

Volume 2, Issue 1
May 2015
ISSN: 1390-9266

LAJC

LATIN-AMERICAN JOURNAL OF COMPUTING

FACULTAD DE INGENIERÍA DE SISTEMAS

QUITO - ECUADOR

Editorial Committee:

PhD. Jenny Torres, National Polytechnic School, Ecuador

PhD. Josafá Pontes, Tokyo Institute of Technology, Japan

PhD. Sergio Luján Mora, University of Alicante, Spain

<http://lajc.epn.edu.ec/>



ESCUELA
POLITÉCNICA
NACIONAL



FACULTAD DE INGENIERÍA DE SISTEMAS

LATIN AMERICAN JOURNAL OF COMPUTING

LAJC

Vol II, Issue 1, May 2015
ISSN: 1390-9266

Published by:
National Polytechnic School
Faculty of Systems Engineering
Department of Informatics and Computer Sciences

Quito-Ecuador

LATIN AMERICAN JOURNAL OF COMPUTING - LAJC

Published by:

National Polytechnic School
Faculty of Systems Engineering
Department of Informatics and Computer Sciences
Ecuador

Editorial Committee:

PhD. Jenny Gabriela Torres Olmedo, National Polytechnic School, Ecuador
PhD. Josafá de Jesús Aguiar Pontes, Tokyo Institute of Technology, Japan
PhD. Sergio Luján Mora, University of Alicante, Spain

Editor in chief:

PhD. Jenny Gabriela Torres Olmedo, National Polytechnic School, Ecuador

Section Editors:

Eng. Hernán David Ordoñez Calero, National Polytechnic School, Ecuador
Ms. Sandra Milena Nazamués Quenguán, National Polytechnic School, Ecuador
Mr. Santiago Alejandro Sandoval Hinojosa, National Polytechnic School, Ecuador

Mailing address:

Escuela Politécnica Nacional, Facultad de Ingeniería de Sistemas
Ladrón de Guevara E11-253, La Floresta
Quito - Ecuador, Apartado Postal: 17-01-2759

Web address:

<http://lajc.epn.edu.ec/>

E-mail:

lajc@epn.edu.ec

Frequency:

2 issues per year

Circulation:

500

EDITORIAL

Dear readers,

We are pleased to present the first issue of the Latin American Journal of Computing – LAJC in 2015. This issue includes five articles covering different aspects of Information Security, Information Systems, Intelligent Systems and Software Engineering.

Latin American Journal of Computing - LAJC is an open access peer-reviewed journal, sponsored by the Faculty of Systems Engineering of the National Polytechnic School of Ecuador, which goal is to bring together researchers and practitioners from academia and industry to present original research papers, technical reviews and state-of-the-art reviews for publication.

We would like to thank all authors, who contributed to the success of the Journal. Special thanks to the reviewers for their contributions to keeping the high quality of the selected papers.

Cordial thanks are due to the Section Editors members for their efforts and the organizational work. Finally, we cordially thank National Polytechnic School for supporting and publishing this volume.

We hope that you enjoy reading this issue and find the articles informative and useful. We highly encourage you to submit your work which fit within the scope of LAJC. Please keep in mind that we have an open call for submissions twice a year.

For detailed instructions on the preparation and submissions of manuscripts, please check the URL below:

<http://lajc.epn.edu.ec/index.php/LAJC/pages/view/call-for-papers>

We will be happy to receive your comments and feedback on our journal.

Best Regards,

PhD Jenny Torres
Editor in chief
LAJC

Latin American Journal of Computing - LAJC Reviewers

We are most grateful to the following individuals for their time and commitment to review manuscripts for Latin American Journal of Computing - LAJC for this edition.

Aguiar Pontes Josafá, PhD. Tokyo Institute of Technology, Japan
Brandão Diego, PhD. Fluminense Federal University, Brazil
Duarte Ferreira Vera Lucia, PhD. Federal University of Pampa, Brazil
Fuertes Díaz Walter, PhD. Army University - ESPE, Ecuador
García Olaya Angel, PhD. Charles III University of Madrid, Spain
Loza Aguirre Edison, PhD. University of Grenoble, France
Luján Mora Sergio, PhD. University of Alicante, Spain
Meliá Beigbeder Santiago, PhD. University of Alicante, Spain
Pousa Federico, PhD. University of Buenos Aires, Argentina
Suntaxi Gabriela, MSc. National Polytechnic School, Ecuador
Torres Olmedo Jenny, PhD. National Polytechnic School, Ecuador

TABLE OF CONTENTS

Hierarchical Path Finding to Speed Up Crowd Simulation

Carlos Fuentes	9 - 20
----------------------	---------------

Realidad Virtual Acústica: El abordaje de las Redes Neuronales Artificiales

José F. Lucio Naranjo, Roberto A. Tenenbaum y Julio C. B. Torres	21 - 26
--	----------------

Propuesta de un Modelo de Planificación y Organización para Implementación de Centros de Servicios tecnológicos en Instituciones Educativas Públicas

Santiago Cordovilla, Christian Madruñero y Bolívar Palán Tamayo	27 - 36
---	----------------

Arquitectura Clúster de Alto Rendimiento Utilizando Herramientas de Software Libre

Leonardo Chuquiguanca, Edyson Malla, Freddy Ajila y Rene Guamán-Quinché	37 - 44
---	----------------

Intrusive and Non-Intrusive Load Monitoring (A Survey) Inference and Learning Approach

Danilo Burbano	45 - 53
----------------------	----------------

Hierarchical Path Finding to Speed up Crowd Simulation using Navigation Meshes

Carlos Fuentes

Abstract—Path finding is a common problem in computer games. Most videogames require to simulate thousands or millions of agents who interact and navigate in a 3D world showing capabilities such as chasing, seeking or intercepting other agents. A new hierarchical path finding solution based on navigation meshes is proposed in order to minimize the complexity in big environments. The method has two steps: Firstly, it creates the hierarchy tree based on a recursive partitioning. Then, the optimal path is found in the hierarchy at certain level. This approach performs much faster path finding calculations than a common A*. These claims are verified on big environments.

Index Terms—A*, crowd simulation, graph algorithms, graph partitioning, hierarchical path finding, HPA*, navigation meshes, path planning, Recast, route planning, shortest path search.

Resumen—La búsqueda de ruta es un problema común en los juegos de computador. La mayoría de los videojuegos requiere simular miles o millones de agentes que interactúan y navegan en un mundo 3D mostrando capacidades tales como persecución, la búsqueda o la intercepción de otros agentes. Se propone una nueva solución de búsqueda de ruta jerárquica basada en mallas de navegación con el fin de minimizar la complejidad en entornos grandes. El método tiene dos etapas: En primer lugar, se crea un árbol de jerarquía basada en un particionamiento recursivo. Después, el camino óptimo se encuentra en la jerarquía a cierto nivel. Este enfoque lleva a cabo cálculos de búsqueda de ruta más rápido que un A* común. Estas afirmaciones son verificadas en entornos grandes.

Palabras claves—A*, algoritmos de grafos, búsqueda de ruta jerárquica, búsqueda de trayectoria más corta, HPA*, mallas de navegación, partición de grafos, planificación de trayectorias, Recast, simulación de multitudes.

I. INTRODUCTION

PATH finding is a common problem in computer games. Most videogames require to simulate thousands or millions of agents who interact and navigate in a 3D world showing capabilities such as chasing, seeking or intercepting

This paper was submitted by corresponding author on 13 February 2015 for review. This work was supported in part by the MOVING research group in the Polytechnic University of Catalonia.

The author, was with Polytechnic University of Catalonia, Barcelona, Spain. He is now an employee in ThoughtWorks (e-mail: cfuentes@thoughtworks.com).

other agents. This behavior is solved using path finding. A* is the most commonly used method to determine the shortest path between two points. It expands the nodes in the graph representation of the environment with the smallest estimated solution cost first; however the cost of this search can grow exponentially with the size of the terrain and the allocated memory is very limited. Therefore, a hierarchical subdivision of the world environment is necessary under those restraints.

Hierarchical path finding has been studied in the last decade, it allows to compute the shortest and optimal route between two locations in large terrains based on a hierarchical graph. This significantly decreases the execution time and memory footprint in crowd simulation environments. Many of these approaches only have one abstract graph to work on. It means, the searches are just done on one abstract level of the environment subdivision. These approaches have been only applied in 3D environments based on regular grids.

This new method proposes a new hierarchical path finding solution for big environments. A navigation mesh is used as abstract data structure to partition the 3D world. Then, a graph representation is extracted and considered it as the level 0 in a hierarchical tree. After, many subdivisions are created by recursively partitioning a lower level graph into a specific number of nodes. The number of nodes is a parameter. The partition is performed until the graph of the latest level cannot be divided. Thus, a particular path planning search can be executed in any level of this hierarchical tree. The higher the level of the hierarchy, the fewer the number of nodes to search in. This approach allows faster path finding calculations than a common A* without any hierarchy.

II. RELATED WORK

Many path finding algorithms have been studied for more than a decade, all of which attempt to balance the inherent tradeoff between two criteria, namely the path planning runtime computation and the resulting path quality.

The computational cost depends on how the scenario is divided (based on grids, waypoints or navigation meshes) and whether the navigation environment is fully static or dynamic, where a complex replanning may be required on the fly in order to get the correct path.

At present, a hierarchy of graphs is applied to reduce the computational calculation known as hierarchical path finding, where the idea is to decompose the search problem into multiple searches on smaller graphs and to cache information about path segments that are shared by many routes. The

higher the level the graph belongs to, the smaller the number of nodes the graph has.

A. Environment Division

In videogames and artificial intelligence fields, the 3D world environment could be characterized as a weighted direct graph. There are well known structures suitable for fast path planning calculation. These spatial divisions allow fast searches for collision-free paths, but which are not necessarily globally the shortest ones. The most common environment representations are regular grids, roadmaps, and navigation meshes (NavMesh).

1) Regular Grids

Regular grids are the simplest way to represent the 3D world. This regular cell grid is commonly used for both global and local paths of agents. The scenario is divided into a two dimensional square cells with the same size. Each cell can have two states: opened or blocked. An opened cell is the space where the character can stand or move in, a blocked one is not accessible for the character because there are obstacles or walls. However, more than two states can be stored in a cell such as density, percentage of occupancy, etc. In the method proposed by Loscos [1], they store more than two states in each cell of a regular grid. Each cell has local information for collision of the environment and agents in order to improve the realism of the simulation. It also can be extended for creating more complex agent-environment behaviors using different layers [2], [3].

A navigation graph can be extracted from the regular grid by using the connectivity information between cells. Each node in the graph represents a cell and an edge the link between two cells. (See Fig. 1).

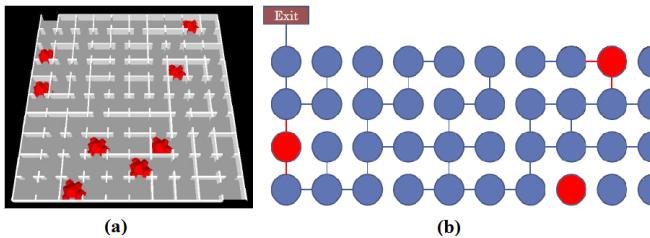


Fig. 1. Regular Grid Extraction with two states: World environment (a). Extracted graph for the upper part of the grid (blue nodes have opened state and red nodes have blocked state) (b).

If the virtual world is small and grid-like, this technique is useful. Also, this approach is ideal for 2D terrains where the height of the character is not relevant. Moreover, the resolution of the grid determines how accurate it represents the walkable space. However, the memory cost becomes unacceptable when the number of cells is extremely high for a regular grid with enough resolution. The idea of partitioning the world representation in cells with big sizes in order to mitigate this problem could affect the path quality.

Other limitation of working on grids is the coarse coverage of underlying terrain, it means that one big cell could cover both walkable and obstacle area. It should

be classified as either walkable or obstacle under some criterion. Also, the paths do not look realistic because of the agent movement is only restricted to do fixed angle turns. Motions created by grid search tend to be unnatural because, for grid searches, a path smoothing step needs to be applied in the post-processing phase resulting in expensive queries. (See Fig. 2).

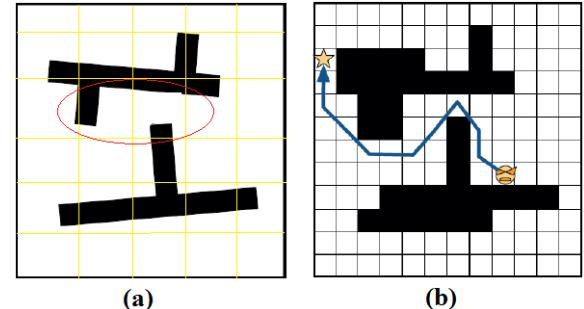


Fig. 2. Limitations of Grid Methods: Walkable and obstacle area inside the same cell (a). Restricted movement (fixed angles) (b).

Therefore, using large grids that are composed for thousands of cells could easily exceed the memory requirements of the current high-end hardware provides. Moreover the preprocessing time of calculating the extracted graph becomes impractically large with the growing number of cells. All of these limitation make a grid representation not suitable for big environments.

2) Roadmaps

The Probabilistic Roadmap Planner (PRM) [4], [5] is a planner that can compute collision-free paths. The PRM consists of two phases: a construction phase (off-line) and a query phase (on-line). In the construction phase, a roadmap is built, it consists of computing a very simplified representation of the free space by sampling configurations at random. Then the sampled configurations are tested for collision and each collision-free configuration is retained as a "milestone". Each milestone is linked by straight paths to its k-nearest neighbors. Finally the collision-free links will form the PRM.

Sampled configurations and connections are added to the roadmap until the roadmap is dense enough. A roadmap is normally represented as a graph in which the nodes correspond to placements of the entity and the edges represent collision-free paths between these placements.

In the query phase, the start and goal configurations are connected to the roadmap. Then, the path can be obtained by a Dijkstra shortest path query. (See Fig. 3).

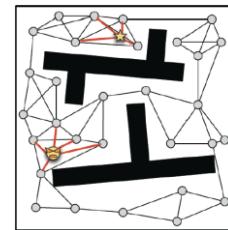


Fig. 3. Probabilistic Roadmap Planner: Connect Start and Goal node to the roadmap.

A roadmap can also use a Voronoi diagram classifier [6]. The Voronoi diagram is one of the most popular structure for spatial partitioning. Given seed points, it will partition the plane in cells such that for each seed there will be a corresponding cell consisting of all points closer to that seed than to any other.

In order to calculate the seeds, a random sample is picked of the entity (placement) in each iteration. Then, the placement is checked whether collision free from the entity is. If so, it is retracted to the Voronoi diagram using binary interpolation. Finally, the edges are also retracted until every part of the edge is at least some pre-specified distance away from the obstacles. (See Fig. 4).

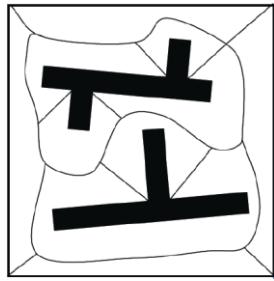


Fig. 4. Voronoi-Based Roadmaps: Retract PRM (nodes and edges) to the medial axis.

Unfortunately, the PRM method leads to low quality roadmaps, consisting of straight line segments that require a lot of time-consuming smoothing in order to be useful for virtual world applications. This is due to the random nature of the PRM method. Also, it drives to larger graphs when many milestones are needed.

3) Navigation Meshes

Navigation Meshes (NavMesh) is a data structure that is specifically designed for supporting path planning and navigation computations. It encodes a convex decomposition of the scene where each convex polygon (nodes) is a walkable area and they are connected using links (edges) that provide the connection between cells for agents to walk through. This representation has few nodes which contain more accurate information about the 3D environment.

The main function of a navigation mesh is to represent the free environment efficiently in order to allow path queries to be computed in optimal times and to support other spatial queries useful for navigation. NavMesh has some properties that are listed below:

- **Linear number of cells.** A navigation mesh must represent the environment with $O(n)$ number of cells or nodes n for efficient path calculations. This is critical for path search to run in optimal times.
- **Quality paths.** A navigation mesh must facilitate the computation of quality paths. At least, locally shortest paths must be provided.
- **Arbitrary clearance.** A navigation mesh must

provide an efficient mechanism for computing paths with arbitrary clearance from obstacles. No pre-computed clearance valued must be known.

- **Representation robustness.** A navigation mesh must be robust to degeneracies in the description of the environment. Each description of obstacles must be handled such as intersections, overlaps, etc.
- **Dynamic updates.** A navigation mesh must efficiently update itself to accommodate dynamic changes in the environment.

An example of the world representations are shown the Fig. 5.¹

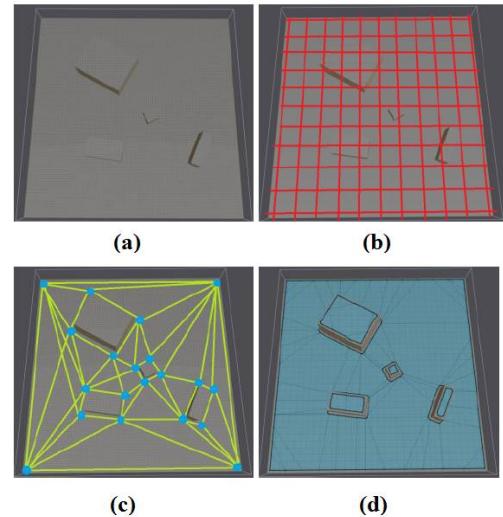


Fig. 5. Representations of the environment division. Empty map a). Regular Grid map b). Roadmap c). NavMesh d).

B. Hierarchical Subdivision

There has been some research recently focused on hierarchical path finding techniques using the A star algorithm such as HPA* [17] which is based on grid maps and clustering. HPA* creates an abstract graph from a grid in order to minimize the complexity of the problem. This abstract graph is built by dividing the environment into squares clusters connected by entrances. Basically, the algorithm has two steps: the pre-processing step where the grids are grouped in a cluster with a user defined size. These clusters will be the nodes of the high level graph. Then, the entrances (connections between two clusters) are placed with one or two transitions.

The clusters are connected with inter-edges with cost 1.0 and the cost of intra-edges are calculated running regular A* [8] searches inside each cluster, for all pairs of abstract nodes that shared the same cluster. The second step is the online search which inserts the start and goal nodes into the abstract graph and searches the optimal path with A* between them. The low level graph is much smaller than the original one.

¹ The 3D model representations were performed using the Recast navigation Tool. [7].

This approach is only based on grids. Finally, HPA* softs the path in an attempt to undo some of the error introduced with the merging of cluster border vertices. However, the final path is still not optimal. Since the abstract graph is much smaller than the original graph, search problems can be greatly simplified by using the abstract graph instead of the low level one. This algorithm is the base for the hierarchical approaches. However it only works on grids.

Another similar method based in HPA* but taking into account the size of the agents and terrain traversal capabilities is Hierarchical Annotated A* (HAA*) [19]. Basically, it is an extension of HPA* which allow multi-size agents to efficiently plan high quality paths in heterogeneous-terrain environments. Each agent has a size property and a capability (ability to walk in a certain type of terrain). The clearance values (See Fig. 6) which is the size of maximum traversal area at each octile (cell in the grid) is calculated for each capability. The clearance values for different type of terrain are shown in the Fig. 6.

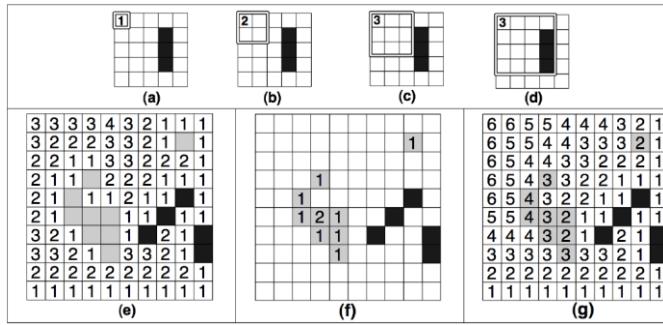


Fig. 6. Clearance Values: (a) - (d) Computing clearance; the square is expanded until a hard obstacle is encountered. (e) - (g) Clearance values for different capabilities.

In order to find the shortest path, an adaptable A* is performed by taking into account the size and capability of each agent. It means the nodes with a clearance value greater than the size will be expanded. The path planning is done over an abstract graph which is created in the same way as HPA* [17]. The entrances between clusters, inter-edges, intra-edges are calculated considering the agent properties.

C. Path finding algorithms

Current state-of-the-art real time path finding algorithms try to improve the performance measures described in (Russell and Norvig, 2003, page 71) in order to guarantee a constant bound on response time. These measures are:

- Completeness: Whether or not a route is found, if one exists.
- Optimality. Whether or not the best path is found.
- Time complexity: Number of iterations to reach the goal.
- Space complexity: Maximum number of nodes stored in memory at each iteration.

In order to do path planning, the representation of the

environment need to be discretized to facilitate efficient path finding queries. Then, efficient planning algorithms need to be developed in order to be able to generate solutions with strict time constraints for extremely large and complex problem domains.

The most known and popular dynamic search algorithm is A* search [8]. It is robust and simple to implement, with strict guarantees on optimality and completeness of solution. Hence, it represents a popular and widely used method for path planning in virtual environments. The A* algorithm uses a heuristic to restrict the number of states that must be evaluated before finding the true optimal path. It guarantees to expand an equal number or fewer states than any other algorithm using the same heuristic. A* may be too slow. Its memory use is also variable and may be high depending on the size of its opened and closed lists and the heuristic function used.

Also, Anytime Planning algorithms find the best suboptimal plan and iteratively improve this plan while reusing previous plan efforts. One of the most popular A* is called Anytime Repairing A* (ARA*) [9]. It performs a series of repeated weighted A* searches while iteratively decreasing a loose bound (ϵ). Then, it iteratively improves the solution by reducing ϵ and reusing previous plan efforts to accelerate subsequent searches. The key to reusing previous plan efforts is keeping track of over-consistent states. ARA* solution is no longer guaranteed to be optimal.

Furthermore, Incremental planning algorithms try to reuse the results of the previous plan calculation in order to reduce the planning effort. It also helps to compute the new plan when there is a small change in the environment. One common replanning method is presented in [10]. D* Lite performs A* to generate an initial solution, and repairs its previous solution to accommodate world changes by reusing as much of its previous search efforts as possible. D* can correct "mistakes" without re-planning from scratch but requires more memory.

Finally, Anytime Dynamic A* (AD*) [11] combines the properties of D* and ARA* to provide a planning solution that meets strict time constraints. It efficiently updates its solutions to accommodate dynamic changes in the environment. These updates are performed by series of repeated searches by iteratively decreasing the inflation factor. AD* cannot handle dynamic changes in goal.

III. HIERARCHICAL PATH FINDING

This method is based in the HPA* algorithm described in the previous section. In order to apply this approach, an initial discretization is needed of the 3D world, and navigation meshes are the most accurate for this purpose.

One advantage of using navigation meshes in comparison with grid approaches, is that the number of cells is much smaller and thus the initial graph abstraction is smaller.

A. NavMesh division

Many tools have been proposed for a NavMesh subdivision. Recast Tool is an automatic open-source navigation mesh generator toolset for games [7]. The automatic NavMesh generation is done via Watershed Partitioning which creates a robust triangulation without overlaps and holes. This method used by Recast is inspired by the work by Haumont [12]. It is automatic, which means that any geometry can be as an input and it will output a robust mesh. It is also fast which means swift turnaround times for level designers. Recast tool also provides an optimized A* implementation into Detour project classes [7]. Fig. 7 shows the tool.²

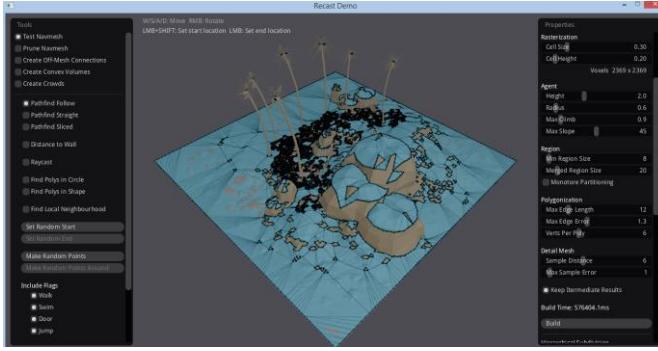


Fig. 7. Recast Tool software. (Model: Tropical Islands (12666 polygons)).

Recast tool generates a NavMesh from a triangle soup. It receives an arbitrary polygon soup with triangles marked as walkable. The reconstruction is done in the preprocessing step of the algorithm and it is divided as follows:

1) Voxelize the polygons

The voxel mold is built from the input triangle mesh by rasterizing the triangles into a multi-layer heightfield. This process makes the method robust against degeneracies of the model (such as interpenetrating geometry, cracks or holes) as well as simplifies the furniture of the scene.

2) Build navigable space from solid voxels

Some simple filters are applied to the voxel mold to prune out locations where the character would not be able to move, for instance: too steep slopes, too low places, etc. This is done by calculating the distance and the slope in each voxel. (See Fig. 8).

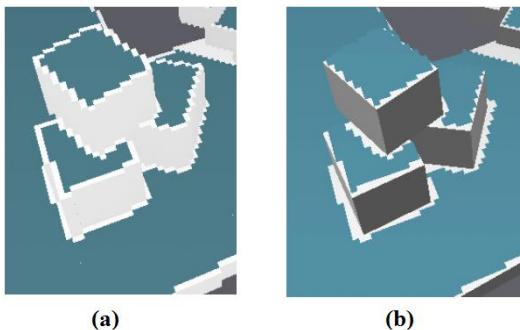


Fig. 8. Build navigable space from solid voxels: Voxelization with walkable cells marked (a) and walkable cells overlayed on top of input geometry

² All the test models were obtained using <http://tf3dm.com/>

3) Build watershed partitioning and filter out unwanted regions

The Watershed Transform [13] finds the catchment basins by building the distance transform of the input areas. It starts from the highest distance one slide at time. Then, it finds any new catchment basins and it fills them with a new ID. Finally, it expands existing regions. The catchment basins become the centers of the regions (see Fig. 9). Then, a filter pass is applied to remove small unconnected regions and merge small regions together. The result is a set of nonoverlapping simple regions that can be used as basis for generating waypoint graphs.

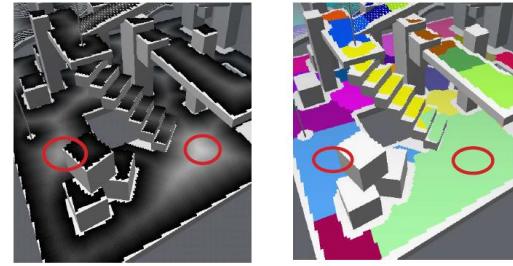


Fig. 9. Build watershed partitioning and filter out unwanted regions: The catchment basins become the centers of the regions. (White areas represent lower region).

