device*. The primary user manages the service users' group that is using the service.

B. General Requirements

G1. The system supports ubiquitous service for live video transmission and video on demand (VOD) applications.

G2. The system attempts to provide QoS as close as possible to the preferred quality (see requirement Q1 below), with minimal user intervention and with minimal overhead for network resources.

G3. Scalability requirements: The system supports up to A potential agents per user. The system supports up to B activations of service per minute. The system supports up to C simultaneous active services. A user may have up to D active services and up to E paused services simultaneously. The parameters A, B, C, D and E can be extended by simple hardware extension.

C. Electronic Image Files (Optional)

U1. Service is supplied via one client at a specific time.

U2. The system provides a convenient user interface.

U3. The user needs to configure a set of agents for each service for each client (device).

U4. For a service, at least one device (client) and one agent must be registered.

U5. The user can alter its set of agents at any time using a convenient interface.

U6. For each service specified in requirement G1 the system supports the following operations: Start, Stop, Pause, and Resume.

U7. The "Start" operation is used for service activation the first time as well as for service re-activation after a "Stop" action. It can be used for inactive service only.

U8. The operation "Stop" is used for termination of the active or paused service.

U9. The operation "Pause" is used for temporary halts of an active service, up to a predefined timeout. The timeout can be interrupted by user operation ("Resume"). Otherwise the service is terminated.

U10. The "Resume" operation is used to continue the service after the "Pause" operation, depending on the type of service, assuming that the timeout has not expired.

U11. The outcome of the "Resume" operation for VOD

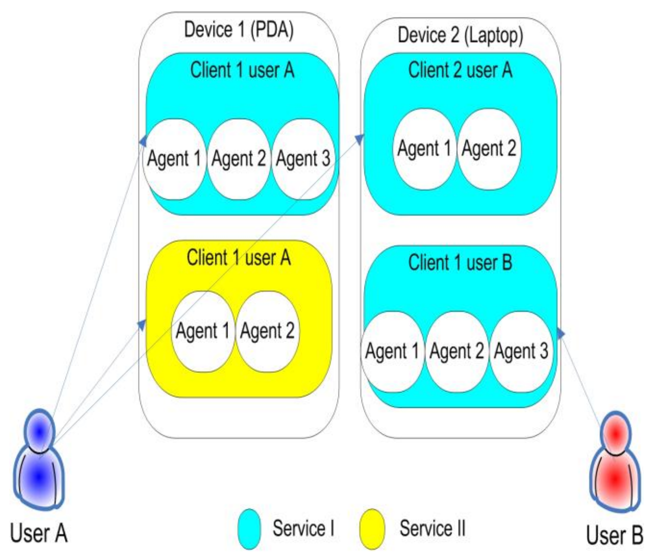


Fig. 2. User-Device-Client-Agent-Relationship. A user can have several devices and can be subscribed to several services. For each (service, user, device) triple there is one client. Each client can manage several agents. Two users can share a device, for the same service or for different services.

application is continuation of the video transmission from the point where it was paused. For the Live Video application it corresponds to resumption of the on-line video at the current time.

U12. For an active service, at least one agent is active.

U13. Changes of agents, without a change in client (switching, adding and subtract agents) are done transparently without the user's intervention.

U14. A paused service can be resumed from any subscribed client (device continuously).

U15. There can be only one active service per device.

U16. A user can have several active services, on several devices.

U17. New service activation on a device that already has another service active on it, is subject to the approval of the primary user of the active service and is equivalent to pausing/ termination of the previous service.

D. QoS requirements

Q1. The user can define *Preferred QoS parameters* and *Required QoS parameters* per service. In addition, the system has default values for these parameters per service.

Q2. The system aims to provide the user with QoS as close as possible to the preferred QoS, with minimal user intervention. If this is impossible the system aims to provide QoS above the required QoS. If the QoS falls below the required level the user is informed and the service is terminated.

Q3. The system provides continuous service for the user as long as there is at least one active client and active agent capable of providing QoS above the required level.

Q4. The QoS parameters are defined for each service separately, including the following characteristics as a minimum: bandwidth, delay and jitter.

Q5. When the QoS parameters are higher than the required level, but below the preferred thresholds, the system attempts to improve the service in the following way: (1) if the preferred QoS can be achieved using the current client (by changing agent/s), the change is performed transparently; (2) if the preferred QoS can be achieved only by another client (device), the transition can be performed, subject to the user's approval; (3) if the preferred QoS cannot be achieved using any other client (device), the system provides the best QoS possible via the current client (device).

Q6. If the QoS is below the required threshold, the system tries to improve the QoS via the current client. If this is impossible, the user is advised to move to another device. If the required QoS cannot be met the service is terminated.

Q7. For each user and active service, the system maintains a set of potential agents as a function of QoS parameters, environmental changes, user preferences, and agent availability.

E. Multiple User Requirements

M1. An active service has one primary user.

M2. There can be several secondary users for an active service.

M3. A user can join a service on a specific device which is managed by a different primary user, subject to both users' approval.

M4. A secondary user can disjoin a service; this action is equivalent to pausing or stopping the service to the specific user.

M5. A primary user can be replaced by another user, subject to both users' approval. The previous primary user becomes a secondary user in this case.

F. Privacy Requirement

P1. A user cannot access information on services that another user is subscribed to, even if they co-exist on the same device (see Fig. 2), unless the user is a primary user who is aware of the secondary users of the same service.

III. System Specifications

In this section we elaborate on the technical problems and provide specifications and algorithms for the system. The listed specifications provide a feasible solution for the pre-defined requirements. This section is divided into the following sub-sections: service state machine and QoS specification, user interface specification, agent state machine, multi user specification and user mobility issues are discussed in the final sub section.

A. Service State Machine and QoS Specifications

According to the requirements, each service can be in one of three states per user: Not Active, Active (A, B, C and D sub-states) or Paused. Fig. 3 depicts the transitions between the states, showing all the valid transitions, their triggers and actions.

1) From *"Not Active Service" to "Active Service"*: The actions that take place in this case are: (i) the system establishes a connection with the agent that sent the "Start" command before it starts to transmit the content; (ii) A handshake procedure is performed with each user's agents, and a list of available agents is generated. The handshake procedure and the management of the available agent list are described in the agent state machine below.

2) From *"Active Service" to "Not Active Service"*: This transition can occur in two cases: upon a user "Stop" command or if the QoS falls below the required level (see requirements Q2, Q6). In these cases the data flow to/from the client is terminated, a termination message is sent to all

available agents, so they can move to the "not active" state, all session data are dropped on both server and client sides.

3) From *“Paused Service” to “Not Active Service”*: A service can go from "Paused" to "Not Active" in two cases: by a user "Stop" command or by a pause timeout expiration, see requirement U9. In these cases, the system sends a termination message to all clients to turn the available agents to "not active”. Service- session related information is dropped from both the server and client sides.

4) From *“Active Service” to “Paused Service”*: The user can implement this transition in several ways: by sending pause command, activation of another service on the same device, approval of client swapping proposed by the system and switching from primary user to secondary user in the multiple user service mode (see requirements U9, U17, M3). In these cases all the service session data are saved in both server and client. Service paused messages are sent to the active agents, so they can go into the “available” state. Data transmission is stopped. On the server side, additional data are stored, such as last active client, pointer to the last transmitted I frame, last packet sequence number, file offset (the location in the movie for VOD service), and last decoding format in use. In addition, the timer for maximum paused time is activated.

5) From *“Paused Service” to “Active Service”*: In case the service is resumed on the same client it was paused on, all the data exist on both the client and the server sides and the service is simply resumed. Otherwise, if the service is resumed on a different client (device) a format adaptation is performed if needed. The last I frame is sent to the client together with the following P frames and a specific notification for the video player film offset. This enables VOD service to resume from the same point it was paused on the previous device. For live video service, the service is resumed according to the current time.

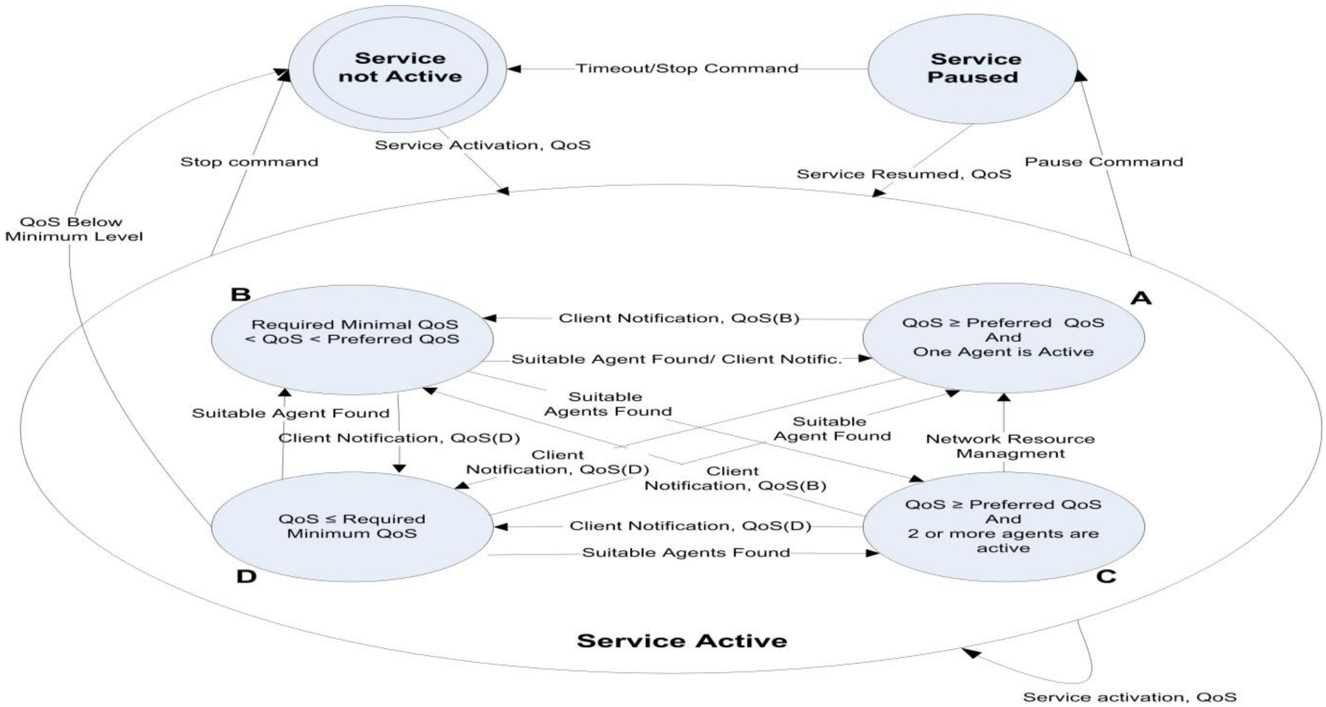


Fig. 3. Service State Machine and QoS management state machine for active service.

6) From *“Active Service” to “Active Service“*: If the "resume" command is initiated on the active client, the command is discarded, otherwise the command is equivalent to pausing the service on a current client and resuming it on a new one.

Other state transitions are server internal and are related to QoS provision. Specifically, it meets requirements G2, U12, U13, Q2, Q3, Q5, Q6 and Q7 above.

The essential server functions are:

- Ensuring an acceptable QoS level via agent management.
- Providing maximum transparency to the user.
- Ensuring efficient usage and minimum overhead of the network resources.

The essential client functions are:

- Providing the user interface to the system.
- Monitoring the QoS parameters for active services and informing the server if needed.

- Combining and synchronizing the data, using [8] or a similar algorithm.

The following information is stored on the server per active user of a specific service:

- The current QoS parameters.
- The current active client and its set of active agents.
- A list of available clients with their available agents.
- Session associated information.
- List of clients and agents used in the last time interval.
- List of forbidden client transitions.

During active service, the server runs the state machine as presented in the "Active Service State" in Fig. 3. In sub-state A the user receives its preferred QoS by one agent; thus the system does not have to improve its QoS or to reduce network resource overhead associated with it.

Sub-state B is characterized by the QoS between the preferred and required levels, thus the service can be continued along with system efforts to improve the QoS according to requirements G2 and Q5. The requirements define the following priorities: QoS, minimum user intervention and network resources (see requirements G2, U13, Q5, Q6); hence, in state B, the following procedure is performed periodically:

```
StateBProc(PreferredQoS, CurrentClient,
       AllAvaliableClients):
Begin
  Bool IsPreferredQoSPossible = false;
  Bool IsPreferredQoSReached = false;
  IsPreferredQoSPossible =
    BestQoS(PreferredQoS, CurrentDevice,
    AllAvaliableClients);
  If (IsPreferredQoSPossible)Then
    Bool IsPreferredQoSReached =
    SwapAgentbyQoS (PreferredQoS, QoSList);
  End
  If (Not IsPreferredQoSReached)Then
       ImproveCurrentClientQoS(CurrentQoS,
              CurrentDevice);
  End
End
```

The function `BestQoS` is described below, it is responsible for generating the list of respective agents and possible QoS by staging a competition between the agents (for details regarding agent competition see [8]). The QoS list includes: (i) Best QoS that can be reached by one agent of the current client (device) (`BestQoSSingleAgentCurrentClient`) and corresponding agent; (ii) Best QoS that can be reached by two or more agents on the current client (device) (`BestQoSMultiAgentCurrentClient`) and corresponding agents; (iii) Similar data about other available clients and agents (iv) Best possible QoS for each client achieved by one or more agents. The inputs to the function are: requested QoS threshold, current device, full list of

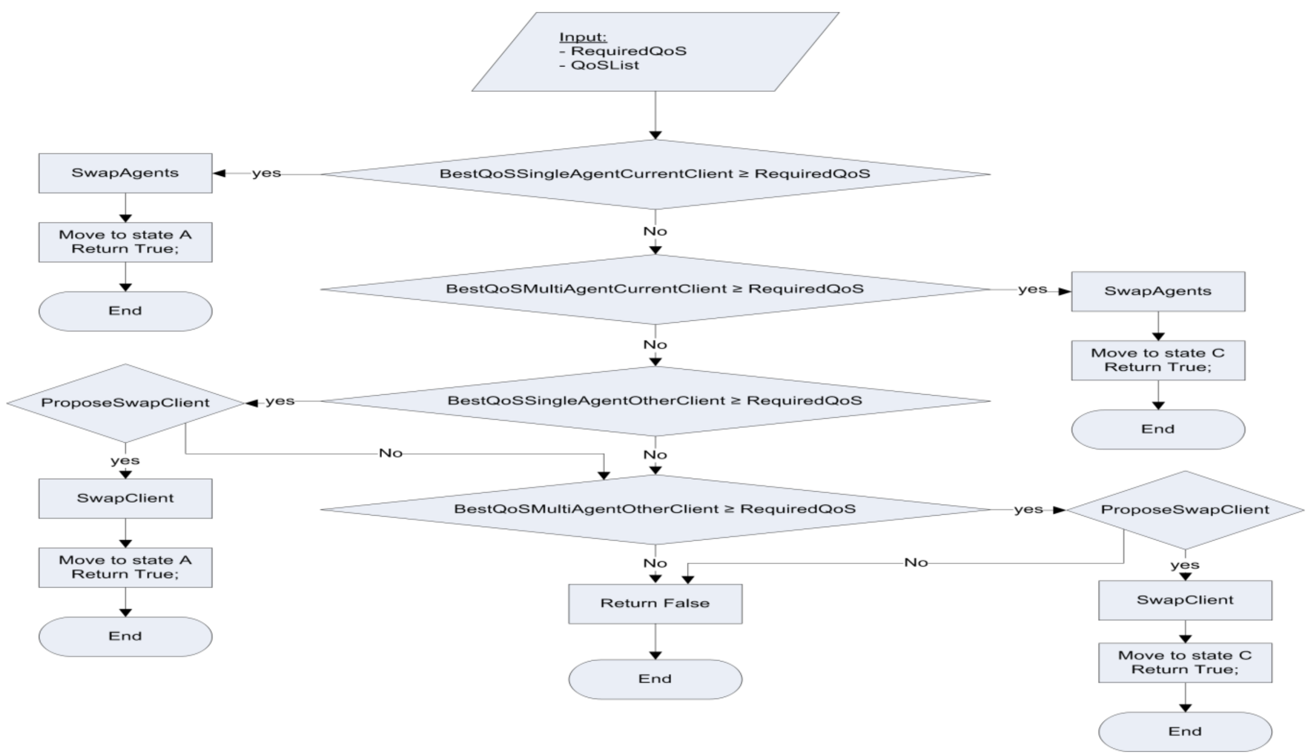


Fig. 4. Swap agent by QoS procedure flow chart

devices and agents.

The pseudo code of the function BestQoS is as follows:

```
BestQoS(QoSThreshold,CurrentClient,
                  AllAvaliableClients):
Begin
Competition(Current Client’s Agents);
Set BestQoSSingleAgentCurrentClient, Agent;
Set BestQoSMultiAgentCurrentClient, Agents;
If((BestQoSSingleAgentCurrentClient ≥
  QoSThreshold)Or
  (BestQoSMultiAgentCurrentClient ≥
  QoSThreshold)) Then return true;
/* Needed QoS can’t be reached by current
client, rest of the clients are checked */
Competition(Available Clients);
Set BestQoSSingleAgentOtherClient, AgentList;
Set BestQoSMultiAgentOtherClient,
AgentMatrix;
maxQoS = max(BestQoSSingleAgentOtherClient,
BestQoSMultiAgentOtherClient);
return (maxQos  ≥ QoSThreshold)
End
```

When agent/s that supply the required QoS is/are found, the system attempts to swap to this/these agent/s. This is performed by the SwapAgentbyQoS() procedure. The flow chart of this procedure is presented in Fig. 4. This procedure scans the possible agents according to a pre-defined order (requirements G2, U13, Q5, Q6). The first choice is one agent on the current client, then, multiple agents on the current clients, finally single and multiple agents of other clients. This procedure updates the QoS state machine, as required.

The system should introduce the user to the full list of other clients that can provide the preferred QoS. Client switching is always subject to user approval. In addition, the system should maintain a list of clients whose transitions are forbidden (see user interface specification for details), to avoid undesirable proposals to the user (requirement G2, Q5).

If the preferred QoS cannot be reached the system should try to improve the QoS on the current client (requirement Q5), to the best possible QoS. This is performed by the ImproveCurrentClientQoS() procedure. It uses the current QoS and current client as inputs. This procedure does not affect the QoS state machine, since after its termination the QoS level is still between the preferred and the required thresholds. The pseudo code of this procedure is as follows:

```
  ImproveCurrentClientQoS(CurrentQoS,
     CurrentClient)
  Begin
  MaxQoS
=max(BestQoSSingleAgentCurrentClient,

BestQoSMultiAgentCurrentClient);
   If  ((MaxQoS > CurrentQoS) And
  (MaxQoS==BestQoSSingleAgentCurrentClient))
  Then
    SwapAgents(CurrentAgents, NewAgent);
  Else if (MaxQoS > CurrentQoS)
  Then
    SwapAgents(CurrentAgents, NewAgents[]);
  End
```

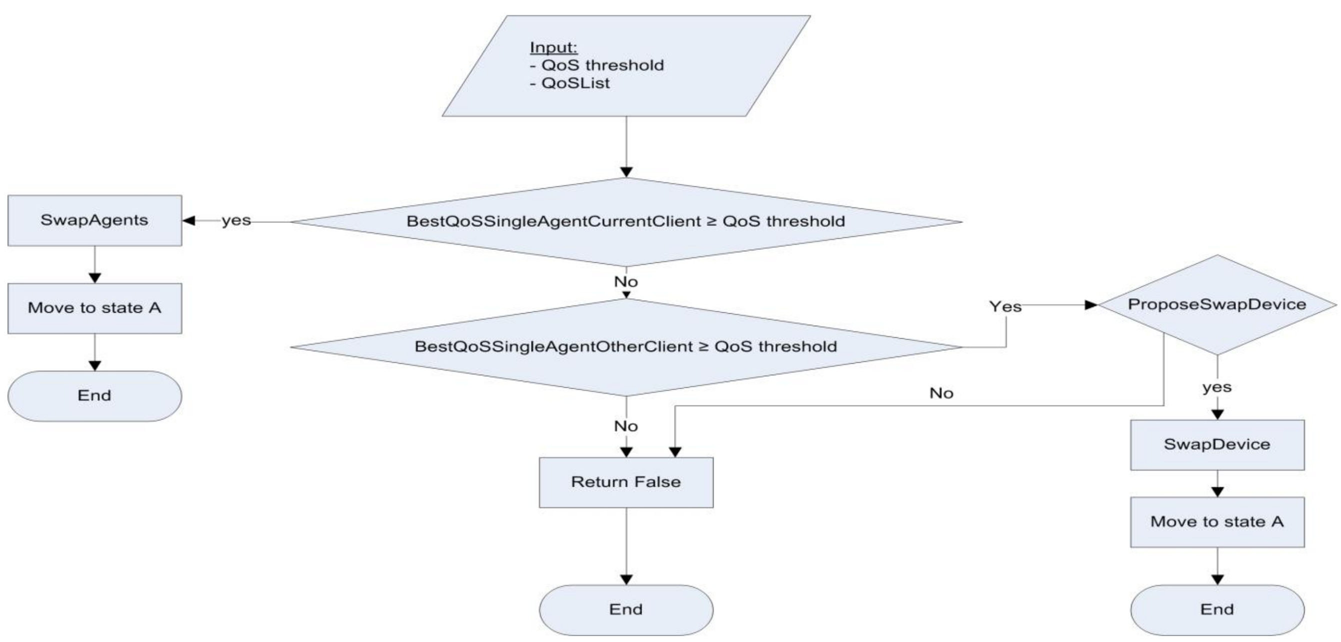


Fig. 5. Swap agent by overhead procedure Flow Chart

In sub-state C the QoS is equal or higher than the preferred QoS, but is supplied by more than one agent. In this case the system should try to reduce network resource overhead, by attempting to supply the preferred QoS using one agent if possible (requirement G2). The system should try to reduce the number of active agents. This is done by the `SwapAgentbyOverhead()` procedure that runs periodically. The flow chart for this procedure is presented in Fig. 5. The procedure updates the QoS state machine accordingly.

In sub-state D the QoS is below the required threshold. In this case the system attempts to improve the QoS; if this attempt fails the service is terminated (see Q2 and Q6.). The pseudo code of this procedure is as follows:

```
  StateDProc():
  Begin
  Bool IsPreferredQoS, IsRequiredQoS = false;
  IsPreferredQoS=SwapAgentbyQoS(PreferredQos)
;
  If (PreferredQos) Then
      Move to State A;
      Return;
  IsRequiredQoS                                =
SwapAgentbyQoS(RequiredQos);
  If (!IsRequiredQoS) Then
      Notify user of service termination;
      Exit Active Service State Machine;
      Service state = Not Active;
      Terminate service;
  Else
    ImproveQoSCurrentClient;
    Move to State B;
  End
```

B. Agent State Machine

In this sub-section we introduce the agent state machine (Fig. 6). As mentioned earlier, for active service, one or more agents can be used for service supply (they are in the “active” state), while the other agents for this service are in the “available” state. Upon service activation the system sends activation messages to all user clients (for the current service). The clients instruct their agents to move to the “*available state*”. The clients and agents must respond to the activation message. This process is referred to as the “handshake procedure”. The outcome of this “handshake procedure” is a list of all available agents for the current service. In order to keep the available agents list updated, all agents must send (by client) keep-alive messages. When a client recognizes that agent/s becomes available after unavailability (for example device turn on) a notification is sent to the system. If the service is active, an activation command to the agent/s is sent. Similarly, when agents become unavailable for active or paused service, they should inform the system if possible. Once an agent does not send a keep-alive message for a specific period of time it is removed from the available agents list (“*not active*” state). Agent that has completed the handshake procedure is regarded as in *available* or *active* state until service termination.

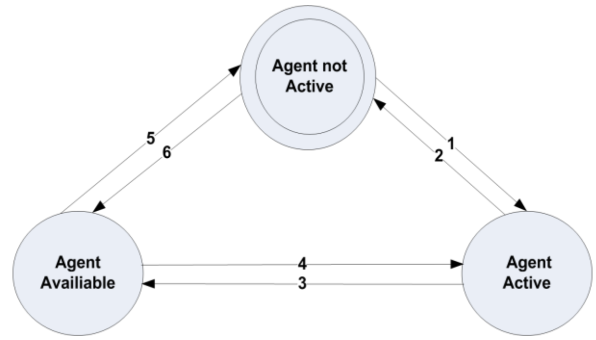


Fig. 6. Agent State Machine

Below are the detailed agent state transitions of Fig. 6.

1) From “*Not Active” to “Active”*. This transition occurs only upon service activation on the specific device.

2) From *“Active” to “Not Active”*. This transition takes place when the agent was one of the agents that provided the service and the service is terminated due to a user's command or due poor QoS (see sub-section QoS specification above).