4) Trace and simplify region contours

It searches the contours by finding a starting point to start tracing (region edge cell). Then it traces around the boundaries of the regions. The cell corner points which will form the polygon and the neighbor region ID are stored. Finally, the contours are simplified using Ramer-Douglas-Peucker algorithm [14]. The algorithm finds initial segments and locks vertices which are between two different regions, if the region is not connected, it locks two extreme vertices. The algorithm iterates through all simplified segments and subdivides the segment at the point with maximum distance error between the vertex and the segment. The initial vertices allow later to find common edges between the polygons. (See Fig. 10). The result is a set of simple polygons. (See Fig. 11).

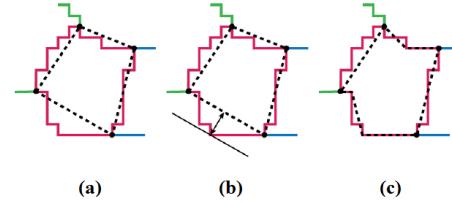


Fig. 10. Ramer-Douglas-Peucker algorithm: Initial vertices at region edges (a); Find vertex with maximum error, and subdivide (b); Iterate until certain error criteria is met (c).

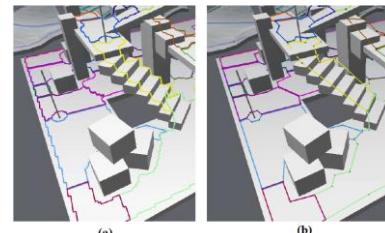


Fig. 11. Contours: Traced contours (a); Simplified contours (b).

5) *Triangulate the region polygons and build triangle connectivity*

Recast uses a modified algorithm from Computational Geometry in C for the triangulation of the polygons. The final step is to combine the triangles and find edge connectivity. The resulting polygons are finally converted to convex polygons which makes them perfect for path finding and spatial reasoning about the level. See Fig. 12.

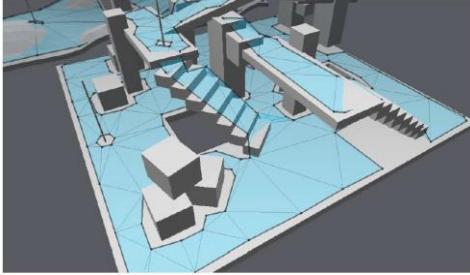


Fig. 12. Triangulation.

These five steps are illustrated in the Fig. 13. Recast is suitable for complex indoors and outdoors scenes with many levels. However, it generates walkable areas where the agents cannot even reach (see Fig. 14). This is due to Recast only taking into account the characteristics of the geometry enclosed by the voxels. It does not consider the connectivity between potentially walkable polygons. This is a problem because the resulting navigation mesh is considered as the initial graph located in the lowest level of the hierarchy. All of these unreachable regions are deleted during the preprocessing step in this approach.

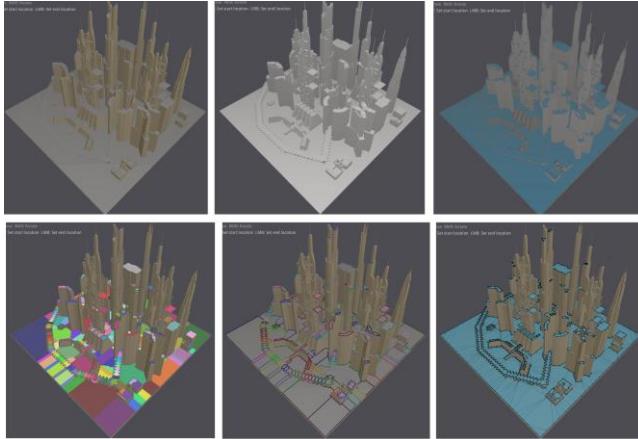


Fig. 13. Recast steps (from left to right): Input mesh and the five steps of Recast process. (Model: Scifi City (2090 polygons))

Once the spatial partition has been done by the Recast tool, the creation of the initial graph is performed. The method has mainly two steps:

- Hierarchical Subdivision. This is done in the preprocessing part. The hierarchy of graphs is created. The graph in the lowest level is given by the Triangulation in the Recast tool.
- Path finding computation. Given any level in the precomputed hierarchy, the path finding is calculated over the graph in that level.

B. Hierarchical Subdivision

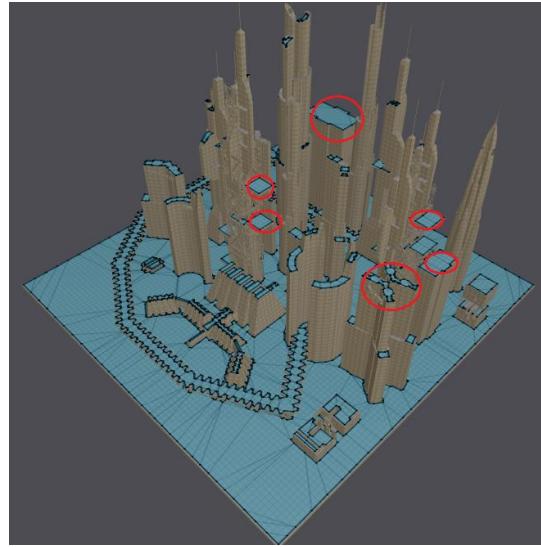


Fig. 14. Recast tool: Unreachable walkable areas. (Model: Scifi City (2090 polygons)).

The first step is to build the framework for hierarchical searches that is defined as a tree of graphs. The lowest graph of the hierarchy ($G_0 = (V_0, E_0)$) is computed by searching the polygons in the Recast triangulation. Each polygon becomes a new node of the graph. For each near polygon that shares the common border, an edge is created between them. See Fig. 15.

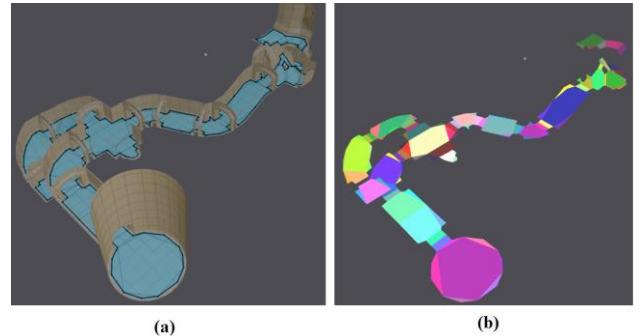


Fig. 15. Hierarchical Subdivision: The NavMesh triangulation of the model (a). The graph of the lowest level (level 0) (b). Nodes are painted in different colors. Edges connect a node with its neighbors (Model: Dungeon (120 polygons)).

Once the lowest level graph is created, the upper levels of the hierarchy are recursively built by partitioning each level until it reaches either the minimum number of the nodes in a graph or a certain threshold (maximum number of levels). The number of nodes which will be merged in each step is defined by the user.

In order to obtain an efficient subdivision of each graph, the k -way multilevel algorithm (MLkP) [15] is used to reduce the size of the graph by collapsing vertices and edges. This algorithm is faster than others multilevel recursive bisection algorithms. The process is described as follows:

First of all, a series of successively smaller graphs is derived from the input graph, this is called "coarsening

phase". Here, the size of the graph is successively decreased. Each graph is constructed from the previous graph by collapsing together a maximal size set of adjacent pairs of vertices. In order to have good partitions, the weight of a new vertex should be equal to the sum of its previous vertices. Also, the new edges are the union of the edges of its previous vertices to preserve the connectivity information in the coarser graph. The coarsening phase ends when the coarsest graph has a small number of vertices or if the reduction in the size of successively coarser graphs becomes too small.

After the coarsening phase, a k -way partitioning of the smallest graph is computed (initial partitioning phase). It is performed by using a multilevel bisection algorithm [15]. Each partition contains roughly $|V_0|/k$ vertex weight of the original graph. The division is done by Kernighan–Lin (KL) partitioning algorithm [16] which finds a partition of a node into two disjoint subsets of equal size, such that the sum of the weights of the edges between those subsets is minimized.

Finally, in the uncoarsening phase, the partitioning of the smallest graph is projected to the successively larger graphs by refining the partitioning at each intermediate level. It assigns the pairs of vertices that were collapsed together to the same partition as that of their corresponding collapsed vertex. After each projection step, the partitioning is refined using various heuristic methods to iteratively move vertices between partitions as long as such moves improve the quality of the partitioning solution. The uncoarsening phase ends when the partitioning solution has been projected all the way to the original graph.

The three phases of the multilevel paradigm are

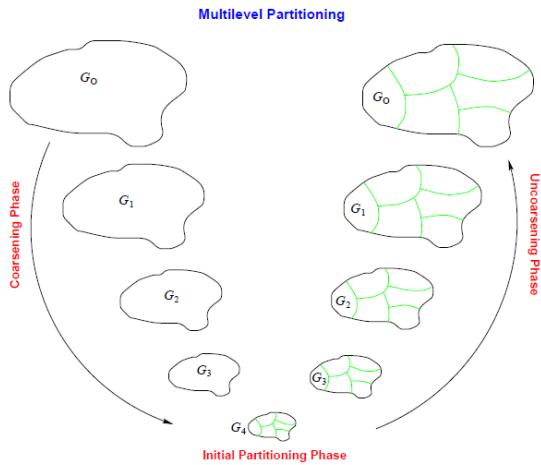


Fig. 16. The three phases of multilevel k -way graph partitioning. G_0 is the input graph, which is the finest graph. G_{i+1} is the next level coarser graph of G_i . G_4 is the coarsest graph.

illustrated in Fig. 16.

The procedure allows to have partitions which ensures high quality edge-cuts. An edge-cut of the partition is defined as the number of edges whose incident vertices belong to different partitions. All of this operations make

the algorithm more complex and hard to implement. An external implementation is used for this purpose. It is called METIS library³. It is a software package for partitioning unstructured graphs. It implements a collection of multilevel partitioning algorithms and is free only for educational and research purposes.

The algorithm 1 shows the steps of partitioning for each graph at each level in the hierarchy.

Algorithm 1 Hierarchical subdivision algorithm

```

1: procedure BUILDHIERARCHY(numMergedNodes,levels)
2:   numLevels = 0
3:   repeat
4:     numParts ← parentGraph.numNodes/numMergedNodes
5:     if (numParts < 2) then return
6:     partitions ← METISPartGraphKway(parentGraph.numNodes, numParts)
7:     partitions ← checkPartitions(partitions,parentGraph.numNodes, numParts)
8:     currentGraph ← buildGraph(partitions)
9:     parentGraph ← currentGraph
10:    numLevels ++
11:   until numLevels < levels

```

The iteration is done until either it reaches the maximum number of levels in the hierarchy (variable \$levels\$) or the graph cannot be subdivided. The number of merged nodes per level to create a new partition is defined by the variable *numMergedNodes*. The *PartGraphKway* function splits the parent graph into k parts using a multilevel k -way partitioning. The k parameter is given by the *numParts* variable. This function returns the partitions in which the parent graph has been divided and that will become in the new nodes of the current graph. These partitions need to be checked before being part of the new graph. It means that for each partition, its subnodes must be linked and must have edges. Otherwise the current partition will not be taken into account for the next iterations. The new graph is created in the *buildGraph* function. The algorithm 2 illustrates the steps required to build a graph for each level.

Once the partitions are established, the new nodes and edges between partitions are created. Each partition has a set of portals which depends of the number of edges. A portal is the middle point in a common edge between two partitions. So, for each pair of portals in the partition, an A* is calculated between them in order to get the cost and the shortest path. This is called an IntraEdge. Each partition has stored the subpath and cost for reaching from one portal to another. In the Fig. 17, the partitions, portal and intraedges are illustrated.

³ METIS has been developed at the Department of Computer Science and Engineering at the University of Minnesota and is freely distributed. Its source code can be downloaded directly from <http://www.cs.umn.edu/~metis>, and is also included in numerous software distributions for Unix-like operating systems such as Linux and FreeBSD.

Algorithm 2 Build Graph algorithm

```

procedure BUILDGRAPH(partitions)
2:   for  $i \leftarrow 1, partitions$  do
      currentGraph.AddNode( $i$ )
4:   for  $i \leftarrow 1, partitions$  do
      nodes  $\leftarrow$  findNodes( $i$ )
6:   for node  $\leftarrow 1, nodes$  do
      for neighbour  $\leftarrow 1, node.numEdges$  do
8:       if  $node.partition \neq neighbour.partition$ 
then
      currentGraph.AddEdge(node.parentNode,
neighbour.parentNode)
10:      portals.Add(node)
      for  $j \leftarrow 1, portals$  do
12:        for  $k \leftarrow 1, portals$  do
          if  $j \neq k$  then
14:            cost  $\leftarrow$  findPath( $j, k$ )
            currentGraph.AddIntraEdge(node,  $j, k$ ,
cost)
16:      portals.Clear()
  
```

An example of the hierarchical subdivision step is shown in the Fig 18.

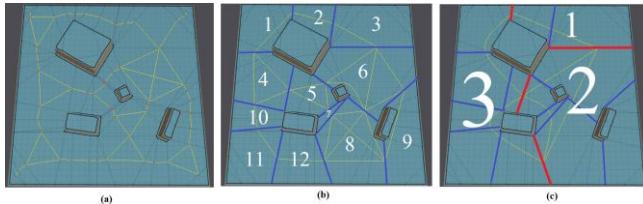


Fig. 17. Hierarchical Subdivision: (Simple map, $numMergedNodes = 5$, $levels = 5$). Portals are presented with red dots. IntraEdges are painted with yellow lines. Partitions are exposed with black, blue and red separation lines respectively. Level 0 = 76 nodes (a), Level 1 = 12 nodes (b), Level 2 = 3 nodes (c). (Model: Simple Map (76 polygons)).

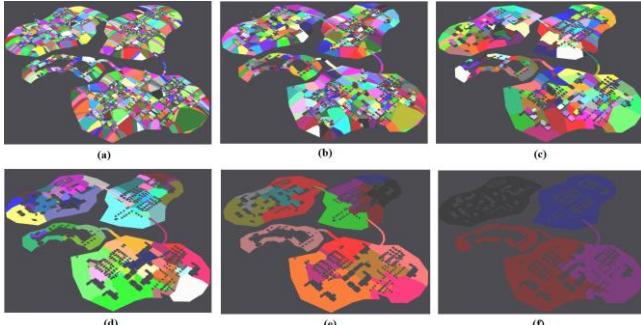


Fig. 18. Hierarchical Graphs: (City Islands, $$numMergedNodes = 3$, $levels = 10$$). Level 0 = 5151 nodes (a), Level 1 = 1469 nodes (b), Level 2 = 316 nodes (c), Level 3 = 72 nodes (d), Level 4 = 17 nodes(e), Level 5 = 4 nodes (f). (Model: City Islands (5515 polygons)).

C. Path finding Computation

Once the hierarchical subdivision is created, the path finding computation can be done at any level of the hierarchy. The second step consists in searching the shortest path from the start node (S) to the goal node (G) in a specific graph. This online search brings a performance improvement as any graph in the hierarchy is smaller than the one in level 0. (See Fig. 22).

The path finding computation has the following five phases:

1) Find S and G at a certain level

The first phase of this step gets S and G in a certain level in the hierarchy. This level is specified by the user in the Recast Tool. The algorithm receives an initial and end positions in the NavMesh environment. Then, S and G nodes are obtained in the graph at level 0 by searching for their positions. Finally, their parents are recursively searched by passing through all the levels in between until reach the desired level. If S and G nodes are in the same partition, a normal A* is run between them and the path finding is completed.

2) Connect S and G to the graph

To be able to search for paths in a graph at certain level, Start and Goal nodes have to be part of the graph. A temporal Start node is connected to each portal in the partition that contains it. Then, an A* is computed between S and the center of each portal in the partition. The path nodes and costs are stored for each portal. Finally, a new intraedge is added between the start node the portal inside the current partition. This step is repeated for G in its respective partition. (See Fig. 19).

For each search, S and G should change and the cost of inserting and deleting them is added to the total cost of finding a solution.

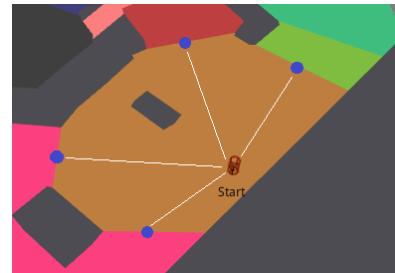


Fig. 19. Connect Start node to the graph: Blue circles are portals of the orange partition. White lines are the computed intraedges. Gray polygons are obstacles or no walkable areas.

3) Search for a path between S and G at the highest level

Once the S and G are temporally linked to the graph, an A* search is performed in the current graph. The time execution becomes faster because the number of nodes is significantly smaller than the graph at level 0.

4) Obtain optimal subpaths

The path planning computation gives all the partitions which are part of the optimal solution. For each of them, the path nodes are recursively got for each lower level until the lowest level is reached in the hierarchy. At the end, the full path is obtained to go from S to G at the level 0.

5) Delete temporal nodes

The nodes S , G and their intraedges are eliminated from the current graph.

The preprocessing step is shown in the Fig. 20, where a hierarchical subdivision has been applied on a map model.

The S and G positions are denoted in white letters. In this sample, the shortest path is found from S to G at level 2. The online step is shown in Fig. 21. The start and goal nodes are connected to their respective portals and run a common A* at level 2. Finally the sub paths are found until level 0 is reached.

The complete path from the start to the end position in the navigation mesh is achieved in a faster way than computing a

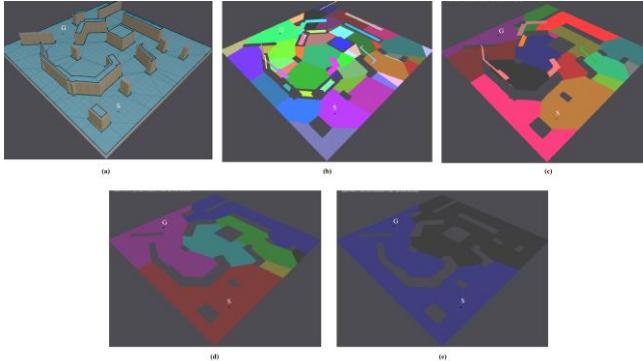


Fig. 20. Hierarchical Subdivision (From (a) to (e)). The start and goal nodes are written in white.

normal path finding at the level 0.

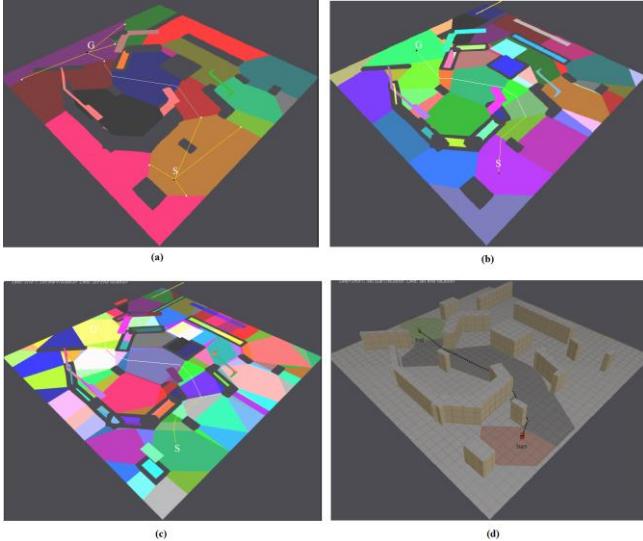


Fig. 21. Path finding Computation: S and G are linked to their partitions at level 2 (a). Sub paths are calculated until level 0 is reached. (Level 1 (b) and Level 0 (c)). The final result of the shortest path between S and G at level 2 (d).

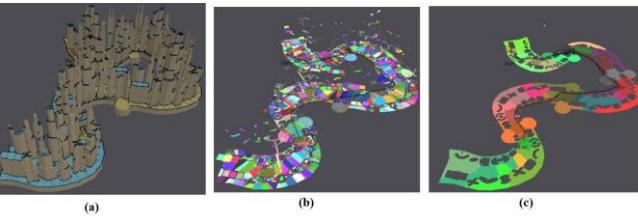


Fig. 22. Path finding Computation: (Serpentine Islands, $numMergedNodes = 4$, $levels = 10$) Path finding at level 0 (3908 nodes) (a). Path finding at level 0 where each node has a different color (b). Path finding at level 3 (28 nodes) (c).

IV. RESULTS AND DISCUSSION

The results were obtained in different models with a variety of sizes to measure the improvement achieved with the proposed method. The comparison is based on the speed time for calculating path finding between the start and goal node. Furthermore, the analysis is focused on how the time taken for each of the steps affects the total time. Also, the impact is explored by varying the number of merged nodes for the different levels on the performance of path finding with the suggested method. The performance results have been tested on an Intel® Core™ i7 processor with NVIDIA® GeForce® 610M graphics card and 8 GB of RAM.

A. Performance Test

The performance tests are based on: the number of nodes in each level of the hierarchy and the measure of the execution time (milliseconds) of path queries.

1) Number of Nodes

The number of resulting nodes is compared in each level in the hierarchy as the number of merged nodes is increased from one level to the next one. It is expected that the higher the level, the lower the number of nodes. As an example, the Fig. 23 shows the results obtained for the Sirus City. The chart shows number of nodes falls steadily over the upper levels in the hierarchy until either one partition cannot be divided any more or the maximum level has been reached.

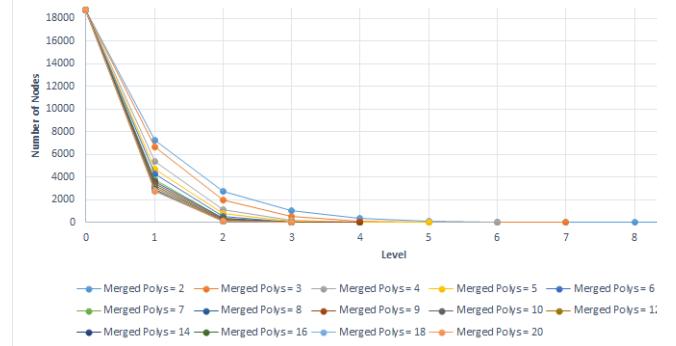


Fig. 23. Level vs Number of Nodes (Sirus City).

The division does not only depend on the `numMergedNodes` parameter but every time that a partition is created. This partition is checked whether it has connections (edges) with other partitions. If not, then this new nodes are not taken into account for the next level, as they cannot be any further merged. Thus, the consecutive subdivisions do not have an exact segmentation depending exclusively on with the `numMergedNodes` parameter.

2) Time Execution

Firstly, the total time of performing a path finding computation was analyzed at different levels against the execution time of performing path planning in level 0 (i.e without any hierarchy).

The sample model is a map with 2615 nodes in its initial graph. This line chart compares the total execution time per level in the hierarchy. Each line represents the number of merged polys from one level to the next one. To begin, the Fig. 24 shows the computation time over City Colony.

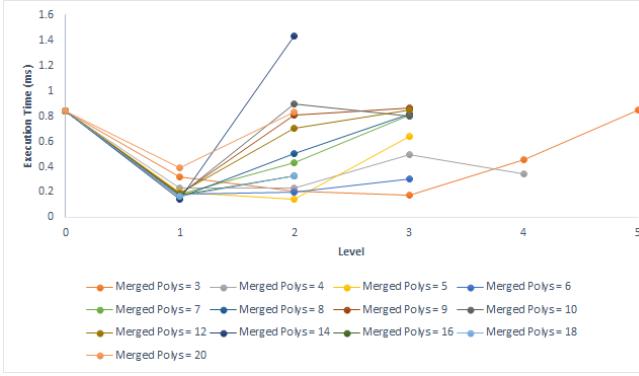


Fig. 24. Level vs Execution Time (City Colony).

It can be seen clearly that the execution time dramatically falls in the first level of the hierarchy. Afterwards, there is a slowly improvement depending on the number of merged nodes 2, 4, and 5 at level two. The other cases have a gentle upward trend but still lower than the computation time at level 0. The highest value reached is 1.43ms with $\text{numMergedNodes}=14$ at level two. In contrast, the fastest time is 0.142ms (six time faster) for the case of 5 merged nodes in level 2. (See Fig. 25).

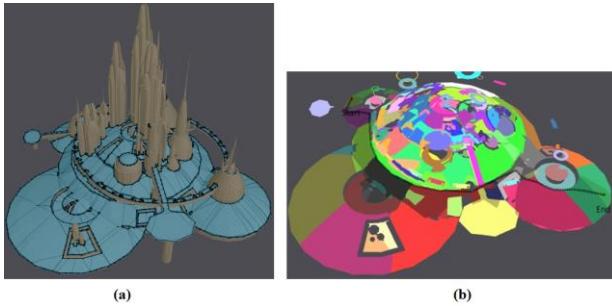


Fig. 25. City Colony model (a). Shortest path ($\text{numMergedNodes} = 5$, $\text{level} = 2$) (b)

To better understand where the bottlenecks of our algorithm appear, the partial times are compared for computing the entire online step for each level in the hierarchy. The online search has been divided in five steps in order to analyze the partial times in this process. See algorithm 3.

3) Find S and G at certain level

The Fig. 26 illustrates the time of getting the partitions to which Start and Goal nodes respectively belong to a certain level. Those partitions are obtained by recursively searching S and G positions in the upper levels of the hierarchy until they reach a predetermined level. Overall, the time has a gradual rise in the entire hierarchy. This was expected as the higher the level we want to reach, the more time is consumed. Notice that the total time for this step is

Algorithm 3 Online Search Pseudo-code

```

procedure ONLINESEARCH(startNodeId, startNodePos,
endNodeId, endNodePos, level)
3:   Find  $S$  and  $G$  at certain level:
       $sNode \leftarrow \text{getNode}(startNodeId)$ 
       $gNode \leftarrow \text{getNode}(endNodeId)$ 
6:   if  $\text{level} == 0$  then
      path  $\leftarrow \text{findPath}(startNodeId, startNodePos,$ 
       $endNodeId, endNodePos, 0)$ 
      return path
9:   Connect  $S$  and  $G$  to the graph:
      linkStartToGraph( $sNode$ )
      linkGoalToGraph( $gNode$ )
12:  Search for a path between  $S$  and  $G$  at the highest
      level:
      tempPath
      findPath( $sNode.id, startNodePos, gNode.id, endNodePos,$ 
      level)
      Obtain optimal subpaths:
15:  for all  $\text{subpath} \in \text{tempPath}$  do
      path  $\leftarrow \text{getSubpath}(\text{subpath}, \text{level} - 1)$ 
      Delete  $S$  and  $G$ :
      deleteNode( $sNode$ )
      deleteNode( $eNode$ )
      return path
    
```

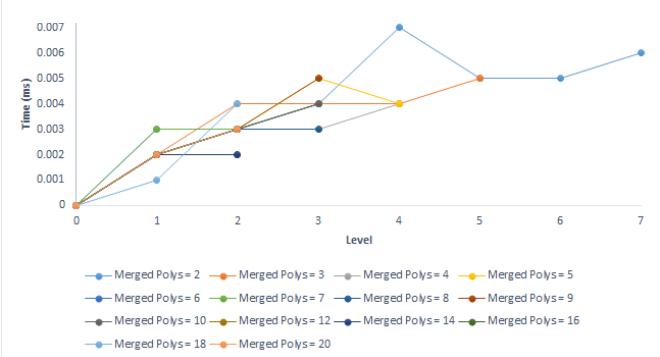


Fig. 26. Get S and G Time vs Level

not significant in the total time (values less than 0.005 ms).

4) Connect S and G to the graph

The chart shows the time of connecting S and G to each of the partitions in their respective partitions. S and G are linked by performing an A^* from those nodes to each portal in the current partition. Then, the path nodes and

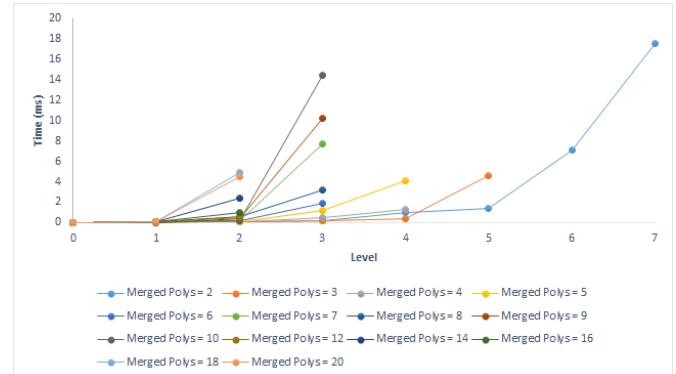


Fig. 27. Link S and G Time vs Level

cost are stored by adding an intraedge.

The figure shows an upward trend throughout the levels of the hierarchy. It exposes a particular strong growth in the highest levels due to the number of portals being bigger in the upper levels. (See Fig. 27). This particular behavior is due to the fact that we have to execute as many A* searches as the number of portals the partition has.

5) Search for a path between S and G at the highest level

Regarding the path finding calculation, an A* computation is faster when the searching is done in a higher level in the hierarchy. A regular A* is performed between the start and goal nodes at certain level. The Fig. 28 shows the general gradual decline for all the cases. The lower the number of nodes, the faster the exploration is. The number of nodes in a specific level highly depends on how many nodes were merged in its lower level. For instance, for the case when *mergedPolys* = 2 and *level*=10, the number of nodes was 22. The partial time was 1.333ms. When *mergedPolys* = 10 and *level* = 3, the

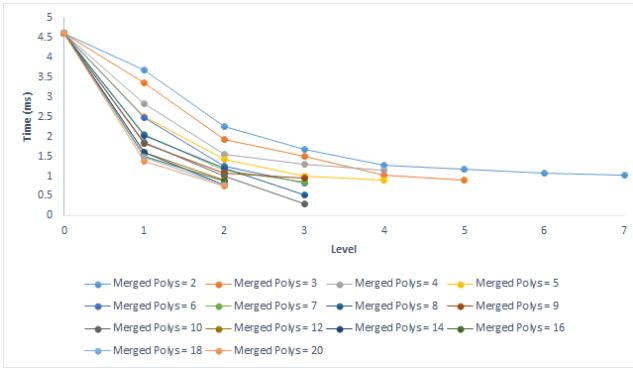


Fig. 28. A* Time vs Level

number of nodes was 11 with a partial time 0.287ms.

6) Obtain optimal subpaths

The chart 30 shows the time of getting the subpaths for each level. The subpaths are obtained by recursively get the stored paths in each nodes of the lower levels until we reach the level 0. Those subpaths have the nodes which become the optimal path at level 0.

As an overall trend, the time of getting subpaths increased fairly slowly until the penultimate level of the hierarchy. Then, this time gradual decline in the highest level. This is due to there are no nodes between S and k. Also, there are not intermediate paths between them. Fig.

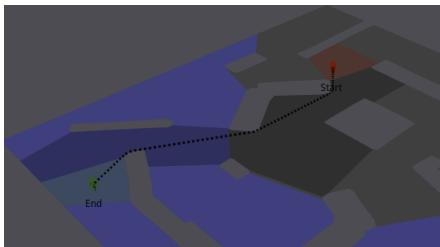


Fig. 29. Obtain optimal subpaths. The map model has two nodes at level 3. No intermediate paths are calculated.

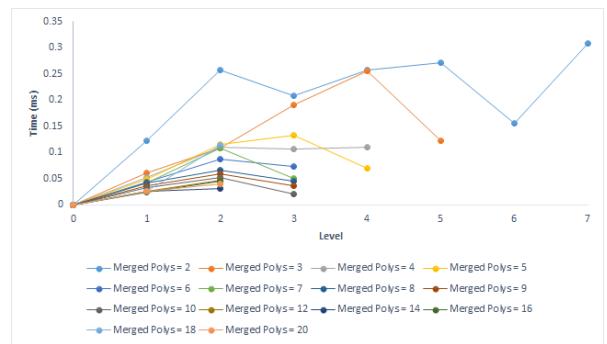


Fig. 30. Time vs Level.

29 illustrates this scenario.

Therefore, the execution time is really low and strictly depends on the map environment and the number of merged nodes.

7) Delete S and G

Deleting temporal S and G nodes have an insignificant time compare to other partial times during the online process.

V. CONCLUSION

With the booming growth of video games, there is a great demand on path finding algorithms. In this method, a new hierarchical path finding framework is presented to speed up crowd simulation for large 3D environments. The approach has two steps: Preprocessing and online search. Preprocessing step builds the hierarchy of levels whereas that online process deals with the path finding search. The approach has a tree hierarchy of graphs where the searching can be performed at any level.

The main contributions of this approach are:

- A path planning algorithm for arbitrary graph types that could be applied in any kind of 3D world representation.
- A recursive partition of a graph based on reducing the connector edges.
- A hierarchy of graphs to find the fastest time execution for path planning.

The evaluation has shown a significant improvement in path finding time execution. The method has better results when the path planning is performed in big world representations (5 or 6 times faster than A*). For small models, a common A* is enough. The trade-off between the chosen level and the size of partitions is important. Also, the approach shows better performance in non-widespread environments.

The framework presented in this research was inspired by HPA* approach but it also provide multi-level search and present a new algorithm that works over any kind of environment division.

VI. FUTURE WORK

Despite the improvements, there is still a large amount of work that could be done to obtain either fast path finding

searches or good path quality. Some of the enhancements that could be incorporated with the current framework. For instance, the improvement of linking Start and Goal node to the current graph. This is an important issue to address in the future. A way of reducing link-time would be to replace A* with a version of Dijkstra's algorithm that does not flush the pool of visited vertices between searches. Also, those nodes could be somehow stored in order to reduce the time execution of connecting and deleting.

It would also be interesting to study if some steps of the online search could be parallelized using a GPU implementation. For example, the process of linking S and G could be performed in separated threads for each portal as well as getting the subpath for each partition. Also, this approach could be extended to work under dynamic environments where the replanning could be done only at certain level of the hierarchy.

ACKNOWLEDGMENT

The research reported in this document/presentation was based on the master thesis of the author. The research, views and conclusions contained in this document are those of the author and his advisor. The author would like to thank to Nuria Pelechano Gomez for her tutoring during his studies and research in the Polytechnic University of Catalonia.

- [13] Roerdink, Jos B.T.M. and Meijster, Arnold., *The Watershed Transform: Definitions, Algorithms and Parallelization Strategies.*, Fundam. Inf. Forum, vol. 22, pp. 303 - 312, Sep 2003.
- [14] David P. Luebke., *A Developer's Survey of Polygonal Simplification Algorithms.*, IEEE Computer Graphics and Applications., vol. 21, pp. 24 - 35, May 2001.
- [15] G. Karypis and Vipin Kumar., *Multilevel k-way Partitioning Scheme for Irregular Graphs.*, Journal of Parallel and Distributed Computing., vol. 48, pp. 96 - 129, Jan 1998.
- [16] Kernighan, B.W. and Lin, S., *An Efficient Heuristic Procedure for Partitioning Graphs*, The Bell Systems Technical Journal, vol. 49, 1970.
- [17] Adi Botea and Martin Müller and Jonathan Schaeffer, *Near optimal hierarchical path-finding.*, Journal of Game Development., vol. 1, pp. 7 - 28, 2004.
- [18] Joseph O'Rourke, *Computational Geometry in C.*, 2nd ed, Computational Intelligence and Games, 2008. CIG '08. IEEE Symposium On, pp. 258 - 265, Dec. 2008.
- [19] Daniel Harabor and Adi Botea., *Hierarchical path planning for multi-size agents in heterogeneous environments*, CIG, 2008.



Carlos Fuentes received the Master degree in Computer Graphics and Virtual Reality from Polytechnic University of Catalonia, Barcelona in 2014. From 2013 to 2014, he was a research assistant with the Moving Research Group in UPC. His research interest includes crowd simulation and autonomous behavior. He is currently working in ThoughtWorks Ecuador.

REFERENCES

- [1] Loscos, Céline and Marchal, David and Meyer, Alexandre, *Intuitive Crowd Behavior in Dense Urban Environments using Local Laws.*, Theory and Practice of Computer Graphics, 2003. Proceedings, pp. 122 - 129. Jun 2003.
- [2] F. Tecchia, C. Loscos, Y. Chrysanthou, *Visualizing Crowds in Real-Time*, Computer Graphics forum, pp. 753 - 765, Dec 2002.
- [3] Franco Tecchia and Céline Loscos and Ruth Conroy and Yiorgos Chrysanthou, *Agent Behaviour Simulator (ABS): A Platform for Urban Behaviour Development*, In GTEC'2001, pp. 17-21, 2001.
- [4] Lydia Kavraki and Petr Svestka and Jean-claude Latombe and Mark Overmars, *Probabilistic Roadmaps for Path Planning in High-Dimensional Configuration Spaces*, Robotics and Automation, vol. 12, pp. 566 - 580, Aug 1996.
- [5] Nieuwenhuisen, D.; Kamphuis, A.; Mooijekind, M.; Overmars, M.H., *Creating Small Roadmaps for Solving Motion Planning Problems*, IEEE International Conference on Methods and Models in Automation and Robotics, pp. 531-536, 2005.
- [6] R. Geraerts and M.H. Overmars, *Automatic Construction of High Quality Roadmaps for Path Planning*, Utrecht University: Information and Computing Sciences, 2004.
- [7] M. Mononen. (2015, April 8), *Navigation-mesh Toolset for games*, GitHub Recast and Detour, 2014. Available: <https://github.com/memononen/recastnavigation>.
- [8] P. E. Hart, N. J. Nilsson, and B. Raphael, *A formal basis for the heuristic determination of minimum cost paths*, IEEE Transactions on Systems, Science, and Cybernetics, vol. 4, pp. 100 - 107, July 1968.
- [9] Maxim Likhachev and Geoffrey J. Gordon and Sebastian Thrun., *ARA*: Anytime A* with Provable Bounds on Sub-Optimality*, Advances in Neural Information Processing Systems 16, 2004.
- [10] Koenig, Sven and Likhachev, Maxim. *D*Lite*, Eighteenth National Conference on Artificial Intelligence, pp. 476 - 483, 2002.
- [11] Maxim Likhachev, David Ferguson , Geoffrey Gordon, Anthony (Tony) Stentz, and Sebastian Thrun., *Anytime Dynamic A*: An Anytime, Replanning Algorithm*, Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS), Jun 2005.
- [12] Haumont, D. and Debeir, Olivier and Sillion, François X., *Volumetric Cell-and-Portal Generation*, Comput. Graph. Forum, 2003.

Realidad Virtual Acústica: El Abordaje de las Redes Neuronales Artificiales

Acoustic Virtual Reality: The Artificial Neural Networks Approach

José F. Lucio Naranjo, Roberto A. Tenenbaum y Julio C. B. Torres

◦

Resumen— En este trabajo se presenta un nuevo abordaje para obtener las respuestas impulsivas biauriculares (BIRs) para un sistema de aurilización utilizando un conjunto de redes neuronales artificiales (RNAs). El método propuesto es capaz de reconstruir las respuestas impulsivas asociadas a la cabeza humana (HRIRs) por medio de modificación espectral y de interpolación espacial. Para poder cubrir todo el espacio auditivo de recepción, sin aumentar la complejidad de la arquitectura de la red, una estructura con varias RNAs (conjunto) fue adoptada, donde cada red opera en una región específica del espacio (gomo). El error de modelaje en el dominio de la frecuencia es investigado considerando la naturaleza logarítmica de la audición humana. A través de la metodología propuesta se obtuvo un ahorro del tiempo de procesamiento computacional de aproximadamente 62% en relación al método tradicional de procesamiento de señales utilizado para aurilización. La aplicabilidad del nuevo método en sistemas de aurilización es reforzada mediante un análisis comparativo de los resultados, que incluyen la generación de las BIRs y el cálculo de un parámetro acústico biauricular (IACF), los cuales muestran errores con magnitudes reducidas.

Palabras clave— Realidad Virtual Acústica. Aurilización. Redes Neuronales Artificiales. Respuestas Impulsivas Biauriculares. Simulación Numérica de Acústica de salas.

Abstract— This work presents a new approach to obtain the Binaural Impulse Responses (BIRs) for an auralization system by using a committee of artificial neuronal networks (ANNs). The proposed method is capable to reconstruct the desired modified Head Related Impulse Responses (HRIRs) by means of spectral modification and spatial interpolation. In order to cover the entire auditory reception space, without increasing the network's architecture complexity, a structure with multiple RNAs

This paper was submitted on the LAJC on February 2015

José F. Lucio Naranjo, Departamento de Informática y Ciencias de la Computación, Facultad de Ingeniería de Sistemas, Escuela Politécnica Nacional, Quito, Ecuador; (email: jose.lucio@epn.edu.ec)

Roberto A. Tenenbaum, Laboratório de Instrumentação en Dinâmica, Acústica y Vibraciones – LIDAV, Departamento de Modelagem Computacional, Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, Brasil; (email: ratanenbaum@gmail.com)

Julio C. B. Torres, Departamento de Expressão Gráfica, Escola Politécnica, Universidade Federal de Rio de Janeiro, Rio de Janeiro, Brasil. (email: julio@poli.ufrj.br)

(committee) was adopted, where each network operates in la specific reception region (bud). The modeling error, in the frequency domain, is investigated considering the logarithmic nature of the human hearing. It was observed that the proposed methodology obtained a computational gain of approximately 62%, in terms of processing time reduction, compared to the classical signal processing method used to obtain auralizations. The applicability of the new method in auralization systems is validated by comparative analysis of the results, which includes the BIR's generation and calculation of one binaural acoustic parameter (IACF), showing very low magnitude errors.

Index Terms— Acoustic Virtual Reality. Auralization. Artificial Neural Networks. Binaural Impulse Responses. Room Acoustics Numeric Simulation.

I. INTRODUCCIÓN

EL reciente desarrollo de los sistemas computacionales ha revolucionado la vida del ser humano. Inúmeras aplicaciones informáticas han sido introducidas en las más diversas actividades humanas, que van desde soluciones particulares para problemas específicos, hasta herramientas generales de apoyo en la vida cotidiana de las personas. Entre todas estas, los sistemas de realidad virtual han recibido especial atención los últimos años, por el hecho de ser considerados verdaderas interfaces entre el ser humano y los computadores. La realidad virtual procura recrear la percepción sensorial humana y sus aplicaciones pueden darse en representaciones de recintos, simuladores de vuelo, juegos electrónicos, simuladores en parques de diversiones, entre muchas otras.

En el caso particular de la realidad virtual acústica, el objetivo es la aurilización. Ella consiste en sintetizar un sonido, a partir de datos medidos y/o simulados, capaz de estimular al oyente de tal forma que cause la sensación de inmersión acústica en un ambiente (simulado o real). Tales ambientes pueden ser desde una simple sala de aula, un gran teatro, una planta industrial o hasta un espacio urbano. Esa señal sonora debe ser reproducida a un ser humano en un ambiente libre de reflexiones sonoras como, por ejemplo, en una cámara anecóica (utilizando filtros de atenuación de diafonía), o a través de audífonos ecualizados. De esa forma, puede sintetizarse la sensación de espacialidad del sonido

pudiendo simular la habilidad humana de distinguir la posición de la fuente sonora dentro del ambiente virtual [1].

La llave de la aurilización es la respuesta impulsiva biauricular (BIR – del inglés, Binaural Impulse Response) para un determinado par fuente / cabeza humana virtual. Computacionalmente, la BIR es generada a partir de un programa de simulación, el cual, al finalizar la fase de cálculo del modelo del ambiente, tendrá almacenadas información (energética, direccional y tiempo de llegada) de todos los frentes de onda que alcanzaron el receptor. El espectro energético de cada una de las frentes de onda incidentes es utilizado para modificar el espectro de un par de respuestas impulsivas asociadas a la cabeza humana (HRIR – del inglés, Head Related Impulse Response), una función para cada oído. El par de HRIRs es escogido de una base de datos, considerando la dirección de incidencia del frente de onda. Sin embargo, debido a la naturaleza discreta de tales bases de datos, es altamente improbable que el frente de onda golpee al receptor en la dirección donde existe una función medida, siendo necesario un procedimiento de interpolación para evitar efectos acústicos no deseados, tales como cliques u otras alteraciones sutiles [2].

Después es necesario adicionar cada una de las HRIRs modificadas en el canal correspondiente (izquierdo y derecho) insertando un atraso correspondiente al tiempo de llegada del frente de onda al receptor. Ese proceso (interpolación de HRIRs, modificación con el espectro de la frente de onda y sumatorio con atrasos para montaje de la BIR) es repetido para cada uno de los frentes de onda que alcanzan el receptor. Tales iteraciones constituyen el llamado *método tradicional* (MT) para producir la BIR, el cual, a pesar de los diversos algoritmos de procesamiento de señales disponibles, requieren de costo computacional de procesamiento relativamente alto. EL presente trabajo presenta una nova abordaje para la generación eficiente de las BIRs aplicadas en procesos de aurilización en sistemas de simulación computacional acústica. La metodología consigue inclusive reducir el costo computacional (i.e. tiempo de procesamiento) para generar HRIRs modificadas, minimizando, al mismo tiempo, cualquier error en el modelaje, de manera que este sea imperceptible al ser humano. En este abordaje, son utilizadas redes neuronales artificiales (RNAs). La ventaja de usar estos sistemas está, principalmente, en la capacidad de las RNAs de adquirir, almacenar y utilizar información experimental con base en el funcionamiento de las neuronas biológicas y en la estructura de procesamiento masivamente paralela del cerebro humano.

Esa estructura consigue que un sistema de RNAs, con una arquitectura correctamente escogida y debidamente entrenada, responda con alta rapidez y precisión, siendo, por tanto, una óptima alternativa para el procesamiento de sistemas de realidad virtual acústica. Una red puede almacenar información relevante de las HRIRs (eliminando la necesidad de bancos de datos e interpolaciones) y, por otro lado, su rapidez de respuesta la credencia como una alternativa eficiente frente al método tradicional de procesamiento de

señales, que involucra transformadas de Fourier y otras multiplicaciones complejas.

II. DESCRIPCIÓN DEL MODELO DE LAS RNAs

Según Tebelskis [3], una red neuronal artificial consiste en un número potencialmente alto de unidades de procesamiento simple, llamadas *neuronas artificiales*, las cuales, mediante una red de pesos excitadores o inhibidores (pesos sinápticos), influencian el comportamiento de otras neuronas.

Debido a sus propiedades, las RNAs han sido aplicadas en una amplia variedad de campos. Algunos de los más recurrentes incluyen aplicaciones en interpolación de funciones [4], reconocimiento de patrones [5], predicciones de comportamiento [6] etc. En lo que se refiere a aplicaciones en acústica, las RNAs han sido utilizadas en reconocimiento del habla [3][7] y apenas en los últimos años comenzaron ser utilizadas para la extracción de información relevante de las HRTFs, [8] y para personalización de HRTFs de acuerdo con medidas antropomórficas humanas [9].

Con la debida preparación, las RNAs pueden aprender a generar las salidas/respuestas de un sistema cualquiera (lineal o no). Para eso, un conjunto de datos de entrada y sus correspondientes salidas (del sistema que actúa como *profesor*, o sea, que entrena la red) son agrupados en *pares de entrenamiento*. Tales pares son presentados a la red durante un proceso llamado de *aprendizaje supervisado con retropropagación* [10].

En este trabajo, después del proceso de entrenamiento, la metodología propuesta fue capaz de reconstruir las HRIRs modificadas, las cuales eran originalmente generadas por medio de modificación espectral y de interpolación espacial. Con el objetivo de cubrir todo el espacio auditivo de recepción, sin aumentar la complejidad de la arquitectura de la red, una estructura con múltiples RNAs fue adoptada, donde cada red opera en una región específica del espacio (gomo), como muestra la Fig. 1.

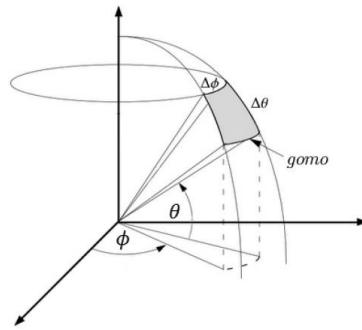


Fig. 1. Área de operación de la red (gomo) limitada en azimut por ϕ y $\phi + \Delta\phi$ y en elevación por θ y $\theta + \Delta\theta$.

Los tres principales factores que influencian en la precisión del modelo — arquitectura de la red, ángulos de abertura del área de recepción y atrasos de las HRIRs — fueron

investigados y una configuración óptima fue determinada [11]. Al final, la arquitectura utilizada en cada una de las redes es mostrada en la Fig. 2. El área de recepción alrededor de la cabeza fue dividida en 1898 gomos y los atrasos de las funciones fueron preservados.

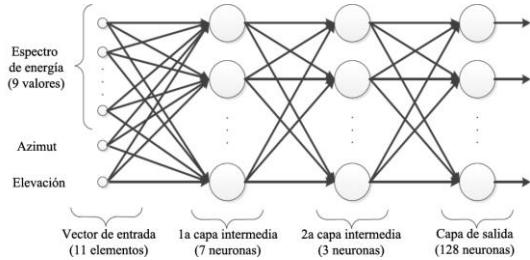


Fig. 2. Arquitectura básica utilizada por las redes neuronales artificiales con dos capas intermedias.

La entrada de cada RNA está dada por 11 elementos. Dos de ellos representan la dirección de incidencia de las frentes de onda, dada por los ángulos de azimut y elevación. Los nueve valores restantes representan el espectro de energía por banda de octava (63 Hz – 16 kHz). La salida está compuesta por 128 valores, cada uno representando una muestra temporal de las HRIR modificadas. Tales muestras permiten almacenar la información significativa de las funciones [12]. La arquitectura escogida, mostrada en la Fig. 2, tiene dos capas intermedias, la primera con siete neuronas y la segunda con tres neuronas. Esta alternativa presentó el mejor balance entre precisión y costo computacional. Cabe señalar que fue obtenida una ganancia computacional (reducción del tiempo de procesamiento) de, aproximadamente, 62% en relación al método tradicional de procesamiento de señales utilizado para aurilización.

III. RESULTADOS NUMÉRICOS

El escenario escogido para la validación de los resultados fue la sala utilizada en la comparación internacional de simuladores acústicos Round Robin 3 [13][14]. La mencionada sala fue reconstruida en el código computacional RAIOS (Room Acoustics Integrated and Optimized Software), tal como muestra la Fig 3. El icosaedro rojo representa una fuente sonora omnidireccional activa, el icosaedro gris es una fuente sonora inactiva y las tres esferas verdes representan receptores o cabezas artificiales.

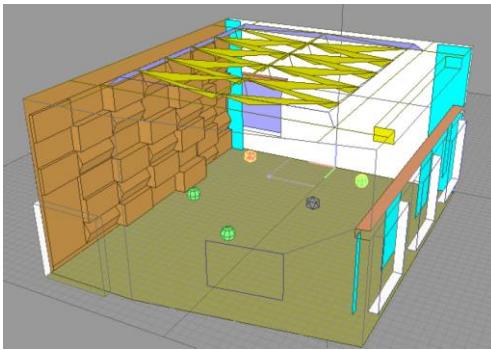


Fig. 3. Estudio de grabación de música en el PTB simulado en el código computacional RAIOS.

En la Fig. 4 pode-se observar la orientación de un de los receptores. EL receptor en cuestión está orientado de manera que el oído derecho está expuesto al sonido directo oriundo de la fuente.

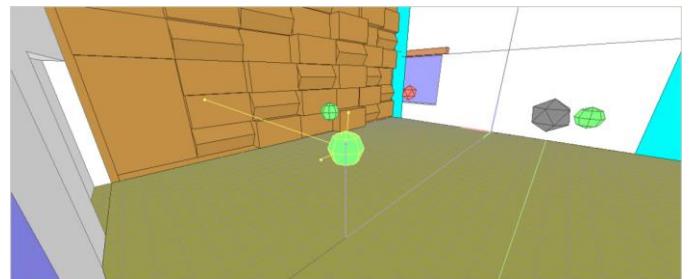


Fig. 4. Orientación de los receptores bi-auriculares.

En primera instancia, la validación fue realizada tomando como referencia los rayos acústicos captados en el receptor durante la simulación acústica, los cuales tuvieron los sus espectros modificados naturalmente pelas características de la propia sala. Con tal información fueron generados resultados con los dos abordajes, el método tradicional (MT) y el método de las redes neuronales artificiales (RNA), relativos al oído derecho.

En la Fig. 5, puede apreciarse el espectro energético de un rayo cuya dirección incidencia es dada pelos ángulos $\phi = 352^\circ$, 9° y $\theta = -15^\circ$, 8° .