3) From *“Active” to “Available”*, this transition occurs in the case of agents/client switching for QoS reasons (see QoS specification sub-section above) or a service state change from active to paused (see service state machine sub-section above).

4) From *“Available” to “Active”*, this can occur if the agent is selected by the system for service transmission for QoS reasons or when the service state moves from “Paused” to “Active” by command from this agent.

5) From *“Available” to “Not Active”*, this transition takes place for available agents of a service when the service state is changed to “Not Active” (from “Active” or “Paused” states).

6) From *“Not Active” to “Available”*, this transition takes place for all agents upon service activation command.

In “Not Active” state the agents have no connection to the system. In the “Available” state the agents are connected to the system and keep-alive messages are exchanged as described above. An agent is “Active” if it is one of the agents that supply the service.

C. User Interface Specifications

User interface has to meet requirements G3, U3, U4, U5, U6, Q1 and P1. In order to meet privacy requirements (P1) the interface to a service should be a remote web page and it should be password protected, for example.

The actions that can be performed by the user interface are:

1) *Join the service:* First, the user has to register and to accept a username and password (requirement P1) to access the system. The next step is setting the preferred and required QoS (requirement Q1). The user can choose the default values. The use of these values is described in the QoS specification sub-section above. Then, the user should register the devices; this process is described below.

2) *Client and agent management:* Client and agent management is divided into three sub-processes: client/agent addition, client/agent subtraction and client transition management. As a guideline the user does not have to address agents. The user chooses to alter the client setting; the system requests the data (list of connection interfaces) from the client and presents the options to the user. These procedures can be implemented at any time according to requirement U5.

(i) *Client/Agent addition*: When subscribing a device or agent to a service, a user can choose to connect from the current device ("Add this device") or to identify the subject device by a unique identification, an IP address for instance. After the device is chosen the system queries the device about possible agents (connection interfaces). All possible agents are presented to the user and the user chooses which agents to install. In case of agent addition the user should choose "modify client settings" and the system will return all the options. After the user confirms the choice the profile at the client side is updated. If the number of potential agents exceeds the A parameter from requirement G3, the system displays the message to the user and client/agent addition is aborted.

(ii) *Client/Agent subtraction*: The procedure can be performed from any device. In agent subtraction, as mentioned above, user chooses "modify client settings" and the system shows all agents. In the case of client subtraction the system presents the user list of all of its clients (devices) and the user chooses the client/s to subtract. The system does not allow subtraction of the last client/agent (requirement U4). After the user confirms the choice the profile at both server and client sides is updated.

3) *Client transition management:* In the QoS specifications sub-section, it was noted that in case of insufficient QoS or inefficient use of network resources the system can suggest switching clients to the user. As mentioned earlier the system should try to minimize the use of this option, to meet the requirement for minimal user intervention (G2). For this reason there should be a minimum period of G1 minutes between two successive proposals to switch devices. Additionally, the system should store recently (for G2 minutes) used agents and clients, in order to avoid "ping-pong" transitions between clients. For the user's convenience the system should avoid proposing invalid transitions. For this reason the system stores a list of forbidden client transitions, per user. The user should manage this list. This list can be updated in two ways: by the user interface or when the system displays the list of optional clients for transition, the user can assign "Do not propose this transition again" to one or more clients. The system also should not propose a switch to a device in active service (of any user) to comply with requirement U15. In summary, client transition can be proposed to a user if: this is the first proposal for G1 minutes and there exists a potential device for transition that has not been used in the last G2 by this user and the service, does not appear on the forbidden transitions list, does not run an ubiquitous service to any user and meets the QoS threshold. Upon the user's approval of the transition client, a swapping procedure is performed. Client swapping is similar to pausing and resuming the service on a new client; the only difference is that the "pausing" is triggered by the user's approval of the client switching and not by the pause command. Pausing and resuming service is described in the service state machine sub-section above. Further work could be done to elaborate the list of proposed clients for transition, by studying users' preferred transitions, for example.

4) *Activation, stopping, pausing and resumption of a service:* These actions are described in the service state machine and the QoS sub-sections. For privacy (requirement P1) the activation and resumption of a service should be password protected operations.

5) *Multiple user service management:* The interface is described in the multiple user service management sub-section below.

6) *Disjoin the service:* The user should choose to unsubscribe to the service. After the user confirms the choice, all associated session information on both the client and server sides are removed.

D.Multiple User Specifications

One aspect of ubiquitous service is that several users may start to consume a service together and then want to continue to consume it separately at a different time and place. The requirements for multiple users (M1 – M5) specify this case.

In order to support the multiple user mode the system and the client need to support the following actions:

- Add user to service user group.
- Remove user from service user group.
- Primary user substitution.

All of the above only apply to active service.

1) *Add user to service user group:* To add a user to a service group, the primary user should choose the "Add user to the current service" option and specify the user. The system asks for the new user's confirmation, according to

requirement M2, by requesting a password. Once the subject user is approved to join the service, the system performs the following steps: the user is added to the service's user group, a connection is established with all the users' clients to move agents to the “available” state and to generate a list of available agents. The service state for the new user is "paused". Note that the user must be subscribed to the service, but the device can only be subscribed by the primary user.

2) *Remove user from service group:* Confirmation by the subject user is performed as above. If the primary user wants to leave the service a new primary user should be chosen according to requirement M5 (below). Then the departing user should chose to stop or to pause the service (requirement M3).

3) *Primary user substitution:* The following steps should be performed: (i) upon user request to change primary user, the system should generate a list of potential primary users. The new primary user can be a user who is in the service's user group and subscribed the current device to the service. If there is no such user, the primary user change cannot be done. (ii) The list of potential primary users is displayed to the current primary user and he/she should choose the new primary user. (iii) The last step is to stage the competition between agents of the new primary user on the current client, and to choose active agent/s for the service [8]. The previous primary user becomes the secondary user and stays in the service's user group.

E. User Mobility

User mobility is a focal issue in ubiquitous service. In this sub-section we show that the mobility issue is resolved in our system. User mobility can have two negative effects: the user needs to switch devices and/or the QoS degrades due to coverage or load issues. If the user needs to switch devices this should be done by pausing the service on the old client and resuming it on the new one. The case of QoS degradation is discussed above. Thus our specifications resolve the user mobility issue without having to take any location associated actions, by enabling a high level of customization and addressing QoS as a general issue that is not related solely to mobility.

IV. Performance Evaluation

In the performance evaluation of our method we consider the following additional competing methods: single transmission (that is, the way live video is transmitted today) and simple (not controlled or coordinated) multiple transmission of 2-5 agents. In simple multiple transmissions, the agents transmit the video streams without coordination with the server and the client joins the streams using the simple join function of the minimal arrival time. For every packet sequence number, the client considers the first instance to arrive. That is, the resulting arrival times are the minimum arrivals times of every packet.

Due to the short distance between the agents, we cannot assume that they are statistically independent. Therefore, our method for evaluating the suggested solution is by measurements of real traffic, rather than theoretical analysis. First, we transmit video using several agents in various conditions in order to collect the data. The transmission of the agents was done using LU60 of LiveU [11], using one to five cellular modems connected to three different cellular networks. Each agent has a different connection to the internet. Next, a feasible solution for splitting and joining is implemented. We record the received data with LiveU's server (LU1000) [11] and also using 'Wireshark' software [12]. We collect data which is relevant to parameters such as delay, jitter and retransmission ratio. Therefore, we record for each packet in each transmission from each agent the

TABLE III
Packet Loss Ratio

Process	Average	Worst case
1 agent	2.56%	17.0%
2 agents	0.04%	1.22%

TABLE IV
Average Percentages of Jitter Condition Violation

Cond.	Best (A)	Best (B)	Best (C)	1 agt.	2 agt.	3 agt.	4 agt.	5 agt.
>0	66%	79%	83%	80%	73%	70%	69%	69%
>1	61%	75%	79%	76%	69%	66%	64%	64%
>2	58%	71%	75%	75%	67%	63%	61%	60%
>3	55%	67%	71%	73%	65%	60%	57%	56%
>4	45%	49%	51%	69%	56%	48%	41%	37%
>5	42%	44%	46%	67%	54%	44%	37%	31%
>6	37%	37%	38%	65%	50%	39%	31%	24%
>7	35%	35%	35%	63%	48%	37%	28%	22%
>8	33%	33%	33%	60%	45%	34%	26%	20%
>9	26%	27%	28%	53%	38%	29%	22%	17%
>10	13%	14%	16%	39%	26%	19%	15%	12%
>11	7%	8%	11%	32%	20%	14%	11%	9%
>12	6%	7%	10%	30%	18%	13%	10%	8%
>13	5%	7%	9%	29%	17%	12%	9%	8%
>14	5%	7%	9%	28%	17%	12%	9%	8%
>15	5%	6%	9%	28%	16%	11%	9%	7%
>16	5%	6%	8%	27%	16%	11%	8%	7%
>17	4%	6%	8%	26%	15%	10%	8%	6%
>18	4%	5%	7%	25%	14%	10%	7%	6%
>19	4%	4%	6%	21%	12%	8%	6%	5%
>20	3%	3%	4%	16%	9%	6%	5%	4%
>25	2%	2%	2%	11%	6%	4%	3%	3%
>30	2%	1%	1%	9%	5%	4%	3%	3%
>35	2%	1%	1%	8%	4%	3%	3%	2%
>40	2%	1%	1%	8%	4%	3%	3%	2%
>45	2%	1%	1%	7%	4%	3%	3%	2%
>50	2%	1%	1%	6%	3%	2%	2%	2%

Packet Sequence Number and Time of Arrival. Finally, we evaluate the method's potential performance using the data collected at the beginning.

The recording is done throughout the day including both peak (busy hour) and off-peak hours. Each experiment consists of 5 samples of video transmissions using one to five simultaneous agents. The experiment is repeated 9 times with long video files (about 15 minutes, 30,000-65,000 packets each). In addition, the experiment is repeated twice with short video files (five minutes) to validate that the observed statistic behavior also fit short transmissions. Overall, the recording trace includes statistics of ~ 6 million real packets.

The analysis was performed three times, with jitter requirements of 13 msec. ("Condition A"), 25 msec. ("Condition B"), and with jitter requirement of 40 msec. ("Condition C"). The considered performance parameters are overhead factor, packet loss ratio and jitter.

Regarding the average overhead, processes with one agent naturally have no overhead (factor 1), processes with two agents have overhead of factor 2 (every packet is transmitted twice), and so on. Table I summarizes the overhead factor of our method relative to a single agent process. Each line in the table specifies the average overhead factor, the observed minimum overhead factor and the observed maximum overhead factor. The best k process with parameter 13 has an average overhead factor of 3.08, the best k process with parameter 25 has an average overhead factor of 1.91 and the best k process with parameter 40 has an average overhead factor of 1.65. The differences in the overhead factors are due to the fact that fewer competitions are generated when the requirement from the jitter is less demanding. In a competition all the potential 5 agents transmit two segments, thus, the overhead increases with the number of competitions. Interestingly, the observed minimum overhead factor of the best k process with parameter 40 is 1.09 which is an excellent result. In this observation, only 46 competitions were generated by the algorithm (92 segments) out of 4000 segments in the total transmission and all other 3908 segments were transmitted by a single agent only (97.7%).

To understand the source of the overhead results, Table II plots the percentages of segments transmitted by a number of agents in each algorithm. For example, when using the best k process with parameter 40, an average of 83.7% of the segments were transmitted by only one agent, 0.1% of the segments were transmitted by exactly two agents and 16.2% of the segments were transmitted by all 5 agents (during competitions). Table II illustrates that the best k algorithms with parameters 25 and 40 chose most of the time to use only one transmitting agent, but kept replacing it when its performance decreased. These insights imply that

TABLE I
OVERHEAD FACTOR

Process	Average	Minimum observed	Maximum observed
Best k (A)	3.08	2.05	3.93
Best k (B)	1.91	1.10	2.79
Best k (C)	1.65	1.09	2.72

TABLE II
AVERAGE PERCENTAGES OF TRANSMITTING AGENTS

Process	% using 1 agt.	% using 2 agt.	% using 3 agt.	% using 4 agt.	% using 5 agt.
Best k (A)	45.5%	3.2%	0.0%	0.0%	51.3%
Best k (B)	77.0%	0.4%	0.0%	0.0%	22.6%
Best k (C)	83.7%	0.1%	0.0%	0.0%	16.2%

the overhead can be reduced significantly by developing a different mechanism to replace/select the transmitting agents other than a competition.

Regarding the packet loss ratio, all best k processes and all multiple transmission processes with 3 agents or more have 0.0% average packet loss ratio. The measurements of the processes with one and two agents are presented in Table III. Generally, the networks are reliable and usually the packet loss ratio is very low. However, a very high packet loss ratio of up to 17% packet loss ratio was observed for a single agent in some cases. Naturally, using additional agent reduces the packet loss ratio dramatically, and using more agents or more sophisticated algorithms reduce the packet loss ratio to 0.0.

In order to evaluate the impact of the statistics on the actual user experience we study the function 1-CDF (Cumulative Distribution Function). It represents the average percentages of times that the arrival process violates the corresponding jitter condition. Table IV presents these jitter statistic of the competing methods. Each line describes the average percentages of times that the arrival processes violate the corresponding jitter condition. That is, the packets inter-arrival time is larger than the specified threshold. For example, in line number twelve, the jitter condition is "smaller than 11", and the process "best k with parameter 25" violates this condition in 8% of the samples on average while the process that uses simple multiple transmissions of three agents violates this condition 14% on average. As can be seen from this table, starting from jitter condition "smaller than 11" the best k processes with parameters 13 and 25 outperform the other processes with a significant small number of condition violation. The best k process with parameter 40 behaves very similar to the process with five multiple agents starting from jitter condition "smaller than 13". Note that all best k processes perform at least three times better than single transmission.

All above mentioned results imply that there is no need to require a strong performance condition to improve the performance significantly. The performance of the algorithm under different performance requirements is similar. However, the overhead increases with the performance condition strength. Thus, selecting normal to weak performance condition is recommended.

V. FURTHER WORK AND CONCLUSION

This article described a novel approach to ubiquitous multimedia service - client/server architecture. The system incorporates detailed requirements, specifications and algorithms. We address all known issues related to ubiquitous service: QoS management, efficient usage of network resources, limited overhead, simultaneously usage of a device by a number of users, user mobility and user interface. An additional advantage of our system is that it is network independent, and thus can use any RAT technology. Obviously, it can coexist with other ubiquitous service architectures.

REFERENCES

[1] P. Bellavista, A. Corradi, and L. Foschini, "MUM: a middleware for the provisioning of continuous services to mobile users", In IEEE Sym.. on Computers and Communications, Vol. 1, pp. 498-505, 2004.

[2] P. Bellavista, A. Corradi, and L. Foschini, "Context-aware handoff middleware for transparent service continuity in wireless networks", Pervasive Mob. Comput. vol. 3(4), pp. 439-466, August 2007.

[3] P. Bellavista, A. Corradi, and C. Giannelli, "Mobility-aware management of internet connectivity in always best served wireless scenarios", In Mobile Networks and Applications, Vol. 14, pp. 18-34, Kluwer Academic Publishers, 2009.

[4] Y. Cui, K. Nahrstedt, and D. Xu, "Seamless User-Level Handoff in Ubiquitous Multimedia Service Delivery" Multimedia Tools Appl. vol. 22(2), pp. 137-170, February 2004.

[5] R. Ferrus, O. Sallent, and R. Agusti, "Interworking in heterogeneous wireless networks: comprehensive framework and future trends", IEEE Wireless Communications, vol. 17(2), pp. 22-31, April 2010.

[6] Yong-Ju Lee, Choon-Seo Park, Jin-Whan Jeong, Hag-Young Kim and Cheol-Hoon Lee, "UMOST : Ubiquitous Multimedia Framework for Context-Aware Session Mobility", Int. Conf. on Multimedia and Ubiquitous Engineering (MUE 2008), pp. 3-8, April 2008.

[7] K. Nahrstedt, D. Xu, D. Wichadakul, and B. Li, “QoS-aware middleware for ubiquitous and heterogeneous environments”, IEEE Comm. Magazine, vol 39(11), pp. 140-148, November 2001.

[8] R. Nossenson and O. Markowitz, “Using coordinated agents to improve live media contents transmissions", Int. Conf. on Systems and Networks Communications (ICSNC 2011), pp. 167-170, Oct. 2011.

[9] M. Takemoto, T. Oh-ishi, T. Iwata, Y. Yamato, Y. Tanaka, K. Shinno, S. Tokumoto, and N. Shimamoto, "A service-composition and service-emergence framework for ubiquitous-computing environments", International Symposium on Applications and the Internet Workshops, pp. 313-318, 2004.

[10] M. Takemoto, H. Sunaga, K. Tanaka, H. Matsumura, and E. Shinohara, "The Ubiquitous Service-Oriented Network (USON) - An Approach for a Ubiquitous World Based on P2P Technology", in Proc. Peer-to-Peer Computing, pp.17-24, 2002.

[11] LiveU web site: http://www.liveu.tv/ accessed: January 2012.

[12] Wireshark web site: http://www.wireshark.org/ accessed: January 2012.

[13] R. Nossenson, O. Yudilevich and O. Markowitz, "Client-Server Architecture and Algorithms for Ubiquitous Video Service", The 6th International Conference on Multimedia and Ubiquitous Engineering (MUE 2012), Madrid, Spain, July 2012.

Ronit Nossenson holds a PhD on stochastic models for web servers and an M.Sc. on incremental connectivity of a graph, both from the computer science department at the Technion, Israel institute of technology. She has over 12 years of experience in modeling and analysis of web traffic. Since 2005, she specializes in optimization and trouble shooting of data traffic over cellular network, with tier-one operators' real traffic analysis. Currently, Ronit is the head of the Communication M.Sc. program in the Computer Science department of the Jerusalem College of Technology.

Orit Yudilevich received the BA degree in computer science and mathematics from the Tel-Aviv University and has recently (2011) received her M.Sc. degree in computer science from the Interdisciplinary Center Herzliya. She is now with the Israel Defense Force.

Omer Markowitz received the BA degree in computer science from the Tel aviv-Yaffo Academic College in 2004 and has recently (2011) received his M.Sc. degree also in Computer science from the Interdisciplinary Center Herzliya.

Automatic Multimedia Creation Enriched with Dynamic Conceptual Data

Angel Martín, Haritz Iribas, Ion Alberdi and Naiara Aginako
Digital Tv & Multimedia Services, Vicomtech-IK4 GraphicsMedia.net, Donostia-San Sebastián, Spain

***Abstract* — There is a growing gap between the multimedia production and the context centric multimedia services. The main problem is the under-exploitation of the content creation design. The idea is to support dynamic content generation adapted to the user or display profile. Our work is an implementation of a web platform for automatic generation of multimedia presentations based on SMIL (Synchronized Multimedia Integration Language) standard. The system is able to produce rich media with dynamic multimedia content retrieved automatically from different content databases matching the semantic context. For this purpose, we extend the standard interpretation of SMIL tags in order to accomplish a semantic translation of multimedia objects in database queries. This permits services to take benefit of production process to create customized content enhanced with real time information fed from databases. The described system has been successfully deployed to create advanced context centric weather forecasts.**

***Keywords* — Context-aware Multimedia, Dynamic Multimedia Production, Content Driven Database, SMIL**

I. Introduction

In the recent years, motivated by the rapid rise of the Internet, new challenges have arisen from the increasing amount of audiovisual data that are becoming available. Furthermore, the traditional multimedia production has experienced a deep change due to the arrival of new communication paradigms and the evolution and convergence of new technologies. In this context, distributed multimedia presentations are getting more and more popular for web users. SMIL (Synchronized Multimedia Integration Language) [1] is a W3C standard, designed for describing multimedia presentations which combine audio, video, images, text or any other media.

When the creation is concerned, in order to reduce time and cost of multimedia production, both the possibility of working on them collaboratively and reusing existing content are mandatory. For reutilization purpose the use of meta-information has become a necessity [2]. The increase of multimedia content, and the growing difficulty to search, filter and manage such content, require an effective and efficient multimedia search and retrieval system [3]. In addition, the Web [4] may also be employed as a medium to connect a group of authors, which may be geographically distributed, empowering users to be more creative and fostering social interactions enabling cooperative production [5]. One of the main advantages of team working is to gain efficiency in time to solve complex problems, in order to reach a better final result. Currently, the majority of authoring tools goes online enabling real time creative collaboration processes through cooperative frameworks.

Audiovisual producers gather different multimedia contents and graphic objects composing more attractive and enriched contents in order to engage the audience while consumers are willing to embrace new technologies inherited from other application areas to get richer experiences [6]. However, content-aware research is usually placed at the end of the production process, instead of integrating the solution within the production chain where the different multimedia objects involved are still isolated. This fact facilitates multimedia analysis workout through exploiting content knowledge inside the description of the creation process, in terms of timeline and spatial schedule of multiple objects in the scene [7]. A drawback related to the majority of solutions is the interoperability, they operate over private formats that hide the creation process and makes not feasible to exploit this information for other purposes. Another major trend of context-aware multimedia research that could take advantage of multimedia creation descriptors focus on dynamic generation of multimedia presentations.

This paper proposes an automatic dynamic content generator platform based on SMIL standard. Regarding automatic generation of video content in order to address the gap between visualization layer and dynamic data retrieval for databases, we have designed standard-based definitions. The main benefit is the optimization of the costs of the producing process, as well as saving time. In our context, the creation of multimedia presentations, the simplest way for that is the reusability of previously created multimedia videos and templates into new ones. Besides, in this

DOI: 10.9781/ijimai.2012.175

automatic generation of content, we are able to insert new 2D-3D objects as the spatial relation among different objects of the scene is defined in the production process. Our work has been tested in a weather forecast platform.

The rest of the paper is organized as follows. In Section II, we introduce the related work involved in the development, while in Section III we explain the implementation which includes architecture and the main features of our platform. In section IV we state the validation of the platform in the meteorological domain. Finally, we give some conclusion remarks in Section V.

II. RELATED WORK

We can classify the related work section into several broad categories: collaborative SMIL authoring tools; SMIL players; and the metadata exploitation for search and retrieval and Content Management System (CMS).

There are several advantages of using the web standard SMIL instead of tailoring a custom XML. First, the standard compliance eases the interoperability with other solutions or platforms and future developments. Furthermore, it provides access to a huge audience through web browsers. There are several authoring tools for developing multimedia presentations in SMIL. Some of the research solutions like H-SMIL-Net [8], LimSee2 [9] and SMILAuthor2 [10] provide SMIL authoring interface for multimedia objects. They are designed to interpret the temporal behavior of a multimedia presentation, associate hyperlinks with media objects, enable interactivity, and describe the layout of the presentation on a screen. Other authoring tools [11], [12], add to these works the opportunity to work collaboratively to make the SMIL file enriching the final multimedia presentation with different points of view, while reducing the cost and production time. Our approach lives on the previously works aggregating support to additional features described below.

The players can be divided into two categories. On the one hand, the specific SMIL players like GRiNS [13], Ambulant [14] and SmilingWeb [15]. SmilingWeb is cross-platform player for SMIL 3.0 presentations contained in web pages. On the other hand, it should be mentioned that some of the most important media players support SMIL, such as Real Player, Totem, QuickTime and Windows Media Player. However, none of them supports 3D animations and virtual environments rendering that enhance the experience generated by our platform.

The amount of multimedia content that has to be managed has already become unaffordable without a good CMS. This requires an extensive use of multimedia metadata [16]. The SMIL standard provides extra information of the different multimedia objects involved in a content using metadata [17]. Some studies have analyzed how SMIL metadata can apply to the indexing and abstracting of multimedia documentation [18]. Besides, it can be seen how SMIL metadata is used to store information of different templates in databases in some papers [19]. It facilitates the search and retrieval of these templates, making the system more efficient to reuse content. Here, the proposed platform gathers the human knowledge on semantic tags through the multimedia objects contextualization performed in the authoring tool enabling the connection with databases that collect real time information.

The platform described in this paper integrates a SMIL collaborative authoring tool, a multimedia presentations player and a metadata tag aggregator to facilitate the reuse of contents, saving the information of templates in a database.

In addition, we have created a system that using ontologies allows retrieving data from different databases dynamically, so multimedia presentations can be automatically generated using the generated SMIL templates. It should be mentioned that the platform produces videos from these multimedia presentations. Other features provided by our approach include the possibility to add an avatar or virtual character, that makes the created videos more friendly, but also an integrated text to speech converter.

III. IMPLEMENTATION

The architecture of our platform is based on SMIL standard and Gstreamer [20] open source multimedia processing framework. Our approach has extended the semantic meaning of some tags of SMIL in order to perform dynamic content creation while keeping standard compliance. It is achieved by bridging semantic context connections of the different contents involved, such as multimedia elements defined in the outline and scheduled in the layout, with data stored in a database. Once the required data are retrieved from the database according to semantic matching algorithms, Gstreamer is responsible for the creation of the video designed in the SMIL template.