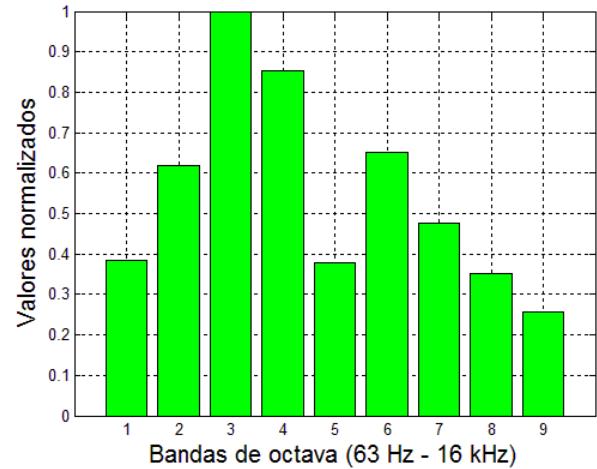


Fig. 5. Espectro de energía del rayo con dirección de llegada ($\phi = 352^\circ$, 9° , $\theta = -15^\circ$, 8°).

Puede observarse, en la Fig. 6a, que la función generada por el método RNA, en el dominio del tiempo, se adapta a la función generada con el MT con pequeños desvíos indicados en los gráficos de error. En las Figs. 6b y 6d, las diferencias entre las curvas de los espectros de amplitud y fase, obtenidas vía RNAs y el método tradicional, se tornan relevantes solamente a partir de los 7 kHz. Los errores más elevados

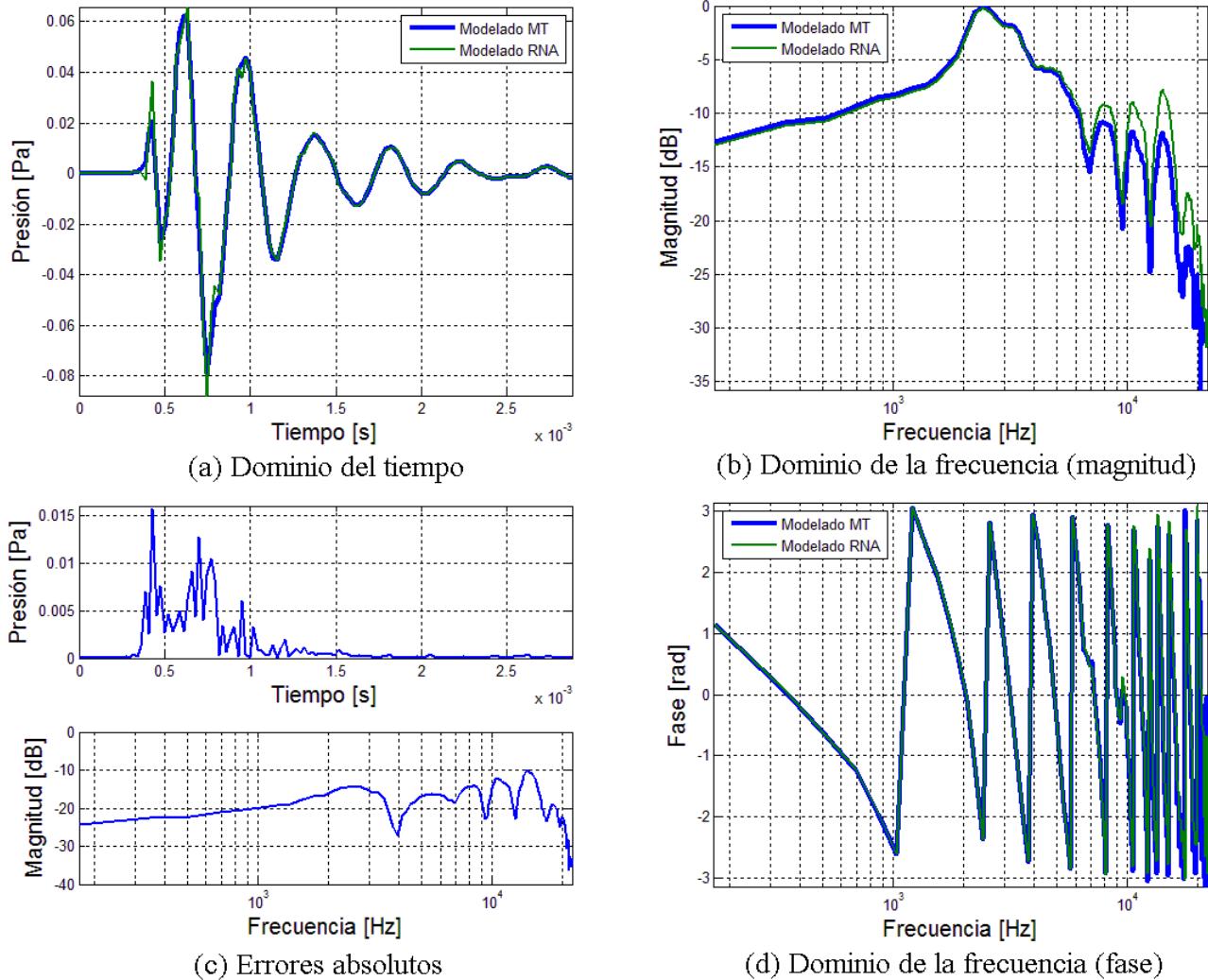


Fig. 6. (a,b,c,d). Resultados generados pelas abordajes RNA y MT, correspondientes al oído derecho, obtenidos para un rayo con dirección dada pelos parámetros $\theta = -15,8^\circ$ y $\phi = 352, 9^\circ$. En (c), se indican los errores absolutos.

fueron identificados, principalmente en altas frecuencias (encima de 10 kHz), como muestra la Fig. 6. Sin embargo, la mayoría de la energía de las HRIRs modificadas está concentrada en la banda de frecuencia entre 2 y 6 kHz, en la cual el modelo presentó desvíos inferiores la -15 dB. El motivo del desvío está relacionado con la capacidad de generalización de la RNA, la cual, aunque permite generar resultados para entradas que no participaron del proceso de entrenamiento, causa que pequeñas variaciones en las funciones sean desconsideradas, o sea, justamente en las componentes de alta frecuencia.

En una segunda fase de la validación, fue generada la BIR del receptor considerado usando el abordaje de las RNAs. Para tener un resultado comparativo, la BIR también fue generada utilizando el MT. La Fig. 7 presenta los resultados obtenidos. Naturalmente, respuestas impulsivas así presentadas parecen ser gráficamente idénticas.

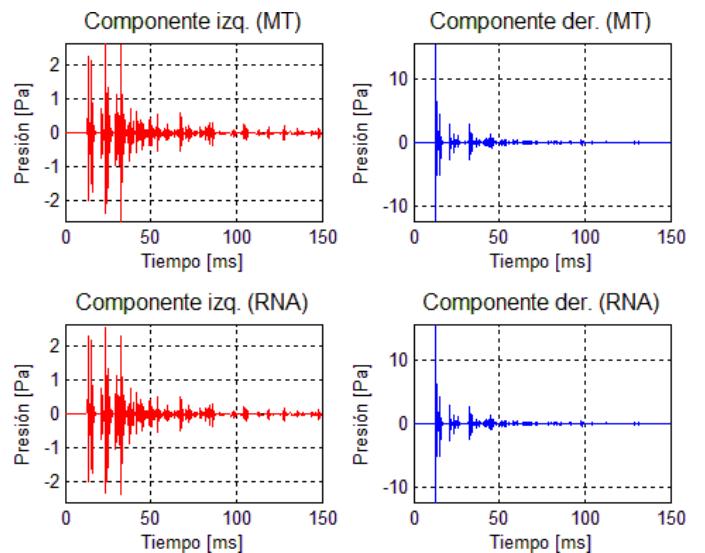


Fig. 7. BIRs generadas con el MT y la técnica de las RNAs.

En la tercera es última validación, fue realizado un test que consideró la correlación existente entre los señales de los dos oídos para poder caracterizar mejor las diferencias entre las BIRs generadas por cada método. Si las funciones de correlación entre los oídos izquierdo y derecho fueran semejantes para ambos métodos, entonces el método RNA habrá preservado las características espaciales y biauriculares, conforme el método tradicional.

En ese sentido fue utilizado el parámetro de calidad acústica bi-auricular conocido como función de correlación cruzada interauricular (IACF - del inglés Inter-Aural Cross-correlation Function)[15]. Los resultados obtenidos son presentados en la Fig. 8.

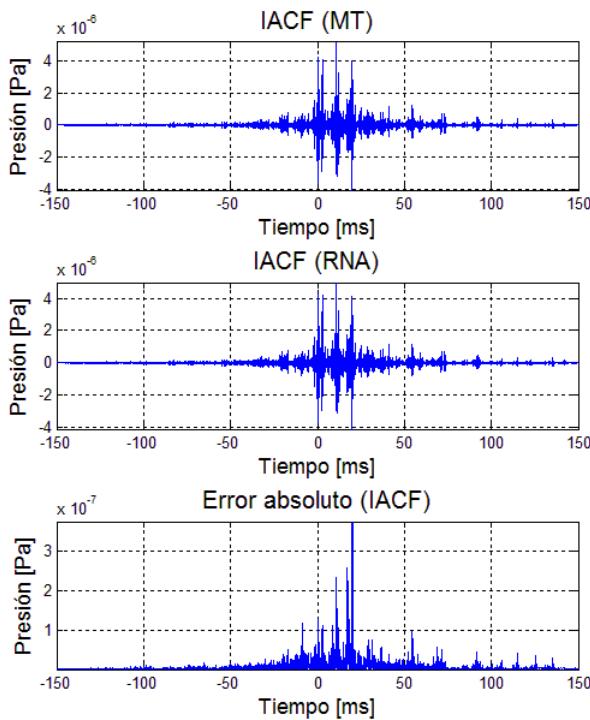


Fig. 8. IACF generadas con las BIRs obtenidas con MT (superior) y la técnica RNA (medio) y el error absoluto entre ellas (inferior).

EL gráfico superior muestra la función generada con la BIR obtenida con el MT, el gráfico intermedio presenta los resultados obtenidos con la técnica RNA y el gráfico inferior presenta el error absoluto entre el IACFs del MT y del RNA. Tales resultados están en concordancia con la orientación que favorece el oído derecho, tanto con el MT como con las RNAs. Los máximos valores alcanzados en lo que se refiere al error absoluto para el IACF tienen una orden de grandeza de 10^{-7} .

IV. CONCLUSIONES

Fue presentado un nuevo abordaje para la generación eficiente de las respuestas impulsivas biauriculares (BIRs). En ese intuito, un conjunto de redes neuronales artificiales (RNAs) fue utilizado para substituir procesos de interpolación

y la modificación espectral aplicados a las respuestas impulsivas asociados a la cabeza humana (HRIRs).

Los resultados numéricos mostraron que el procesamiento de señales tradicional puede ser reemplazado por un conjunto de RNAs cuya salida presenta errores próximos a -17 dB en relación al método tradicional. Tanto los errores absolutos entre las funciones, como los resultados comparativos utilizando el parámetro de calidad acústica biauricular IACF, muestran que un conjunto de redes, cada una con una arquitectura reducida (apenas dos capas intermedias con 7 y 3 neuronas respectivamente), es capaz de reemplazar al método tradicional, computacionalmente desventajoso.

Los resultados obtenidos a partir de rayos acústicos generados pelo simulador acústico RAIOS, los cuales no fueron considerados durante entrenamiento, muestran que la generalización de la red funciona con errores de la orden de -15 dB. Sin embargo, el precio de una buena capacidad de generalización es la concentración de los desvíos en altas frecuencias.

En la práctica, las BIRs generadas con el método tradicional o con las RNAs no presentan diferencias significativas. El parámetro biauricular IACF presentó resultados que estaban en concordancia con los parámetros escogidos para la simulación. Más aún, el error absoluto calculado entre los IACFs generados pela técnica clásica y la metodología propuesta alcanzan valores de la orden de 10^{-7} . Eso confirma numéricamente la precisión del modelo propuesto.

REFERENCES

- [1] BLAUERT, J., 1997. Spatial Hearing. The MIT Press, Cambridge.
- [2] WERSNYI, G., 2009. Effect of emulated head-tracking for reducing localization errors in virtual audio simulation IEEE Transactions On Audio, Speech, And Language Processing, Vol. 17, n. 2, pp. 247–252.
- [3] TEBELSKIS, J., 1995. Speech Recognition using Neuronal Networks. Ph.D. Thesis - Carnegie Mellon University, Pittsburgh.
- [4] MINGO, L. F.; GIMÉNEZ, V.; CASTELLANOS, J., 1999 Interpolation of boolean functions with enhanced neuronal networks. In: Second Conference on Computer Science and Information Technologies. CSIT'99. Yerevan, Armenia. p. 17–22.
- [5] BISHOP, C., 2005. Neuronal Networks for Pattern Recognition. Oxford: Oxford University Press.
- [6] ALMEIDA, F., PASSARI, A., 2006. Aplicação de redes neurais na previsão de vendas no varejo. Revista de Administração - RAUSP, v. 41, n. 3, p. 257–272.
- [7] HOLMES, J., HOLMES, W., 2001. Speech Synthesis and Recognition. 11 New Fetter Lane, London: Taylor & Francis. 213–218 p.
- [8] HARASZY, Z., IANCHIS, D., TIPONUT, V., 2009. Generation of the head related transfer functions using artificial neuronal networks. 13th WSEAS International Conference on CIRCUITS, p. 114–118.
- [9] HU, H.; ZHOU, L.; MA H. & WU, Z., 2008. HRTF personalization based on artificial neuronal network in individual virtual auditory space. Applied Acoustics, v. 69, n. 2, p. 163–172.
- [10] RÜMELHART, D., MCCLELLAND, J., the PDP Research Group, 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press, 328–330 p.
- [11] LUCIO NARANJO, J.F.L., 2014. Inteligência Computacional Aplicada na Geração de Respostas Impulsivas Biauriculares e em Aurilização de Salas. Ph.D. Thesis, Universidade do Estado do Rio de Janeiro, Nova Friburgo.
- [12] TORRES, J.C.B., PETRAGLIA, M.R., AND TENENBAUM, R.A., 2004. An Efficient wavelet-based HRTF for auralization Acta Acustica united with Acustica, Vol. 90(1), 108–120.

- [13] TENENBAUM, R.A., CAMILO, T.S., TORRES J.C.B. AND GERGES, S.N.Y., 2007a. Hybrid method for numerical simulation of room acoustics with auralization: Part 1 - Theoretical and numerical aspects. *J. Brazilian Soc. Mech. Sci. Engin.*, Vol. 29(2), 211–221.
- [14] TENENBAUM, R.A., CAMILO, T.S., TORRES J.C.B. AND STUTZ, L.T., 2007b. Hybrid method for numerical simulation of room acoustics: Part 2 - Validation of the computational code RAIOS 3. *J. Brazilian Soc. Mech. Sci. Engin.*, Vol. 29(2), 222–231.
- [15] VORLÄNDER, M., 2008. Auralization: Fundamentals of Acoustics, Modeling, Simulation, Algorithms and Acoustic Virtual Reality Springer, Berlin



José Francisco Lucio Naranjo nació en Quito - Ecuador, el 01 de junio de 1979. Obtuvo el grado de ingeniero en Sistemas Informáticos por la Pontificia Universidad Católica del Ecuador (2005). En 2010 y en 2014, obtuvo, respectivamente, sus grados de M.Sc. y de Ph.D. en Modelado

Computacional en el IPRJ, Universidad del Estado de Río de Janeiro, Brasil. Su principal campo de estudio está enfocado en técnicas de simulación numérica acústica aplicadas en la generación de realidad virtual aplicando redes neuronales artificiales.

El Dr. Lucio está actuando como investigador y profesor titular de informática en la Escuela Politécnica Nacional del Ecuador (EPN). También ha actuado como profesor sustituto en la Universidad del Estado de Río de Janeiro (UERJ), tutor presencial en el Programa de Posgrado en Educación a Distancia en la Universidad Federal Fluminense, profesor curricular en la Universidad de las Américas (UDLA) y profesor auxiliar en la Universidad Central del Ecuador (UCE). También ha sido miembro asociado de la Sociedad Brasileña de Acústica (SOBRAC).

Instrumentación en Dinámica, Acústica y Vibraciones (LIDAV) en el Programa de Postgrado en Modelado Computacional, en la Universidad de Estado de Río de Janeiro, en Nova Friburgo, Brasil. Es autor de tres libros sobre dinámica: Dinâmica, 1997, en portugués; Fundamentals of Applied Dynamics, 2004, en Inglés, publicado por Springer-NY; y Dinâmica Aplicada, 2006, también en portugués. Él es también el autor y coautor de más de 150 artículos publicados en revistas y congresos.

El Profesor Tenenbaum es miembro honorario y fundador de la Sociedad Brasileña de Ciencias Mecánicas e Ingeniería, ABCM; un miembro fundador de la Sociedad Brasileña Acústica, SOBRAC; Es miembro de la Acoustical Society of America, ASA; y la Sociedad Internacional de problemas inversos en Ciencias e Ingeniería, ISIPSE, entre otros.



Julio Cesar Boscher Torres nació en Río de Janeiro, Brasil, el 20 de diciembre de 1971. Obtuvo su grado de Ingeniero Eléctrico en 1993 y, en 1998 y 2004, respectivamente, obtuvo su M.Sc. y su Ph.D., en Ingeniería Eléctrica, de la Universidad Federal de Río de Janeiro (COPPE / UFRJ). Tiene experiencia en Ingeniería Eléctrica, que abarca los siguientes temas: procesamiento de señales, ondas, aurilización, acústica de salas y simulación acústica. Otras áreas de interés son la música, grabación en estudio, stereophotogrammetry y visión 3D y simulación. Desde 2004, trabaja en la UFRJ como profesor asociado, impartiendo cátedra e investigando en las áreas mencionadas. El Prof. Torres es un miembro de IEEE.



Roberto A. Tenenbaum es Ingeniero Mecánico por la Universidad Federal de Río de Janeiro, en 1972; obtuvo su M.Sc. en el campo de Termoelasticidad por la COPPE, Universidad Federal de Río de Janeiro, en 1975, y su D.Sc. en Ciencias en el campo de la acústica en la misma universidad, en 1987. Su principal campo de interés es la dinámica, acústica y vibraciones.

Ha trabajado como ingeniero en la Compañía de Tecnología Nuclear, en Río de Janeiro, de 1973 a 1974. Luego, como profesor e investigador en el Laboratorio de Acústica y Vibraciones (LAVI) perteneciente al Programa de Postgrado de Ingeniería Mecánica, en la Universidad Federal de Río de Janeiro, RJ-Brasil, de 1974 a 2004. Actualmente, se desempeña como profesor e investigador en el Laboratorio de

Propuesta de Modelo de Planificación para Implementación de Centros de Servicios Tecnológicos en Instituciones Educativas Públicas

Planning Model Proposal for Implementation of Technological Service Centres in Public Educational Institutions

Santiago Cordovilla, Christian Madruñero y Bolívar Palán Tamayo

Resumen—Las TIC brindan la oportunidad de expandir el acceso a la educación, fortalecer la educación digital, y elevar la calidad educativa, permitiendo que la enseñanza y el aprendizaje sean un proceso continuo y cotidiano. No obstante integrar efectivamente las TIC en el sistema educativo es un proceso complejo que implica revisión del currículo y pedagogía, las competencias docentes y el financiamiento.

En Ecuador se experimenta un avance tecnológico, principalmente por el acceso a la información y comunicación. En la última década la extensión de la Internet permite que las TIC alcancen su mayor popularidad hasta el presente sin ser aún suficiente.

Con el fin de reducir el desconocimiento de las TIC principalmente en la población en edad escolar se plantea un modelo en el que se estructura un enfoque lógico para el diseño de Centros de Servicios Tecnológicos (CST) que tendrían como objetivo proporcionar beneficios tecnológicos a población estudiantil.

Palabras clave—Centro de Servicio Tecnológico, Diseño, Educación, Estrategia, Implementación, Institución Educativa Pública, Modelo, Población, Servicios, Tecnologías de la Información.

Abstract—TIC provides the opportunity to expand access to education, to strengthen digital education, and to raise the quality of education, to allowing teaching and learning are ongoing, everyday process. However effectively integrate TIC in education is a complex process involving review of curriculum and pedagogy, teaching skills and financing.

In Ecuador experienced us a technological advance, particularly for access to information and communication. In the last decade the extent of the Internet allows TIC reach their highest popularity this without even enough.

Este trabajo está avalado en parte por el Ministerio de Educación del Ecuador.

Santiago Cordovilla, actualmente trabaja en el departamento de tecnología de la Presidencia de la República, Quito, Ecuador (e-mail: corlubu@gmail.com).

In order to reduce the lack of TIC mainly in the school-age population propose a logical design model for the Technology Services Center (TCS) approach would aim to provide technological benefits to student population.

Index Terms—Design, Education, Implementation, Information Technology, Model Town, Public Educational Institution, Services, Strategy, Technology Service Center.

I. INTRODUCCIÓN

LAS tecnologías de la información y las comunicaciones (TIC) que incluyen principalmente la radio, la televisión, y la Internet se han promocionado como herramientas de alta efectividad para el cambio educativo.

Las TIC brindan la oportunidad de expandir el acceso a la educación, fortalecer la pertinencia de la educación en un lugar de trabajo cada vez más digital, y elevar la calidad educativa ayudando a que la enseñanza y el aprendizaje en la cotidianidad sea un proceso continuo. Sin embargo, la integración efectiva de las TIC en el sistema educativo es un proceso complejo que implica no sólo la implantación de la tecnología, sino también el currículo y la pedagogía, la disposición institucional, las competencias docentes y la financiación a largo plazo.

En Ecuador se ha experimentado un avance importante respecto al acceso a tecnología, el mismo que se traduce en oportunidades para todos los sectores especialmente el educativo. Estas oportunidades permiten fortalecer la educación y sus distintos procesos, tales como la educación a distancia, el aprendizaje mediante las nuevas tecnologías de información y comunicación en instituciones de educación regular, la capacitación en línea para los docentes, entre otros.

Christian Madruñero, actualmente trabaja en el Banco del Estado (e-mail: cmadru@gmail.com).

Bolívar Palán Tamayo, Departamento de Informática y Ciencias de la Computación, Facultad de Ingeniería de Sistemas, Escuela Politécnica Nacional, Quito, Ecuador. (e-mail: bolivar.palan@epn.edu.ec).

En la última década la extensión de la Internet y la popularización de los centros tecnológicos permiten que las TIC alcancen su popularidad, dejando de ser solo accesibles para algunos grupos de personas privilegiadas y convirtiéndose en una herramienta importante en la toma de decisiones.

Se plantea un modelo que permite estructurar un diseño de un Centro de Servicios Tecnológicos (CST) que brinde beneficios tecnológicos y reduzca el desconocimiento de las TIC en la población en edad escolar.

Los CST son promovidos como una respuesta a los problemas de la brecha digital en el que amplios sectores de la sociedad no tienen acceso a las TIC y, por tanto son excluidos de los beneficios socio-económicos como la telemedicina, teletrabajo, comercio electrónico, acceso a redes sociales, búsqueda de información en la Internet, entre otras.

En este modelo propuesto, el diseño de CST se basa en el entendimiento de que si bien el acceso a las TIC es importante, por sí misma, no es suficiente. Los CST deben permitir el uso de las TIC de manera que ofrezca beneficios tangibles a sus usuarios, y por lo tanto se constituyan en herramientas para el desarrollo profesional de cada individuo.

En consecuencia, este es un modelo que proporciona una guía para la implantación de CST considerando la optimización de recursos tecnológicos. En principio definiendo el lugar más adecuado conforme la densidad poblacional, acceso a la Internet, recursos físicos disponibles y la brecha digital existente. Y en lo posterior estableciendo un procedimiento que permita el uso eficaz de los CST en el tiempo que los mismos se encuentren disponibles fuera de los horarios de clase de las instituciones educativas a las que pertenezcan.

II. ACCESO A LAS TIC EN LA EDUCACIÓN PÚBLICA

Al presente, la educación constituye uno de los ejes principales de las políticas de desarrollo del Ecuador según lo estipula su constitución en su artículo 26 [1]. Esta política no es única en el país; en general en América Latina los esfuerzos se han concentrado en ampliar la cobertura a la educación primaria y secundaria, y para esto se ha incursionado en el uso de nuevas Tecnologías de la Información y Comunicaciones (TIC). [2]

De acuerdo a informes presentados por el Sistema de Información de Tendencias Educativas en América Latina (SITEAL) la cobertura a la educación primaria y secundaria se ha expandido desde el siglo pasado [3]. Incluso, según estudios realizados por la UNESCO, en la década del 2000 al 2010 la escolaridad de América Latina se incrementó en el nivel primario a un 95% (Fig. 1) [4].



Fig. 1 Tasa neta ajustada de escolarización en la enseñanza primaria. Fuente: UNESCO [5]

El estado ecuatoriano en cuestión de educación enmarca sus metas en el Plan Decenal de Educación (PDE) 2006-2015, aprobado el 26 de noviembre de 2006 en la consulta popular. El PDE determina las políticas de acción que orientan el trabajo a realizarse en un periodo de 10 años [6].

Entre las políticas establecidas en el PDE se cita el “*Mejoramiento de la infraestructura y el equipamiento de las Instituciones Educativas*”.

Esta política implica la construcción de nuevas Unidades Educativas y/o el repotenciamiento de otras tantas. Dentro de esta acción se ha previsto la implantación de laboratorios de informática para el uso de los estudiantes y de ser factible para el uso de las comunidades [7].

A. Uso de las TIC en la Educación

En la educación como muestra la Tabla I el uso de las TIC se ha convertido en un imperativo para mejorar la eficiencia y la eficacia en todos los niveles y en contextos formales y no formales. La educación, incluso en la etapa escolar, tiene que proporcionar instrucción en tecnologías informáticas.

Un conocimiento técnico profundo más una actitud positiva hacia esta tecnología son los requisitos esenciales para los ciudadanos de éxito de las próximas décadas [8].

TABLA I

USO DE TIC EN INSTITUCIONES EDUCATIVAS PÚBLICAS POR PROVINCIA

Provincia	No	Si	Total
AZUAY	549	617	1166
BOLIVAR	611	148	759
CAÑAR	353	216	569
CARCHI	195	156	351
CHIMBORAZO	1096	312	1408
COTOPAXI	506	323	829
EL ORO	413	457	870
ESMERALDAS	1090	256	1346
GALAPAGOS	6	18	24
GUAYAS	2903	1864	4767
IMBABURA	203	343	546
LOJA	1140	396	1536
LOS RIOS	1585	338	1923

MANABI	3145	768	3913
MORONA SANTIAGO	792	142	934
NAPO	312	113	425
ORELLANA	513	62	575
PASTAZA	474	72	546
PICHINCHA	653	1752	2405
SANTA ELENA	126	194	320
SANTO DOMINGO			
DE LOS TSACHILAS	346	213	559
SUCUMBÍOS	579	108	687
TUNGURAHUA	288	360	648
ZAMORA CHINCHIPE	376	84	460
ZONA NO DELIMITADA	80	12	92
TOTAL	18334	9324	27658

Fuente: Ministerio de Educación [9]

Las TIC son utilizadas en distintos ámbitos tanto de investigación como en actividades diarias. Debido a la creciente importancia de la computadora, los estudiantes, no pueden permitirse el lujo de mantenerse al margen de este medio potencial.

B. Ventajas del uso de las TIC en la Educación

Las TIC engloban todos aquellos aparatos que tienen que ver con el procesamiento de la información para una mejor y eficaz comunicación.

El hardware y el software como el retroproyector, la televisión, el radio, los computadores y programas informáticos se utilizan en el proceso educativo. Sin embargo hoy en día las TIC se centran en el uso de tecnologías informáticas para el procesamiento de los datos y la comunicación. En este contexto, las ventajas en la educación se pueden enumerar de la siguiente manera:

- Acceso rápido a la información.
- La fácil disponibilidad de datos actualizados.
- La conexión de las regiones geográficamente dispersas.
- La atención personalizada al aprendizaje individual.
- Acceso a una gama más amplia de medios de comunicación.
- Mayores oportunidades de aprendizaje.

C. Perspectivas de uso de las TIC en la Educación

El desarrollo de nuevas tecnologías interactivas tiene un impacto en los aspectos de la enseñanza y el aprendizaje. Esto es más evidente en el caso de las tecnologías interactivas que fascinan al público amplio, como la Realidad Virtual (RV) y los juegos de computadora y video.

La RV implica un entorno digital en tres dimensiones, multisensorial, inmersivo e interactivo, esto ha dado lugar a que el público imagine como la tecnología va a dominar la forma de interactuar en nuestro trabajo, la educación y el ocio. La RV es una tecnología que tradicionalmente se asocia principalmente con el juego y el entretenimiento. Por razones didácticas, la RV ha sido propuesta como un avance tecnológico que tiene el poder para facilitar el aprendizaje. Puede trabajar para los

educadores, como una herramienta para ayudar a los estudiantes a interactuar, inmersos en un ambiente de aprendizaje en un entorno basado en la tecnología.

La naturaleza visual de la realidad virtual y la manera intuitiva en la que los estudiantes pueden controlar y manipular objetos virtuales se cree que son los dos ingredientes principales que apoyan el aprendizaje en entornos virtuales. La investigación conjunta entre realidad virtual y la educación es un campo relativamente joven, pero en los últimos años ha mostrado un crecimiento considerable [10].

Considerando la prioridad que tiene la educación y la cantidad de recursos que dispone el estado se requiere un modelo de gestión que permita identificar aquellos sitios en los que se pueda implementar centros que brinden el acceso a las tecnologías de la información y comunicación tanto a estudiantes como a la comunidad en general.

D. Conceptualización de un centro de servicios tecnológicos para la educación.

Se puede definir el centro de servicios tecnológicos (CST) como: “*Un lugar físico en donde se reúnen las personas para satisfacer sus necesidades tecnológicas*”. Desde un punto tecnológico social el proyecto InforCauca (2000-2003) financiado por el Centro Internacional de Investigaciones para el Desarrollo define infocentro o telecentro comunitario como: “*gente comprometida con su entorno, gente que lucha, que aprende, que visiona, que sufre y que goza de errores y aciertos, gente que ve en las nuevas tecnologías de la información y comunicación una herramienta muy útil para utilizarla en beneficio de todos.*” [11]

1) Objetivos del CST.

Entre los principales objetivos de un CST presentados por el sitio web de la Unión Internacional de Telecomunicaciones con su programa: Connect a School [12] se puede citar:

- Facilitar un centro de conocimiento para las comunidades.
- Hacer realidad el acceso global a la información a través de la Internet.
- Promover la venta de productos locales a través del comercio electrónico.
- Proporcionar información relevante del gobierno tales como desastres naturales, situaciones políticas, etc.
- Fomentar la participación de las instituciones educativas en la red de CST para lograr la habilitación de servicios a los estudiantes promoviendo la cooperación, la inclusión social e impulsando el desarrollo de la institución.
- Promover en las instituciones las ventajas y oportunidades que ofrecen para su desarrollo y calidad de vida las Tecnologías de la Información y Comunicaciones y los servicios disponibles en los CST.
- Proveer a la ciudadanía de servicios públicos en línea (24/7) para las áreas de salud, ambiente, educación, gobierno y economía, que puedan ser accedidos de manera sencilla por la ciudadanía y los estudiantes.

- Impulsar la organización institucional para la participación activa de los estudiantes en la formulación, ejecución y evaluación de nuevos proyectos y servicios entorno al CST.
- Fortalecer la organización institucional existente, incorporándola a los servicios y proyectos planificados por los CST.
- Diseñar programas de capacitación y entrenamiento dirigidos a las comunidades educativas, organizadas para desarrollar sus capacidades, habilidades y destrezas en el uso y aplicación de las TIC.
- Promover el diseño y la creación de una red estudiantil comunitaria nacional de cooperación y prestación de servicios educativos entre las instituciones que se benefician a través de los CST.
- Impulsar las relaciones de cooperación y la sinergia entre el Ministerio de Educación, las comunidades educativas y demás actores sociales (públicos y privados) participantes, para dar cumplimiento a los compromisos adquiridos.
- Propiciar espacios de concertación entre los actores aliados y la comunidad educativa, con el fin de incentivar el acceso y la apropiación de la cultura tecnológica requerida para su desarrollo integral.
- Impulsar el proceso de transferencia progresiva de la gestión de los CST a las comunidades educativas organizadas.

En resumen los CST son creados como una respuesta a los problemas de la brecha digital, en el que amplios sectores de la sociedad no tienen acceso a las TIC. Al existir esta brecha una parte de la ciudadanía son excluidos de los beneficios socio-económicos que dicho acceso conlleva. Al proporcionar la forma de crear un CST, se ayuda a superar esta disparidad.

III. DISEÑO DEL MODELO DE PLANIFICACIÓN Y ORGANIZACIÓN DEL CST.

El diseño del modelo planteado se origina tomando como referencia al modelo “Percolator” desarrollado en el Departamento de Ingeniería Computacional y Eléctrica de la Universidad de Trinidad y Tobago por Mallalieu y Sean Rocke como un marco metodológico en propuesta para dar soluciones TIC a comunidades excluidas tecnológicamente. [13]

Los CST deben ser organizados de tal manera que permitan el uso de las TIC de una manera que ofrezcan beneficios tangibles a la ciudadanía y, son por lo tanto las herramientas para el desarrollo de sus usuarios. Los objetivos son identificados desde el principio, y luego se expresan en términos de mejora de los flujos de información, las relaciones de asociación y opciones de tecnología, lo que finalmente se dará a través de la implantación, operación y evaluación del CST.

En el proceso se determinó dos grandes subprocessos; en el uno se realiza un diseño lógico de lo que se espera. Posteriormente se define un diseño físico del mismo. Como resultado del proceso del diseño lógico se define los servicios que el CST va a prestar. El proceso se representa en la Fig. 2.

La primera acción a realizarse en la planificación del CST, es considerar el diseño lógico y luego el diseño físico que tendrá el mismo.

Dentro del diseño lógico, se tiene la estrategia de desarrollo (diseño y modelos de CST) y la estrategia de información (interesados, socios, socialización). Luego en el diseño físico se deberá considerar la estrategia de implantación (dónde, cómo, etc.), la estrategia de sostenibilidad (planes y financiamiento) y finalmente la estrategia de evaluación.

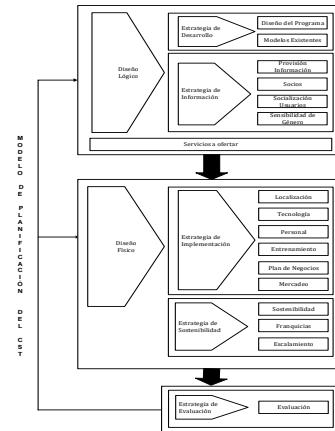


Fig. 2 Pasos para diseño del CST. Fuente: Autores

En la Fig. 2 se analizan que las alianzas entre los actores constituyen un factor importante, porque coadyuvan a lograr la articulación y colaboración en la implantación de los CST. Así también es importante realizar la identificación de los requerimientos de las comunidades según su ubicación y de acuerdo a ello los proyectos que requieren ser ejecutados para implementar los CST.

Al inicio de la 1^a fase (Fig. 3) se estima la premisa que, la información recopilada se encuentre desordenada, imprecisa y no muy clara; por lo que será necesario ordenarla y clasificarla. Paralelamente la información se documenta considerando:

- Numerar los requerimientos identificados.
- Determinar la factibilidad de atender estos requerimientos. De no ser factible se indican las razones del porqué no se puede atender el mismo. En caso de ser factible se procede a clasificar los requerimientos por su tipo.
- Verificar que los requerimientos no estén repetidos o que existan similares.
- Si el requerimiento implica características especiales, se divide en varias partes para su tratamiento.

Una vez hecho lo anteriormente señalado, se procede con la 2^a fase (Fig. 4), en esta se deberá considerar la estrategia de implantación (dónde, cómo, etc.) y la estrategia de sostenibilidad (planes y financiamiento).

A. GUÍA DE APLICACIÓN PARA IMPLANTACIÓN DEL CENTRO DE SERVICIOS

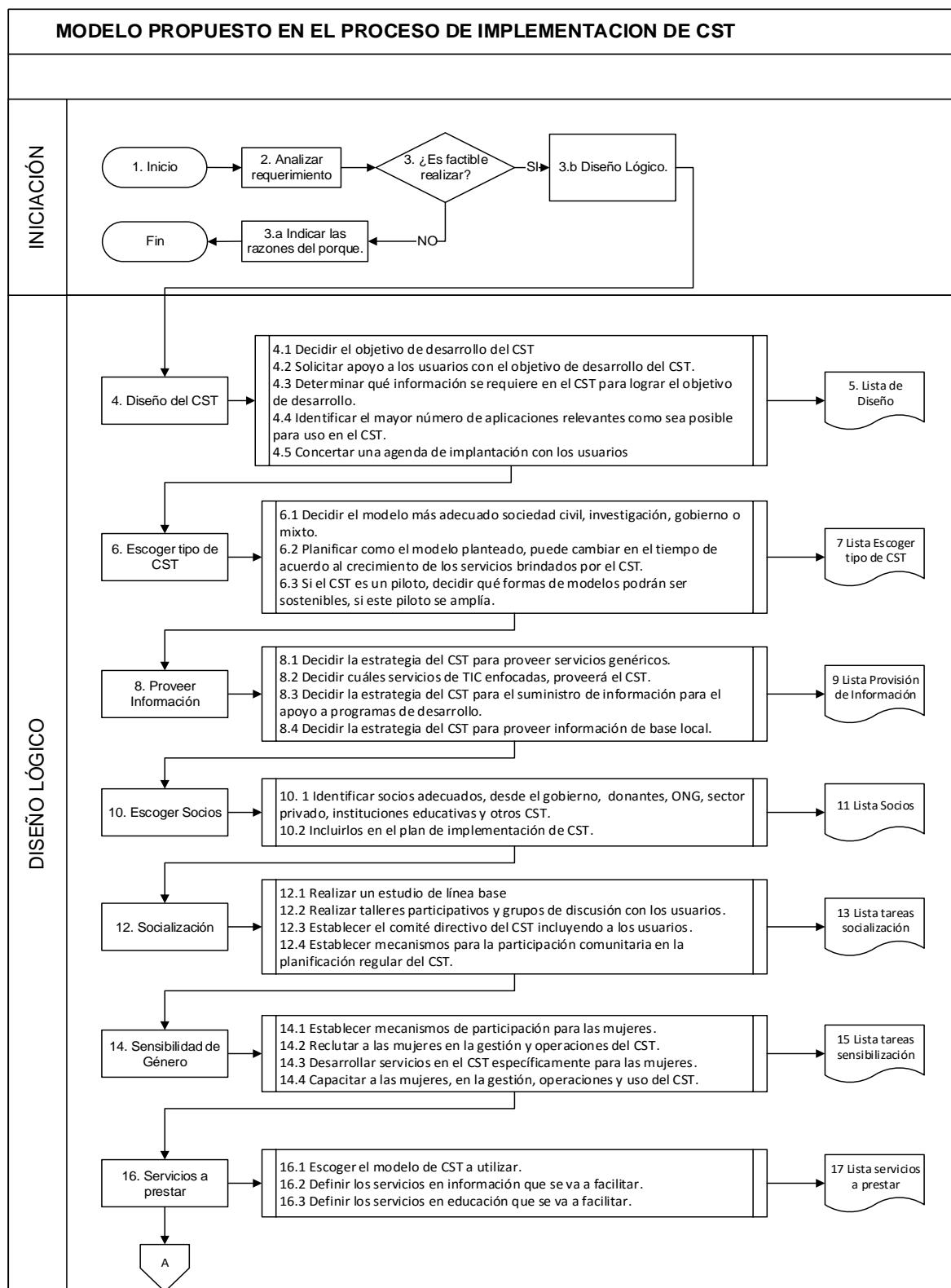


Fig. 3 Modelo para el proceso de implementación del CST.

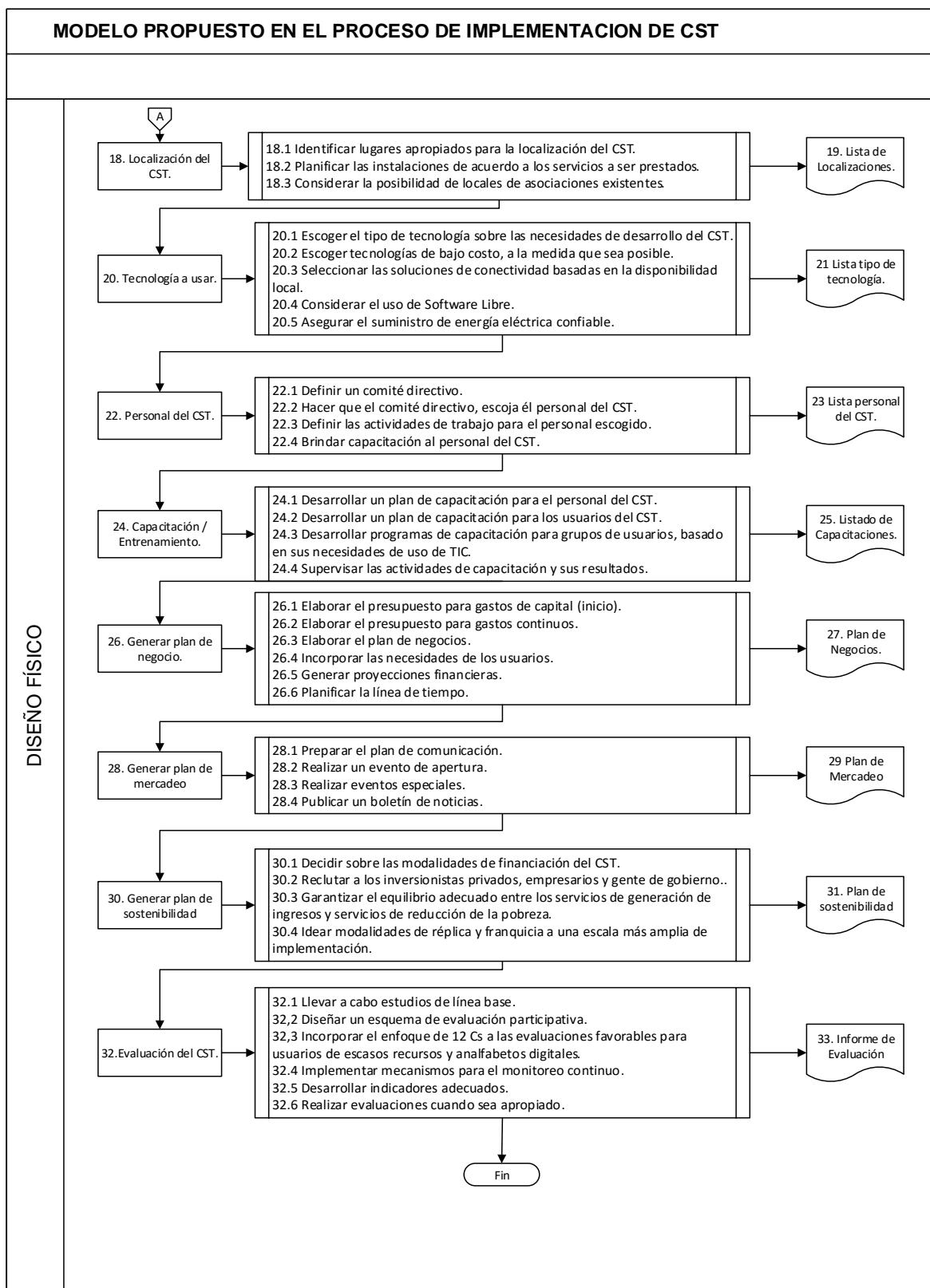


Fig. 4 Modelo para el proceso de implementación del CST (continuación).

IV. EVALUACIÓN DEL MODELO PROUESTO.

Para este caso de estudio se consideró oportuno trabajar con el distrito Pasaje-Chilla-El Guabo por encontrarse en la provincia de El Oro (Ecuador), una de las provincias con uno de los más bajos indicadores de acceso a la Internet, e implantación de distritos y circuitos por zona [14].

Ubicado al sur del Ecuador (Fig. 5), el distrito seleccionado corresponde a los cantones Chilla, El Guabo y Pasaje. El cantón Chilla se ubica en la Hoya de su mismo nombre, en el ramal montañoso de la cordillera de los andes, mientras que los cantones de Pasaje y El Guabo limitan con la provincia del Azuay y el océano Pacífico respectivamente.



Fig. 5 Ubicación Geográfica Pasaje-Chilla-El Guabo.

El distrito Pasaje-Chilla-El Guabo tiene una extensión de 1.400 km² aproximadamente. En este distrito se registra variedad de microclimas lo que implica diferentes actividades productivas centradas principalmente en la agricultura, la ganadería y los cultivos de ciclo corto.

En el cantón Chilla se ubica una única parroquia de su mismo nombre. El cantón El Guabo está dividido en cinco parroquias: El Guabo, Barbones, La Iberia, Tendales, y Río Bonito. Finalmente Pasaje se conforma por cuatro parroquias urbanas: Loma de Franco, Tres Cerritos, Ochoa León y Simón Bolívar, y seis parroquias rurales: La Peña, Buena Vista, Casacay, El Progreso, Caña quemada y Uzchcurrumi.

La población del distrito suma 125.229 habitantes aproximadamente, de los cuales un 58,11% corresponde a la población ubicada en el cantón Pasaje; otro 39.91% está asentada en el cantón El Guabo, y el 1,98% restante de la población habita en el cantón Chilla.

Para aplicar el modelo de planificación y organización orientado a la implantación de CST en instituciones educativas públicas del distrito Pasaje-Chilla-El Guabo, se basó en los pasos listados a continuación, conforme a la Fig. 2.

1. Diseño Lógico.

- Estrategias de Desarrollo.
 - Diseño del programa.
 - Modelo del CST.
- Estrategia de Información

- Provisión de Información
 - Socios, Socialización de usuarios, sensibilización de género.
2. Diseño Físico
- Estrategia de Implantación.
 - Localización Física.
 - Tecnología.
 - Diseño Físico
 - Personal.
 - Entretenimiento.
 - Plan de Negocios.

De las instituciones educativas públicas existentes en Pasaje-Chilla-El Guabo, se seleccionó la Escuela Luis Ángel León Román por los siguientes condicionantes:

- No dispone de laboratorio de informática.
- Tiene acceso a la Internet.
- En su infraestructura cuenta con cerramiento.
- El número de estudiantes (210), justifica la implantación de un laboratorio.

1. Diseño del Programa

El primer paso consiste en realizar el levantamiento del diseño del programa, para lo cual se utiliza, el formulario 1.

FORMULARIO 1: DISEÑO DEL PROGRAMA

DISEÑO LÓGICO DEL CST / ESTRATEGÍA DE DESARROLLO / DISEÑO DEL PROGRAMA		
Fecha elaboración:		
Elaborado por:		
Zona:	Distrito:	
1. Objetivo de la implantación del CST		
2. Los usuarios están de acuerdo con la implantación del CST?		
3. Información requerida en el CST		
Descripción:		
Nro.	Pregunta	Respuesta
1	¿A quién o a quienes está previsto que el CST brinde servicio?	
2	...	
4. Aplicaciones relevantes para el CST		
5. Agenda de implantación del CST		
Cronograma		
Observaciones:		

2. *Modelo del CST.*

El CST debe brindar servicios simples como: facilitar el acceso a la Internet a los estudiantes, contactarse con familiares, encontrar en un foro ayuda para su trabajo de la escuela.

Este punto es uno de los más importantes ya que en él se decide cuál es el tipo de CST más adecuado para implementar.

Esto influye en la planificación, en el tiempo y los recursos que se necesita para implantarse.

3. *Provisión de Información*

Las principales fuentes de información son el INEC, el Ministerio de Educación del Ecuador y el Instituto Geográfico Militar (IGM). La información requerida para este trabajo corresponde a la localización y equipamiento de las instituciones educativas fiscales de educación regular.

4. *Socios, Socialización de usuario y sensibilidad de género.*

Una vez definido la pre-factibilidad en que se desarrolla el análisis del requerimiento se continúa con la recopilación de información correspondiente a los involucrados con el CST, para este caso de estudio el socio será el Gobierno el mismo que se encargará de socialización y la sensibilidad de género.

5. *Localización Física*

Distrito Pasaje-Chilla-El Guabo

6. *Tecnología*

El desarrollo de nuevas tecnologías, los cambios organizativos e incluso los cambios en la sociedad pueden imponer nuevas exigencias a los CST. A pesar de estos cambios se considera que el CST debe caracterizarse por tener confiabilidad y flexibilidad. Es decir, confiabilidad en los servicios que presta y flexibilidad que permita su modificación constante

7. *Diseño Físico del CST.*

Como se muestra en la Fig. 6, las dimensiones establecidas para el espacio en el que funcionará el CST se enmarcan en 7.50m de largo por 7.00m de ancho. Dispone de una entrada de 1.8m de ancho.

El CST debe estar preparado para adaptarse a las nuevas tecnologías. En virtud de ello se ha previsto disponer de los planos eléctricos, de telecomunicaciones, de tuberías conforme fueron instalados y no como fueron planificados.

En el diseño del CST se consideró sistemas de audio, control ambiental, seguridad, televisión, alarmas, voz y datos, y aunque no todos los sistemas se prevé sean instalados, al menos se estima hacerlo posteriormente.

Instalaciones Eléctricas

La instalación eléctrica es uno de los puntos más importantes en el diseño del CST ya que proveerá de energía a los equipos de cómputo y a los demás equipos como impresora, proyector y equipos de telecomunicación. Adicionalmente se requiere que la instalación contemple la conexión a tierra individual de cada uno de estos equipos. La alimentación eléctrica de los equipos

en primera instancia será a través de corriente alterna. La energía eléctrica que alimenta los equipos del CST llega a través de una línea dedicada desde el distribuidor independiente de la energía que alimenta a los demás sistemas.

Iluminación

Para el CST se ha considerado un sistema de iluminación que garantice un entorno agradable para el usuario y facilite al docente o expositor el uso del área para la enseñanza. Por ello se ha dividido el área de trabajo en sub áreas de igual medida de tal manera que cada área disponga de una misma intensidad de luz evitando la reflexión o proyección de sombras.

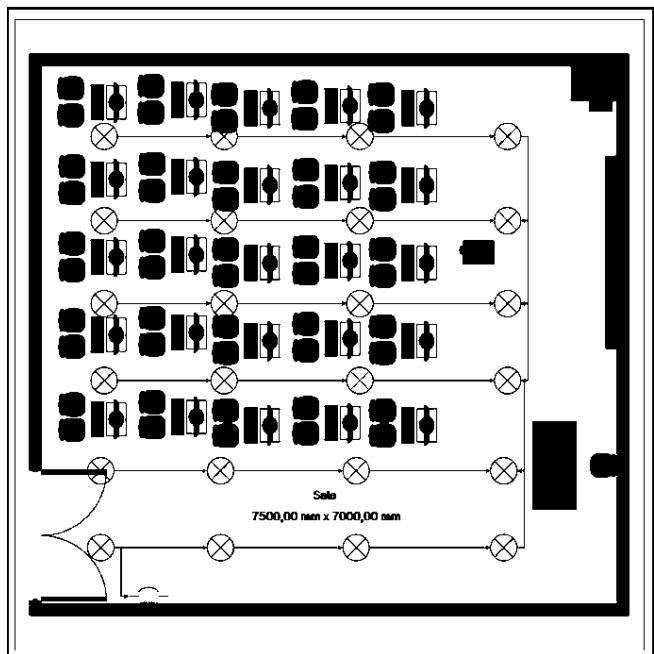


Fig. 6 Diseño Físico del CST.

8. *Personal*

Conforme la estructura organizativa de las Instituciones Educativas Públicas se prevé un docente de informática con una carga horaria distribuida entre la labor pedagógica en la institución y la colaboración con la comunidad.

Sin embargo es probable que su tiempo de trabajo fuera de las horas clase sea mínimo, por lo que se plantearía seleccionar un responsable a tiempo completo.

9. *Entrenamiento*

Para el entrenamiento se ha previsto realizar 3 talleres a cargo de los responsables. Los talleres corresponde a:

- Inducción. En él se da a conocer el proyecto, sus tareas, el inventario de bienes; es decir un panorama general del trabajo.
- Socialización del manual de procedimientos. En el manual se especifica los procedimientos generales que tiene que realizar durante su jornada laborar en el CST. Entre los más relevantes se cita:
 - Recepción de material de oficina.
 - Control de inventario.

- Registro de usuarios y de uso de los computadores.
- Registro de ingresos y egresos, en el caso de brindar servicios de impresión, escáner, etc.
- Registro de uso de impresora.
- Plan de mantenimiento preventivo.
- Plan de mantenimiento correctivo.

10. Plan de Negocio, Mercadeo y Sostenibilidad

Para este caso de estudio no se registra un plan de negocio, mercadeo y sostenibilidad ya que se prevé el financiamiento por parte del gobierno.