Meanwhile, as we described in the previous section, the SMIL document itself helps the platform to facilitate the work among different people. SMIL describes not only the play out outline but also the production process so it is perfect to trace the activity of each team member in the cooperative creation of multimedia contents. Regarding concurrency and consistency, the SMIL structure centered on parallel or sequence relations between the different elements involved eases the required control. This means, each member can place a multimedia content inside a sequence without timing constraints or in parallel with others establishing just the z-order, from background to foreground, and the position in the layout. The changes are updated in the timeline and the different lines where concurrent contents are represented in order to provide the team current situation awareness. The authoring process driven by a visual representation of the timeline and the different lines available for each content helps to build a consistent SMIL document that depicts the visual designed outline and layout.

Taking into account that the present research in collaborative environments focuses on authoring tools and the required communication and concurrency control on top of XML modules based on Web Services (WS), we have developed a Service Oriented Architecture (SOA) which integrates WS communication and control solutions. However, we will focus on the real novelty around dynamic content generation with data retrieved from databases.

A. *Extending SMIL specification*

For our approach, SMIL document capabilities potential in terms of content management is twofold depending on the metadata granularity. On the one hand, SMIL provides suitable features in terms of CMS where the metadata, that describe the document content, enable indexing of a large library of generated documents while metadata also provide recommendation solution basis in a similar way that MPEG-7 [2] does but easing web publishing and processing. On the other hand, SMIL document templates also support creation of dynamic multimedia reports and contents through the semantic definition of the different multimedia contents involved. This means, aggregated metadata can be exploited in order to achieve dynamic generation of multimedia contents enabling temporal and spatial visual templates that can be fed with real time retrieved data from databases in order to perform updated and customized videos according to user preferences [21]. This way, our developed platform bridges the gap between data visualization and data storage leveraging data available from databases.

However, no tags for this kind of circumstances are conceived at all in SMIL. So, we use existing tags within a new semantic context in order to keep backwards compatibility with current players. These tags would ease the mapping translation from concepts within the SMIL document, defined like a template, to database (DB) queries.

We use the standard <meta> and <metadata> tags as they are suitable from the point of view of aggregating semantic meaning for the different visual objects scheduled in the document while keep fully standard compliance.

Three different semantic layers have been defined:

1) *First layer:* Metadata related to the general document template suitable for ownership and semantic keyword indexing based CMS.

2) *Second layer:* Semantic keywords defining Primary Key (PK) field name and related fields from DB tables.

3) *Third layer:* Multimedia objects and DB fields mapping.

The different roles of the added attributes are scheduled in the Fig. 1.

The meta-information included in the first layer is stored in a database for more exhaustive searches, and to facilitate reuse of content. In the depicted example the SMIL document represents a “weather report” template to show the “weather figures forecast”.

Not just semantic keywords are defined for the general SMIL document, but also for different multimedia objects involved in the second layer. This way advanced CMS indexing is supported easing advanced search of objects inside produced contents by the authoring tool.

The subtemplate structures, defined in the second layer, provide a gathered semantic scheme of required fields from the DB (“attribute”) around a PK (“subtemplate_key”) where the keywords (“content”) provide semantic concepts related to DB table fields. In the example depicted the concept employed as PK is “city” and the required fields to fulfill the necessary data are “city name”, “forecast icon” and “wind speed”. So, these definitions establish the items in order to infer the SQL database queries.

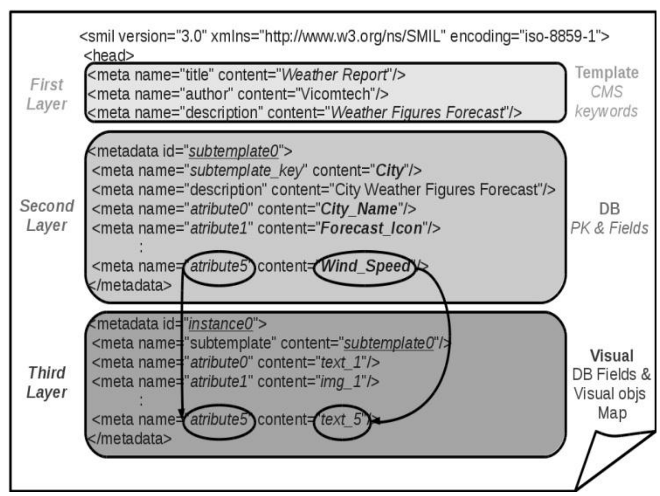


Fig. 1. Semantic data for dynamic performance.

Finally, in the third layer, multimedia object IDs involved in the SMIL document (“content”) are mapped to the DB fields (“attribute”). So contents from layer 2 and 3 are linked through the attributes correspondence. Again, according to the Fig. 1, the values of the layer 2 attributes like “wind speed” retrieved from the fields of different DB tables for a specific PK, in this case a “city”, would be linked to a layer 3 text (“text_5”) respectively. This way the aggregated metadata enable the DB middleware to translate layer 2 structure to DB queries such as SQL

$$select\ f(X)\ from\ T(Y)\ where\ f(Y) = target\_P$$

where:

- f(X) is the result of semantic matching of keyword X with field names from the DB.
- T(Y) is the result of semantic matching of keyword Y with table names from the DB.
- target_PK is the target primary key.

In our example X would be “wind speed”, Y “city” and target_PK the city-ID whose forecast parameters will be introduced in the resulting multimedia forecast report.

The definition of subtemplates in the second layer permits the reutilization of subtemplates when creating new multimedia contents in order to reuse a previous design within the same template document. The subtemplate structure has been conceived like a “Class” of object oriented programming languages while instance structure has been designed as an instance of object oriented programming languages. Therefore, the same dataset definition could have different visual instances within the same template depending on the PK value provided. For example, according to the elements defined in the SMIL document depicted in the Fig. 1 if the designer is interested in create a multimedia forecast for several cities, he should just copy the existing instance and add as many as he wants changing the “instance id” and keeping the “subtemplate id” reference.

B. Dynamic multimedia platform architecture

Once we have defined the extension to SMIL standard in order to both CMS features to reuse the content and make a platform that can create videos with dynamic data automatically; in this section, we give an overview of the developed platform based on Gstreamer to render the result.

This platform provides a solution for:

- User edition through the authoring tool to create both multimedia templates and contents.
- Applications and services that request the automatic generation of a certain content according to the context of the user or periodic time scheduled events.

Fig. 2 shows the general architecture of the approach where individual building blocks are described below:

1) *Authoring tool:* Here, users can work collaboratively in the implementation of the multimedia content. Moreover, to facilitate user work, and remove learning curve of the SMIL language, we have developed a multimedia edition tool based on SMIL standard that supports aggregation of metadata tags according to an ontology in order to accomplish DB incrustation.
2) *Context Scheduler:* Third party application or user context aware service that wants to provide user-centric multimedia content.
3) *DB Translator:* Semantic translator from layer 2 metadata and specified PK values, from the authoring tool or a context scheduler, to DB queries. The translator engine is deployed on top of OWL-API and Pellet solutions [23]. The response is the result of perform the query.
4) *Template DB*: SMIL document templates in a database organized around layer 1 metadata.
5) Template Manager: CMS that manages the library of SMIL document templates.
6) *Content DB:* Multimedia DB with contents for specific applications such as weather forecast reports.
7) *Gstreamer Render:* Multimedia processing platform able to mix and orchestrate contents both online and offline contents (Content DB).
8) *SMIL Parser:* Web requests and responses processor to create a content in SMIL format.
9) *Server:* Web service manages two kind of requests:

a) *Available multimedia templates:* Request includes target topics. So, level 1 keyword based multimedia template search is performed by the Template Manager. The response includes all the recommended templates and their layer 2 and 3 data to be translated to DB queries.

The designed workflow starts when the Authoring Tool for human active creation or Context Scheduler for automatic multimedia generation, requests through a WS some available multimedia templates to provide a multimedia content to a specific user. Here context is performed by means of some semantic keywords that define the multimedia target type.

The Template Manager retrieves SMIL documents exploiting information from layer 1 matching with the request keywords. The response includes exclusively the layer 2 and 3 data of the filtered SMIL templates.

b) *Multimedia content creation:* Request includes level 3 with real DB values for each multimedia item. The generation of the video according to the involved SMIL templates and the data provided is performed. The response is the URL with the generated video.

In case there is a request for multimedia content creation from dynamic data, DB Translator takes charge of processing layer 2 data to transform them into DB queries. Afterwards layer 3 elements are filled by Context Scheduler or Authoring Tool with the figures and values resulting from the DB queries. Here it is important to highlight that values from layer 3 linked to a text object will produce text rendering. While if it is related to image or video objects it refers an URL to an existing content. Last but not least, a text value referred to an audio will unleash a Text To Speech (TTS) audio generation.

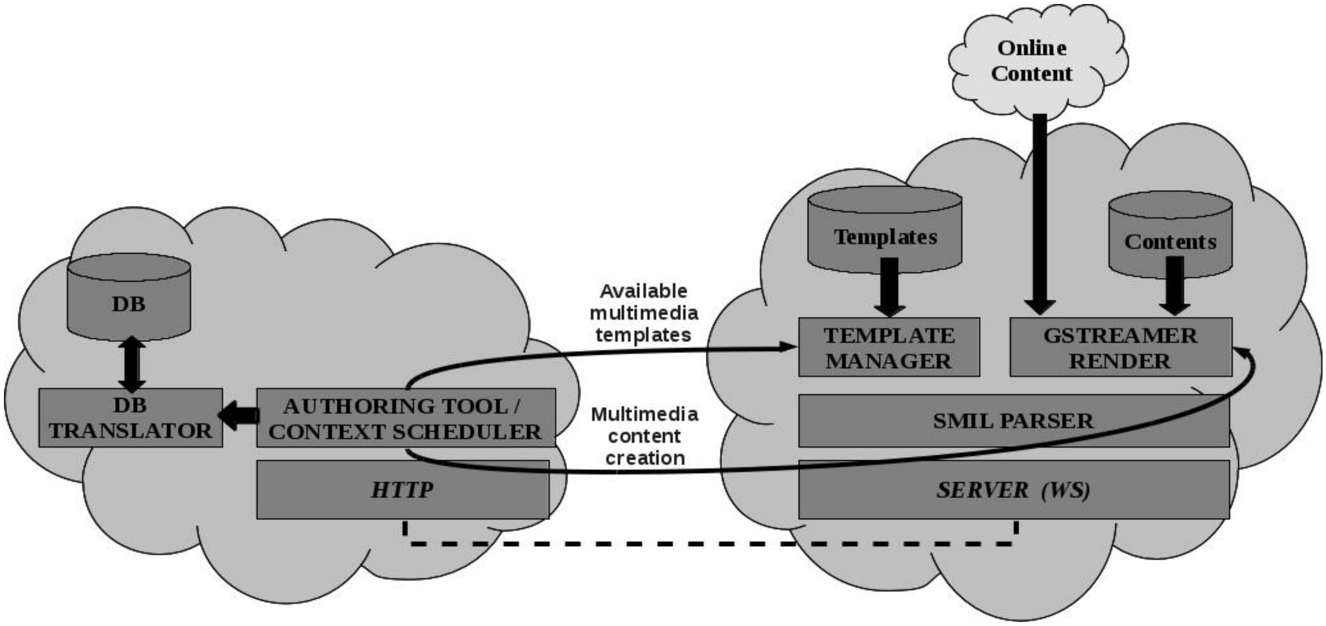


Fig. 2. Architecture for dynamic multimedia creation.

c) Design of ontology: We have designed an ontology in order to define the concept relations between the dynamic content saved in the databases. Therefore, it is possible to manage the information of different databases in order to extract more exact information. The template editor has a limited access to the definition of DB queries so is necessary to implement anarchitecture in order to develop a more powerful tool.

In the meteorological context, we have defined an ontology drawing all the relations between the parameters of weather stations. Consider a real weather station, it contains several meteorological or environmental sensors that measure different forecast parameters. But only measures are relevant, we have to consider more information as sensor model, sensor picture, geolocalization... The definition of the attributes of the concepts is highly relevant.

Ontology permits inserting PK in a total transparent manner from the user point of view. When creating a new template, user inserts a meteorological station ID and the system, using the defined ontology and its relations, shows the possible parameter definitions.

Attributes give the capacity of adequate the presentation of the parameters to the user. Therefore, depending on the nature of each concept, different rules are applied to new objects.

Once all the required data of the desired templates form the final fulfilled SMIL document, it is employed to request the multimedia SMIL driven creation through the WS.

For the process of creating the video, we have developed a SMIL Parser that reads the SMIL document and creates a DOM [22] representation. It extracts the data into a format that we can easily use to create a Gstreamer pipeline. After the platform generates the pipeline with the marks of time and space indicators, GStreamer module executes the pipeline. The GStreamer Render generates the video results and responds with the URL of the video. The Gstreamer core module, gnonlin is in charge of managing the temporal and spatial positions of the different multimedia elements that make up the output video.

To sum up, we have designed a context scheduler that according to time, user preferences and context defines the required PKs in the automatic generation case; or the user himself, in case of on demand authoring, to later perform the renderization request by inserting the retrieved data from the DB inside the corresponding instance of the SMIL document. In order to accomplish our objectives, we have developed all the modules defined in our platform which permits the total control of the automatic content creation.

C. Content-aware production inline

Multimedia applications' next move focuses on 3D-2D fusion and user centric and context-aware that needs of advanced multimedia analysis algorithms to achieve realistic experience and automatic creation of dynamic contents customized to the user context and preferences.

The availability of isolated visual objects within the production chain eases the introduction of 2D-3D objects through the <z-index> SMIL attribute associated to each object. We have exploited the intrinsic information from the production process to remove issues associated to research solutions usually placed at the end of the production chain, such as segmentation, clustering, depth map and more in particular for 3D-2D fusion T-unions, optical flow, partial occlusions, etc. As we keep information of the production process in the SMIL document, we manage to make the later fusion of 3D-2D objects and effects realistically.

During the design process of our content generator platform, we have defined a module to control the z-index of the objects completing a scenario. Therefore, during the creation process of new dynamic content we are able to insert new 2D-3D data to enrich the already defined scenario making the content production affordable.

D. Authoring Tool

The authoring tool to create new objects has been designed as a video editor adapted to user needs. This video editor is divided in two parts. On the one hand, we have the usual editor that permits the mix of audios, videos, texts, images and avatars in the easiest way possible for the multimedia items creators. On the other hand, the collaborative editor permits the reuse of old video templates and videos templates that other collaborators made, and the blend of them for creation of new material.

The whole editor has three main areas structure. On the right there is a preview and options area, on the left there is an elements list and bellow those areas, there is an edition area with the buttons that are necessary for it.

- Elements List

On the top-left side of the interface there is a list of elements with the content library that will be used for the creation of the templates. With the below buttons we can add and remove elements, with the buttons plus and minus respectively, to the list so that the user could use them in the edition.

- Edition Lists

In the video editor there are three elements lists, in the first list are audio elements, in the second list are background elements and the speeches of avatars and in the third list lessons main elements.

The user can manage intuitively all the elements; add, erase and change elements position and duration with the buttons placed at the bottom of the interface. It is so easy to use that e.g. if the user wants to add an element to any list, the system identifies the type of the element and adds to the corresponding list, if it is an audio element it adds in the first list while if it is a video or an image the user just has to define if he wants to add it as a background element or as a main element of the learning object.

- Edition in Collaborative Web Editor

SMIL has different tags for the definition of timing process and the most important for our collaborative system is <par> and <sec> tags.

The elements inside the tag <par> are reproduced all in parallel being grouped to support complementary pedagogical information. The tag <sec> permits the reproduction of elements in sequence enabling sequential outline construction. It is really important because permits to reproduce an element after other without an explicit definition of the start time. The system knows that as soon as one element ends next one must start, maintaining the established order. The user just has to define the order of the elements and when saves it, the system takes the information of the different files and saves in a resultant file.

To sum up, according to the strategy of keeping the developed platform simple, experts can create new objects through defining new multimedia templates or modifying existing ones through a desktop editor application described before. Afterwards the created objects are pushed automatically to the server.

IV. VALIDATION

To test the approach tailored for weather forecast production, validation work has been mainly driven by semantic translation of multimedia objects in database queries by checking correct connection of meteorological concepts, defined by metadata, to real time acquisition databases. SMIL rendering compliance was also performed, where we have created different templates with all kinds of items, including pictures, videos, texts and audios.

Using the authoring tool, we have designed a wide range of meteorological report templates in which different weather maps have been generated as depicted in the Fig. 3. On the one hand, we have produced weather maps where the multimedia elements are not interrelated with semantic information layer. On the other, we defined maps with the items related to weather stations from which data are obtained. In this case tests performed achieve multimedia results automatically adapted to geographical user context and preferences. We also proved the correct implementation of the defined ontologies concepts by retrieving the requested parameters using semantic high-level concepts as weather stations.

As mentioned above, templates are very useful because allow us to create new presentations each day with the same structure, but with new data. These data are acquired from the database, using the weather stations as a PK identifier. In other words, when creating a new video, system translates the metadata attached to a multimedia template item, for example that visualizes the atmospheric pressure, and it takes the value automatically from the database of a specific station according to the user context system.

After testing, we have seen that when introducing elements, the definition of <z-index> in SMIL document has allowed us to use 3D elements, such as avatars, in a direct manner. We resolve the problem of occlusions as the depth of the objects in the scene is already defined.

About the metadata aggregation, the platform makes even easier the use of templates and databases. Access to a database via the web platform makes it more intuitive for inexperienced users. User can provide human understandable meteorological concepts that would be translated to real time acquired atmospheric measures. This way the user just has to provide a geographic context to establish the database PK. After selecting this information, the system automatically creates the queries for the database so that the user does not have to worry about anything. In addition to manually creation of videos, it has proven the periodic creation of videos in an automatic manner.

Fig. 3. Platform produced weather report.

This testbed based in the weather information platform does not have real time performance restrictions or high scalability requirements, but it faces dynamic contexts describing a solution extensible from time and localization features to context user-centered multimedia applications. The automation of this process has enormously reduced the cost of producing new contents for weather forecasts and the time needed for the creation of new videos which implies that new content can be updated with higher frequency.

V. CONCLUSION

Creating new multimedia contents and keeping them as a template in order to ease reutilization, decreases dramatically the amount of digital content replicated in edition platforms. Moreover, the capability of creating this content collaboratively permits new envisaged creation possibilities and sharing of content. The SMIL documents considered as templates also ease the reutilization of production work according to reuse and recycle philosophy of green computing minimizing the set of needed resources.

The redefinition of the metadata tag of SMIL standard permits the declaration of objects in SMIL presentations. Giving a semantic meaning to these metadata, reveals the possibility of creating new updated multimedia content, including dynamic data, that automatically minimizes the production costs of lots of contents. This property becomes crucial in repetitive processes which must be updated with new data, for example in weather reports.

The proposed format to describe the content outline makes feasible to exploit this information for other purposes related to content-aware research, such as auto-summaries, 2D-3D fusion or visual near-duplicated research. This one may be relevant for rights holders to provide control of their content enabling copyright policies and content management through automatic processes to identify, claim, and apply policies to partial or entire content.

Future work will include more complex relations in terms of database structures that will also need of enhanced semantic translators from context based keywords to existing databases. We also plan to take benefit of the multimedia object segmentation provided by the SMIL templates in order to boost visual analytics applications improving the search algorithms.

REFERENCES

[1] W3C, SMIL Specification, http://w3.org/AudioVideo
[2] Cornejo, C. M., Ruiz-Rube, I., Dodero, J. M. “Semantic Management of Digital Contents for the Cultural Domain”, Recent Trends in Information Reuse and Integration, Springer, pp 211-226, 2012
[3] HongJiang Z., "Multimedia search: Past and current approaches", IEEE 10th Workshop on Multimedia Signal Processing, 2008
[4] Olson, G.M, Zimmerman, A., and Bos, N., Eds., "Scientific Collaboration on the Internet“ MIT press, Cambridge, MA, 2008
[5] Sung ,M.Y., Lee D.H., “A Collaborative Multimedia Authoring System”, Lecture Notes in Computer Scien-ce, Volume 3033/2004, pp.311-318, 2004
[6] Jansen, J., Cesar, P., Bulterman, D., Stevens, T., Kegel, I., and Issing, J., “Enabling Composition-Based Video-Conferenfing for the Home”, IEEE transactions on Multimedia, IEEE, Volume 13, pp. 869-881, 2011
[7] R. Kernchen, S. Meissner, K. Moessner, P. Cesar, I. Vaishnavi, M.u Boussard and C. Hesselman, “Intelli-gent Multimedia Presentation in Ubiquitous Multide-vice”, IEEE Computer Society, pp. 52-62, April-June 2010
[8] S. Bouyakoub , A. Belkhir, “H-SMIL-Net: A Hierarchical Petri Net Model for SMIL Documents”, Proceedings of the Tenth International Conference on Computer Modeling and Simulation, p.106-111, 2008
[9] Deltour, R., Layaida, N., Weck, D., “LimSee2: A Cross-Platform SMIL2.0 Authoring Tool”, The European Research Consortium for Informatics and Mathematics – ERCIM News. n. 62 , 2005
[10] Chun-Chuan, Y., Chen-Kuei, C, Yung-Chi, W., “Extension of Timeline-based Editing for Non-deterministic Temporal Behavior in SMIL2.0 Authoring,” Journal of Information Science and Engineering, vol. 24, no. 5, pp. 1377-1395, 2008
[11] Sung, M. Y., Lee, D. H. “A collaborative authoring system for multimedia presentation”, Proceedings of the IEEE International Conference on Communications, 1396-1400, 2004
[12] El Saddik, A., Rahman, Md.A., Hossain, M.A., “Authoring Multimedia Objects in Collaborative Ambient Intelligent Virtual Environment”, IEEE International Workshop on Haptic Audio Visual Environments and their Applications, 2005
[13] Bulterman, D.C.A., Hardman, L., Jansen, J., Mullender, K.S., Rutledge, L., "GRiNS: A GRaphical INterface for creating and playing SMIL documents" WWW7 Conference, Computer Networks and ISDN Systems, vol. 30, pp. 519–529, 1998
[14] Bulterman, D.C.A., Jansen, J., Kleanthous, K., Blom, K., Benden, D., “Ambulant: a fast, multi-platform open source SMIL player”, Proceedings of the 12th annual ACM international conference on Multimedia, 2004
[15] O. Gaggi, L. Danese. “A SMIL player for any web browser”, Proceedings of International Conference on Distributed Multimedia Systems, p 114-119, 2011
[16] Garcia, R., Celma, O. “Semantic Integration and Retrieval of Multimedia Metadata”, 5th Int.Workshop on Knowledge Markup and Semantic Annotation, 2005
[17] Bulterman, D., Rutledge,L., SMIL 3.0 Flexible Multi-media for Web, Mobile Devices and Daisy Talking Books
[18] Gil, B., Pérez, J. R., Sánchez, R., “Gestión de documentación multimedia con Synchronized Multimedia Integration Language (SMIL).” // Scire: representación y organización del conocimiento. 10:2, p 87-97, 2004
[19] Jiang, D., Liu, Z., Zhong, L., "A smil-based multimedia system for mobile education", 2nd IEEE International Conference on Broadband Network & Multimedia Technology, IC-BNMT '09, p 724-728, 2009
[20] Gstreamer web page, http://gstreamer.freedesktop.org/
[21] Steele, R., Lubonski, M., Ventsov, Y., and Lawrence, E., “Accessing SMIL-based Dynamically Adaptable Multimedia Presentations from Mobile Devices”, Proc. Of Int. Conf. on Inf. Technology: Coding and Compu-ting, ITCC04, 2004
[22] W3C, DOM Specification, http://www.w3.org/DOM/

[23] OWL api engine website, http://owlapi.sourceforge.net/

A. Martin. Angel Martin is with the Department of Digital Tv & Multimedia Services, Vicomtech-IK4 GraphicsMedia.net, Spain. He received his engineering degree in 2003 from University Carlos III of Madrid, Spain. He has published many papers on multimedia technology in many conferences. His research interests are in the areas of multimedia services, interactive multimedia and rich media contents creation and transmission.

In 2003 he starts his Phd in the Communications and Signal Processing Department of the University Carlos III of Madrid in the video coding research area. At the same time he collaborates with the company Prodys developing a standard MPEG-4 AVC/H.264 codec for DSP architectures, and web applications for coding management of IP surveillance cameras. In 2005 he starts to work on Telefonica I+D in projects related to multimodal interactive service projects for Home Network environments and web services. Later, also in Telefonica I+D, goes deeper into image processing area in terms of 3D video and multiview coding. From 2008 to 2010 in Innovalia Association as R&D Project consultant developing projects and applications for smart environments, ubiquitous and pervasive computing in the artificial vision area. Currently he is on Vicomtech managing and developing R&D projects around multimedia content services.

I. Alberdi Ion Alberdi is a research assistant at Vicomtech-Ik4 in the area of Digital TV & Multimedia Services. He graduated in Computer Engineering in 2011 from UPV-EHU where he undertook his undergraduate dissertation based on analysis of cartographic imagery at the Italian Università degli Studi di Padova. His research interests are in the areas of multimedia services and computer vision.