A. Análisis y discusión de los resultados.

El análisis consiste en la validación del modelo y permite establecer la efectividad del mismo. Para ello se ha realizado una comparación entre lo propuesto y el resultado obtenido. En este caso se toma en cuenta la siguiente información:

- Proyección de usuarios del CST atendidos en el lapso de un mes.
- Proyección de usuarios capacitados en el CST en el lapso de un año.
- Proyección de reducción de la brecha digital en un año.

B. Usuarios atendidos en el CST

Considerando la capacidad máxima del CST planificada (40 horas, 20 personas por hora, se atendería a 800 personas por semana). Conforme a la malla curricular de Educación General Básica estipulada en el Acuerdo Ministerial 041 de 2014, en la que se indica que un estudiante debe tomar dos horas de computación, la capacidad se reduce a 400 personas.

C. Usuarios capacitados en el CST

En el caso de que se pudiera contar con un proyecto que permita la capacitación de los usuarios del CST de manera focalizada en sus necesidades o al menos que se brinde cursos de conocimientos básicos en TIC, se esperaría disminuir 400 personas del total de la demanda.

D. Reducción de la brecha digital.

Si con la implantación de CST se puede alcanzar un uso generalizado de las TIC como es el uso de la Internet, los elementos multimedia, los computadores, la comunicación vía Internet sea por voz IP o redes sociales, esto indicaría claramente que la brecha digital ha disminuido. No obstante no se podría definir si se ha reducido la brecha hasta no implementar el CST y evaluar sus resultados en un corto plazo.

Finalmente se considera que el modelo puede ser mejorado con la implantación en varios distritos y circuitos educativos donde se observe un analfabetismo digital mayor, lo que permitiría determinar las ventajas y desventajas del modelo propuesto.

V. CONCLUSIONES Y TRABAJO FUTURO.

La propuesta del modelo de diseño de los CST, tiene un impacto importante, por lo menos en el corto plazo; para ciertos

individuos supondrán ventajas y, desventajas para otros. En la actualidad, la adopción diferencial de la Internet ha creado una brecha digital entre los países desarrollados y en desarrollo. La provisión de CST ofrece una estrategia para reducir la brecha digital, pero es sólo una parte relativamente pequeña.

El uso de las TIC en el sector educativo permite abordar los problemas del desarrollo de las comunidades. La experiencia y el aprendizaje a partir de dos décadas de uso, ha dado lugar a una mejor comprensión de lo que funciona, lo que no, y por qué las TIC permiten el desarrollo y la disminución de la brecha digital, en los grupos educativos en edades comprendidas entre 3 y 17 años.

Las propias TIC están situadas dentro de los contextos y las condiciones sociales y políticas reales. Al mismo tiempo, la capacidad transformadora de las TIC se ha convertido en un motor de cambio, y los países que han comprendido las tecnologías y los arneses para satisfacer tanto las metas de crecimiento y desarrollo están cosechando los beneficios de dicha transformación.

En las manos del gobierno, las TIC son herramientas poderosas. En las manos de los pobres, las TIC tienen un efecto de empoderamiento. Y cuando los gobiernos utilizan las TIC para conectar a los pobres, se producen resultados dramáticos. Un teléfono móvil en las manos de una mujer pobre rural es el mejor símbolo de los resultados espectaculares que se producen cuando se utilizan las TIC para ayudar en los procesos de desarrollo.

El uso de la geo estadística para escoger dentro del territorio ecuatoriano donde proceder a implementar los CST, permite tener datos más certeros de a que distritos de acuerdo a la oferta y demanda de servicios tecnológicos es indispensable atender.

El modelo de diseño de los CST planteado en esta tesis, permite un conocimiento más profundo de la realidad en cuanto a acceso a las TIC poseen las diferentes instituciones educativas y cuál es la demanda de servicios comunitarios como el aprendizaje a distancia, y el gobierno electrónico, requieren.

La validación del modelo en el distrito Pasaje – Chilla – El Guabo permitió conocer las necesidades tecnológicas que requieren, para poder disminuir el analfabetismo digital, ya que la demanda de estas necesidades es alta pero la oferta es baja.

En las manos del gobierno, las TIC son herramientas poderosas. En las manos de los pobres, las TIC tienen un efecto de empoderamiento. Y cuando los gobiernos utilizan las TIC para conectar a los pobres, se producen resultados dramáticos. Un teléfono móvil en las manos de una mujer pobre rural es el mejor símbolo de los resultados espectaculares que se producen cuando se utilizan las TIC para ayudar en los procesos de desarrollo.

El uso de las TIC para abordar los problemas del desarrollo es la disciplina emergente para el desarrollo. La experiencia y el aprendizaje a partir de casi dos décadas de experimentación ha dado lugar a una mejor comprensión de lo que funciona, lo que no, y por qué las TIC permiten el desarrollo y la disminución de la brecha digital.

A. Trabajo Futuro

En los futuros CST no se dispondrá de equipamiento sino únicamente de conectividad; no se verá a un docente que expone el taller, por ejemplo sobre las funcionalidades de un software o determinadas habilidades electrónicas – siguiendo formatos de clase magistral, sino que éste actuará como dinamizador, resolverá dudas, atenderá a la diversidad del grupo, establecerá grupos de trabajo y su participación en los talleres cobrará una dimensión de facilitador / de dinamizador y de transmisor de conocimiento y no tanto de contenido [15].

Los usuarios de los CST se caracterizarán por ser activos, autónomos en su aprendizaje y tener conexión móvil permanente fuera de los centros, accediendo a contenidos audiovisuales, lecturas, presentaciones, podcasts, etc. En definitiva, la adquisición del conocimiento base se producirá fuera de los CST, en remoto y no se perderá tiempo en todo aquel contenido que se puede encapsular.

Los usuarios al acceder tendrán excelentes políticas de seguridad que garantizarán el acceso desde diferentes dispositivos a las plataformas y espacios virtuales que ofrecen los CST. Quizás como beneficio indirecto los propios usuarios incrementarían su percepción y actitud ante la Ciberseguridad. Accederán a entornos personales de aprendizaje, contenidos divulgativos sobre las TIC y muchas otras opciones centradas más en los servicios que en los propios equipamientos.

VI. REFERENCIAS

- [1]. Aguirre, P. D. (2012). Texto comentado a la Constitución de la República del Ecuador. Loja, Loja, Ecuador: DYKINSON, S.L.
- [2]. UNESCO, Lugo María Teresa. (2011). La matriz TIC. Una herramienta para planificar las Tecnologías de la Información y Comunicación en las instituciones educativas. Buenos Aires: UNESCO.
- [3]. SITEAL. (2011). Informe sobre tendencias sociales y educativas en América Latina. Buenos Aires: IIEP-UNESCO/OEI.
- [4]. UNESCO. (2011). Informe de Seguimiento de la EPT en el Mundo 2011. Accedido 10-07-2013, desde UNESCO: www.efareport.unesco.org
- [5]. UNESCO. (2011). Panorámica regional: América Latina y el Caribe. Informe de seguimiento de la EPT en el mundo 2011. París.
- [6]. Ministerio de Educación. (2006). Documentos Legales y Normativos. Accedido 10-08-2013, desde www.educacion.gob.ec/documentos-legales-y-normativos/
- [7]. Ministerio de Educación. (2006). Documentos Legales y Normativos. Accedido 10-08-2013, desde www.educacion.gob.ec/documentos-legales-y-normativos/
- [8]. Instituto de aprendizaje abierto y a distancia (IDOL). (2014, Enero 19). Universidad de Mumbai. Accedido desde http://www.mu.ac.in/myweb_test/ma%20edu/ICT%20-%20Edu..pdf
- [9]. AMIE Ministerio de Educación. (2013). Indicadores Educativos calculados con datos del Archivo Maestro de Instituciones Educativas (AMIE). Quito.
- [10]. ROUSSOU, M. (2000). Immersive Interactive Virtual Reality and Informal Education", Proceedings of User Interfaces for All: Interactive Learning Environments for Children. Athenas: s/e.
- [11]. Figueroa, E. C. (2003). Telecentros comunitarios. Cali: InforCauca.
- [12]. Conectar la Escuela, Conectar una Comunidad. (2014, Julio 16). www.connectaschool.org
- [13]. Mallalieu Kim, R. S. (2005). Selecting Sustainable ICT Solutions for Pro-poor Intervention. Trinidad and Tobago: Department Of Electrical And Computer Engineering The University Of The West Indies, St. Augustine.
- [14]. SENPLADES. (2012). Secretaría Nacional de Planificación y Desarrollo. Recuperado 23-10-2013, desde <http://www.planificacion.gob.ec/>
- [15]. PRIETO Paco, A. M. (2012). Telecentros 3.0 y la Innovación Social en la Sociedad Red. Andalucía: Guadalinfo.

VII. BIOGRAFÍAS



Santiago Cordovilla Rubio nació en Ambato-Ecuador, el 19 de Mayo de 1978. Se graduó de la Universidad Politécnica Salesiana en Quito donde obtuvo el título de Ingeniero en Sistemas. Obtuvo también el título de Magíster en Gestión de las Comunicaciones y Tecnologías de la Información en la Escuela Politécnica Nacional en Quito-Ecuador. Su experiencia laboral incluye la General Motors del Ecuador, Comercial Etatex. Sus campos de interés de especialización incluyen Software Libre, Tecnologías de la Información, Seguridad Informática, Ethical Hacking, Informática Forense, Operaciones y Procesos, Buenas Prácticas de manejo de proyectos (Prince, PMP). Actualmente se desempeña como soporte tecnológico en la Presidencia de la República del Ecuador.



Christian Madruñero Mueses nació en Quito-Ecuador, el 30 de Octubre de 1980. Se graduó de la Universidad Politécnica Salesiana en Quito donde obtuvo el título de Ingeniero en Sistemas. Alcanzó también el título de Magíster en Gestión de las Comunicaciones y Tecnologías de la Información en la Escuela Politécnica Nacional en Quito-Ecuador. Laboró en el Programa de las Naciones Unidas para el Desarrollo, Comercial Etatex y actualmente se desempeña como Especialista en el Ministerio de Educación del Ecuador. Sus campos de interés de especialización incluyen Tecnologías de la Información, Operaciones y Procesos, Buenas Prácticas de manejo de proyectos y Educación.



Bolívar Oswaldo Palán Tamayo, Ingeniero en Electrónica y Telecomunicaciones. Escuela Politécnica Nacional. Magíster en Ciencias de la Computación e Informática. Escuela Politécnica Nacional. Diplomado en Habilidades de Gestión Administrativa. Universidad Técnica Particular de Loja. Profesor del Departamento de Ciencias de la Computación e Informática de la Escuela Politécnica Nacional. Socio Consultor en Tecnologías de la Información y Comunicación. PALAN TAMAYO CONSULTORES – PATCO CIA. LTDA. Gerente General.

Arquitectura Clúster de Alto Rendimiento Utilizando Herramientas de Software Libre

High Performance Cluster Architecture Using Free Software Tools

Leonardo Chuquiguanca, Edyson Malla, Freddy Ajila y Rene Guamán-Quinché

Resumen—En este artículo se presenta los resultados obtenidos de implementar una arquitectura clúster de tipo beowulf clase II con hardware convencional y herramientas de software libre, lo que permitió convertir un laboratorio utilizado para la enseñanza de programación de sistemas informáticos, en una sala de altas prestaciones mediante el uso de software: MPICH, Mosix, Blender, Hadoop y Ganglia Monitoring System.

El método de resolución de problemas permitió: la comprensión del problema, diseñar una estrategia de solución y lograr la implementación de la arquitectura clúster con un rendimiento similar al de un supercomputador, optimizando el tiempo de procesamiento en: simulación de procesos, reconocimiento de patrones, renderizado de imagen y video, análisis de grandes volúmenes de datos Big Data, cifrado de códigos, evaluación de algoritmos, etc.

Las pruebas permitieron el procesamiento de millones de operaciones por segundo alcanzando una eficiencia del 85.2% de la capacidad total del clúster. Este logro importante facilita también la investigación académica en otros campos donde se requiera procesar grandes volúmenes de información y obtener resultados en un corto tiempo.

Palabras clave—Clúster, Middleware, Mosix, MPICH, Beowulf.

Abstract— In this article the results of implementing a beowulf cluster type II architecture, with conventional hardware and free software tools is presented, allowed to convert a laboratory used for teaching programming systems, in

Este artículo fue enviado para revisión el 11 de febrero de 2015. La implementación de la arquitectura clúster de alto rendimiento utilizando herramientas de software libre fue apoyada por la Carrera de Ingeniería en Sistemas de la Universidad Nacional de Loja, como proyecto de grado.

L. Chuquiguanca graduado de la carrera de Ingeniería en Sistemas de la Universidad Nacional de Loja actualmente administrador de sistemas en ProtelcotelSA SA. Quito-Ecuador (e-mail: lrchuquihuancav@unl.edu.ec)

E. Malla graduado de la carrera de ingeniería en Sistemas de la Universidad Nacional de Loja Actualmente Ingeniero en Sistemas del Gobierno Autónomo Descentralizado Municipal San Felipe de Oña, Azuay – Ecuador (e-mail: ejmallab@unl.edu.ec)

F. Ajila docente Investigador de la Escuela de Ingeniería Industrial, Facultad de Mecánica de la Escuela Superior Politécnica de Chimborazo (e-mail: freddy.ajila@espoch.edu.ec)

R. Guaman-Quinché docente Investigador de la Carrera de Ingeniería en Sistemas Informáticos de la Universidad Técnica de Manabí (e-mail: eguaman@utm.edu.ec)

a room with high performance, through a software: MPICH, Mosix, Blender, Hadoop and Ganglia Monitoring System.

The method using for troubleshooting allowed; understanding the problem, and to design a solution strategy and achieve implementation of the cluster architecture with a performance like a supercomputer by optimizing the processing time: process simulation, pattern recognition, image and video rendering, and an analysis of Big Data, encryption codes, evaluation algorithms, etc.

The tests allowed the processing of millions of operations per second reaching an efficiency of 85.2% of total capacity of cluster. This important achievement expedite also the academy investigation in others realms where required to process large volumes of information and get results in a short time.

Index Terms— Cluster, Middleware, Mosix, MPICH, Beowulf.

I. INTRODUCCIÓN

MULTITUD de aplicaciones dentro de la investigación científica requieren de una gran demanda de potencia de cómputo que solo pueden ser cubiertas por supercomputadores, que por su alto costo y los presupuestos reducidos para investigación para investigaciones académicas y científicas no pueden ser adquiridas por Instituciones de Educación Superior y centros de investigación, las inversiones en computación deben resultar rentables en cuanto a escalabilidad y rendimiento.

La informática para investigaciones, denominada computación de alto rendimiento usa potentes herramientas y procesos de computación para generar datos en investigaciones académicas avanzadas. Con un clúster de computación de alto rendimiento, los centros de investigación pueden obtener la velocidad y potencia de una costosa supercomputadora a una fracción del costo y con menos riesgo de sufrir tiempos de inactividad prolongados.

Una de las soluciones más fiables comparadas a la adquisición de supercomputadores es la implementación de un clúster de alto rendimiento, formado por hardware convencional y herramientas de software libre que unidos a una red de alta velocidad [1], [2], [3] ofrecen ventajas significativas en términos económicos y de escalabilidad comparadas al utilizar un único ordenador convencional, proporcionando así

resultados con un margen de error mínimo en tiempos relativamente bajos [4], [5].

Existen algunas implementaciones de clúster en centros de investigación y en instituciones de educación superior, uno de ellos se lo diseñó para la ejecución de un modelo de predicción climática, (Grupo de Ciencias de la Tierra y del Ambiente de la Dirección de Investigación (DIUC) de la Universidad de Cuenca, Ecuador. El objetivo de la implementación del clúster dentro del Grupo CTA, es fomentar a la investigación científica en áreas que requieren alta capacidad computacional, optimizando la utilización de los recursos de hardware disponibles, para contar con una herramienta potente y capaz de solventar, en lo posible esas necesidades. Se tiene como base el sistema operativo, GNU/Linux; el sistema de administración del clúster y el estándar utilizado en este proyecto para el desarrollo de software es PVM y MPI como sistemas de programación paralela, el middleware Mosix contiene lo necesario para ejecutar una aplicación que permitirán enlazar el entorno de programación paralela de la capa superior con compiladores paralelos MPI, PVM, GNU, JAVA, etc; software y herramientas de administración, que facilitan la gestión de archivos, balanceo de nodos, la imagen del sistema operativo, que ofrece a los usuarios el acceso unificado a los recursos del sistema [6].

Otro caso de éxito se basó en el diseño e implementación de un clúster de cómputo de alto rendimiento en el Centro de Investigación en matemáticas de la universidad de Guanajuato, México. El clúster de alto rendimiento denominado “El Insurgente” fue diseñado y construido específicamente para aplicaciones de cómputo científico que requieren grandes volúmenes de datos en memoria RAM, programación distribuida con un híbrido de MPI y OpenMPI, usando redes de bajo costo. De manera lógica, cada nodo del clúster tiene una parte de hardware y otra de software. El hardware está compuesto por procesadores, memoria, interfaz de red y discos duros entre otros. En cuanto al software, el nivel bajo corresponde al sistema operativo, el medio consiste en las librerías de paralelización y el alto está representado por la aplicación que se desea ejecutar en el clúster [7].

La implementación de la arquitectura clúster se realizó en un laboratorio de la Carrera de Ingeniería en Sistemas de la Universidad Nacional de Loja, formado por 15 computadores personales con características homogéneas; las principales configuraciones que se realizaron son: instalación y configuración del sistema operativo Debian “wheezy”, configuración e instalación del middleware Mosix [8], [9] en el kernel de Linux para obtener una imagen única de sistema orientada a la computación distribuida, para la programación paralela se utilizó MPICH [10], como una de las aplicaciones más implementadas del estándar Message Passing Interface (MPI). Para el proceso de renderizado se usó el software Blender [11]; para el análisis de datos (Big Data) [12], se implementó el framework Apache Hadoop [13]. Además se configuró el acceso remoto utilizando claves públicas a través del protocolo Secure Shell (SSH), para la configuración y compartición del sistema de ficheros en red se utilizó el protocolo Network File System (NFS) que trabaja a nivel de capa de aplicación según el modelo OSI, se realizó el diseño de la topología de red y la configuración del sistema de monitorización con Ganglia Monitoring System [14]. Todo ello

permitió mejorar el tiempo de cómputo para calcular algoritmos o procesamiento de información.

La estructura del artículo es la siguiente:

La Sección II presenta como Marco Teórico, material bibliográfico sintetizado referente a arquitecturas clúster y la computación de alto rendimiento. La Sección III detalla la metodología de resolución de problemas utilizada para el desarrollo del proyecto. La Sección IV describe el proceso de implementación de la arquitectura clúster. En la Sección V se presentan los resultados obtenidos en diversas pruebas utilizando herramientas de software libre. Finalmente, la Sección VI se establece las conclusiones a las que se ha llegado al término del proyecto.

II. MARCO TEÓRICO

A. Arquitectura clúster de alto rendimiento

Un clúster es un conjunto de nodos de bajo costo conectados entre sí a través de una red de comunicaciones de alta velocidad, que operan bajo software que actúa como un sistema único de administración, responsable de distribuir las cargas de trabajo entre los nodos, de forma automática y transparente al usuario como si se tratara de un único ordenador [3], [4], [5].

De manera lógica, cada nodo del clúster está formado por hardware y software. El hardware está compuesto por las partes de un ordenador convencional. En cuanto al software, el nivel bajo corresponde al sistema operativo, el nivel medio consiste en las librerías de paralelización y el nivel alto está representado por la aplicación que se desea ejecutar; una aplicación se ejecuta en el nodo maestro, el sistema operativo y las librerías de paralelización se encargan de ejecutar copias de este programa en los nodos esclavos [8], [9], [10]. De las distintas arquitecturas disponibles, se ha elegido la arquitectura clúster de tipo beowulf, como se ilustra en la Fig. 1 la misma que está compuesta por hardware convencional y herramientas de software libre.

Un clúster de alto rendimiento es utilizado principalmente con fines académico-científicos, su objetivo principal es proporcionar altas prestaciones de capacidad de cómputo superior a los que pudiera ofrecer un ordenador común. Este tipo de arquitecturas son una alternativa a la utilización de grandes y costosas supercomputadoras [11], [12]

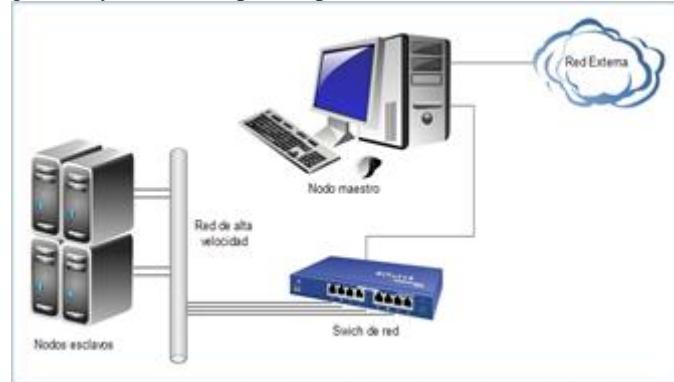


Fig 1. Clúster de computadoras formado por hardware convencional

B. Clasificación de los clústers

Existen dos tipos de clúster según la arquitectura de las computadoras que lo conforman:

Clúster homogéneo: Todos los nodos tienen las mismas características técnicas de hardware y software. Son idénticos y por lo tanto la capacidad de procesamiento y rendimiento de cada nodo es la misma.

Clúster heterogéneo: Al contrario de los clúster homogéneos, los nodos son completamente distintos en cuanto a hardware y software.

C. Importancia de la computación de alto rendimiento

La programación paralela se origina por las limitaciones de las aplicaciones secuenciales; integrando varios procesadores para llevar a cabo sus funciones, la programación paralela permite resolver problemas que requieren más memoria o mayor velocidad de cómputo. También existen razones económicas, pues el precio de los ordenadores secuenciales no es proporcional a su capacidad computacional, mientras que la conexión de varios procesadores utilizando una red nos permite obtener un aumento de prestaciones prácticamente proporcional al número de procesadores con un coste adicional mínimo [15].

Además con el uso del clúster se logra reducir el tiempo de resolución de problemas computacionales, o bien resolver problemas que no cabrían en la memoria de un solo procesador secuencial. Y para esto es necesario utilizar sistemas de altas prestaciones y algoritmos paralelos que utilicen estos sistemas de manera eficiente [16]

III. METODOLOGÍA

Durante el desarrollo de esta investigación, se utilizó la metodología de resolución de problemas que se organiza en siete etapas descritas a continuación:

1. **Identificación del problema.** Esta fase comprendió el estudio de la revisión bibliográfica y casos de éxito del funcionamiento de los clúster, en centros de investigación e instituciones de educación superior con la utilización de software libre y hardware convencional. Se analizó también la situación actual del laboratorio de cómputo de la Escuela de Ingeniería en Sistemas de la Universidad Nacional de Loja (CIS-UNL) para identificar si existían procesos que requieran procesar grandes volúmenes de información a altas velocidades y no se disponía de un supercomputador.
2. **Explicación del problema.** En el laboratorio de la CIS de la UNL no existían procesos que requieran grandes velocidades de procesamiento pero por la falta de un sistema computacional de alto rendimiento los docentes y estudiantes estaban privados de realizar proyectos académicos relacionados con el procesamiento de grandes volúmenes de información y se desconocía que con la utilización de hardware convencional y software libre se podían obtener grandes beneficios económicos, comparados con la adquisición de supercomputadores que realicen tareas dedicadas a la computación de alto rendimiento.

3. **Idear estrategias alternativas de intervención.** Para idear alternativas que permitieron solucionar el problema mencionado se realizó el análisis de los recursos técnicos de hardware, software y redes con las que cuenta el laboratorio de la CIS de la UNL, siendo el punto de partida para la implementación de la arquitectura clúster.

4. **Decidir la estrategia.** Una vez realizado el análisis de la situación actual del laboratorio de la CIS de la UNL, se procedió a realizar la búsqueda de las herramientas de software libre, la elección del middleware, la elección de la arquitectura clúster, el diseño de la topología y el direccionamiento de red.

5. **Diseño de la intervención.** En esta fase se determinó los tiempos de implementación del proyecto para el laboratorio CIS de la UNL y las actividades a cumplir en los plazos establecidos, logrando con éxito la culminación del mismo.

6. **Desarrollo de la intervención.** En esta fase se estableció la instalación y configuración de cada uno de los aplicativos de software libre mencionados anteriormente.

7. **Evaluación de los logros.** La evaluación de los logros obtenidos se la realizó aplicando pruebas de procesamiento a distintos proyectos, logrando evidenciar tiempos mínimos de ejecución al utilizar los recursos de la arquitectura clúster.

IV. PRUEBAS

Se realizaron diversas pruebas en la arquitectura clúster, para determinar la capacidad de procesamiento alcanzado, ejecutando aplicaciones que requieren alta capacidad computacional. Para realizar las pruebas a los siguientes proyectos se utilizó 1, 5, 10 y 15 nodos respectivamente.

A. Proyecto 1: Cálculo del valor aproximado de pi utilizando librerías MPI

El proyecto evalúa el algoritmo utilizado para calcular el valor de pi, utilizando librerías de paso de mensajes MPI.

TABLA I. RESULTADOS CÁLCULO DEL VALOR APROXIMADO DE PI UTILIZANDO MPI

Número de nodos	Número de procesadores	Tiempo de ejecución (seg)
1	8	0.959801
5	40	0.321487
10	80	0.228983
15	120	0.136832

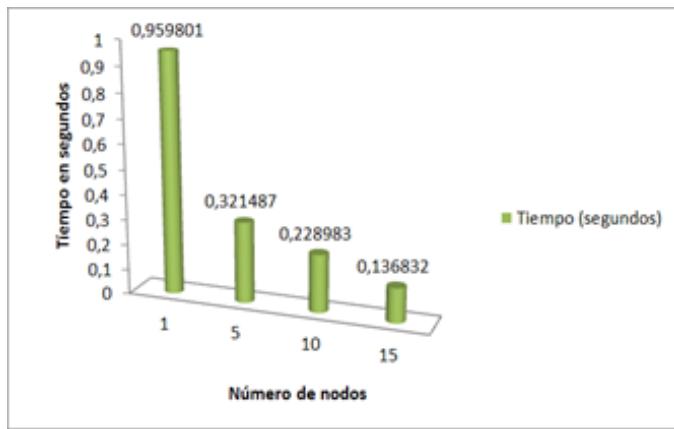


Fig 2. Tiempos de ejecución cálculo del valor aproximado de pi.