H. Iribas. Haritz Iribas is is with the Department of Digital Tv & Multimedia Services, Vicomtech-IK4 GraphicsMedia.net. He graduated in Technical Telecommunications Engineering in 2008 at the Public University of Navarre (UPNA-NUP). Between the 2006 and 2010 he teached sciences at the private schools "Laguntza" and "International college". Since July of 2010 he has been working in Vicomtech-IK4. His research interests are in the areas of multimedia services, interactive multimedia and computer vision.

N. Aginako. Naiara Aginako is with the Department of Digital Tv & Multimedia Services, Vicomtech-IK4 GraphicsMedia.net, Spain. Between 2000 and 2005 she studied Ingenieria Superior de Telecomunicación (Computer Engineering applied to Telecommunication) at the University of the Basque Country (UPV-EHU). She was cooperating with the Electronics and Communications Department of the above mentioned Faculty between 2003 and 2005 where she undertook her undergraduate dissertation.

Since October of 2005 she has been developing and leading R&D projects in Vicomtech-IK4. She undertook her postgraduate course in Informatics Engineering at the University of the Basque Country (UPV-EHU) during 2008-2010. Since 2006, she also teaches junior level courses at the University of the Basque Country at the Electronics and Telecommunications Department.

Accessing Wireless Sensor Networks Via Dynamically Reconfigurable Interaction Models

Maria Cecília Gomes, Hervé Paulino, Adérito Baptista, and Filipe Araújo,
CITI/Departamento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal

***Abstract* — The Wireless Sensor Networks (WSNs) technology is already perceived as fundamental for science across many domains, since it provides a low cost solution for environment monitoring. WSNs representation via the service concept and its inclusion in Web environments, e.g. through Web services, supports particularly their open/standard access and integration. Although such Web enabled WSNs simplify data access, network parameterization and aggregation, the existing interaction models and run-time adaptation mechanisms available to clients are still scarce.**

Nevertheless, applications increasingly demand richer and more flexible accesses besides the traditional client/server. For instance, applications may require a streaming model in order to avoid sequential data requests, or the asynchronous notification of subscribed data through the publish/subscriber. Moreover, the possibility to automatically switch between such models at runtime allows applications to define flexible context-based data acquisition. To this extent, this paper discusses the relevance of the session and pattern abstractions on the design of a middleware prototype providing richer and dynamically reconfigurable interaction models to Web enabled WSNs.

***Keyword* — *Web Enabled* Wireless Sensor Networks, Dynamic Interaction Models, Design Patterns**

I. Introduction

Simulation applications for unexpected but extreme events like large-scale flooding, hurricanes, severe droughts, etc., demand the access to different types of data collected across wide scale geographic areas, and for long periods of time. Only large amounts of diverse data support more precise information extraction and knowledge, concerning a better evaluation of complex events of this kind.

Wireless Sensor Networks (*WSNs*), in particular, offer a good low cost solution for such large-scale environmental monitoring since they comprise a high number of sensor devices deployed throughout the geographic area to be evaluated. Current WSNs may include different types of sensors, spanning from simple and static devices to increasingly complex mobile devices. WSNs allow hence the development of more or less elaborated applications [1] to which the interaction with the real world is a pressing requirement. Such includes not only more traditional applications like the ones mentioned above, but WSNs also allow the surge of novel ones. This is the case of the *Participatory Sensing* area [2] where applications like urban traffic management or virtual communities' support typically rely on data acquisition and dissemination through mobile devices (e.g. using sensors embedded in private cars and mobile phones).

Nevertheless, one disadvantage of WSNs is still their low-level limited interfaces. To this concern, high-level abstractions have been used to simplify WSNs access, allowing their representation as data streams, databases, or through mobile agent models, for instance. Likewise, abstracting WSNs as Web services [3][4] allows their inclusion in Web environments, e.g. in the context of business processes. Namely, the service paradigm via standard Web technologies supports a uniform and simple access to WSNs, their parameterization and aggregation, and the systematic access to collected data.

A service-based access to WSNs also allows their integration with very different systems, since the service paradigm provides a uniform access to, and aggregation of, distinct entities. One example may be the seamless integration of WSNs providing online, almost real-time, data acquisition with Cloud-based applications consuming that data. In fact, and considering the perceivable trend on making everything accessible as a service (*XaaS*), the service concept may provide a powerful but simple abstraction for heterogeneous systems' access, interaction, and integration, may those systems be *Web enabled WSNs*, *Internet of Things* entities (*IoT*) [5][6], *Grid* or *Cloud* computing services (for standardization efforts in this area see [7]), etc.

Nevertheless, the access to those types of services may have requirements behind the traditional request/response interaction, demanding therefore dynamic/richer interaction models [8]. For instance IoT entities having one single client (the owner) may be interfaced through a stateful Web service. Cloud computing services, in turn, may interface stateful resources or long running activities which need to be inspected in terms of resource consumption, dynamic requirements, or overall cost [9][10]. Considering specifically Web enabled WSNs, sensors may have to be inspected/interrogated (e.g. in terms of sensor autonomy evaluation and sensing frequency) and also be modified (e.g. sensor parameterization).

DOI: 10.9781/ijimai.2012.176

Additionally, sensing data may have to be acquired with different *QoS* depending on contextual information (e.g. sensing data streaming on an emergency situation versus periodic data notification for sensors' autonomy preservation).

WSNs accesses may consequently be modeled as Web services interfacing *stateful resources* [11][12] requiring the realization of a dynamic/variable state which has to be kept consistent along several message exchanges between a service and each one of its clients [13]. This is captured in the *Web Services Resource Framework* (*WSRF*) norm [14], which had its origin in the context of Grid computing [13] in order to represent the access to typical, long running, *High Performance Computing* applications. Consequently, such Web enabled WSNs may benefit from richer/dynamic interaction models for sensor data acquisition and dissemination that however are not generally available in current solutions.

The following sections describe the dimensions concerning such limitations and propose a solution towards richer interactions for Web enabled WSNs access. Subsequent Sections IV and V describe, respectively, the implementation architecture and an application scenario. The conclusions are described in the final section as well as future work.

II. Problem Dimensions

Consider an emergency application for a critical area prone to cyclic wild fire situations. In order to more accurately calculate a fire ignition probability [15], simulation applications in this domain benefit from consuming almost real-time/online sensing data provided by different types of WSNs deployed in the area, e.g. temperature, humidity and wind characteristics' monitoring. Under normal conditions, temperature data acquisition from a single type of sensors may be enough. However, in the presence of draught weather conditions, more precise temperature data may be needed, e.g. collected from different sensors at different heights. Moreover, if a fire ignition does occur, different types of data like wind velocity and direction are also needed.

In case a client uses a traditional request/reply interaction model to collected sensing data, several independent client requests are necessary in order to process enough quantities of different data. One solution is to support the collective data processing and dissemination as a single interaction action, similarly to what happens in the *mashups* concept. Moreover, the supporting system should allow the dynamic selection of those data sources at runtime. Additionally, and due to the low autonomy of typical sensing devices, the *QoS* in terms data acquisition rate and delivery should be low under normal situations, e.g. winter time for the fire application, and high in emergency situations.

Other requirements may also be considered. For instance, critical data may be needed not only to the fire simulation's execution, but also to the firemen deployed in the area. Namely, these may be using mobile devices for their coordination and relevant data may now depend upon their geographical location (e.g. data collected at the vicinities of the firemen's position). Likewise, if the mobile devices have already a low battery level, a data stream cannot be processed anymore, but sporadic data delivery is required instead.

Whatever the clients' perspective, the most adequate solution is to provide flexibility on data sources' dynamic selection and aggregation, and also in terms of the data acquisition rate. Such may be supported through selecting an adequate interaction model between the service and its clients, at some point in time. The supporting system should also provide their dynamic modification based on context data. For instance, a *Streaming* model is preferable for a continuous sensing data delivery; a *Producer/Consumer* is necessary if there are data delivery requirements; and a Publish/subscriber model is more adequate whenever low rate data transmission is enough.

Having defined such a (more or less) complex monitoring scenario for different Web enabled WSNs data sensing, its reuse for related clients/applications may also be useful. For example, the described scenario could be used in the context of a similar tornado simulation application for the area. Likewise, in case additional firemen corporations are deployed into the affected location, the contextual information perceived by the former firemen should be quickly and easily shared to the new ones.

Contextual information sharing may also support the coordination of relevant agents, e.g. considering that emergency protocols have to be precisely defined and known both by the actors in the field and authority entities. Based on a common context, emergency support systems may hence enforce some forms of pre-defined automatic dynamic reconfiguration capabilities concerning the evolution of a critical event. Such rules may be incorporated in those systems and be automatically triggered in face of particular events, e.g. sensing data values collected in a problematic area may trigger a switch from normal to an emergency situation.

Therefore, it is our opinion that richer/dynamic interaction models are necessary on accessing Web enabled WSNs and that they should be captured allowing their sharing among different clients and reuse for similar situations. In the following, we propose a novel session-based abstraction to represent and contextualize such dynamic interaction models.

III. Proposed Solution

The conceptual view of the proposed solution is depicted in Fig. 1. The middleware layer hides the details inherent to accessing Web enabled WSNs and provides an interaction context to clients, either individual or to a set (e.g. clients which may benefit from sharing a particular interaction). The solution is based on a) the *Session* concept to capture dynamic rich interactions with Web enabled WSNs, and on b) the *Pattern* concept to implement a confined, structured, and well defined mechanism for dynamic reconfiguration within a session context.

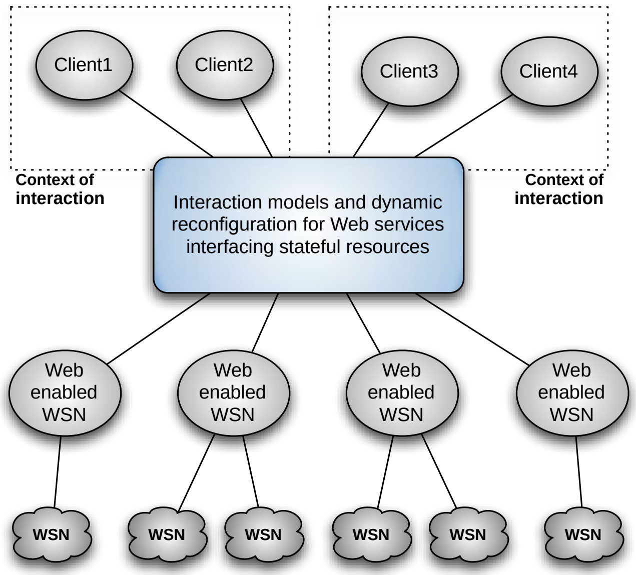


Fig. 1. Conceptual view.

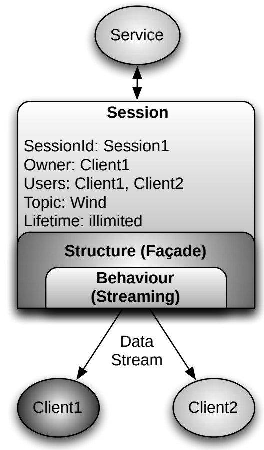


Fig. 2. Session abstraction with a Stream-based interaction model.

A. *Sessions Capturing Dynamic Interaction Models*

The *Session* concept represents the interaction context of a set of users accessing the same Web enabled WSN services, as well as the dynamic reconfiguration features possible within that context. A session includes:

1) The identification of the data sources plus the particular interaction model in use at some point in time for data dissemination. All client accesses within this session's context obey the semantics of that interaction model, which defines the service/users' data and control flow dependencies. Basic interaction models are Client/Server, Publish/Subscriber, Streaming, and Producer/Consumer. Fig. 2 depicts an example for a Wind data source, whose data is disseminated through a Streaming interaction model.
2) Management information, such as a unique session identifier used by new clients to join the session; the identifiers of the session's current members; the identifier of the *session's owner*, the sole that can perform explicit dynamic reconfigurations and terminate the session; and the session's life time limit which when expired causes the session's termination and the consequent notification of all its members. If this time is *unbounded*, the owner must explicitly request the termination. The session in Fig. 2 has two clients and an unbound lifetime limit.
3) The possible adaptation mechanisms consisting of structured and context-based dynamic reconfigurations. These depend on the characteristics of the WSNs service/users interaction context and may also be pre-defined:
   - *The interaction's* context includes:
     i. The context of the service client (e.g. a mobile device with limited autonomy or progressing to a different geographic area).
     ii. The interaction medium between the Web enabled WSN service and its user (e.g. the characteristics of the supporting communication networks).
     iii. The Web enabled WSN service's context (e.g. services representing relevant data sources like temperature or humidity sensing data whose critical values have to be acknowledged).
   - *System evolution results from on-demand/pre-defined interaction models' dynamic modifications*. Users may explicitly require dynamic reconfigurations, or these may be automatically triggered by the runtime system based on pre-defined rules and upon change detection of the cited interaction context.

B. *Pattern-based Dynamic Interaction Models*

Within a session's context, the *pattern concept* is used both

- To implement the interaction model in use by all clients belonging to the session at some point in time; and
- To provide a structured dynamic adaptation mechanism ruling a session's evolution.

Implementing the Session's Interaction Model

Patterns underlie an interaction model's implementation in the context of a session. Such is accomplished following the ideas in [16] where pattern abstractions in the form of parameterized *Pattern Templates* capture structure and behavior with separation of concerns, allowing their flexible composition.

The implementation of a particular interaction model is based on the composition of one or more structural patterns with a behavioral pattern. *Structural Patterns* capture a session's "static view" in terms of the structural dependences/relations among its members (e.g. a Façade or a pipeline) without specifying any restrictions in terms of data or control flows.

The "dynamic view" is defined, on the other hand, by *Behavioral Patterns* like *Producer/Consumer*, *Streaming*, *Publish/Subscriber*, and so on. These characterize the dependencies in terms of data and control flows among a session's members, as well as their role concerning the behavioral patterns' semantics (e.g. roles of *producer* and *consumer* when considering the Producer/Consumer pattern).

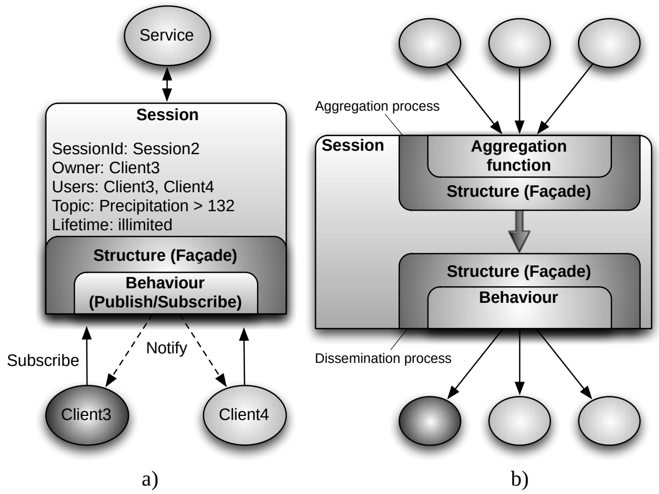


Fig. 3. Session's Interaction Models implemented as the composition of structural and behavioral patterns.

The left-hand side of Fig. 3 (a) presents the composition of a Façade structural pattern with the Publish/Subscriber behavioral pattern. The Façade captures the common interface for data dissemination to all clients in the session, and the behavior defines how that data is disseminated to session's clients.

Different interaction models enable the presentation of data flows with distinct quality services at different points in time. This allows diversity on accessing Web enabled WSN services/data sources, as well as for their modification when convenient. For example, the use of a Client/Server model to inspect a data source versus a Publish/Subscriber model to receive asynchronous event notifications.

The right-hand side of Fig. 3 (b), in turn, presents the implementation of an *Aggregation* model in the context of a session, which consists on the aggregation, and possible processing, of multiple data sources, and their dissemination. Such is supported by a hierarchical structure, namely a two-staged process (a two stage pipeline structure) consisting of an aggregation phase and a dissemination phase. Both phases must present the same behavior, for instance, an aggregation of streams must be disseminated according to the Streaming behavior. The logic used to combine the multiple data sources is defined in the form of an *aggregation function* parameterized upon the model's definition. This approach accommodates the definition of application-specific stream processing techniques to filter the data, compute statistics, and so forth.

Structured Dynamic Adaptation

The pattern abstraction also supports a structured dynamic adaptation mechanism dependent on the current state of the interaction's context. As a result

a) Each pattern can be directly reconfigured at runtime, both in the dimensions of structure and/or behavior (e.g. to replace a behavior by another one);
b) The adaptation/evolution of the system may be represented as a pre-defined sequence of patterns captured as a state machine (see Section IV).

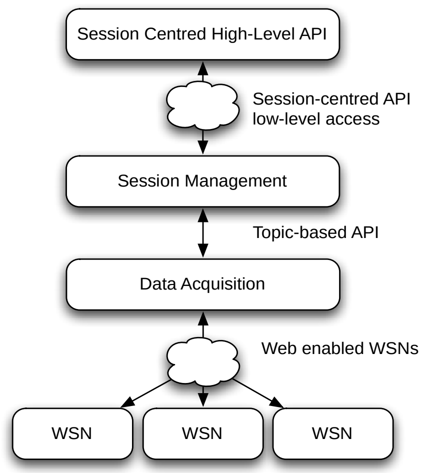


Fig. 4. Overall architecture.

IV. A Middleware for WSNs Adaptable Access

The proposed middleware implements the concepts described in the previous section providing rich and dynamic interaction models for Web enabled WSNs. It is implemented as a Web accessible platform upon which sessions can be shared by multiple geographically dispersed users.

The middleware's architecture, depicted in Fig. 4, follows a multi-tier model that cleanly separates the multiple concerns of the system, such as presentation, logic and data access. From a bottom-up perspective, the layers that compose the middleware are:

- ***Data Acquisition:*** interacts with Web enabled WSNs, the data sources, providing a topic-based API. Upper layers can hence associate topics to data sources or define restrictions on those same sources. For example, a topic may refer to a stream of data produced by a given service or only to the items of the stream that obey a given condition (e.g. subscription of precipitation levels above 132 units, as depicted in the left-hand side of Fig. 3 (b)).
- ***Session Management*****:** implements the session abstraction, supplying tools for session creation/termination; session management, ranging from membership accounting to parameter configuration (e.g. lifetime specification); and possible dynamic reconfiguration mechanisms. Since a session may comprise geographically dispersed members, this layer exposes a simple Web service interface intended to be used by higher-level language APIs.
- ***Session-Centered High-Level API*****:** provides a high-level session-centric interface for the cited capabilities.

The remainder of this section will further detail the *Session Management layer*, the core of the middleware, and the *Session-Centered API* used in the example of Section V.

A. Session Management Layer

A session hosts a single behavior/interaction model to which all of its clients are automatically bound. This behavior must be defined when the session is created but may also change in time, as a response to a reconfiguration action. The client that creates a session is titled its *owner* and is the sole with

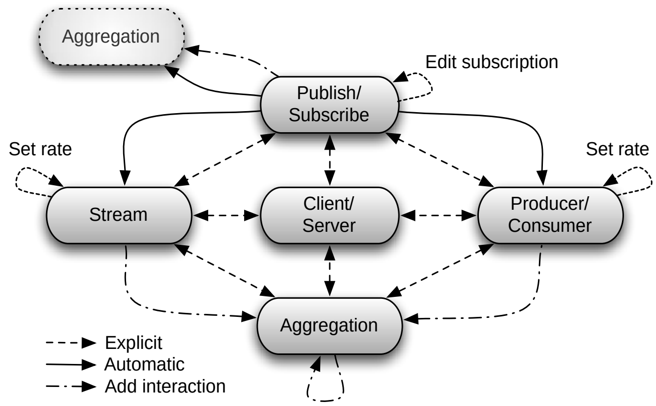


Fig. 5. Reconfiguration state machine: explicit reconfigurations.

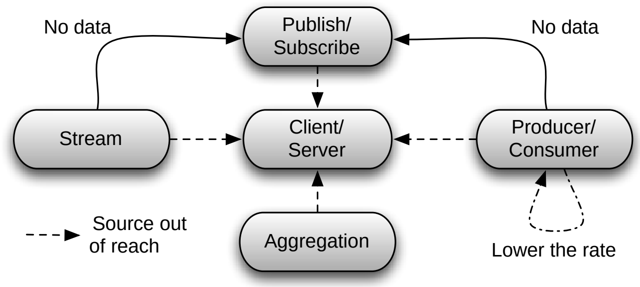


Fig. 6. Reconfiguration state machine: implicit reconfigurations.

permissions to perform reconfiguration actions that have a session-wide impact. The other members must comply with the session's current configuration, and adapt to any consummated reconfiguration or leave. The composition of one or more structural patterns with a behavioral pattern provides the framework upon which sessions are implemented, as described in Section III.B.

Pattern-based Dynamic Reconfiguration

The reconfiguration mechanisms featured in the middleware have the purpose of adapting, in the context of a session, the way a particular client or a set of clients (the session's members) interacts with a set of Web services.

The separation of the *session*, *structure*, and *behavior* concepts, and the way they are combined to support session execution, cleanly evidences the responsibility of each one. For instance, the session contextualizes the overall interaction; a new client joining an existing session is captured as a structural reconfiguration independent from the behavior (i.e. the new client has the same behavior as the other existing clients in the same session); the replacement of the session's interaction model in use is captured as a behavior reconfiguration independent from the structure (all clients in the session are notified of a new behavior ruling data dissemination).

Additionally, the reconfiguration actions can be characterized as implicit (automatically triggered by the middleware) and explicit (requested by a client). Orthogonally, their scope may be confined to the tuning of the current interaction model, or have a session-wide impact, replacing the current model altogether. The conjunction of all the reconfigurations supported by the middleware defines a state machine whose description follows.

Explicit Reconfigurations

Valid reconfiguration requests may be issued by any member of any session, at any moment in time. Their purpose is twofold: to tune or to replace the current interaction model. Tuning requests are model dependent, and must conform to the currently active reconfiguration interface. For instance, setting the data rate is only available in the *Streaming* and *Producer/Consumer* models.

The remainder requests have a broader impact and thus have their semantics bound to the role of the client in the session. Only a reconfiguration request issued by the session's owner may encompass the entire session. The other members are notified of such reconfiguration and will have to adapt to the new configuration or leave the session. Requests issued by some other member than the owner do not affect the target session. It is the client that is moved to another session fulfilling the required parameters. If no such session exists at the time, it is created on the fly.

Fig. 5 illustrates the transitions of the state machine that are triggered by explicit requests. The ones that actually perform a state transition have been divided into three categories:

- ***Explicit***: an explicit *reconfigure* request.
- ***Automatic***: reconfiguration actions that, when in the scope of a *Publish/Subscribe* model, can be programmatically associated to a particular topic subscription. As soon as the middleware receives a notification on that topic it automatically reconfigures the client, according to its role in the session (owner or regular member).
- ***Add interaction:*** addition of new data sources to the session. This reconfiguration forces the interaction model to become an aggregation, being that the dissemination model is inherited from the current configuration, e.g. adding a new source to a stream will result in the aggregation of two streams.

Implicit Reconfigurations

These constitute responses to changes in the context of the client, the service, or their communication channel. Their purpose is to ensure that the data flow between a session's sources and clients is adjusted according to the session configuration parameters and the ability of the sources to meet these requirements.

Fig. 6 presents the transitions of the state machine dedicated to this type of reconfigurations. Three scenarios are handled:

- ***Session out of reach***: this transition is triggered whenever the data source is no longer reachable. The session's clients are notified of the incident and from that point on they will only able to interact with the source through the Client/Server model. Naturally, as long as the source is out of reach, any request will return an error message.
- ***No data***: when in the context of the Stream and Producer/Consumer interaction models, the absence of

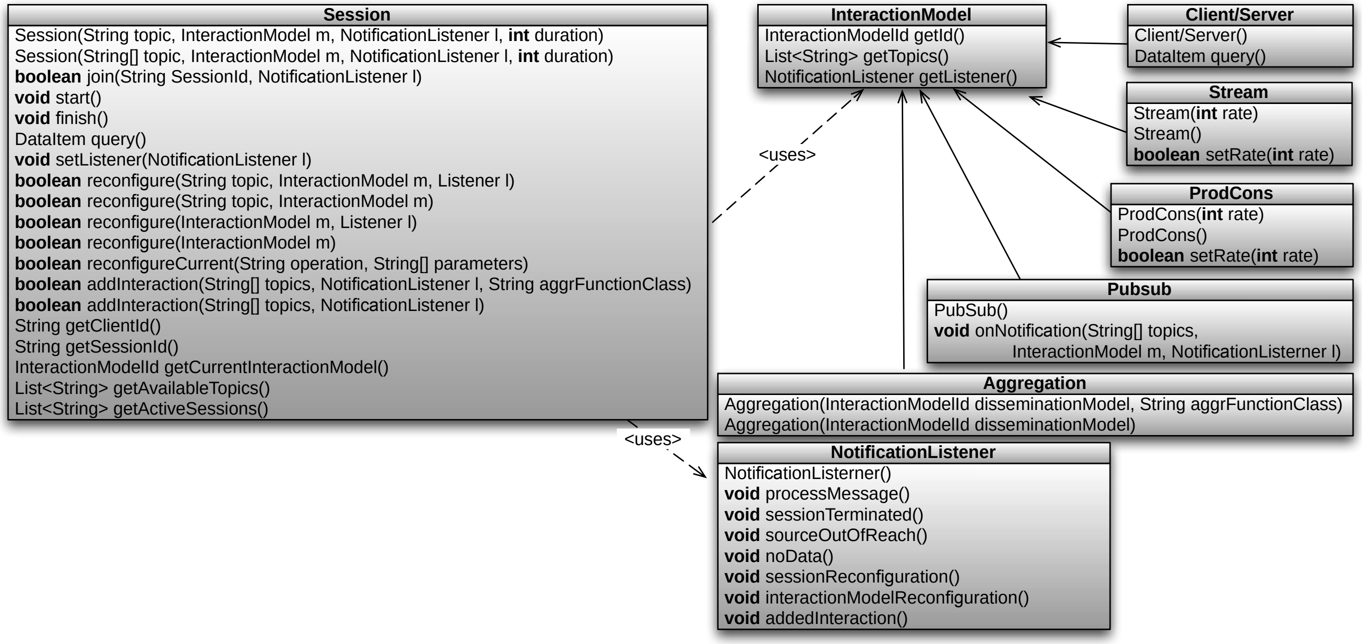


Fig. 7. API's simplified class diagram.

new data items causes the session to be reconfigured to Publish/Subscribe. Clients are notified of both the data stream's interruption and resuming.

- ***Lower the rate:*** the Producer/Consumer interaction model enables clients to consume data-streams at their own pace, which may be significantly slower or faster than their production rate. To support such feature, the middleware buffers data items on both ends of a client connection. In this context, the *Lower the rate* transition is triggered whenever the buffer that resides on the client end detects that it is no longer able to consume the data at the current pace. As the name implies, the reconfiguration lowers the rate to which the data items are sent to that particular client. Thereby, this reconfiguration targets a single connection, and not the whole session.

B. The Java Session-Centered API

A high-level session-centric API has been developed for the Java language. It exposes all of the middleware's features, providing the means for applications to create, destroy, join and reconfigure existing sessions. Moreover, it specifies how an application can process incoming data items and react to consummated reconfigurations. Fig. 7 showcases a simplified version of the API's class diagram.

Creating and Joining Sessions

Session are instances of the Session class that can be parameterized with the topic(s) of the data sources, an interaction model (the default is *Client/Server*), a listener to handle incoming data (more on this ahead), and a duration in minutes (the default is *unbound*). All interaction models share a common interface (InteractionModel) but provide specific reconfiguration interfaces (the methods of each class). The ability to join existing sessions is provided by the join() method. It requires the identifier of the session to be joined and the listener to handle incoming data. The inquiry of which sessions and topics are currently active is possible through methods getActiveSession() and getAvailableTopics(), respectively.

Reconfiguration Requests

Three methods are provided for requesting explicit reconfigurations: reconfigureCurrent(), reconfigure() and addInteraction(). The first empowers the tuning of the current interaction model, while the remainder two instantiate the *explicit* and *add interaction* transitions of Fig. 5, respectively.

Handling Incoming Data and Notifications

A special handler that we refer as *listener* must process all the data received in the scope of a session. This handler must subtype abstract class NotificationListener and implement methods to process the reception of new application data items (processMessage()) and of all possible exceptions and reconfiguration notifications (the remainder methods).

V. Application Scenario

The application scenario chosen to illustrate some capabilities of our proposal belongs to the domain of the *Data Driven Applications and Systems* (DDDAS) [21]. These applications are characterized by the need to dynamically incorporate sensing data into a running simulation. Inversely, the simulation should also be able to dynamically parameterize how such sensing data is collected (e.g. restricting data acquisition to the most affected areas in order to reduce data processing). Our example describes only a partial scenario in the context of a fire monitoring and simulation application, as introduced in section II. Namely, a session contextualizes the dynamic aggregation of sensing data collected in a critical area, and typical clients to this session are fire workers and a fire evolution simulation. These clients may hence share the

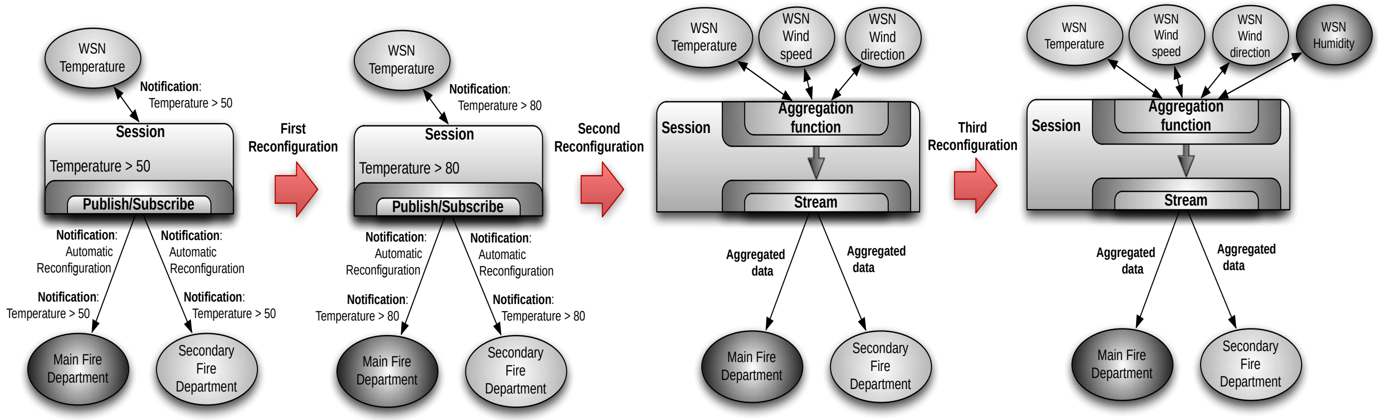


Fig. 8. Dynamic reconfigurations in a session context.

same context both in terms of collected data and the used interaction model for that data dissemination; additionally, all clients in the session are notified of the same dynamic reconfiguration events. The example in the next sub-section describes only the perspective of the fire workers.

A. Wildfires notification application

Consider a fire detection application supporting a fire department responsible for a critical geographical area prone to recurrent wild fire events. The department is interested on receiving a notification whenever the temperature in the area rises to values above 50º Celsius. Furthermore, when this happens, a dynamic reconfiguration should cause a switch from an *alert state* scenario to a *critical state* contemplating the raise of the temperature above 80º. Based on this last notification indicating a probable imminent fire ignition, the next step requires on-line (almost real time) data acquisition on wind-speed and direction, besides temperature. Such different data types should also be aggregated according to user's defined criteria.

In case a secondary fire department is appointed to fight a fire in the same area, the application should provide them with access to the same data as the main fire department. Furthermore, if during the fire fighting period the main fire department decides to add another source of data, e.g. "Humidity", in order to gain more precise information about the conditions in the terrain, this has to be acknowledged by the secondary fire corporation as well. Fig. 8 represents such modifications in the context of a session capturing this application scenario.

B. System dynamic evolution

The first image in Fig. 8 (on the left-hand side) depicts a session created by the middleware including:

1) The interaction context between the session clients, namely the *Main Fire Department* (the session's owner) and the *Secondary Fire Department* (the auxiliary corporation).
2) The available data sources accessible in the session, i.e. a Web enabled WSNs acquiring temperature data.
3) The interaction model in use is the *Publish/Subscribe* being the subscription topic: temperature values above 50, which defines an alert state.
4) The dynamic reconfiguration rules.

Namely, if such temperature value of 50 is observed, a user-defined dynamic reconfiguration takes place (*First reconfiguration* in 8) modifying the subscription topic. The fire departments are now interested in being notified when the temperature reaches 80º or above which indicates a critical situation. Note that the interaction model is left unaltered, and thus both departments are notified of this event.

On such scenario, another automatic dynamic reconfiguration (*Second reconfiguration*) is triggered to build an aggregation of multiple data sources. In the face of a critical situation, temperature data inspection is not enough, and new data sources on wind speed and direction are dynamically added to the session context. Data collected from different types of Web enabled WSNs may hence be aggregated in the context of the session and processed according to a user-defined aggregation function. Moreover, for a precise evaluation of the fire situation (e.g. if a fire ignition is imminent or has already occurred), a continuous data flow from the sensor devices monitoring the area is now mandatory. Such is also depicted in the new configuration, where the interaction model used for both the aggregation and dissemination stages is the *Streaming* model.

Finally, the case when additional data sources are still needed, e.g. on humidity values, is illustrated by the *Third reconfiguration*. The aggregation model remains as the underlying interaction model but a new Web service interfacing WSNs has been added, allowing the definition of a different aggregation function for processing all the types of incoming data.

```
// Listeners
Listener ownerListerner1 = new TemperatureListener50();
Listener ownerListerner2 = new TemperatureListener80();
Listener ownerListerner3 = new
TemperatureWindSpeedWindDirectionListener();
// Interaction models
Aggregation fire =
          new Aggregation(InteractionModel.STREAM,
                     "fire.HPAggregationFunction");
PubSub critical = new PubSub();
critical.onNotification({"Temperature", "WindSpeed",
                   "WindDir"}, fire, ownerListerner3);
PubSub alert = new PubSub();
alert.onNotification( "Temperature>80", critical,
                   ownerListerner2);
// Session creation
Session s = new Session("Temperature", alert,
                     ownerListerner1);
s.start();
…
// Later, during the fire fighting
Listener ownerListerner4 = new HumidityListener();
s.addInteraction("Humidity", ownerListerner4);
```

Listing 1: Main Fire Department's session

Note that this session's context captures the subordination, in the field, of the *Secondary Fire Corporation* to the *Main Fire Department* in terms of relevant collected data and the associated response. For instance, Listing 1 sketches the creation of a session with a *Publish/Subscribe* interaction model used to notify temperature values. When those values exceed a minimum threshold, the above critical situation is established and a pre-defined dynamic modification takes place ("Second reconfiguration" in Fig. 8). This reconfiguration is defined by the session owner (*Main Fire Department*) and consists on an *Aggregation* of streams on temperature, wind direction, and wind speed values, as depicted in Listing 1.

To share the same session context and subsequently be notified of the same events as the owner - including dynamic reconfigurations - the *Secondary Fire Corporation* has to know this session's identifier and join it as specified in Listing 2.

```
Session s = Session.join(sessionId, new ClientListener1());
```

Listing 2: A new fire corporation joins the session

In order to acknowledge and handle the events occurring in the context of the session, the above *Secondary Fire Corporation* (or other novel clients joining the session at some point in time) has to implement the `ClientListener1` handler as it is sketched in Listing 3. The disclosed methods handle the reception of data items, displaying them in a user interface (gui), and define a new listener able to process the reconfigurations possible in a new session's state.

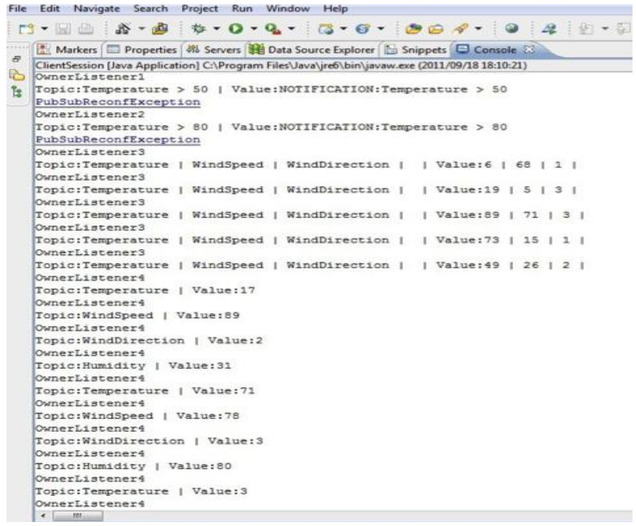


Fig. 9. Output of the main fire department - owner.

```
public class ClientListener1 extends AbstractListener {

 public void processMessage(DataItem msg) {
  gui.display(msg.getTopic() ,msg.getContents());
 }
 public void InteractionModelReconfiguration(
           ReconfException n) {
  gui.displayNotification(n.getTopic(), n.getReason());
  getSession().setListener(new ClientListerner2());
 }
}
```

Listing 3: Sketching the implementation of ClientListener1

C. Example output

In the context of the previous particular scenario Figs. 9 and 10 illustrate the reception, on both fire departments, of the data values and notifications disseminated in the context of the session. The display of data values complies with the following format:

Topic*: subscribed_topic* | Value*: received_value*

while the display of notifications adheres to format:

Topic*: subscribed_topic* | Value*:*NOTIFICATION*:reason*

Reconfiguration notifications are disseminated to all members of a session, including its owner, which pre-defined the reconfiguration request. This approach entails a uniform way to react to a given notification, regardless of the member's role in the session.

Fig. 10 also shows the situation when the Main Fire Department requests a novel stream on humidity values (third reconfiguration on Figure 7) but an aggregation function is not supplied in the invocation of `addInteraction()` (last line of Listing 1). As a consequence, the messages are no longer aggregated by the middleware, who simply forwards them. Such behavior can be observer on both figures from the point the fourth listener takes action.

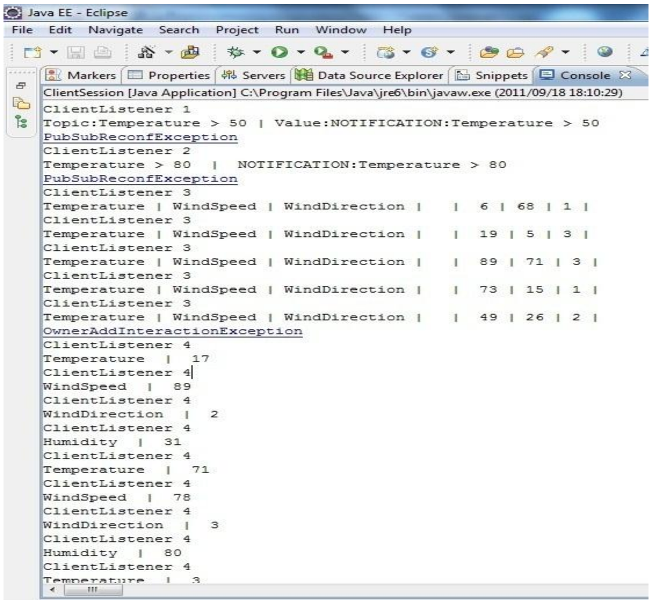


Fig. 10. Output of the second fire department – participative user.

VI. Related Work

To the best of our knowledge, existing middleware platforms for Web enabled WSNs do not address client-WSN interaction model's dynamic reconfiguration concerns nor provide a session abstraction to capture and reuse such dynamicity. Among those platforms we highlight:

52º North [12], the most known implementation of the Sensor Web Enablement (SWE) [3], a set of models and Web service interfaces proposed by the Open Geospatial Consortium for the Web integration of sensor systems. The models focus on the description of sensor systems and their capabilities to collect and process observations, while the services address the collection, storing, and dissemination of sensor reading and alerts (notifications).

Global Sensor Network (GSN) [20], which aims at building a sensor Internet by connecting virtual sensors, abstracting data-streams originating from either a WSN or from another virtual sensor. SQL queries can be performed on top of these virtual sensors.

SenSer [4], a generic middleware for the remote access and management of WSNs, being the latter virtualized as Web services, in a way that is programming language and WSN development platform independent. Its distinguishable properties include the ability to filter the acquired data and to submit WSN reprogramming requests.

As for the presence of the pattern concept on system's dynamic adaptations, the work in [17] presents one solution for self-adaptability of service-generated data streams targeting problems such as data loss or delays associated with communication networks disruptions. However, interaction models are not present as explicit configuration options considering service interactions as described in our proposal. Although the cited approach does implement a (sophisticated) *Producer/Consumer* interaction model, such is restricted to the support system (i.e. it is not explicitly visible at the point-to-point interaction level between a service and its user). Furthermore, in [17] there is no reference to the possibility of dynamically adding new data flow consumers or additional data sources, as we proposed in the session's context.

Some other works use reconfigurable *Architectural Patterns* for adaptable system's definition [18]. The architecture of the *Publish/Subscriber* pattern, for instance, allows the reconfiguration of publishers, subscribers, and subscribed events. The *Master/Slave* pattern also allows the addition of new slaves to optimize task execution [19]; such is also incorporated in our solution within the context of a session. In spite of such reconfigurable system architecture definition, these works do not provide a session capturing an interaction's context, nor a pattern-based system evolution based on pre-defined rules conform to those pattern's semantics.

Finally, our solution is based on the work by [16] which, however, does not provide a session abstraction to contextualize and reuse dynamic interaction models, nor implements a state machine for pattern-based system evolution.

VII. Conclusions and Future Work

Current applications relying on WSNs for large-scale environment monitoring require adequate abstractions for network access and parameterization, and sensing data acquisition. However, applications also increasingly request the seamless integration of WSNs in heterogeneous and dynamic complex systems, what is possible via the service concept. Moreover, the access to sensing data requires richer interaction models besides the traditional synchronous request/reply model, for example the *Publish/Subscribe* and *Streaming* models. Based on such Web enabled WSNs this work proposes a *session* abstraction in order to capture, contextualize, and reuse diverse richer dynamic interaction models to those services.

A session embodies the common interaction characteristics relating a set of users accessing the same service at some point in time, and all perceive the same events occurring meanwhile in the session's context. A session also contextualizes the possible dynamic adaptations both in terms of the service, the communication medium, or the clients' contexts. For instance the sensing data, the data transfer rate, or a client's mobile device autonomy, may all trigger the modification of the interaction model. Furthermore, both the interaction models and the rules for their dynamic adaptation rely on the *pattern* concept and depend on individual pattern semantics. The system's evolution is captured in a state machine based on pre-defined pattern-based rules. Being well defined, such per-pattern reconfigurations allow adaptation automation and contribute to limiting, to some extent, the impact of the dynamic reconfiguration upon the overall system.

The performance evaluation in terms of the overhead of one additional middleware layer between a Web enabled WSN and its users (SenSer platform [4]) is one point that unfortunately is missing in this paper but which will be studied in the near future. Likewise, more application scenarios are needed in order to evaluate the expressiveness of the model.

Nevertheless, it is our opinion that such novel session-based abstraction opens several interesting further developments concerning the inclusion and aggregation of diverse WSNs sensing data in different domains. For instance the aggregation of session-based interactions may be captured in the form of workflow dependencies and be used in ambient intelligence contexts and participatory sensing applications. Furthermore, the proposed middleware's deployment in a Cloud computing platform may provide clients a ubiquitous and reliable access to sessions. These cases are already under development.

ACKNOWLEDGMENT

The authors would like to thank Professors Omer Rana and José Cardoso e Cunha for the initial ideas on pattern and service abstractions, which ultimately conduced to this work.

REFERENCES

[1] C. F. García-Hernández, P. H. Ibargüengoytia-González, J. García-Hernández, and J. A. Pérez-Díaz, "Wireless sensor networks and applications: a survey," International Journal of Computer Science and Network Security, vol. 17, no. 3, pp. 264 –273, 2007.

[2] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, R. A. Peterson, H. Lu, X. Zheng, M. Musolesi, K. Fodor, and G.-S. Ahn, "The rise of people-centric sensing," IEEE Internet Computing, vol. 12, pp. 12–21, July 2008.

[3] M. E. Botts, G. Percivall, C. Reed, and J. Davidson, "OGC Sensor Web Enablement: Overview and high level architecture," in GeoSensor Networks, Second International Conference, GSN 2006, Revised Selected Papers, Lecture Notes in Computer Science, S. Nittel, A. Labrinidis, and A. Stefanidis, Eds, vol. 4540, Springer, 2008, pp. 175–190.

[4] H. Paulino and J. R. Santos, "A middleware framework for the Web integration of sensor networks," in Sensor Systems and Software - 2nd International ICST Conference, S-CUBE 2010, Revised Selected Papers, Lecture Notes of the ICST, Springer-Verlag, 2011, pp. 75–90.

[5] T. Kindberg, J. J. Barton, J. Morgan, G. Becker, D. Caswell, P. Debaty, G. Gopal, M. Frid, V. Krishnan, H. Morris, J. Schettino, B. Serra, and M. Spasojevic, "People, places, things: Web presence for the real world," MONET, vol. 7, no. 5, pp. 365–376, 2002.

[6] ITU, "ITU Internet report 2005: The Internet of Things," International Telecommunication Union, Tech. Rep., 2005 [Online]. Available: http://www.itu.int/osg/spu/publications/internetofthings/

[7] L. Badger, T. Grance, R. Patt-Corner, and J. Voas, "Cloud computing synopsis and recommendations (draft), NIST special publication 800-146," Recommendations of the National Institute of Standards and Technology, Tech. Rep., 2011 [Online]. Available: http://www.nist.gov/itl/csd/20110512_cloud_guide.cfm

[8] A. Baptista, M. C. Gomes, and H. Paulino, "Session-based dynamic interaction models for stateful Web services," in Exploring Services Science - Third International Conference, IESS 2012, Lecture Notes in Business Information Processing, Springer-Verlag, 2012.

[9] M. Creeger, "Cloud computing: An overview," ACM Queue, vol. 7, no. 5, p. 2, 2009.

[10] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Commun. ACM, vol. 53, pp. 50–58, April 2010.

[11] T. Kobialka, R. Buyya, C. Leckie, and R. Kotagiri, "A Sensor Web middleware with stateful services for heterogeneous sensor networks," in Intelligent Sensors, Sensor Networks and Information, 2007, ISSNIP 2007, 3rd International Conference, 2007, pp. 491–496.

[12] C. Stasch, A. C. Walkowski, and S. Jirka, "A geosensor network architecture for disaster management based on open standards." in Digital Earth Summit on Geoinformatics 2008: Tools for Climate Change Research, 2008, pp. 54–59.

[13] I. F. at al., "Modeling stateful resources with web services v. 1.1," Computer Associates International, Inc., Fujitsu Limited, Hewlett-Packard Development Company, International Business Machines Corporation and The University of Chicago, Tech. Rep., 2004 [Online]. Available: http://www-106.ibm.com/developerworks/library/wsresource/ws-modelingresources.pdf

[14] OASIS, "Oasis web services resource framework (WSRF) TC," [Online]. Available: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsrf

[15] F. Darema, "Dynamic data driven applications systems: New capabilities for application simulations and measurements," in International Conference on Computational Science (2), 2005, pp. 610–615.

[16] C. Gomes, O. F. Rana, and J. Cunha, "Extending grid-based workflow tools with patterns/operators," Int. J. High Perf. Comput. Appl., vol. 22, pp. 301–318, August 2008.

[17] V. Bhat, M. Parashar, M. Kh, N. K, and S. Klasky, "A self-managing wide-area data streaming service using model-based online control," in in Proc. 7th IEEE Int. Conf. on Grid Computing, 2006, pp. 176–183.

[18] M. C. Huebscher and J. A. McCann, "A survey of autonomic computing degrees, models, and applications," ACM Comput. Surv., vol. 40, pp. 2–25, August 2008.

[19] M. Aldinucci, M. Danelutto, and P. Kilpatrick, "Towards hierarchical management of autonomic components: A case study," in Proceedings of the 17th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, PDP2009, IEEE Computer Society, 2009, pp. 3–10.

[20] K. Aberer, M. Hauswirth, and A. Salehi, "A middleware for fast and flexible sensor network deployment," in Proceedings of the 32nd International Conference on Very Large Data Bases, Seoul, Korea, September 12-15, 2006, 2006, pp. 1199– 1202.

[21] F. Darema, "Dynamic data driven applications systems: A new paradigm for application simulations and measurements," In Int. Conf. on Computational Science, volume 3038, Springer LNCS, 2004.

Dr. Maria Cecília Gomes, PhD, is an Assistant Professor at the Departamento de Informática (Computer Science Department) of the Faculdade de Ciências e Tecnologia/Universidade Nova de Lisboa. She is a member of the Research Center for Informatics and Information Technologies (CITI), and member of the management committee and national PI of *ARTS: Towards Autonomic Road Transport Support Systems* (COST Action TU1102). Her current research interests include autonomic systems particularly applied to road transportation support systems, service-oriented computing, and models.

Dr. Hervé Paulino, PhD, is an Assistant Professor at the Departamento de Informática (Computer Science Department) of the Faculdade de Ciências e Tecnologia/Universidade Nova de Lisboa. He is a member of the Research Center for Informatics and Information Technologies (CITI), on which he is the principal investigator of the SABLE (Service Abstractions for Parallel Computing) research team, and also a collaborator member of the Research Center for Research in Advanced Computing Systems (CRACS). His current interests are on the fields of service-oriented computing and, concurrent and parallel programming.