De la prueba realizada se deduce que utilizando un solo nodo, el tiempo de ejecución es de 0.959801 segundos, mientras que al utilizar los 15 nodos de la arquitectura el tiempo es de 0.136832 segundos, como se observa en la Tabla. I, obteniendo una disminución de 0.822969 segundos de tiempo de procesamiento, lo que se interpreta como una eficiencia de aproximadamente 88%, como se ilustra en la Fig. 2.

B. Proyecto 2: Cifrado de códigos con John The Ripper (JTR) y MPI.

Las pruebas se realizaron sobre uno de los ficheros más importantes de Linux como lo es shadow, este fichero almacena información cifrada de las contraseñas de cada una de las cuentas de usuario del sistema operativo. La contraseña utilizada tiene una longitud de 8 símbolos, y consta de 62 combinaciones de caracteres (26 letras del abecedario mayúsculas + 26 letras del abecedario minúsculas + 10 dígitos), por lo que se realizaron cerca de 218 trillones de posibles combinaciones.

TABLA II. RESULTADOS CIFRADO DE CÓDIGOS CON JOHN THE RIPPER (JTR) Y LIBRERÍAS MPI

Número de nodos	Número de procesadores	Tiempo de ejecución (horas)
1	8	09:48:34
5	40	07:21:22
10	80	03:06:57
15	120	00:27:24

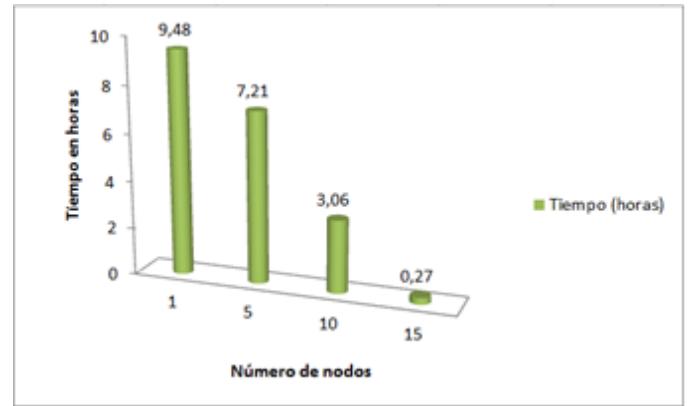


Fig 3. Tiempo de ejecución de cifrado de código utilizando John The Ripper

El tiempo de ejecución al descifrar el fichero shadow, en un solo nodo es de 9.48 horas, mientras que al utilizar los 15 nodos de la arquitectura el tiempo de la decodificación es de 27 minutos, como se detalla en la Tabla II obteniendo una disminución de 9.21 horas de tiempo de procesamiento, lo que se interpreta como una eficiencia de aproximadamente 97%, ver Fig. 3.

En la Fig. 4, se observa el balanceo de carga al ejecutar el proceso de cifrado de código utilizando John The Ripper y librerías MPI en los 15 nodos de la arquitectura clúster.

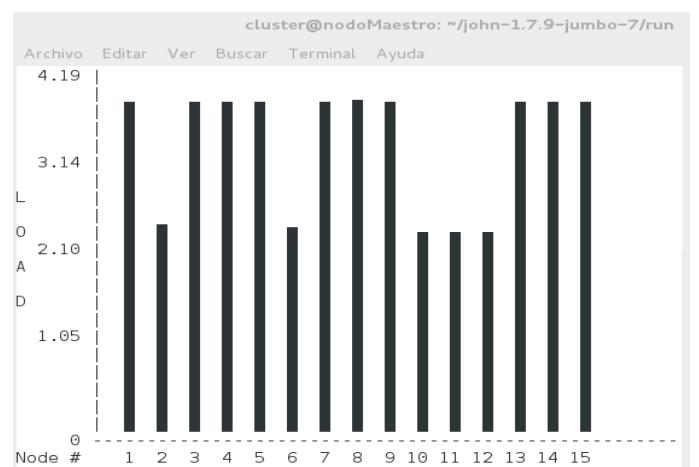


Fig 4. Balanceo de carga en el clúster al ejecutar MPICH+JTR

C. Proyecto 3: Renderización de imagen y video con Blender

Este proyecto trata sobre el armado final de un cubo de rubik que está formado por 115 frames, una vez terminado el renderizado se obtiene como producto final, un video con una duración aproximada de 11 segundos y un tamaño total de 18.5 MB.

TABLA III. RESULTADOS RENDERIZACIÓN DE IMÁGENES Y VIDEOS CON BLENDER

Número de	Número de	Tiempo de
-----------	-----------	-----------

nodos	procesadores	ejecución (min)
1	8	15:40
5	40	09:13
10	80	04:06
15	120	01:31

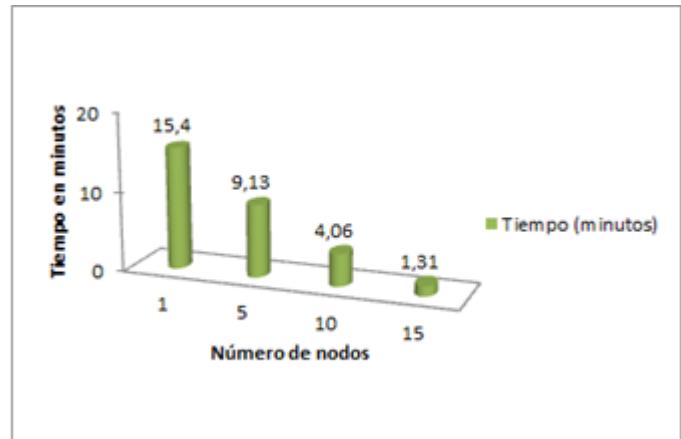


Fig 5. Tiempo de ejecución de renderización de imágenes y videos con Blender.

El tiempo de ejecución utilizando Blender para renderizar el proyecto en un solo nodo es de 15.4 minutos, en cambio utilizando los 15 nodos de la arquitectura el tiempo de renderizado es de 1.31 minutos, ver Tabla III, obteniendo una disminución de 14.09 minutos de tiempo de procesamiento, alcanzando un rendimiento de aproximadamente 92%, como se observa en la Fig. 5.

D. Proyecto 4: Análisis de datos (Big Data) con Apache Hadoop

Para esta prueba se utilizó un dataset sobre la medición de la calidad del aire en España en las estaciones de Castilla y León, en el que se encuentra información recopilada desde el año 1997 hasta el año 2013, el mismo que se encuentra disponible en el siguiente enlace <http://goo.gl/jZO6OT>. El fichero de extensión .csv, fue modificado a nuestros requerimientos con lo que se logró obtener un archivo que contiene más de 16 millones de líneas y un peso mayor a 950 Mb.

TABLA IV. RESULTADOS DE ANÁLISIS DE DATOS (BIG DATA) CON HADOOP.

Número de nodos	Número de procesadores	Tiempo de ejecución (seg)
1	8	92
5	40	78
10	80	54
15	120	32

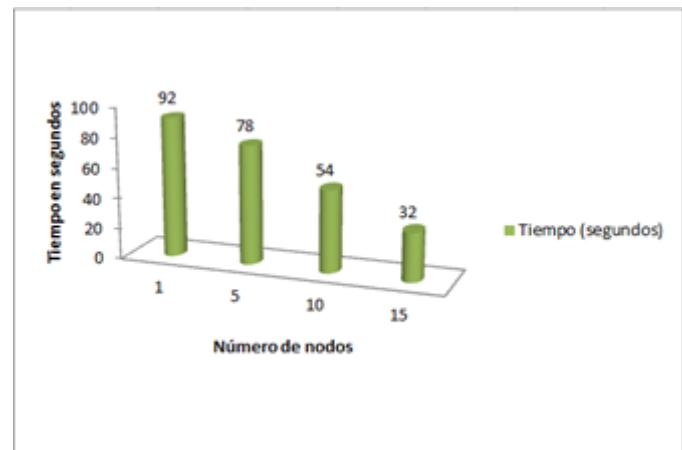


Fig 6. Tiempo de ejecución de análisis de datos (Big Data) con Hadoop

Utilizando el framework Apache Hadoop para el análisis de datos (Big Data), se deduce que utilizando un solo nodo, el tiempo de ejecución es de 1 minuto con 32 seg, mientras que al utilizar los 15 nodos de la arquitectura el tiempo es de 32 segundos de tiempo de procesamiento, como se detalla en la Tabla IV obteniendo una disminución de 60 segundos de diferencia, lo que se interpreta como una eficiencia de aproximadamente 74% de aceleración de los procesos relacionados al análisis de datos, ver Fig. 6.

E. Proyecto 5: Pruebas de compresión de música con MOSIX

En el proyecto realiza la compresión de 32 canciones en formato .wav a un formato de ficheros de audio .flac, haciendo uso de Mosix para la migración automática de los procesos en cada uno de los nodos del clúster. El tamaño de los archivos antes de realizar la compresión es de 5.2 Gb, luego de realizar el proceso el tamaño es de 1.9 Gb.

TABLA V. RESULTADO DE LA EJECUCIÓN, PRUEBAS DE COMPRESIÓN

Número de nodos	Número de procesadores	Tiempo de ejecución (min)
1	8	22:16
5	40	11:23
10	80	09:34
15	120	07:45

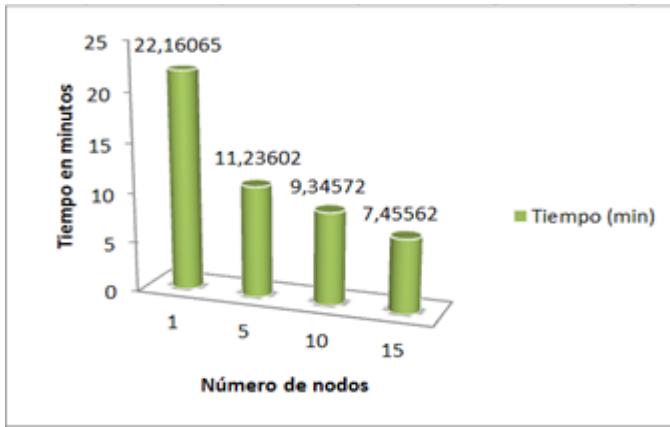


Fig 7. Tiempo de ejecución de pruebas de compresión de música con Mosix

De la prueba realizada con el middleware Mosix para la compresión de los ficheros de audio se deduce que utilizando un solo nodo, el tiempo de ejecución es de 22.16065 minutos, mientras que al utilizar los 15 nodos de la arquitectura el tiempo es de 7.45562 minutos, ver Tabla V, obteniendo una disminución de 14.70503 tiempo de procesamiento, lo que se interpreta como una eficiencia de aproximadamente 75%, ver Fig. 7.

F. Monitorización de la arquitectura clúster

Gracias al sistema de monitoreo Ganglia Monitoring System, tenemos registradas las actividades de cada uno de los nodos de la arquitectura clúster, como se observa en la Fig. 10 permitiendo recolectar métricas como: ocupación de los procesadores, uso de memoria, espacio en disco, etc [15].

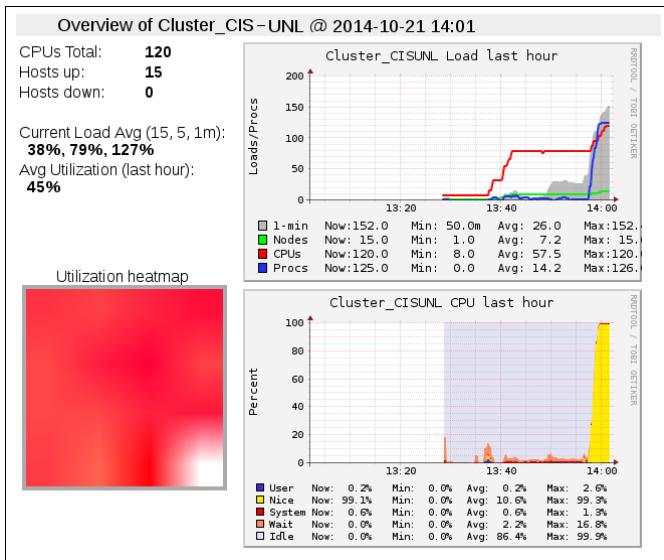


Fig 8. Monitorización arquitectura clúster de alto rendimiento CIS-UNL

V. RESULTADOS

A. Esenario real

Los resultados se obtuvieron en base al siguiente esenario real:

El diseño lógico de la arquitectura clúster fue la siguiente:

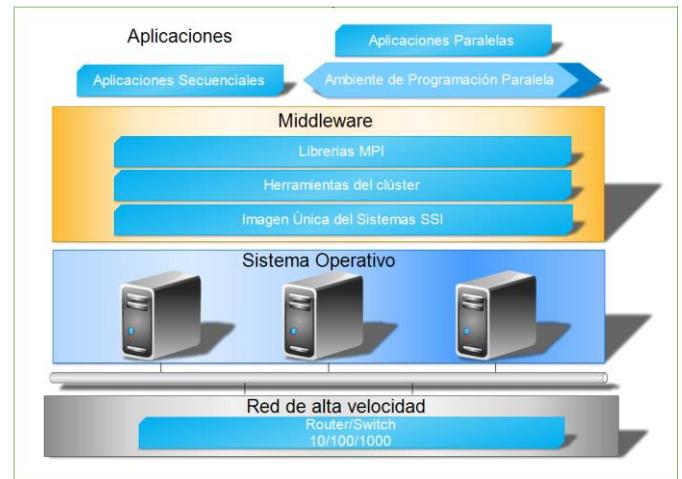


Fig 9. Diseño lógico arquitectura clúster de alto rendimiento CIS-UNL

Para la capa inferior o la red de datos de alta velocidad se utilizaron:

Tarjetas de red PCI Adapter 802.11 que soportan velocidades de hasta 1000 Mbps; un router CISCO LINKSYS EA4500 Giga bit Ethernet con una velocidad de transmisión de 450 Mbps y un Switch: CISCO Catalyst 2960 de capa 2 con 48 puertos Gigabit Ethernet.

La topología de la red implementada fue de tipo estrella ya que con esta topología no es necesario que los nodos esclavos tengan conexión a internet, para poder hacer uso de este servicio en el nodo maestro se instalaron dos tarjetas de red, una para conectarse a la red LAN en la cual están enlazados todos los nodos que componen el clúster y la segunda tarjeta para la conexión a Internet. Ver Fig. 10.

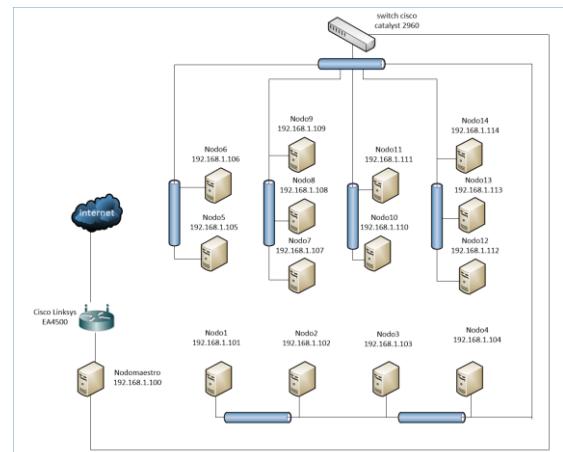


Fig 10. Diagrama de topología de la arquitectura clúster CIS-UNL

Para la capa del sistema operativo y el hardware se utilizó:

El Sistema operativo Debian 7.5 “Wheezy” (64 bits). El hardware del clúster de procesamiento fue armado con 15 computadoras personales Corei7-2600 de 3.64 GHz, memoria RAM 4 GB de 1333 MHz y disco duro de 500 GB con 7200 rpm. Al final los recursos hardware del clúster de procesamiento son los indicados en la tabla VI.

TABLA VI. RECURSOS TOTALES DE LA ARQUITECTURA CLÚSTER

Recursos	Descripción
Capacidad en memoria RAM	60 GB
Número de procesadores	120 de 3.4 GHZ
Capacidad en disco duro	7.32 TB

Para el Middleware se utilizó herramientas de software libre como:

Middleware Mosix 3.4.0.12, Compiladores: GCC 4.5.2. Soporte para C, C++, MPICH 3.1.2, como implementación del estándar Message Passing Interface (MPI), John The Ripper 1.7.9-jumbo-7.

Para la capa superior o de aplicaciones se utilizó herramientas de software libre como:

Blender 2.72b, Apache Hadoop 2.5.1, Ganglia Monitoring System 3.3.8.

B. Evaluación de resultados

Los principales resultados obtenidos son los que se indican en las tablas II, III, IV, V y VI; es fácil deducir entonces que el clúster de procesamiento implementado con los recursos hardware y software antes mencionados cumple con las expectativas académicas para el procesamiento de grandes volúmenes de información logrando con ello los resultados esperados al finalizar el proyecto.

Al no contar con un supercomputador para realizar mediciones de tiempos de procesamiento de grandes volúmenes de información, no fue posible realizar un estudio comparativo con los resultados obtenidos del clúster de procesamiento armado con hardware convencional y software libre por lo que asumimos que el clúster mencionado tiene características técnicas de procesamiento similares a las de un supercomputador.

VI. CONCLUSIONES Y TRABAJOS FUTUROS

La arquitectura clúster tipo beowulf es la más óptima en ambientes universitarios ya que se implementó utilizando hardware convencional y herramientas de software libre, con lo cual se obtiene un ahorro generalizado en costes de administración, mantenimiento y monitorización, comparadas

con la adquisición de un supercomputador. Se debería armar una arquitectura clúster de procesamiento con al menos 15 computadores personales con 8 procesadores de 3.6 GHz cada uno para tener una arquitectura eficiente, ya que al realizar la ejecución de varios tipos de algoritmos complejos como: cifrado de código, renderización de imágenes, análisis de datos (Big Data), compresión de ficheros de audio, esta arquitectura clúster alcanzó un rendimiento del 85.2%, en relación a la capacidad de procesamiento de un solo computador convencional. Además la arquitectura clúster de alto rendimiento fue diseñada para proporcionar capacidad de procesamiento de grandes volúmenes de datos, lo que permite realizar investigaciones en campos como: minería de datos, Big Data, sistemas de gestión de base de datos, evaluación de estructuras de datos, evaluación de algoritmos, programación paralela, reconocimiento de patrones, entre otros. En esta arquitectura clúster se puede ejecutar una amplia gama de aplicaciones de cómputo científico, para la evaluación de métodos numéricos y técnicas de optimización, solucionando problemas principalmente de ciencia e ingeniería.

VII. AGRADECIMIENTOS

Los autores agradecen a la Universidad Nacional de Loja, el Área de la Energía, las Industrias y los Recursos Naturales no Renovables, y a la Carrera de Ingeniería en Sistemas por el apoyo técnico y logístico desinteresado para la culminación exitosa del proyecto. De igual manera los autores agradecen el apoyo académico desinteresado de los docentes investigadores de la Escuela Superior Politécnica de Chimborazo y de la Universidad Técnica de Manabí.

VIII. REFERENCIAS

- [1] G. Cáceres, “Estrategia de implementación de un clúster de alta disponibilidad de N nodos sobre linux usando software libre”, 2012.
- [2] L. M. Santos Jaimes, S. Peñaloza, and E. R. Cruz Cruz, “Cluster implementation of a prototype for the resolution of a particular problem”, Journal Article, vol. 1 2010.
- [3] N. Pérez Otero, S. Méndez, C. V. Ayusa, M. I. Aucapíña, and V. J. Lopez, “Aplicaciones del cómputo de altas prestaciones”, in XI Workshop de Investigadores en Ciencias de la Computación, 2013.
- [4] D. Zhao, K. Qiao, and I. Raicu, HyCache+: Towards Scalable High-Performance Caching Middleware for Parallel File Systems”, in IEEE/ACM CCGrid, 2014.
- [5] J. d. J. R. Quezada, S. B. Rionda, J. M. V. Félix, and I. A. M. Torres, “Diseño e implementación de un clúster de cómputo de alto rendimiento”, Acta Universitaria, vol. 21, pp. 24-33, 2011.
- [6] R. Gualán, A. Vásquez and O. Vega, “Una primera aproximación a la implementación de un clúster para la ejecución de un modelo de predicción climática”, 2012.
- [7] R. Quezada, S. B. Rionda, J. Felix, and I. Torres, “Diseño e implementación de un clúster de cómputo de alto rendimiento”, Acta Universitaria, vol.21, pp. 24-33, 2011.
- [8] A. Barack and A. Shiloh, “The MOSIX Cluster Operating System for Distributed Computing on Linux Clusters, Multi-Clusters and Clouds”, 2014.
- [9] M. C. O. System, “Administrator’s, User’s and Programmer’s Guides and Manuals”, July 2014.
- [10] R. Latham and A. J. Pe, “MPICH Installer’s Guide”, Mathematics and Computer Science Division Argonne National Laboratory, 2014.
- [11] Blender, “Blender is a free and open source 3D animation suite” [Online]. Available: <http://www.blender.org/>. [Accessed: 05-Nov-2014].
- [12] R. Serrat Morros, “Big Data: análisis de herramientas y soluciones”, 2013.

- [13] Apache, "Hadoop - Apache Hadoop 2.5.1." [Online]. Available: <http://hadoop.apache.org/docs/r2.5.1/index.html>. [Accessed: 13-Nov-2014].
- [14] G. M. System, "Ganglia Monitoring System" [Online]. Available: <http://ganglia.sourceforge.net/>. [Accessed: 04-Dec-2014].
- [15] B. Otero, R. Astudillo, and Z. Castillo, "Un esquema paralelo para el cálculo del pseudoespectro de matrices de gran magnitud," Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería, 2014.
- [16] A. Sheharyar and O. Bouhali, "A Framework for Creating a Distributed Rendering Environment on the Compute Clusters", arXiv preprint arXiv:1401.0608, 2014
- [17] R. Bhatnagar and J. Patel, "Performance Analysis of A Grid Monitoring System-Ganglia", International Journal of Emerging Technology and Advanced Engineering, vol. 3, pp. 362-365, 2013.



Leonardo Chuquiguanca, Administrador de Sistemas en la empresa Protelcotsa S.A. Ingeniero en Sistemas graduado en la Universidad Nacional de Loja - Ecuador, Activista de software libre, Administrador de servidores, redes y telecomunicaciones, Seguridad de información. Provincia de Loja, Ciudad Loja, Ecuador, 2015.



Edyson Malla Ingeniero en Sistemas del Gobierno Autónomo Descentralizado Municipal San Felipe de Oña, Azuay – Ecuador (Marzo 2015 hasta la actualidad). Ingeniero en Sistemas graduado en la Universidad Nacional de Loja – Ecuador (Febrero 2015). Conocedor de software libre, redes, telecomunicaciones, análisis y diseño de sistemas. Provincia de Loja, Ciudad Loja, Ecuador, 2015.



Freddy Ajila, Docente Investigador de la Escuela de Ingeniería Industrial, Facultad de Mecánica de la Escuela Superior Politécnica de Chimborazo ESPOCH (Desde Octubre 2014 hasta la actualidad). Profesor de Sistemas Operativos, Arquitectura de Computadoras y Estructuras de Datos de la Carrera de Ingeniería en Sistemas de la Universidad Nacional de Loja – Ecuador (Abril 2013 – Julio 2014). Magíster en Telemática graduado en la Universidad de Cuenca – Ecuador (Agosto 2011). Ingeniero en Informática graduado en la Universidad Técnica Particular de Loja – Ecuador (Junio del 2006). Activista de software libre, Administrador de servidores, redes y telecomunicaciones. Provincia de Chimborazo, Ciudad Riobamba, Ecuador, 2015.



Rene Guaman Quinché Docente Investigador de la Carrera de Ingeniería en Sistemas Informáticos de la Universidad Técnica de Manabí. Experto en tecnologías y accesibilidad Web. Magíster en Sistemas Informáticos Avanzados en la Universidad del País Vasco y candidato a Doctor por la Escuela Politécnica Nacional. Provincia de Manabí, Ciudad Portoviejo, Ecuador, 2015

Intrusive and Non-Intrusive Load Monitoring (A Survey)

Inference and Learning Approach

Danilo Burbano

Abstract—There is not discussion about the need of energy conservation, it is well known that energy resources are limited moreover the global energy demands will double by the end of 2030, which certainly will bring implications on the environment and hence to all of us.

Non-Intrusive load monitoring (NILM) is the process of recognize electrical devices and its energy consumption based on whole home electric signals, where this aggregated load data is acquired from a single point of measurement outside the household. The aim of this approach is to get optimal energy consumption and avoid energy wastage. Intrusive load monitoring (ILM) is the process of identify and locate single devices through the use of sensing systems to support control, monitor and intervention of such devices. The aim of this approach is to offer a base for the development of important applications for remote and automatic intervention of energy consumption inside buildings and homes as well.

Appliance discerns can be tackled using approaches from data mining and machine learning, finding out the techniques that fit the best this requirements, is a key factor for achieving feasible and suitable appliance load monitoring solutions. This paper presents common and interesting methods used.

Privacy concerns have been one of the bigger obstacles for implementing a widespread adoption of these solutions.

The implementation of security over these approaches along with fine-grained energy monitoring would lead to a better public agreement of these solutions and hence a faster adoption of such approaches.

Index Terms—NILM; ILM; supervised learning; unsupervised learning; privacy; self-sensing; sensors; machine learning algorithms; smart meter; security

Resumen—La detección y predicción de la somnolencia es clave para la implementación de vehículos inteligentes destinados a prevenir accidentes en carreteras. Existen varios enfoques para crear este tipo de vehículos. En este artículo se analiza el enfoque de visión por computador, donde dispositivos embebidos son usados conjuntamente con técnicas de aprendizaje automático (i.e. inteligencia artificial) y reconocimiento de patrones para implementar soluciones para la detección del nivel de fatiga de un conductor de un vehículo. La mayoría de investigaciones en este campo basados en visión por computador se enfocan en el análisis del parpadeo de los ojos del conductor, esta solución combinada con patrones adicionales como el reconocimiento del bostezo o el

movimiento de cabeza constituye ser una solución bastante eficiente. El primer paso en este enfoque es el reconocimiento del rostro, para lo cual el uso del algoritmo AdaBoost muestra resultados precisos en el proceso de extracción de características, mientras para la detección de somnolencia el uso de clasificadores como el Support Vector Machine (SVM) muestra también resultados precisos.

Un componente básico en la tecnología de visión por computador es el uso de una base de datos de imágenes espontáneas acorde al Facial Action Coding System (FACS), con la cual el clasificador puede ser entrenado. Este artículo presenta un prototipo sencillo para detección de somnolencia, en el cual el método de Viola-Jones es utilizado para el reconocimiento de rostros y un clasificador tipo cascada es usado para la detección de ojos cerrados en una secuencia continua de imágenes lo que constituye un indicador de somnolencia.