Adérito Baptista, MsC, was a computer science master student at Faculdade de Ciências e Tecnologia/Universidade Nova de Lisboa. He is currently a software developer at Novabase, Portugal.

Filipe Araújo, MsC, was a computer science master student at Faculdade de Ciências e Tecnologia/Universidade Nova de Lisboa. He is a record management systems and R&D developer at Quidgest, Portugal.

Dynamic, ecological, accessible and 3D Virtual Worlds-based Libraries using OpenSim and Sloodle along with mobile location and NFC for checking in

Rubén González Crespo, Sergio Ríos Aguilar,
Pontifical University of Salamanca in Madrid, Spain
Roberto Ferro Escobar, Nicolás Torres
Distrital University Francisco José de Caldas, Bogotá, Colombia

***Abstract* — This paper proposes the implementation of a 3D virtual library, using open platforms such as OpenSimulator and Sloodle, applied to the integration of virtual learning environments. It also proposes their application to the creation of open libraries to share and disseminate the new dynamic nature of knowledge, in the understanding that 3D virtual worlds may contribute to the future of libraries as part of green initiatives to achieve an ecologic and sustainable planet.**

***Keywords* — 3D Virtual Library, OpenSim, Sloodle, NFC, GPS check-in, Grid.**

I. Introduction

TODAY, one of the most relevant topics in computer-based content management research efforts is the way of presenting and visualizing content, stepping ahead from simple static web delivery of documents, books and publications via HTTP on the Internet. Nowadays, users demand better ways of accessing new sources of information, and an evidence of this is the current evolution of Digital Libraries, making huge efforts in order to make large collections of documents available to general public.

Anyway, current prevailing paradigms show that changes are needed in the way these contents are presented to the information consumers, who are looking to interact with Web contents in a more friendly manner using richer interfaces. And provided that recent evolution in office and domestic broadband access to Internet has improved the performance of networks interconnections and hence the globally perceived speed of Internet traffic, this change begins to come true thanks to virtual worlds simulators like Open Simulator [1].

This simulator was created under the guidelines of Secondlife, released in 2007 by Linden Labs [2], and placed under the GPL license [3]. OpenSim meets the general client-server model and has a modular structure written in C# and supports multiple viewers, such as the HIPPO viewer [4] , RealXtend viewer [5] and the Secondlife official viewer itself

The dynamic nature of Web 3.0 [6] combines artificial intelligence techniques, geospatial mobility, mashups, interactivity, multimedia and 3D transformations to build a semantic database able to learn. And everything is accomplished in 3D virtual worlds, where the user can be embodied in avatars that allow travelling to and wandering many worlds constantly arising, built using freely available design tools. Secondlife was initially used as a game in a virtual world built on a commercial server, and after a few years, due to the many worlds that can be created, Linden Labs decided to release the code to allow for an open construction of virtual worlds and private islands by programmers.

One of the most promising research areas related to 3d virtual worlds technology is the study of the integration of these 3d virtual worlds with current virtual learning environments and content management systems, given the powerful emergent properties arising from such a combination.

Another prevalent trend is the use of mobile broadband in order to access any kind of information using a personal mobile device, and one promising technology now close to reaching the necessary critical-mass is Near Field Communications (NFC). This technology provides a proximity-based interface to a wide array of services, including mobile check-ins, in special check-ins to virtual worlds using smart tags deployed at strategic places in the real world.

Also, when NFC is not available, and we need to promote (or restrict) the access to the virtual world Library from within certain known premises, it is always possible to get the real-world location of the user proving his/her presence via location based services (ie GPS, A-GPS…etc), used as an alternative check-in method. Current Android and iOS mobile devices (>90% market share) provide us with location services with enough accuracy (meeting E-911 and E-112 strict requirements with a wide margin)

This paper proposes the fundamental basis for the development of a 3d virtual world based Library using OpenSim, and facing the integration of a virtual learning environment (VLE, LMS) such as Moodle, with the help of the Sloodle platform.

 DOI: 10.9781/ijimai.2012.177

Fig. 1. View of an open virtual world simulator with geospatial database connectivity and information sharing in real time.

II. Tools for CREATING METAVERSES

A. OpenSim 3D modeller

OpenSim is a 3D application server that uses the same standards found in SecondLife to communicate with their users. This modeller emerges from the analysis of the structure of SecondLife viewer's source code, released early in 2007 by Linden Labs under GPL. It was reverse engineered in order to create a 3D application server, and is characterized by being free software, having a modular structure written in C # and its support for multiple viewers; to date, OpenSim is in version 0.7.4

Among the most promising features of OpenSim is the ease with which universities can customize their LMS (Learning Management Systems) [8] integrating the legacy systems with OpenSim databases and creating content that suit the needs and methodologies of each institution.

OpenSim is being currently used to design, create and easily implement 3D virtual worlds by a huge base of programmers around the world. With this tool, it is possible to create content targeted to different areas, such as e-learning, virtual tours, simulations, among others. Moreover, being a free software tool leads itself to low-cost development and allows developers to contribute to the development of new features.

B. Moodle for Web courses

Moodle (Modular Object Oriented Dynamic Learning Environment) is an open source software package for creating courses and web sites based in PHP and designed to work with different database engines, especially MySQL [9]. Moodle aims to help educators to create and easily manage online learning communities, and is designed to foster the construction of knowledge by students.

Its first version was published in August 2002 and to date the current version is 2.3. Moodle currently has over 6.856. 000 courses offered through 72.164 registered sites in 223 countries, with Spain holding the second place in World Ranking of registered courses, and Colombia in the eighth position.

C. Sloodle project

The SLOODLE project (Simulation Linked Object Oriented Dynamic Learning Environment) [10] is a successful code that integrates multi-user virtual environments like Second Life and OpenSim with learning management platform MOODLE. Sloodle provides a wide range of tools (some still in development) that enable virtual learning worlds and immersive education, and are fully integrated with a web-based learning management system, used by thousands of educators and students worldwide.

D. Ubuntu operating system

Ubuntu is an operating system that uses a Linux kernel and based on Debian [10], is maintained by the British firm Canonical and the developer community. This distribution is intended to improve the end user experience, making it easier to interact with the operating system through an improved GUI. It is one of the most popular distributions today, with approximately a 50% share among Linux distributions.

For the development of this project, it should be noted that Linux is widely known as an operating system designed from the beginning as multithreaded as multiuser, so it does manage very safely the permissions for each user on processes and files, preventing users from accessing each other's information. Ubuntu has default user "root" disabled, so that the system repeatedly asks the user's password to install components and forces him to be especially careful with permissions to modify files and folders.

III. Installation of Metaverse in OpenSim

Below is a brief summary of the steps needed in the Ubuntu distribution in order to assemble a metaverse in OpenSim, and linking it to the Moodle platform.

A. Installing XAMPP Platform

Given that we need a web server to distribute the contents held by Moodle and that it is advisable to configure MySQL as the default database engine for a OpenSim server, we will use the XAMPP [11] Apache distribution, that includes the installation of MySQL. The XAMPP installation files for linux are available on the official website "Apache Friends". Once downloaded the file, you must deploy it uncompressed in the "opt" directory. Finally you must start the server using the command "/opt/lampp/lampp start".

One of the most powerful utilities found in XAMPP is the php-based administration tool for MySQL databases "phpMyAdmin", which can be accessed typing in a browser "http://localhost".

By default, the admin page does not have access control enabled. So, anyone with access to the server is able to change these valuable data. We recommend setting passwords. The XAMPP documentation describes how to do this.

B. Installing the Metaverse world with OpenSim

On the OpenSim download page we can find the necessary files for the installation of the server. It is advisable to mount the server on a 32-bit Linux version, since 64-bit versions often have compatibility issues with OpenSim and its dependencies.

OpenSim is designed to run on Microsoft's .NET platform. As the code and scripts used in the metaverse are based on C#, we should install the Mono platform on Linux, which has the libraries necessary to compile such languages. The required files for installing Mono on Linux are available at their official download page or they can be downloaded directly using the terminal installation commands (eg Ubuntu: sudo apt-get install mono-complete). After having installed mono opensim unzipped the file, we must open a Terminal. To start the server, just type "mono OpenSim.exe" command, with root permissions.

If there are no compatibility issues, the server should start gracefully, loading one by one all the necessary dependencies to the metaverse. If this is the first time you run OpenSim on the computer, it does not have any metaverse to begin with, so the simulator requests the creation of one, asking for the following information:

- New region name[ ]: alphanumeric character. Enter the desired name for the region.
- RegionUUID [default_UUID]: UUID. Unique identifier for this region. OpenSim proposes a default UID, so you should just press enter unless you prefer a specific UUID.
- Region Location [1000.1000]: x, y coordinates. It is the point in a coordinate plane you want for the region. Mode must be specified for non-grid overlay on existing regions region. Otherwise you can accept the default one point by pressing the enter key.
- InternalAddress [0.0.0.0]: Wildcard Mask (IP address). This address is actually a wildcard mask that allows you to restrict the range of addresses that can access the server. In most cases, should be left with default value, so that all hosts are allowed to access our server.
- InternalPort [9000]: # port (0-65535). Internal server port to be used for server access.
- AllowAlternatePorts: True or False. This feature is still in experimental stage. Default is false.
- ExternalHostname: ip address. This is the address used by clients to access the server, which should take into account some other considerations that are discussed later. To start, you can leave "SYSTEMIP".
- Do you wish to join an existing region to Udin estate (yes / no)? [Yes]: This is used to attach the newly created region to an existing state. If there is a previously created state which wants to join the region you should write its name, or on the contrary, type " no "and proceed to create one.

The parameters that are required to create one are:

- Estate name [MyEstate]: alphanumeric character. Any name.

The following data creates a new user on the server and will be the first user to access it:

- Estate owner first name [Test]: alphanumeric character. The name to use for the estate owner.
- Estate owner last name [User]: alphanumeric character. The name to be used for the estate owner.
- Password: alphanumeric character. User Password created.
- Email [ ]: alphanumeric character. Email of the newly created user.

Now we have a region, a state and a user. So, we can run the server and it can be accessed from any viewer. These parameters can be modified at any time by accessing the Regions.ini located in the bin/Regions directory within OpenSim. Text file appears as shown in Figure 2.

Fig. 2. View of Regions with location, internal address, port and the name UDIN server, Regions ID it is compose for a hexadecimal number unique in the world.

C. OpenSim settings for different network topologies.

OpenSim is a simulator that runs on a server, ie it is not a local program as a video game in which the user starts a process in his computer and it ends when he finishes using it. The situation resembles that of an interactive website in which the user makes requests to the server and starts or finishes sessions, accesses databases or performs specific actions in which processing occurs on the server (which may be in another location or on the same computer from which you connect the client). If you have a server, you must have a client with the necessary application to access its services. In case you have OpenSim Hippo viewers, like OSV, RealXtend or just the Second Life Viewer, they allow users to connect to servers running the simulator. Because the client and server can be located on the same network or on separate networks, we'll go on providing a brief explanation of the basic configurations of the above parameters, so that clients are able to access to servers.

Depending on the topology, we can have multiple address mappings (static or dynamic –depending on if the server has DHCP enabled or not-)

Changes on the server are made in the file Regions.ini, located in the bin/Regions folder within OpenSim. The only parameters that must be modified to access are "ExternalHostname" and "Internal Address".

It is possible to change the client directly off the display. The only parameter important to change is Login URI (the socket through which you access the server), it also allows viewers to save configurations and to assign names and other descriptive parameters that are not of interest here. Here are some settings for the server and the client, and network topologies.

- *Topology 1: Client and server are hosted on the same computer.*

The client and server configuration shown in Table I:

TABLE I
SETTING PARAMETERS FOR TOPOLOGY 1

Client	InternalAddress	0.0.0.0
	ExternalHostName	SYSTEMIP
Server	Login Uri	http://localhost:port

The network topology is shown in Figure 3.

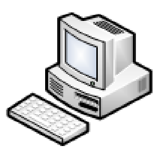


Fig. 3. Topology 1, this computer hosts server and client software, the virtual worlds development builder is inside.

- *Topology 2: Client and server on the same network.*

In this case we have the client and server using the same network serving different IP addresses (usually within a network with private addresses). The configuration is shown in Table II:

TABLE II
SETTING PARAMETERS FOR TOPOLOGY 2

Cliente	InternalAddress	0.0.0.0
	ExternalHostName	SYSTEMIP
Servidor	Login Uri	http:// public_address:server

The network topology is shown in figure 3.

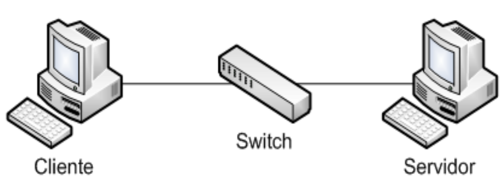


Fig. 4.. Topology 2, the client and server using the same network and serving different IP addresses

- *Topology 3: Client and server connected via the Internet, where the server has a public address.*

This configuration occurs when the server is connected to the internet through a broadband modem provided by an ISP and obtains a public address. The server configuration is the same as in the previous case. On the client, we can use this server's public address as part of the login URI. Figure 5 shows this network topology.

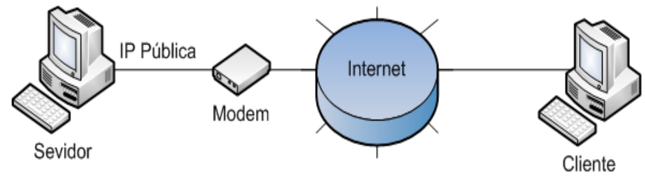


Fig. 5. Topology 3 Client and server connected via the Internet, where the server has a public address over the Web.

- *Topology 4: Client and server connected via the Internet, where the server has a private address.*

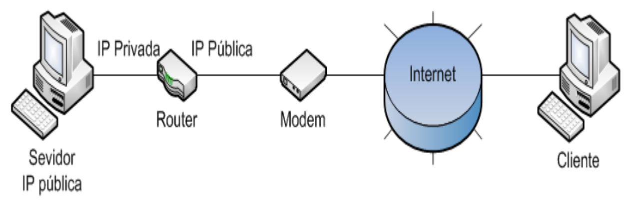


Fig. 6. Topology 4 Client and server connected via the Internet.

This configuration occurs when the server is connected to internet through a router connected to the ISP modem and therefore has a private address, which is translated by the router (NAT) to get out to the Internet. In this case, in addition to the above settings, we should enable port forwarding in the router for transport protocols TCP and UDP.

TABLE III
SETTING PARAMETERS FOR TOPOLOGY 4

Cliente	InternalAddress	private _address
	ExternalHostName	public_ address
Servidor	Login Uri	http:// public_address:port

To allow the client to connect to the server in the latter topology is also necessary to make port forwarding in the router. In current routers this function is usually called "virtual server" and is used to specify which ports redirect the public address to a private address.

D. Construction of objects in OpenSim

Creating and modifying objects in OpenSim is the basis for the construction of a metaverse. The first thing you should do is increasing the size of the island, this is accomplished using the "terrain fill" command long with the number of the desired size. The region may have a maximum size in the grid, because if set to a larger size, it would overlap with other regions in grid mode.

Different tools within OpenSim are:

- Vision (ctrl + 1) is used to change the angle of view, the approach and orbit of the camera.
- Move (Ctrl + 2): With this option you can move the different objects in the region.
- Edit (Ctrl + 3): this is one of the most important tools; this tool can easily modify position, rotation and size of objects.
- Create (ctrl + 4): This tool creates different primitive objects like cubes, cylinders, cones, spheres, prisms, toroids and others.
- Ground (ctrl + 5): With this tool you can raise, level, smooth down or reset the selected field. Using this tool you can also change the properties to specific "plots", including the type of texture that will be used to play media.

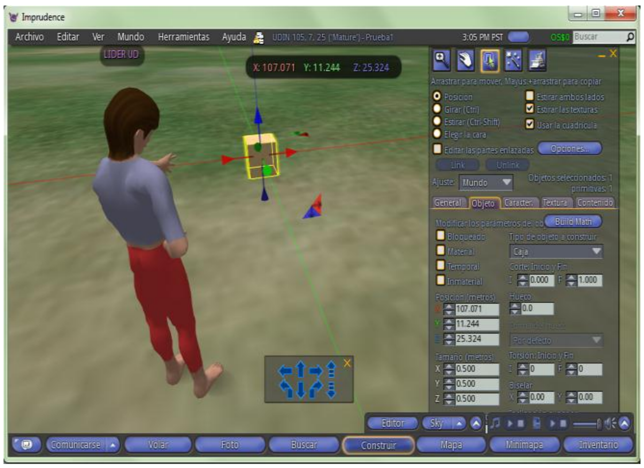


Fig. 7. OpenSim viewer, basic tools menus and tabs for designing objects in the virtual worlds.

When you open any of the aforementioned tools, you can make specific modifications to the objects just created, using the tabs at the bottom of the popup window.

The tabs available in the tools menu are:

- General: here you can make a description and edit the permissions and ownership of objects, and define the action performed by default when an avatar touches it.
- Purpose: In this tab you can set "physical" properties of the object in the metaverse, such as position, rotation, size, or even if the object is material, ghost, etc. The detailed description of the management of client tools for building an OpenSim metaverse are not the subject of this document.
- Features: edit the characteristics associated with the physics engine simulator. OpenSim accepts different physics engines that allow more realistic simulations of objects, but for the purpose of this project are not relevant.
- Texture: each object may also have different textures associated, which create more realistic environments and are also used for multimedia playback within the metaverse.

The complexity of building a metaverse starting from scratch using basic objects and the time it takes will depend on the user's ability using these tools. The detailed description of the metaverse construction is not the subject of this document.

IV. Linking the Metaverse Created with Moodle

If you have a Moodle platform installed on a server accessible to the client, this can be linked to the OpenSim metaverse created using Sloodle tools, so that users can make use some of Moodle functions directly from within the metaverse. For this project we installed Moodle on the same machine hosting an OpenSim server, in order to perform some tests.

A. Installing Sloodle

The files required for Sloodle operation must be deployed on the server and on the user/client machine. The files of the server are available for download at the official project wiki. The decompressed folders must be copied into the following subfolders. It is important to remove any previous version of Sloodle directories installed in Moodle, before copying the new ones.

- Module (required): You must copy the "Sloodle" folder into the "mod" folder within Moodle.
- Block (Optional): You must copy the "sloodle_menu" folder into the "blocks" folder of Moodle.
- Plugin Sloodle object allocation (Optional): You must copy the "sloodleobjet" folder into the "mod/assignment/type" folder within Moodle.

The user files for OpenSim virtual objects are required to make use of the functionality of Moodle within the metaverse. This is called the Sloodle Set. Once downloaded, it must be loaded into the metaverse via the "load oar" command using the command line in the OpenSim server. Keep in mind that this loads a new region, so before you do so, it is advisable to save the region on which you are working.

Once SLOODLE oar has been loaded, the objects appear. Sloodle objects therefore must be stored in the user's inventory. With the above, the modules are installed and Sloodle is ready for use with OpenSim.

B. Moodle in OpenSim

Before you can access Moodle from OpenSim using a user with admin role, you must enter edit mode through the "Enable Editing" button located in the upper right of the course and proceed to click on the "Add activity" drop down menu and then the "SLOODLE Controller". This leads to a page designed to setup the data controller, of which we only need to supply the name, so after filling this field, so you can click Save and return to course.

This controller handles all communications between Moodle and OpenSim and is therefore necessary to use any other Sloodle tools.

Some of the Sloodle tools, such as "Presenter", must have a Moodle driver object that is added the same way as the main controller. These tools added from Moodle page are required for objects that are added to Sloodle metaverse and they have some functionality, otherwise you cannot access the contents of the Moodle database. Other tools such as the "Sloodle Choice" or "Quiz Chair" can interact directly with native Moodle tools as "Choice" or "Quiz" respectively.

In this case the example will be a "Presenter" because you need to make video presentations in the metaverse. The detailed explanation of each one of the Sloodle tools is not the purpose of this paper and is left to the reader. As previously stated, the driver is first added to Moodle using "Add Activity" and then "Presenter". At this point, you must add the metaverse object. All Sloodle client tools can be added through the metaverse's "Sloodle Set", added above the metaverse and kept in inventory.

Before performing any task in Moodle using OpenSim, you must link the SLOODLE set located on the client host with Sloodle Controller located on the server. The steps for doing this are:

- Step 1: Click the SLOODLE Set
- Step 2: Write in Moodle site address, for example http://localhost/moodle or http://200.56.46.103:54661/moodle.
- Step 3: A dialog box appears in the viewfinder. Continue clicking on "Go to page".
- Step 4: To authorize objects Sloodle opens a browser which in turn shows the viewer . It should come with admin role, or shows the course in which you added the controller . Select "authorize previously added to Moodle", and post the form, after which you can close the browser.
- This should be done only once for the main controller Sloodle set, and once for each object you add to SLOODLE set in the metaverse.
- Step 5: Finally you must click again on the Sloodle Set, and then click Download configuration, so that the object has authorized access to the specified Moodle platform.

At this point you have a connection between OpenSim and Moodle but still have no functionality. Functions are added by SLOODLE set objects. The steps to add a Sloodle tool are:

- Step 1: Click the Sloodle set.
- Step 2: Click "Rez an object"
- Step 3: Click the object number to be added to the metaverse, in this case "7" representing the Presenter.
- Step 4: Click the object and then "Go to page" to authorize the new object in Moodle, as explained above for the controller and then close the page.
- Step 5: Click on the object and then select "Download configuration"

If everything is ok, the object is linked to the Moodle platform. These changes are required in Moodle as Sloodle objects do not store information, all the information is in the Moodle database.

- Module (required): You must copy the "Sloodle" folder into the "mod" folder within Moodle.
- Block (Optional): You must copy the "sloodle_menu" folder in the "blocks" folder within Moodle.
- Plugin Sloodle objects allocation (Optional).

V. Library Design for virtual worlds

The use of digital libraries has changed over the years and even more with the advent of the Internet, from that moment the way knowledge is accessed really changed because in cyberspace there are countless sources of information: text, video and more. The change is so important that new generations of students prefer to search the web before going to a library; therefore arises the need for more innovative spaces that encourage children, youth and adults to enter virtual worlds and visit this new interaction spaces and new models of accessing knowledge. The ACL [12] make rules for Distance Learning Library Services, since 2000.

The fundamental basis for the creation of a virtual library are similar to a real repository but the interface is implemented by Sloodle, and therefore it must have intelligent access protocols. This is achieved using Web-based architectures, OWL, XML [13] [14], Ontologies, Semantic Services for RDF presentation [15] , data management, Grid repositories [16], databases compatible with XAMPP, connection to external sources of scientific data, social, geographic and more. All of them can be achieved by using Service Oriented Architectures (SOA) [17].

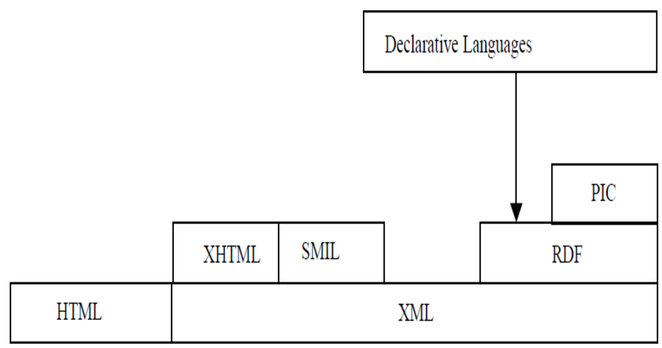


Fig 8. Language Layers on the Web, the Semantic Web, with machine processable information defined for their semantic content. Such semantic interoperability thanks of the W3C, XML/XML Schema and RDF/RDF Schema [13].

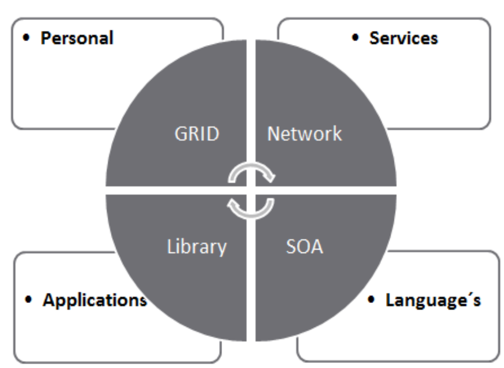


Fig. 8. Example using SOA for building virtual libraries over virtual worlds

One interesting and promising way of accessing the virtual Library at known places where we would like to promote it, is through publicly accessible NFC smart tags. In this scenario, when a user touches (or approaches) these tags with an NFC-enabled mobile phone, an event fires up an app which grabs the necessary URI that allows the user to access the virtual world.

This use of NFC technology can also act as a check-in method, or proof of attendance at known places where the Virtual Library is being promoted.

As seen, it is necessary to involve quite a lot of human resources, as researchers, Engineers, software developers, university students, research groups in the field of network architecture and broadband access to high speed Internet, Grid computing [18] and SOA services and a community of architects builders and 3D metaverse designers [22].

VI. RESULTS

The main result is a metaverse based on the engineering faculty of the "Francisco José de Caldas" Distrital University [19], hosted on a server accessible from the internet, and tested for Video Streaming [20] using OpenSim textures and mobile networks inside the University campus .

In Figure 10 you can see the photo of the real Engineering faculty building, and at the right, OpenSim built the building as proposed for the operation of a virtual 3D library. When the user logs in and reincarnates as an Avatar, is able to visit the site and see all kinds of information, all of the above accomplished by using Sloodle. In the future, we expect to further develop this initiative in collaboration with security staff trained to feed the virtual world.

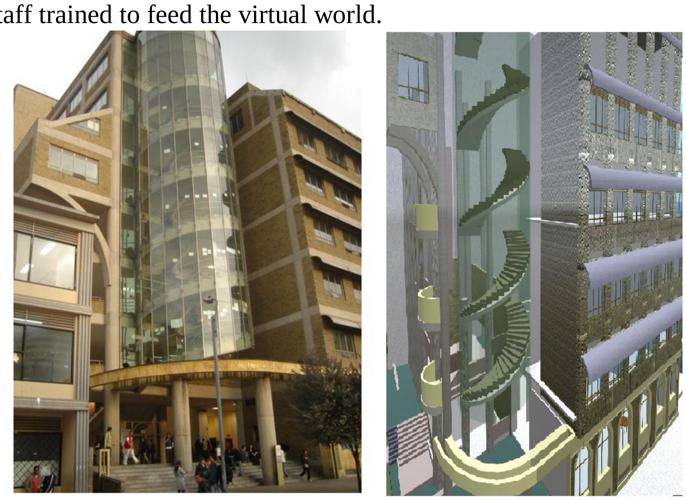


Fig. 9. Metaverse example created in the Distrital University, with access to virtual library, Left real Building, Right virtual 3D Building in Opensim.

Also, there are objects -within the metaverse- linked to the Moodle platform installed on the same server.

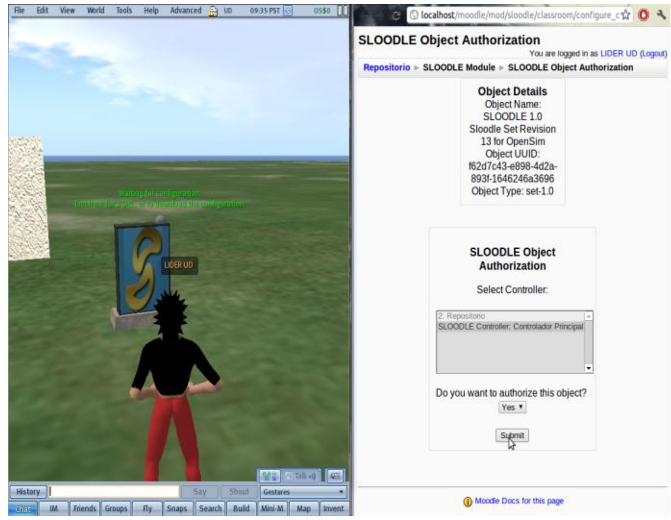


Fig 10. View window Sloodle controller with access from the metaverse.

VII. CONCLUSION

These kind of metaverses may target the educational virtual environments and the dissemination of scientific material, even in simulation labs, thanks to the programming tools here presented, which allow the control of the characteristics of the objects in these metaverses. Different research groups of the Distrital University are working with free tools like OpenSim and finding their application in specific fields of knowledge, and also other universities have already made significant progress: a good example of application is the development of 3D virtual libraries in these worlds.

One promising research area is the mobile broadband access to these virtual digital libraries, just making the knowledge ubiquitous, and special interest should be paid to recent proximity based interfaces such as NFC, which will be prevalent in mobile devices in the very near future, and a invaluable tool for check-ins as a gate between real and virtual worlds.

The use of virtual digital libraries help the environment as they save resources as printed paper, physical visits to sites, reduction in greenhouse gas emissions, electricity, time of travelling, encourages telecommuting from home and an endless resources that help to improve our world.

ACKNOWLEDGEMENTS

This proposal was carried out in the network of advanced technology research in Francisco José de Caldas University in Bogotá Colombia, as part of the Ph.D. in Engineering, within the "LIDER" [21] research group. Special thanks to Dr. José Nelsón Perez for his valuable support to develop these important projects that allow the creation new scientific horizons. With his support, we expect to continue moving forward with new developments in the research group.

REFERENCES

[1] G. P. Fishwick, "An introduction to Opensimulator and virtual environment agent-based M&S applications," in Simulation Conference (WSC), Proceedings of the 2009 Winter, Dec. 2009, pp. 177 –183.64.

[2] Linden Research, Inc, Available: http://lindenlab.com

[3] M. Barbulescu, M. Marinescu, O. Grigoriu, G. Neculoiu, V. Sandulescu, and I. Halcu, "GNU, GPL in studying programs from the systems engineering field," in Roedunet International Conference (RoEduNet), 2011 10th, June 2011, pp. 1 –4.

[4] Hippo OpenSim Viewer, Available: http://mjmlabs.com/viewer

[5] RealXtend Viewer, Available: http://realxtend.org

[6] M. Pattal, Y. Li, and J. Zeng, "Web 3.0: A real personal web! More opportunities and more threats," in Next Generation Mobile Applications, Services and Technologies, 2009. NGMAST '09. Third International, Conference on, Sept. 2009, pp. 125 –128.

[7] T. Kuroiwa and S. Bhalla, "Dynamic personalization for book recommendation system using web services and virtual library enhancements," in Computer and Information Technology, 2007. CIT 2007, 7th IEEE International Conference on, Oct. 2007, pp. 212 –217.

[8] M. Phankokkruad and K. Woraratpanya, "Web services for learning management systems: Communication architecture, in Communications (MICC), 2009 IEEE 9th Malaysia International Conference on, December. 2009, pp. 403 –408.

[9] M. Di Giacomo, "MySQL: Lessons learned on a digital library," Software, IEEE, vol. 22, no. 3, pp. 10 – 13, May-Jun 2005.

[10] S. Spaeth, M. Stuermer, S. Haefliger, and G. Von Krogh, "Sampling in open source software development: The case for using the debian gnu/linux distribution," in System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on, Jan. 2007, p. 166.

[11] D. D. Dvorski, "Installing, configuring, and developing with Xampp," March 2007.

[12] Content and Design of Academic Library Web Sites for Distance Learners: An Analysis of ARL Libraries, Association of College & Research Libraries. Guidelines for distance learning library, December 2001.

[13] Hannes Bohring, Soren Auer, "Mapping XML to OWL Ontologies" June (2005) Available: http://www.informatik.unileipzig.de/~auer/publication/xml2owl.pdf.

[14] Brewster, C. O. (2004). "Knowledge Representation with Ontologies: The Present and Future". IEEE, 72-81.

[15] Simons, Gary F., William D. Lewis, Scott Farrar, D. Terence Langendoen, Brian Fitzsimons and Hector Gonzalez. 2004b. The semantics of markup: Mapping legacy markup schemas to a common semantics. Proceedings of 4th Workshop on NLP and XML: RDF/RDFS and OWL in Language Technology. Association for Computational Linguistics, Barcelona, Spain, July 2004. Online: http://emeld.org/documents/SOMFinal1col.pdf

[16] K. Czajkowski, I. Foster, N. Karonis, C. Kesselman, S. Martin, W. Smith, and S. Tuecke. "A resource management architecture for metacomputing systems". In IPPS/SPDP '98. Workshop on Job Scheduling Strategies for Parallel Processing.

[17] R. Fracchia, G. Jeney, M. Mazzotti, E. Piri. (2011). Optimization of Multimedia over wireless IP links via X-layer design" University of Bologna.

[18] Niaye NM, et. al.. "Performance comparison of hierarchical checkpoint protocols grid computing". International Journal of Interactive Multimedia and Artificial Intelligence. 2012, v 1, n 5, (pp 46-53).

[19] M. Erdmann, D. Fensel, I. Horrocks , M. Klein, S. Melnik. "The Semantic Web on the respective Roles of XML and RDF". IEEE, In Conference publication in the Third Workshop on Workflows in Support of Large-Scale Science, On 17-17 Nov. 2008, pp 1-10.

[20] Francisco José de Caldas Distrital University 1948 in Bogotá Colombia, Available: http://udistrital.edu.co

[21] Julian Camargo, Roberto Ferro,. (2009). "Research Laboratory for Development Electronic and Networks" LIDER-UD". Available: http://201.234.78.173:8080/gruplac/jsp/visualiza/visualizagr.jsp?nro=00000000005315.

[22] Ndiaye, N. M., P. Sens, and O. Thiare, "Performance comparison of hierarchical checkpoint protocols grid computing", International Journal of Interactive Multimedia and Artificial Intelligence, vol. 1, issue Special Issue on Distributed Computing and Artificial Intelligence, no. 5, pp. 46-53, 06/2012 DOI: 10.9781/ijimai.2012.156

M.Sc. Roberto Ferro Escobar, Electronic Engineer, M.Sc. in Networks 2004, became a Member of IEEE in 1999, a Senior Member in 2011, teaching at the "Francisco José de Caldas" Distrital University since 2006 in Bogotá Colombia, ex-manager of the Research Network of Advanced Technology "RITA", former manager program of Electronic Engineering 2009-2011, is currently developing his Ph.D thesis on Computer Engineering from the Universidad Pontifical of Salamanca in Madrid Spain.

Ph.D. Rubén González Crespo, Ph.D. in Computer Engineering from the Pontifical University of Salamanca in Madrid Spain, lecturer in Sciences and Engineering in the same university.

M.Sc. Eng. Sergio Ríos Aguilar. MSc in Telecommunication Systems, Computer Engineer, and PhD Candidate, lecturer in Sciences and Engineering at Pontifical University of Salamanca in Madrid, Spain Researcher on advanced mobile services.

Eng. Nicolas Torres Electronic Engineer became a Member of IEEE in 2010, student at the "Francisco José de Caldas" Distrital University in Bogotá Colombia 2006, Active member of LIDER research group.

Model Innovation of Process Based on the Standard e-commerce International GS1

[1]Giovanny Mauricio Tarazona Bermúdez, [1]Luz Andrea Rodríguez Rojas, [2]Cristina B Pelayo, [2]Oscar Sanjuan Martínez
[1]*Faculty of Engineering*, District University Francisco José de Caldas, Bogotá, Colombia
[2]*Computer Science Department*, University of Oviedo, Oviedo, Spain

***Abstract* — This article focuses on the design and characterization of management model for MSMEs, based on e-commerce and the GS1 international e-com standard. The first part contextualizes electronic commerce and its impact on domestic industry, and briefly describes the B2B e-commerce model used in Colombia. Subsequently the first step to apply the model is presented, which corresponds to the design of a diagnostic methodology that evaluates the technological, technical, commercial and administrative aspects of the organization; after that are exposed the results of the pilot experiment performed on a MSME from Bogota, and finally will be explained the procedures for the implementation of the model.**

***Keyword s*— e-commerce, Technology Management, ICT, Knowledge Management, MSMEs.**

I. Introduction

Today innovation in processes supported by technological advances are a decisive factor in the success of business organizations. [1]It is evident that the use of technology for production operations, administrative or any kind has become a tool rather than necessary, indispensable for achieving the mission of a company purposes. [2] [3]

Rapid technological progress in recent years have not only impacted the industry and production automation, but also in logistics and marketing of products, has opened the field to information technology developments in both software and hardware, [4] coupled with advances in information technology and communications are the backbone of business processes on the web, particularly electronic commerce, which has emerged as an essential tool for successful companies in a global market. [5]

The ability to sell their products to thousands or perhaps millions of people is the obsession of a business as a basic condition for sustainability, and in turn what might seem virtually unattainable for most Colombian MSMEs, given the conditions of a Latin American country with a huge digital divide and developing economy. This is why the implementation of electronic commerce as a factor for innovation in management processes in these companies is postulated as one of the alternatives that can bring major benefits to this productive sector. [6]

It is important to characterize the categories of electronic commerce (Fig. 1. Methods of Electronic Business) and to contextualize the model designed and proposed in this article, within Colombia:

Company-Company

Here are the companies that use a network as a means to make the purchase orders and receive invoices from their suppliers, and with it all other electronic documents for the purchase order cycle to be completed.

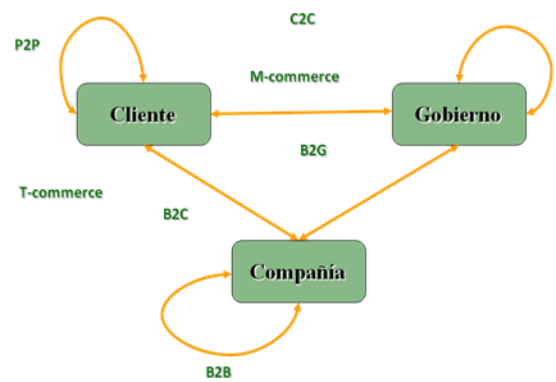


Fig. 1. Categories of Electronic Business

Company-customer

Through the network, the company offers its products to customers who accessed the quantity you need, do hereby numerous products currently marketed.

Company-Administration

Here are the transactions carried out between different companies and government organizations, such as paying taxes, even though this is true in its infancy.

Customer-Administration

Here the user would communicate with state agencies to make their tax payments or social assistance, although this is not instituted yet, it's about time it is.

Customer-Customer

It was one of the first e-business practices appeared in the early 90's web pages are little known and specialized. Your maximum evolution came with the popularization of e-bay that gives the possibility of direct sales between its users. [7]

DOI: 10.9781/ijimai.2012.178

T-commerce: The idea is to let users buy products over the Internet but through their interactive television, instead of over a phone (m-commerce) or through a PC or PDA.

B2B (business-to-business): On the Internet, also known as e-biz, is the exchange of products, services, or information between businesses rather than between businesses and consumers.

Peer to Peer. In a P2P network, the "peers" are computer systems which are connected to each other via the Internet. Files can be shared directly between systems on the network without the need of a central server. In other words, each computer on a P2P network becomes a file server as well as a client.

From this conceptual base, the project proposes applying information technologies, combined with academic concepts and skills in marketing and logistics in order to design and characterize innovative marketing processes for MSMEs producing and distribution of sweets, snacks and canned Bogotá based on the standard e-commerce GS1 International, as a means to leverage the leading commercial chain stores in the country.

II. PROCEDURE FOR PAPER SUBMISSION CHARACTERIZATION OF E-COMMERCE NATIONAL MODEL “B2B”

International GS1 is an organization of 105 institutions in over 150 countries, serving various industrial and economic sectors, by managing logistics standards to promote the efficiency of Value Networks. At home, GS1 Colombia manages the international standards of bar code (barcode), electronic commerce (e-com), synchronization of databases (GDSN) and the Electronic Product Code (EPCglobal).

Given this, it is essential that any model that seeks to facilitate the access of MSMEs to these markets is based on these standards. [8] The following will highlight the main features of the B2B model used in Colombia, approach having as standard documents that are used and their implementation within the business cycle, logistical and financial e-commerce. in the figure. 2. Logistics Cycle entitled B2B e-commerce in Colombia, sample characteristics.

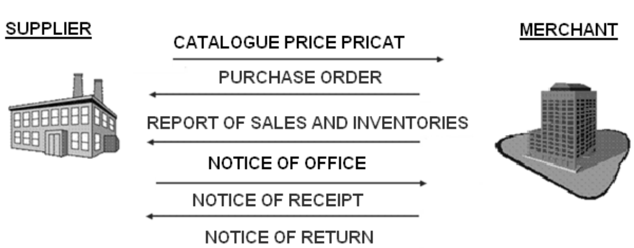


Fig. 2. Logistics Cycle B2B e-Commerce in Colombia.

The first step is to generate a *catalog business* called ***PRICAT*,** arising from the negotiations made between the supplier and dealer in a traditional way. In this negotiation the products are down to sell outlets where available and their sales prices are set.

PRICAT The electronic document should be based on the EDI standard, and should be sent to CABASnet it is the electronic product catalog that Colombia, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Paraguay and Peru, have been developed for the benefit of centralization and synchronization information to clients and suppliers use daily.[9]

Customers, suppliers, distributors and chains currently benefit from having a tool at their disposal to facilitate commercial transactions between business partners, both locally and internationally.

CABASnet who will review the structure of the catalog and sent to the e-commerce customers (chain stores, supermarkets, hypermarkets, etc.) through the interface CEN (Center for electronic business) which is operated by IBC, a group company Assenda S. Carvajal A. In Figure 3 PRICAT document processing, exemplified the expected workflow

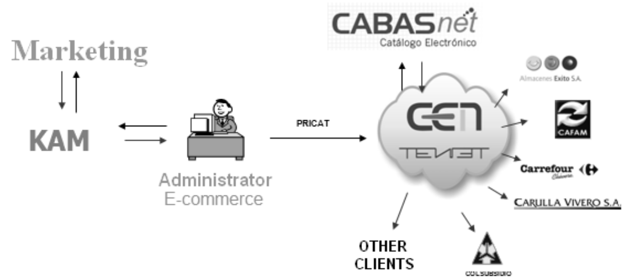


Fig. 3. Document Processing PRICAT

When a product has been coded in-store customer through the application of a PRICAT proceed from the client to generate and transmit a purchase order, the order will obey the expected demand for the product by the end customer in point of sale. The purchase order must be sent fully in the international standard GS1 EDI. [10]

After receiving the purchase order, the supplier shall dispatch in lead time, you cannot advance or delay in delivery of the order. One day before delivery at the point of sale, the producer must send a *notice of release,* electronic document which specifies the types of packaging of goods, total units shipped, the final destination of the goods, among others. Following the dispatch to print a bar code which records the information of the standard. [11]

The *notice of receipt* is an electronic document with the supermarket chains which suppliers indicate the total number of articles (products) that were received from a specific purchase order. This document is the basis for producing and sending invoices electronically from suppliers.

A key part of e-commerce model as the centerpiece of the proposal of innovative processes with large supermarket chains in the country is easy for suppliers to verify the behavior of the demand for their products. This facility provides two tools: a) purchase orders transmitted daily or as often as deemed necessary provider, or b) the sales and inventory information sent weekly supermarket chains through weekly or monthly *sales and inventory report.*

The agents involved in this innovative process serve three (3) main roles: *Manager E-Commerce* acts as the support using tools and techniques to electronic information exchange, manage and control electronic order management and physical, among others. *The commercial agent* is responsible for assembling and transmitting orders at the point of sale, track the behavior of demand and inventories, and maintain close business relationship with customers at point of sale. Finally, the *Key Account Manager* (KAM) is the one in charge of trade negotiations (products, pricing, promotions) with customers, maintain control and review of sales and inventory reports, update and ensure the availability of official price lists and catalogs to customers.

III. DIAGNOSTIC METHODOLOGY

The method of diagnosis is the initial stage of model implementation. This will determine if the company you want to apply is ready and capable to do so. The test uses the following technique is performed in the following stages: 1) Assessment of technological component and 2) Evaluation of organizational factors (technical, administrative and commercial). Assessment of the technological component is based on the model proposed by the Research Group on electronic commerce in Colombia (GICOECOL) as a diagnostic tool that will analyze the technological critical variables inherent in any company wishing to implement a model of electronic commerce

The diagnosis methodology corresponds to a classification matrix (see Figure 4), whereby a company is classified into one of four possible categories, depending on the readiness index that is the same to implement e-business models and progress in implementing these within the organization. [12]

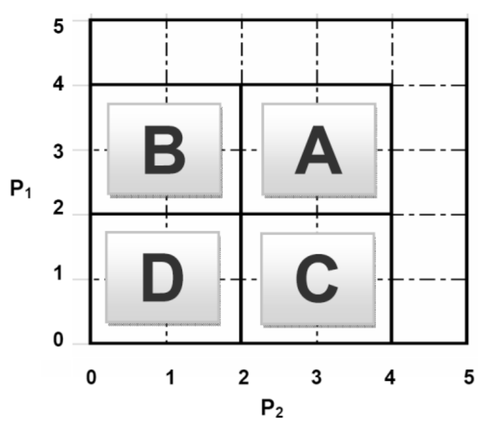


Fig. 4 Matrix diagnostic

Where P_1 represents the *readiness* of the company for the implementation of ICT and / or electronic commerce y P_2 is *the value of progress in implementation* of ICT and / or electronic commerce.

In the original methodology, the first stage evaluates the situation of the company in comparison with the economic sector in which it competes. However, for purposes of this project, modifications were made to the measurements of each of the variables, it is not necessary to compare the enterprise sector in which it competes.

In this vein, the stages for the application of the matrix, adapted to this project are:

1) Internal evaluation of the organization (Y axis).
2) Measurement of levels of implementation of e-commerce models (X axis).

Internal evaluation of the organization aims to quantify the diagnosis of the company regarding its progress in ICTs and electronic commerce. This stage is divided into four (4) steps to get a proper diagnosis quantified: Evaluation of technological resources, Web Portal Assessment, Evaluation, Human Resources and eventually final internal evaluation of the organization.

For evaluation purposes the value zero (0) is the zero impact rating, the value one (1) corresponds to the rating of low impact, impact value two (2) corresponds to the average score, and the maximum value four (4) corresponds to the highest rating. This criterion applies to all assessments.

Evaluation of technological resources (E_{RT}) is done based on the results of a questionnaire that seeks to quantify the computational resources such as computers, internet access, existence of a systems department, purchase of specific software, existence of databases, among other resources. This evaluation will yield a coefficient with value between zero (0) and four (4).

Web Portal Assessment *(EPW)* is based on a second questionnaire in which the variables of the company's Web site, such as relevant content, ease of navigation, interactivity and graphic design among others, are qualified with the degree of compliance low, medium or high. Similarly, this ratio is obtained by correcting the value on a scale from zero (0) and four (4).

For the evaluation of Human Resources *(ERH),* there is the classification of staff competencies in three (3) categories: 1. Those who have basic computer operations; 2. Those who manage their job-specific software, and 3. Those who are competent to Web 2.0 software development capability, as well as the capabilities of category 1 and 2. From this, apply a new questionnaire to each employee of the organization to be assigned to one of these categories. For each category determines the number of employees needed to implement the model, depending on the product portfolio of the company and the number of stores in which you want to sell them. With these two facts, you get a coefficient for each category by dividing the number of employees available among the number of employees needed. Finally, average the three (3) coefficients and the average is multiplied by four (4) for the coefficient of human resources *(ERH)* with a value between zero (0) and four (4).

The final internal evaluation of the organization is calculated by averaging the evaluations of technological resources, Web portal and Human resources, so you will get the grade of zero (0) to four (4) to be called P1. See Equation (1)

$$P_1 = \frac{E_{RT} + E_{PW} + E_{RH}}{3} \qquad (1)$$

Moreover, the second stage to assess the technology component, corresponds to the evaluation of levels of implementation of e-commerce, ie, progress in the use of electronic commerce for business interaction with the different

TABLE I
CUSTOMER REQUIREMENTS OF B2B E-COMMERCE MODEL IN COLOMBIA.

Type of Activity	No	Description of requirement
Commercial	1	Presentation of the units compatible product as the chain of markets, this in order to facilitate the marketing of products focused on the end customer (consumer).
	2	Comply with legal requirements concerning net contents and labeling, it should indicate at least the following information: product name, ingredients, net contents and drained mass according to the international system of units, manufacturer's name and address, lot identification, marking date and instructions for storage, instructions for use, sanitary registration number and expiration date.
Technique	3	Evidence of a solid structure that allows production to meet minimum levels of delivery. The company must have a sufficient production capacity available to ensure an adequate level of service.
	4	Comply with the sanitary conditions laid down in legislation, including health registration for each product, which shall be in force throughout the supply relationship.
	5	Geographic coverage and capacity of local minimum to ensure continuous supply. The company must ensure compliance with the logistic distribution function, therefore must have good carrying capacity sufficient or a logistics operator offering processes of "recorded delivery".
Administrative	6	Updated financial statements. The company must demonstrate a financial capacity that allows you to fully comply with the supply contract.
	7	Legal constitution of company or corporate guiding principles and defined. The act of the company should be governed by a well-defined strategic planning.
	8	Trained and sufficient to fulfill the duties of each role involved in the model.
	9	Management of return policies and restocking of the merchandise required to cause the fulfillment of its due date or damages and damages suffered by it
	10	Clearly defined billing processes comply with all applicable laws.

agents in their environment: Government, other companies, financial sector, customers, employees and suppliers.

This measurement of a new questionnaire, which the company will get a score of thirteen (13) and sixty-five (65) points. The higher the score, the progress in implementation of e-commerce is business. As is known, the evaluation of the company in the matrix of diagnosis is based on a score of one (1) to four (4), therefore it is necessary to correct this score to finally locate the assessment index P2 the matrix.

The second part of the diagnostic assess organizational factors. Of particular interest for our analysis the technical, commercial and administrative, which are directly associated with marketing and logistics components of any company, which together with the technological component, diagnosed earlier, are central to the model. These requirements are described in Table I: Customer requirements of B2B e-commerce model in Colombia

To diagnose the organizational variables, will be based on the requirements that customers have the model in Colombia, ie, chain stores and supermarkets.

For each type of requirement applies a questionnaire, so as to assess compliance with each requirement, and you get a score between zero (0) and ten (10) for each type of activity (commercial, technical and administrative).

After obtaining the coefficients of technical activities (*CAT)*, business (*CC*) and administrative activities (*CAD*), are weighted together with the technological coefficients P_1 y P_2, to obtain a total score consolidated company, which is between zero (0) and two hundred forty (240). Table II describes the diagnostic consolidated matrix

TABLE II
CONSOLIDATED MATRIX DIAGNOSIS

Coefficient	P_1	P_2	Total
Commercial Activities (CC)	$CC*P_1$	$CC*P_2$	Σ
Administrative activities (CAD)	$CAD* P_1$	$CAD* P_2$	Σ
Technical activity (CAT)	$CAT*P_1$	$CAT*P_2$	Σ
	Total		

In the column "Total", is located the summations of the results of the multiplications, as shown in the table, to finally get the total score will be critical for the implementation of the model, it will be the basis for categorizing the companies in three (3) categories.

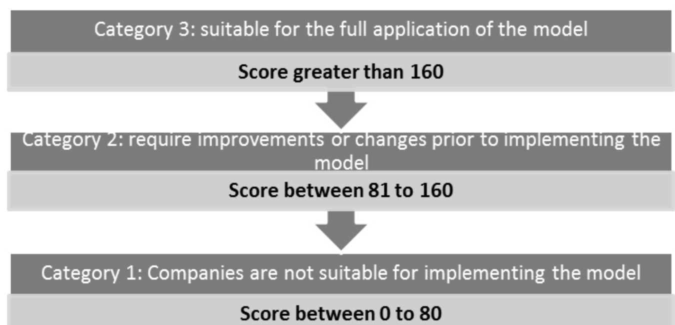


Fig. 5. Final classification of the companies evaluated

IV. PILOT TEST

A pilot test was conducted in Bogota MSMEs. The results of the evaluation of each of the coefficients technological, administrative, technical and commercial summarized below, are present en table III next.

TABLE III
RESULTS OF EVALUATION OF TECHNOLOGY COMPONENTS

Technology Assessment Results Internal Rate (P_1)	
Evaluation of technological resources (E_{RT})	0,75
Web Portal Assessment (E_{PW})	0,00
evaluation of Human Resources (E_{RH})	0,00
Average value P_1	0,25

Results Coefficient Implementation of Electronic Commerce (P_2)	
Score Electronic Commerce (PCE)	13
Final value P_2	0

TABLE IV
RESULTS OF EVALUATION OF BUSINESS

Summary of results evaluation ratio of commercial activities (CC)			
Product	Requirement 1	Requirement 5	CCi
Bocadillo veleño	10	5	7,5
Lonja de bocadillo	10	6,25	8,125
Herpo	10	8,75	9,375
Dulce de breva	0	8,75	4,375
Arequipe industrial	0	5	2,5
Glasse industrial	0	5	2,5
Bocadillo industrial	0	5	2,5
CC			*5,26*

In the tables (IV, V,VI and VII) are present the information obtained of pilot test

TABLE V
RESULTS OF EVALUATION OF TECHNICAL ACTIVITIES

Summary of results of technical activities evaluation ratio (CAT)	
Geographic Coverage Ratio (CCG)	0
Coefficient of Health Registry (CRS)	0
Unused capacity (CNU)	9,57
Coefficient of technical activities (CAT)	*3,19*

TABLE VI
EVALUATION RESULTS OF ADMINISTRATIVE

Summary of results of administrative evaluation rate (CAD)	
Requirement 2: Financial Statements	0
Requirement 7: Legal constitution of company and corporate principles	6,66666667
Requirement 8: Trained personnel and sufficient	0
Requirement 9: Handling return policies	10
Requirement 10: Billing processes defined	0
Average value Coefficient Administrative Activities (CAD)	*3,3*

TABLE VII
SUMMARY FINAL DIAGNOSIS

Coefficient	P_1=0,25	P_2= 0	Total
Commercial Activities (CC)	CC*$P_{1=}$ 1,3169	0	1,3169
Administrative activities (CAD)	CAD* $P_{1=}$ 0,8333	0	0,8333
Technical activity (CAT)	CAT*$P_{1=}$ 0,7976	0	0,7976
Total			*2.9472*

The total weighted score for the firm diagnosis was two point ninety-four (2.94) which places him in category one (1), meaning that the company is not discussed in adequate conditions to implement the model e-commerce now.

The critical factor in obtaining this result are the poor of ICTs in that the company has, so it is recommended that initial investments in this field, also for the poor level of ownership of e-business models in the company is recommended training for employees in the use of computational tools and generate added value to the administrative processes of the company.

V. DESIGN AND IMPLEMENTATION OF THE MODEL CHARACTERIZATION

The model of innovative processes that make use of electronic commerce as tools discussed in this document requires that companies wishing to implement meet the following conditions: Register in organization dedicated to the control and monitoring the quality and safety of food and pharmaceutical products in Colombia (INVIMA) all products to market, obtaining the standard Bar Codes for the products to market, acquisition and certification of EDI mailbox E-com GS1, recruitment or training of skilled personnel for e-commerce platform, CEN Affiliation software transactional EDI transmission and acquisition of specific software to create barcode EAN 14.

Once you have met the requirements for marketing, you can implement the model. The process begins with the creation and transmission of the document PRICAT, and culminates with the delivery and return notices. The fig., 6 describes the activities proposed

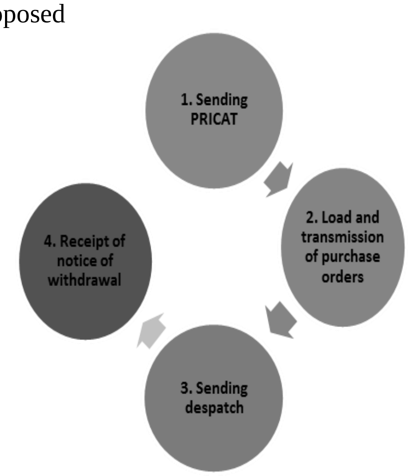


Fig. 6. Marketing process through E-com standard documents of GS1

The procedure for creating and sending the PRICAT consists of the following stages:

1) Implementation of selected software.
2) Sending a letter and attached documents.
3) Dealing with the chain (customer).
4) Development and Shipping PRICAT

Sending PRICAT´s will be necessary for the addition of new references (products), changes in existing prices or the withdrawal thereof. This will be done by the senior management of the organization and the administrator of the e-commerce platform.

After this, proceed with the preparation and transmission load of orders, the following procedure:

1) Load and transmission of purchase order.

2) Download and validation of the data associated with purchase orders.

The agents involved in this stage are the agent and manager of e-commerce platform.

The next step is to prepare and dispatch advices. For the preparation of this document should make reference to the orders received and the total quantities per order to ship. The procedure starts with the implementation of specific software for generating warnings office with a bar code to thereby generate the document itself. The software implementation is performed by the senior management and the completion and submission of dispatch by the administrator of e-commerce platform.

Finally, the procedure for receiving notice of withdrawal is as follows:

1) Receipt of notice.

2) Analysis and monitoring it.

For the preparation of this document should make reference to the orders received and the total quantities per order to ship. The agent involved in the management of refund notices are the administrator of the e-commerce platform and Key Account Manager (KAM).

VI. Conclusion

The government plans for the technological inclusion of micro, small and medium businesses included both in the National Development Plan, as in the plan MSMEs digital 2009, opening the possibility of obtaining the necessary funding for implementation of ICTs that support e-business models, facilitating the acquisition of software and hardware to improve the processes involved in managing the supply chain, resulting in more competitive and efficient companies.

Likewise, certification to international GS1 standards enable companies to restructure their logistics processes, thus becoming a tool to increase their ability to seek new markets and channels for marketing their products.[1*]

Critical to implementing innovative processes supported by e-business models is the appropriation of ICTs by the company and human resource training in the use of these tools, which can be checked in detail by the results obtained in the pilot test, where low scores in these areas not determined the feasibility of implementing the proposed model in the selected company.

For this reason, it is important to note that the implementation of innovative processes supported by the proposed e-commerce model, it is necessary for companies to appropriate knowledge management models that allow the previous experience in the use of technological resources and tools that facilitate the appropriation of concepts and procedures outlined in this document.

References

[1] K. Laudon and L. Jane, Sistemas de Información Gerencial, Pearson Prentice Hall, 2004.

[2] M. Dell, Collaboration equals innovation InformationWeek, Manhasset, 2003, Jan 27.

[2] G. Tarazona and V. Medina, "Generación de Valor en la Gestión de Conocimiento," in *Seventh Laccei Latin American and Caribbean Conference for Engineering and Technology*, San Cristobal, Venezuela, 2009.

[4] D. Teece, "Business Models, Business Strategy and Innovation," *Long Range Planning,* pp. 172 - 194, 2010.

[5] T. Koupluspoulos and C. Frappaolo, Lo Fundamental y lo más efectivo de la Gerencia del Conocimiento, McGraw Hill Iteramericana S.A, 200.

[6] J.-M. Nurmilaakso, "EDI, XML and e-business frameworks: A survey," *COMPUTERS IN INDUSTRY,* pp. 370 - 379, 2008.

[7] F. Hayes, "The story so far: e-commerce," *COMPUTERWORDL,* 2002.

[8] T. Benson, PRINCIPLES OF HEALTH INTEROPERABILITY HL7 AND SNOMED Health Information Technology Standards, London: Springer, 2010.

[9] ediversa, "GUIA DE IMPLANTACIÓN PRICAT D.96A VERSION 2," EDIVERSA, Barcelona, 2006.

[10] HYPERLINK "http://web.invima.gov.co/" http://web.invima.gov.co Consultado el 26 de septiembre de 2012

[11] HYPERLINK "http://www.cabasnet.org/" http://www.cabasnet.org/ Consultado el 15 de agosto 2012

[12] D. Barrera, G. Diana and G. Tarazona, Propuesta Metodologica De Evaluación Previa A La Implementación De Modelos De Comercio Electronico, Bogota: Universidad Distrital Francisco Jose De Caldas, 2008.

Tarazona, Giovanny. Doctoral candidate in systems and computer services for internet in the Oviedo University, Asturias, Spain (2012) - diploma of advanced studies November 2007 Pontifical Salamanca University in campus Madrid Spain, Specialization in computer project (2006), Specialization in engineering software (1999). He is Industrial engineer, Francisco José de Caldas District University, Bogota, DC, (1998). He is Full Time Professor Francisco José de Caldas District University- Faculty of Engineering since July 2003. He is founding member of KAIZEN PBT GROUP Ltda, legal representative, director of several projects, Bogotá, DC, 1996 - December 2011. Director Research Group on Electronic Commerce Colombian GICOECOL Francisco José de Caldas District University, Director of several research projects.

Rodriguez, Luz Andrea: Master Engineering and website design (2012), Specialization in hygiene and occupational health (2010), Industrial Engineer, Francisco José de Caldas District University, Bogota, DC, (2008). She is Full Time Professor Foundation Liberators University (August 2011- May 2012) Area Production. She is full professor in occupational health at INPAHU (August 2010-May 2011). Director project Research "Strategies for Strengthening of Information System Tourism Health in Bogota".

Pelayo G-Bustelo, B. Cristina is a Lecturer in the Computer Science Department of the University of Oviedo. She is Ph.D. from the University of Oviedo in Computer Engineering. Her research interests include Object-Oriented Technology, Web Engineering, eGovernment, Modeling Software with BPM, DSL and MDA.

Sanjuán Martínez, Oscar is a Lecturer in the Computer Science Department of the University of Oviedo. He is Ph.D. from the Pontifical University of Salamanca in Computer Engineering. His research interests include Object-Oriented technology, Web Engineering, Software Agents, Modeling Software with BPM, DSL and MDA.

Service Orchestration on the Internet of Things

Jordán Pascual Espada
Computer Sciencie Department, University of Oviedo, Oviedo, Spain

***Abstract* — On July 27, 2010, Jordán Pascual Espada defended his Master's thesis at Oviedo University (Spain), titled: "Service Orchestration on the internet of things". This Master's thesis is the final part of the Web Engineering Official Research Master belonging to the European Higher Education Area. Jordán Pascual Espada defended his dissertation in a publicly open presentation held in the School of Computer Engineering at Oviedo University, and was able to comment on every question raised by his committee and the audience. The master's thesis was supervised by his advisors, Juan Manuel Cueva Lovelle and Oscar Sanjuán Martínez. The thesis has been read and approved by his thesis committee, receiving the highest rating**

***Keywords* — Virtual Object, Internet of Things, DDTS, Web Services, Smart Phone.**

I. Summary

Internet of things promotes that many physical objects and electronic devices are able to communicate with other similar autonomous. These physical objects must be clearly identified; in most cases they contain a digital part as a microprocessor or a radio frequency tag. Multitude of different physical objects can be part of the internet of things systems, appliances, vehicles, industrial machinery, services machines, etc. Internet of Things offers a new range of possibilities to make life easier for people supporting or automating many daily tasks.

Same time as technology advances more and more objects "things" that were only beginning to appear physically in digital format. Examples are: books, maps, event tickets, airline tickets, calendars, contact cards, electronic purses, etc.. By observing the behavior of these digital objects we find that there is no standard format or recommendations to standardize its use, no mechanism to deal with in a general way, store, share, or processing by other applications that are not familiar with the specific format.

Problems arising from the lack of standardization in digital objects are:

1. Most of the virtual objects have a specific format, these objects are composed of different types of files and structures. The software application that manages the virtual object should be able to decode it, to operate with it. This handicap leads to the need of installing many applications in case we want to operate with different virtual objects. It makes it harder for a company or developer to place in the market their own virtual object, since nobody would be able to decode it without the suitable software.
2. Lack of Communications: Ideally for an object that is part of an IOT system in which various virtual and physic objects interact to automate tasks is that each object must communicate the actions it is able to execute to the other objects (Fig.1). Since there is no standard format way to get actions or services from a giving virtual object, it is very difficult to interact with another application.

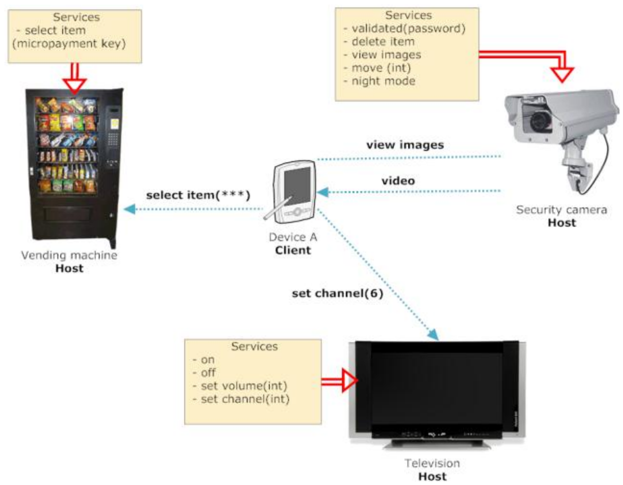


Fig. 1. Service Orchestration scenario.

This research provides a common structure for the construction of digital objects. Which gets all digital objects regardless of their complexity or business logic can be interpreted the same way by any electronic device that has the computing power required (embedded systems, computers, PDAs, mobile phones, etc). The proposal can eliminate the pre-configuration needs and specific software requirements for the manager of different digital objects [1,2].

The proposed model encourages the integration and communication of digital objects, applications, devices and users. Establish a process similar to that followed for integration of physical elements in the network of the Internet of things, where digital objects should offer the possibility of other users or applications might discover them and get their catalog of actions.

The validity of the proposal has been illustrated with the development of several prototypes and specific test that evaluate the most important aspects of the proposal.

DOI: 10.9781/ijimai.2012.179

REFERENCES

[1] Pascual, J, Sanjuán, O., Cueva, J.M, Pelayo B.C., Álvarez, M. & González, A. (2010). Modeling architecture for collaborative virtual objects based on services. Journal of Network and Computer Applications, v 34, n 5, (pp 1634-1647).

[2] Pascual, J, Sanjuán, O., Cueva, J.M, & Pelayo B.C. (2011). Virtual objects on the Internet of things. International Journal of Interactive Multimedia and Artificial Intelligence, v 1, n 4, (pp 23-29).

Pascual Espada, Jordan received a PhD in Computer Science from the University of Oviedo (Spain) in 2011. He is Research scientist at Computer Science Department of the University of Oviedo (Spain), also he assists on the development of cloud computing applications with Elasticbox. B.Sc. in Computer Science Engineering and a M.Sc. in Web Engineering from the University of Oviedo. His research interests include the Internet of Things, exploration of new applications and associated human computer interaction issues in ubiquitous computing and emerging technologies, particularly mobile and Web applications. He has been author of published papers in several journals and recognized international conferences and symposiums

Artificial intelligence methodologies applied to quality control of the positioning services offered by the Red Andaluza de Posicionamiento (RAP) network

Elena Giménez, Antonio José Gil, *Department of Cartographic, Geodetic Engineering and Photogrammetry, University of Jaén, Jaén, Spain*

***Abstract* — On April 26, 2012, Elena Giménez de Ory defended her Ph.D. thesis at University of Jaén, entitled: "Robust methodologies applied to quality control of the positioning services offered by the Red Andaluza de Posicionamiento (RAP) network". Elena Giménez de Ory defended her dissertation in a publicly open presentation held in the Higher Polytechnic School at the University of Jaén, and was able to comment on every question raised by her thesis committee and the audience. The thesis was supervised by her advisor, Prof. Antonio J. Gil Cruz, and the rest of his thesis committee, Prof. Manuel Sánchez de la Orden, Dr. Antonio Miguel Ruiz Armenteros and Dr. Gracia Rodríguez Caderot. The thesis has been read and approved by his thesis committee, receiving the highest rating. All of them were present at the presentation.**

***Keywords* — RTK, Active GNSS network, Robust Statistics, RAP, Reference frame**

I. Thesis Summary

The development of accurate GNSS (Global Navigation Satellite System) navigation techniques in the recent years, particularly GPS (Global Positioning System), has provided a very important contribution in civil applications: intelligent fleet management systems, social networking, tourist information, traffic management, air traffic control, etc. That is why RTK (Real Time Kinematic) positioning is widely used throughout the world. This is due to the fact that choosing between working at real-time and obtaining centimeter accuracy is no longer necessary: using RTK positioning is possible to combine both. In addition, RTK networks materialize and transmit a reference frame to the users. For this reason, in Andalusia, Spain, an active GNSS network has been materialized, Red Andaluza de Posicionamiento (RAP) [4]. Accuracy studies, implementation of working methodologies and the analysis of the goodness of the obtained results are widely studied topics in Geodesy and Cartographic Sciences [8]. Therefore it is necessary to perform, a rigorous analysis of the two main aspects of these networks: real-time positioning and reference frame. Since statistics techiniques are widely used in artificial intelligence [7], robust methodologies are employed in both analysis methods.

As done in an increasing number of studies, robust methodologies were employed in order to analyze the quality of the real-time positioning services, with particular respect to outlier identification and accuracy assessment. Classical statistical methods do not appear adequate because they are mainly based on the hypothesis of normally distributed samples on Least Squares adjustment results; besides, most outlier detection tests are set up for univariate samples. Thus, a method based on the robust computation of Mahalanobis distances was developed, tested and applied to the RAP Network. This method is able to detect outliers in multivariate samples, as proved by its evaluation by comparing the results obtained from randomly generated data with those stemming from other classical methods.

Moreover, the proposed method is effective with multivariate and small samples, without making assumptions about its distribution, so it can be applied in other problems, no necessary within the field of Geodesy [5].

The application of this method in the processing of RTK positions, recorded in real time with a GNSS receiver assisted by the RAP network, for positioning services is detailed shown in the thesis. Besides, parameters like the coverage of the service, the time to fix ambiguities, and the accuracy and precision of the RTK positioning are also analyzed as described in [6].

In order to analyze the reference frame of the RAP, let us note that the reference frame of a GNSS active network is disseminated to users. But the stations of these networks are not static: due to geodynamic phenomena and local processes change their positions. Thus, the correct realization of the reference frame, as well as the study of the position changes in its permanent stations, are crucial issues in the man-agement of RTK networks.

In Spain, the official reference frame is a realization of the European Terrestrial Reference System 1989 (ETRS89), materialized by the REGENTE network. But Andalusia is

DOI: 10.9781/ijimai.2012.1710

located along the contact between the Eurasian and African plates, so some RAP permanent station coordinates are not expected to maintain consistency over time with ERTF89. That is why the analysis of the reference frame becomes more necessary.

First, the daily time series solutions must be computed. Processing was performed with the Bernese 5.0 software [3] using the precise ephemerides available from the International GNSS Service (IGS) and absolute calibrated antenna offset values provided by the U.S. National Geodetic Survey (NGS).

The reference frame study has been performed as described in [1]: possible outliers are removed from the time series using LTS (Least Trimmed Squared) method, and using a stochastic model for a signal, the correlated signal can be separated from the uncorrelated noise. Finally, the coordinates were computed by least-squares.

In order to compare the calculated coordinates with the official coordinates of the permanent stations, they must be given in the same reference frame. If permanent station velocities are provided, the transformation can be computed following Boucher and Altamimi [2]. When these velocities are not supplied, a 7-parameter transformation can be computed using, for instance, EUREF permanent stations.

The obtained results show the evolution of the RAP permanent stations positions. The coordinate transformation between official coordinates and calculated coordinates cannot be described as a rigid motion, as expected considering the mentioned geodynamic phenomena.

Finally, the methods described in this thesis, meant to analyze the positioning services of a GNSS active network, can be used to study the positioning services provided by any GNSS active network.

Acknowledgment

Elena Giménez de Ory received a 4-year PhD grant from the Spanish Ministry of Education and Science in the period September 2007-August 2011. The research was financed by the research project ESP2006-10113 supported by the Spanish Ministry of Education and Science (European Regional Development Fund-ERDF).

References

[1] Y. Leung, J.H. Ma and W.X. Zhang, “A new method for mining regression classes in large data sets”. *IEEE Trans. Pattern Analysis Machine Intell.*, vol 23, pp. 5-21, Jan 2001.

[2] W. Lorenzo, R. González-Crespo and A. Castillo-Sanz, “A Prototype for Linear Features Generalization”. *Int. J. Interact. Multimed. Artif. Intell.*, vol 1, pp. 59-65, Dec 2010.

[3] E. Giménez, M.S. Garrido, M.C. Lacy and A.J. Gil, “Comparing RTK Positioning from Updated REGAM and MERISTEMUM CORS Networks in Southeast Spain”, *J. Appl. Geodesy*, vol 5, pp. 23-35, Apr 2011.

[4] B. Benciolini, L. Biagi, M. Crespi, A. M. Manzino, and M. Roggero, “Reference frames for GNSS positioning services: Some problems and proposed solutions”, *J.Appl.Geodesy*, vol 2, pp. 53-62, Apr 2008.

[5] M.S. Garrido, E. Giménez, J.A. Armenteros, M. C. Lacy and A.J. Gil, “Evaluation of NRTK Positioning Using the RENEP and RAP Networks on the Southern Border Region of Portugal and Spain”, *Acta Geod. Geophys. Hung.*, vol 47, pp. 52-65, Mar 2012.

[6] E. Giménez, M. Crespi, M.S. Garrido and A.J. Gil, “Multivariate Outlier Detection Based on Robust Computation of Mahalanobis Distances, Application to Positioning Assisted by RTK GNSS Networks”. *Int. J. Appl Earth Obs. Geoinf.* 16, pp. 94-100, Jun 2012.

[7] *Bernese GPS Software Version 5.0*, University of Bern, Bern, Switzerland, 2006.

[8] C. Boucher. and Z. Altamimi (May 2011). Memo: Specifications for reference frame fixing in the analysis of a EUREF GPS campaign. Available: http://etrs89.ensg.ign.fr/memo-V8.pdf.

Elena Giménez de Ory is an Associate Professor at the Pontifical University of Salamanca. On 2012 she defended her Ph.D. thesis in Geodetic and Cartographic Engineering at the University of Jaén. Her current research interests include robust statistics and its applications to RTK positioning based on active GNSS networks.

Antonio J. Gil is a full professor of Geodesy at the department of Cartographic, Geodetic and Photogrammetric Engineering at the University of Jaén. He has been the Head of the UJA Geodesy Research Group since 1997. His current research interests include, among others, Global Navigation Satellite Systems and positioning and navigation based on active GNSS networks. He is author of more than a hundred works, articles and conference reports.

The 10th IEEE International Symposium on Parallel and Distributed Processing with Applications

The 6th International Conference on Multimedia and Ubiquitous Engineering

Leganés, Madrid, 10-13 July 2012

Universidad Carlos III de Madrid

www.uc3m.es

ARCOS

www.arcos.inf.uc3m.es