Palabras clave—Somnolencia, Clasificador tipo cascada, método Viola-Jones, FACS, AU

I. INTRODUCTION AND MOTIVATION

ONE of the most important innovations for reduction in the energy wastage is trough monitoring energy consumption. To achieve this, as in [2], research efforts have led to the development of Appliance Load Monitoring (ALM) aimed to perform detailed energy sensing and to provide information on the breakdown of the energy spent. Researches done so far, have split ALM in two approaches namely Intrusive Load Monitoring (ILM) and Non-Intrusive Load Monitoring (NILM), the former referred to a distributed sensing (i.e. inside the household) and the latter to a single point sensing (i.e. outside the household).

NILM examines the specific appliance signatures within the aggregated load data as was initially proposed by Hart [3], whereas the research for ILM approach has been addressed to ubiquitous sensor deployment due to its nature of distributed sensing, thereby allowing to identify NILM with smart meters devices and ILM with smart plugs.

It is important to stand out that there are some projects where the device meter is inside the house but is a single point of sensing at the same time, from an electrical point of view this can be classified as NILM, because are noise-based solution

that monitor the whole current operation of the house, but from a physical point of view this can fall inside ILM approach, for this paper this type of ALM will be classified depending of the propose of the solution, which means NILM for energy consumption and ILM for sensing systems.

NILM as well as ILM share some common characteristics, one of them is the use of machine learning techniques in order to predict the behavior of new appliance and to traduce the raw data (e.g. current waveforms and voltage) in an easy and understandable form, these techniques also allow to made deeper analysis of the electricity consumption in such a way that it is even possible to know the behavior of the consumer, bringing privacy concerns into account.

This paper contribution is to show a general and arranged overview of these approaches, in order to make easy to understand for people not familiar with this terms, the general idea of this solutions, therefore helping those people to know exactly the specific subject to search for, in case of a deeper study is needed.

A main framework has been extracted from these projects for NILM and ILM as well, with focus on the machine learning techniques used for these approaches. This paper also shows the possibility to read the TV content through this techniques (obviously one of the most destabilizing issues that brings privacy concerns) in an attempt to avoid delay in the adoption of these good solutions, because knowing the weakness of this approaches is the starting point for finding a suited solution for this drawback.

The reminder of this paper is organized as follows. First Section 2 gives an overview of the related work about these approaches in energy sensing and inference and learning subjects. Section 3 gives an overview of the NILM and ILM frameworks. Section 4 and 5 introduces the most common machine learning algorithms used for electrical appliance configuration. Section 6 states the privacy concerns about these approaches given real scenarios of security.

II. RELATED WORK

There are several works related to NILM and ILM, in which an assorted of device meters are used for the electrical signal analysis (i.e. energy sensing) and various solutions as well for appliance recognition (i.e. inference and learning).

A. Energy Sensing (Data Acquisition)

As was pointed out NILM is related to smart energy meters (low and high frequency) devices as it can discern devices from the aggregated data acquired from a single point of measurement [2], the aim of measure aggregated load have led to interesting projects like distributed NILM (dNILM) that uses an algorithm capable of running on limited meters with minimal bandwidth while still maintaining a high degree of accuracy [4], others works make use of different technologies like single electricity sensors as Ruzelli et al. that proposed the use of a single wireless energy monitoring sensor attached to the smart meter enabling real-time appliance recognition [6]. Another

interesting work is an iterative way for separating individual appliances from an aggregated load as in [10].

On the other hand most of the work done so far about ILM is related to smart plugs, where technologies such as sensors and wireless are used aimed to get pervasive computing, Elzabadani et al. create a self-sensing space using Radio Frequency Identification (RFID) to recognize a plugged device and Open Services Gateway Initiative (OSGi) platform that facilitates the installation and operation of multiple services (i.e. appliances) on a single device [7]. Wook et al. were one step further and besides RFID, exploited the power draw characteristics of appliances using their unique current consumption as their IDs, along with a self-oriented platform known as Atlas [8]. A quite different work has been proposed by Zufferey et al. using plug-based low-end sensors which can automatically recognize home appliances based on their electric consumption profiles [9]. A very interesting work was made by Patel et al. that present an approach for a low-cost and easy to install power line that uses a single plug-in sensor to detect a variety of electrical events throughout the home [12].

An important research done that take advantage of new technologies as Gupta et al. that uses a noised-based solution through the electromagnetic influence (EMI) generated by switch mode power supplies (SMPS), an innovation found in modern consumer electronics and fluorescent lighting [5], this type of solution should fall inside NILM approach because is a single point of sensing, nevertheless it is also a device use in-home (connected to an outlet) therefore some could see it as ILM, although this can be seen as an hybrid work, in this paper the distinction between ILM an NILM will be made considering the sensing characteristic thus this work is consider in ILM approach.

It is important to stand out that works based in analyze the electrical noise on the power line like [5,12] born as an alternatively solution for ILM due to the high cost of one sensor per appliance, nevertheless this solutions also can be implemented for NILM, because it provides disaggregating electrical information for energy monitoring as claimed by its authors.

B. Inference and Learning (Classify Appliances)

There is an important work made by researches in order to find the best and feasible machine learning techniques used to identify appliance-specific states from the aggregated measurement. Before giving details about this works, it is important to stand out two main important works done over a prerequisite for any machine learning method, which is a database of load measurements, upon inference and learning methods can be applied. Kolter et al. [11] state key properties for such databases based in Reference Energy Disaggregation Data Set (REDD) [13].

According to the literature and works done about machine learning techniques with NILM systems, it is possible to classify in *supervised learning approach* and *unsupervised learning approach*, the former identify manually (i.e. labelling)

the device change-points in a whole home energy signal or the system is trained on individual device power signals, whereas the latter consider the whole home signal without prior information (i.e. not labelling), and automatically separate signals [11].

Supervised learning works can be found in projects like Gupta et al. where events were automatically extracted and sent to a computer with labelling software afterwards inference was made using k-Nearest Neighbors algorithm (k-NN) using a 10-fold cross validation [5]. Ruzelle et al. proposed an approach that empowers users profiling where generated signatures are used to train an Artificial Neuronal Network (ANN) then employed to recognize appliance activities [6]. Kolter et al. monitor each individual circuit in the home focusing in Factorial Hidden Markov Model (FHMM) as a method for disaggregation with Expectation-Maximization (EM) for training [11]. Parson et al. use EM algorithm where data to train as well as appliance model are extracted from the aggregated load, the inference mechanism is done with an extension of the Viterbi algorithm able to iteratively disaggregate individual appliances [10].

Unsupervised learning have interesting works like dNILM approach uses event detection as a part of the learning phase where the meter detect abrupt changes in the power readings in order to classify appliances along with clustering algorithms and dynamic programming [4]. Gonçalves et al. used a genetic and K-means clustering algorithm for state changes and a matching pursuit algorithm to reconstruct the original power signals [15].

ILM approach has limited literature about the machine learning techniques used because in this approach there is often a middleware layer hiding this implementation, but as these solutions identifies each appliance through smart plugs [7,8], this process could be seen as labelling the signal therefore as supervised learning, also as in [9] where appliance recognition is based on supervised learning using two classification algorithms k-NN and Bayesian and in [12] that uses Support Vector Machine (SVM) performing classification by constructing an N-dimensional hyperplane that optimally separates the data into multiple categories. These works helps to infer that the middleware phase also can be seen as the encapsulation of even detection, feature extraction and learning.

III. NILM AND ILM FRAMEWORKS

Before given an explanation about the machine learning techniques used for disaggregation load as well as for learning or classifying the appliances in NLM and ILM approaches, a general overview of the architecture used in these approaches will be made in this section.

A. NILM Framework

The aim is to partition the whole house or building load energy data into its major constituents and estimate their power consumptions based on current voltage measurement of the

main circuit entering a house. Figure 1 illustrates this approach summarized in phases or layers.

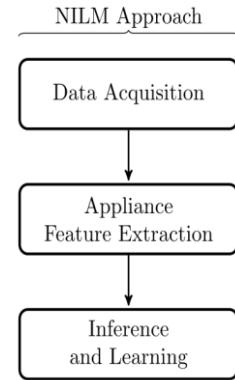


Figure 1. NILM framework

- DATA ACQUISITION phase acquires load measurement at an adequate rate so that distinctive load patterns can be identified. All smart meters nowadays at least have this feature.
- APPLIANCE FEATURE EXTRACTION phase processes the raw data (i.e., voltage and current waveforms) in order to compute the power metrics (e.g., active and reactive power). Afterwards detect events such as appliance state transition (e.g., On to OFF) from the power measurements. An event detection module detects the ON/OFF transition of appliances by analyzing the changes in power levels. A proper goal of this phase would be to get an accurate appliance database or dataset as an input for the next phase.
- INFERENCE AND LEARNING phase analyzes the outputs of the previous phase. This phase can be divided in two approaches: *load identification* and *system training*, the former identify appliance-specific states from the aggregated measurement, the latter aim to training methods and learning of appliance features. As was pointed out in the Section 2, the main methods used for reach this aims are supervised and unsupervised algorithms.

B. ILM framework

Currently the aim of this approach is to pervasive computing (i.e. ambient intelligence, self-sensing spaces), where the main idea is to detect every single appliance with a smart plug, then middleware solutions are used to integrate and coordinate the whole nodes thus achieving a real time status and management of the household. Figure 2 illustrates this approach summarized in phases or layers.

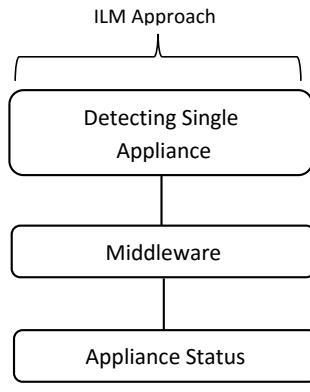


Figure 2. ILM framework

- DETECTING SINGLE APPLIANCE phase where using sensor technologies like smart plugs it is possible to identify the type as well as the location of the device that have been plugged. ILM projects have been used different technologies for this purpose like RFID and X10¹ protocols.
- MIDDLEWARE phase is related to the need of abstraction for every appliance by represent each physical device as a service in order to reach automatic identification and self-integration of appliances. The middleware solutions identified in works made so far are Atlas and OSGi platform. The personalized software used in projects as in [12] could also be considered in this phase.
- APPLIANCE STATUS phase is the final output of ILM, a user friendly interface that shows the real-time status of the household for control, intervention and monitoring.

A better understanding for achieving pervasive computing through ILM approach is illustrated in Figure 3, where there are three main components: i) Home Network, which is a wireless ad-hoc network of small devices. ii) Middleware, which is an orchestrator (i.e. coordinator system) of the nodes which compound the Home Network and provides the services (e.g. web services) to control such devices iii) Internet, the remote access that allows users to interact through an application (e.g. web pages, Android application) with their Home Network (i.e. smart house).

It is important to stand out that this approach originally does not consider devices connected directly to the live wire (i.e. non-plugged devices) such as light on a wall. This problem was stated in [8] although the detailed solution was not mentioned, but considering this, it is possible to infer that the need of measuring the current consumption showed up in such approaches. Thus analyze the signature of the current draw for identify the device with certainly addressed ILM to some machine learning techniques used in NILM and some solutions implemented in NILM that obtained the aggregated load from

the live wire [6,11] can be projected to ILM as was already done in [9,5,12]. Works of machine learning techniques used in such works could be fall inside the Middleware phase.

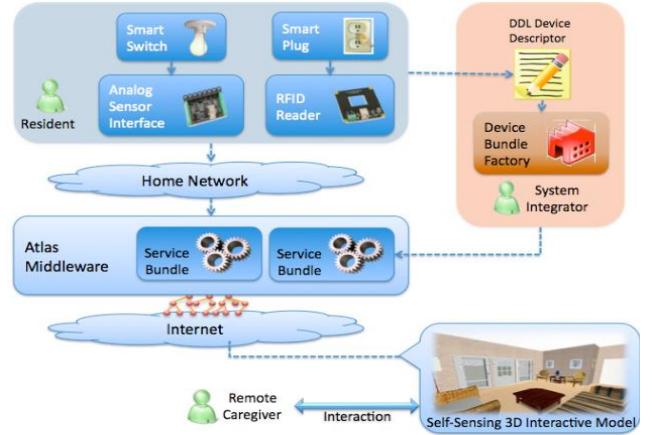


Figure 3. Pervasive computing through ILM approach

IV. DATABASE ESTABLISHMENT

The most important factor of enabling and advancing machine learning methods in any problem is the availability of proper databases.

A good reference is REDD, where the datasets can be categorized in high frequency (15kHz) for whole home data and low frequency (1 Hz) for individual circuit and plug data, the former can be used for NILM and the latter for ILM.

For supervised methods a good practice for establishment of a database would be to have tree dataset: (1) training set, (2) validation set and (3) test set.

New data has to be easily incorporated into the database for developing new methods and evaluating existing ones. On the other hand, the database should also enable the applications and innovations of various methods based on different features. Considering all these factors key properties such as informative, diverse and scalable should be defined in a database.

A. Informative

This property focuses on availability of information in the recordings of a database, such that a database is more informative if the data it contains can serve as training sets and test sets for more methods.

B. Diverse

The performance of a machine learning method is strongly influenced by within category differences and across category differences. A database has to be diverse enough to enable machine learning methods to capture these differences.

C. Scalable

¹ X10 is a protocol for communication among electronic devices used for home automation

New data has to be easily incorporated into the database. As the amount of data increases, the burden of using the database should be kept low, such that data processing, method modification, and comparison between methods before and after adding new data are easy.

A comprehensive database can be established gradually by merging data sets from different sources like single appliance data sets and whole main circuit data sets [13].

V. MACHINE LEARNING TECHNIQUES

As was pointed out in the Section 2 there are several machine learning techniques used for classification and estimation that can be broadly be divided in supervised and unsupervised learning. The former will be explained in subsection B and C, the latter in subsection D, E, F, G.

A. Problem Description

Formally, the aim of NILM is as follows. Given a discrete sequence of observed aggregate power readings $x = x_1, \dots, x_T$, determine the sequence of appliance power demands $w^{(n)} = w_1^{(n)}, \dots, w_T^{(n)}$, where n is one of N appliances.

Alternatively, this problem can be represented as the determination of appliances states, $\mathbf{z}^{(n)} = z^{(n)}, \dots, z_T^{(n)}$, if a mapping between states and power demands is known.

Each appliance state corresponds to an operation of approximately constant power draw (e.g. ‘on’, ‘off’ or ‘standby’) and t represents one of T discrete time measurements [10].

B. Training using Aggregate Data (Supervised Learning)

The most often model used is hidden Markov model (HMM) where data is represented as a sequence of observations, as can be seen in Figure 4.

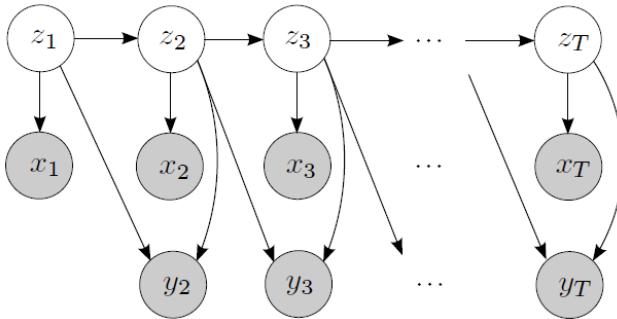


Figure 4. HMM variant. Shaded nodes represent observed variables and unshaded nodes represent hidden variables

z_t represents the state of appliance (e.g. on/off) at an instant time whereas x is an observation sequence (e.g. vector of electrical parameters such as active and reactive power) corresponding to the house aggregate power demand measured by the smart meter and y corresponds to the difference between two consecutive aggregate power readings such that $y_t = x_t - x_{t-1}$.

These derived observations are used to infer the probability that a change in aggregate power, y_t , was generated by two consecutive appliance states [10]. It is important to stand out that the standard HMM model does not consider the y observation.

This model has been translated from a simpler model of state transitions like the one shown in Figure 5a.

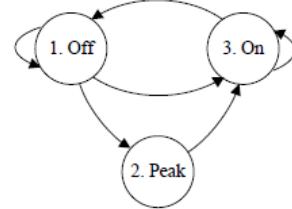


Figure 5a. State transition model

Supervised training methods use prior knowledge of appliance behavior and power demands, the generic model of an appliance type consists of prior over each parameter (i.e. active and reactive power) of an appliance’s model. This prior is simply an expected normal operation of a device, as represented in Figure 5b.

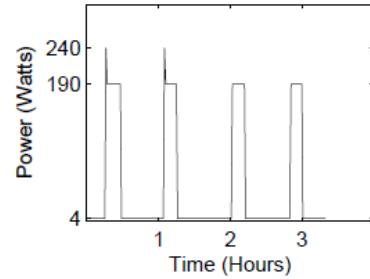


Figure 5b. Prior of appliance power

An appliance prior should be general enough to be able to represent many appliance instances of the same type (e.g. all refrigerators) [10]. From this information (Figure 5a and Figure 5b), values for the appliance’s prior can be inferred as shown in Figure 5c.

z_t	μ	σ^2
1	4 W	10
2	240 W	150
3	190 W	100

Figure 5c. Emission density

These prior models could be determined in a number of ways. Most directly, a domain expert could estimate an appliance’s emission density using knowledge of its power demand available from appliance documentation. Alternatively, these parameters may be constructed by generalizing data collected from laboratory trials or other sub-metered homes [10].

As supervised method requires a clean dataset for training the classifier, in order to achieve this many projects focus on exploit periods during which a single appliance turns on and off without any other appliances changing state, because this allow to have a signature in the aggregate load which is unaffected by all other appliances apart from the baseline load.

Figure 6 shows an example of the aggregate power demand. From hours 1 to 3 it is clear that only the refrigerator is cycling on and off. This period can be used to train the refrigerator appliance model, which is then used to disaggregate the refrigerator's load for the whole duration [10].

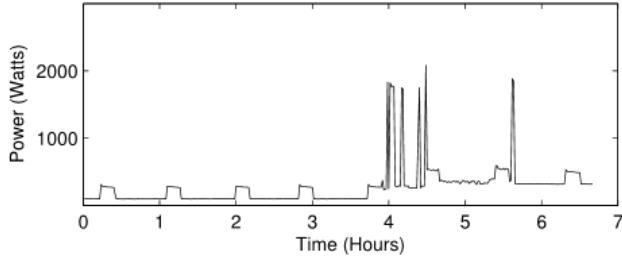


Figure 6. Example of aggregated power demand

C. Energy Disaggregation (Supervised Learning)

Subtracting the refrigerator's load will consequently clean the aggregate load allowing additional appliances to be identified and disaggregated. This subtraction can be achieved by running the EM algorithm on small overlapping windows of aggregate data.

The disaggregation task aims to infer each appliance's load given only the aggregate load and the learned appliance's parameters (Figure 5c). After learning the parameters for each appliance which we wish to disaggregate, any inference mechanism capable of disaggregating a subset of appliances could be used, for instance k-NN, EM, Viterbi algorithms. In general these algorithms can be found in machine learning literature, thus research about machine learning techniques that best fit the requirements of this type of aggregated load data has been the focus of research about NILM.

D. Event Detection Algorithm (Unsupervised Learning)

As a part of both the learning phase and the monitoring phase, the first goal of the meter is to detect abrupt changes in the power readings which correspond to loads changing state. Thus, the event detector produces an output whenever there is sufficient change from the current power value and by ignoring minor changes, meter readings collected every second can be represented in a compressed form. The input is a series of power data tuples like {time stamp, real power value}.

The algorithm uses a series of tuples as the input and creates a series of tuples with the same format for the output. These output tuples represent sudden changes in power, produced when an appliance changes state. Figure 7 illustrates the output

produced by running the event detector algorithm against lab data.

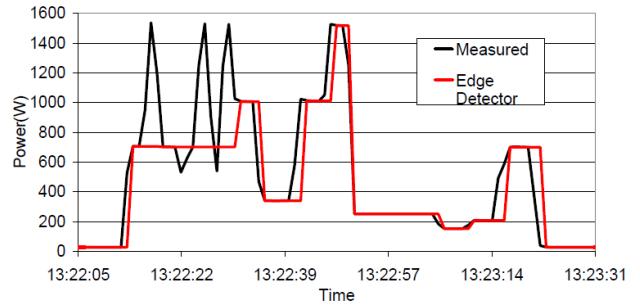


Figure 7. Event Detection Algorithm

E. The Learning Phase (Unsupervised Learning)

Power changes must be identified in order to classify what large appliances are present and group each power change with the respective load. Once the loads are identified, they are placed in a static table and returned to the meters to be used during the real-time phase.

Considering that the most basic load is on/off model, the first step to analyze the data on the backend controller is to establish clusters of on and off events in order to begin identifying what appliances are being seen by the meters.

F. Clustering Algorithm (Unsupervised Learning)

This algorithm accomplishes the establishment of clusters of on and off events by taking an interval of data and grouping like events by their respective power changes. The algorithm retrieves the first element of the event list, determines the power change, and then searches the rest of the array for matching events by assigning the first event to a new cluster. The algorithm also returns how many times each cluster event was turned on or off during the collection period.

It is important to stand out that in real world scenarios, not all loads can be explained by just on/off states, therefore some solutions for accurately identify present loads is building finite state machines (FSM) based upon genetic algorithms with dynamic programming, which means that the table returned to the meters should be a dynamic one.

G. When to Learn (Unsupervised Learning)

The schedule for learning consists of both reactive learning initiated from the meters and proactive learning initiated from the backend controller. When either detects the need for learning it informs the other and the meters begin to transmit data until the controller indicates that it has sufficient information to build a static table. Figure 8a shows an example of a static table rebuilt after a new appliance has been detected as is illustrated in Figure 8b.

Appliance	#States	State 1 Real Power (W)		State 2 Real Power (W)	
		Mean	Std	Mean	Std
(A) Dryer	2	5594	200	2510	90
(B) Oven 1	1	5733	65	NA	NA
(C) Oven 2	1	2215	74	NA	NA
(D) Garage	1	961	55	NA	NA
(E) Toaster	1	1509	45	NA	NA

Figure 8a. Static table

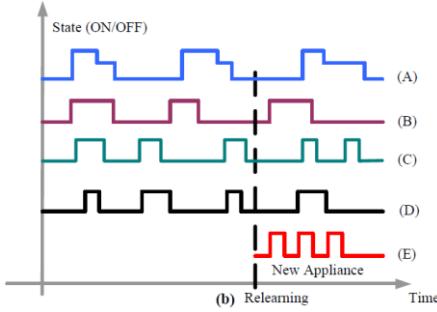


Figure 8b. Appliance States

H. Comparison of load disaggregation algorithms

As was stated in different sections of this paper, there are several algorithms used for the task of disaggregation from the aggregated load, this techniques can be applied to NILM and ILM approach as well, in order to make easy further research of this approaches a summarized and categorized list of this algorithms is done in Table I [2].

TABLE I. Disaggregation Algorithms

Learning Algorithm	Accuracy %	Training S ^a /U ^b	Real-Time
SVM	75-98	S	Yes
Bayes	80-99	S	Yes
HMM	75-95	B	No
Neural Networks	80-97	S	Yes
k-NN	70-90	S	Yes

a. Supervised. (S)

b. Unsupervised. (U)

c. Both (B)

VI. PRIVACY CONCERN

Concerns about NILM and ILM approaches showed up since smart meters are able to become surveillance devices that monitor the behavior of the customers, which could lead to unprecedented invasions of consumer privacy as in [14] where issues about security and privacy are shown.

A. Data Transmission

An experiment was done with a real provider of smart meters, Discovery². All data are transmitted (Figure 9a) to the servers hosted by Discovery, and then customers are able to access these data via a web-browser.

```
POST /api/w.html HTTP/1.1
Content-Type: application/x-www-form-urlencoded
Host:85.214.93.99
Content-Length:851
Content-Type:application/x-www-form-urlencoded
version=0.9&identity=[REDACTED]&msg=228601&values=[{"meterdata":"00000285.9823514*kWh","tckdelta":"00000285.9822239*kWh","seconds":"399511319.61"}, {"meterdata":"00000285.9824793*kWh","tckdelta":"00000285.9823514*kWh","seconds":"399511321.61"}, {"meterdata":"00000285.9826075*kWh","tckdelta":"00000285.9824793*kWh","seconds":"399511323.61"}, {"meterdata":"00000285.9827358*kWh","tckdelta":"00000285.9827358*kWh","seconds":"399511325.62"}, {"meterdata":"00000285.9829915*kWh","tckdelta":"00000285.9828636*kWh","seconds":"399511329.62"}, {"meterdata":"00000285.9831196*kWh","tckdelta":"00000285.9829915*kWh","seconds":"399511331.62"}, {"meterdata":"00000285.9832476*kWh","tckdelta":"00000285.9831196*kWh","seconds":"399511333.62"}]
&now=399511335.65
```

Figure 9a. Captured communication between smart meter and server

The energy consumption data are saved in a text file format, while being transferred to the central servers. The unencrypted data can be easily hacked out; in addition none of the data are signed. The identity (highlighted in black in Figure 9a) of any smart meter is immediately revealed when the data are being transmitted to the central servers and could be used by an attacker to send different power consumption data to the server.

A java-script based application requests the data from the Discovery server and offers the visualization of the profile. A typical profile example can be seen in Figure 9b.

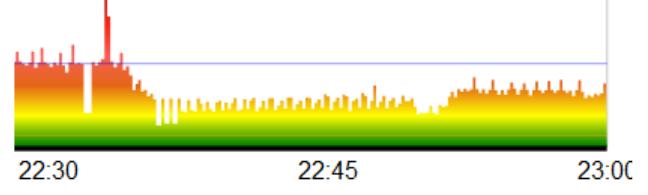


Figure 9a. Captured communication between smart meter and server

B. TV/Film Detection

The test was made with an LCD television which uses the display technology to produce colored images. The core of the content identification program detailed in [14] is the power consumption prediction function. The input of the function is the multimedia content; the output is power usage prediction as would be displayed by a smart meter.

The first step is to measure the power consumption for a series of pictures consisting of elementary shades. The additive RGB color notation with one byte (i. e. values 0–255) is used that increase the brightness from black to white running over 254 shades of gray. The experiment shows that maximum power consumption is reached with rather dark pictures (e. g., RGB 32-32-32) but this also depends on the television user settings. The next step is to extract frames from the movie and calculate the brightness of each frame.

As the experimental results were obtained with a smart meter operating on a two-seconds interval, calculations were made with an average value of power consumption for a number of consecutive frames adding up to two seconds of a movie, which means 50 frames for a movie with a typical 25 frames per second (fps) rate. Figure 4 shows the final result.

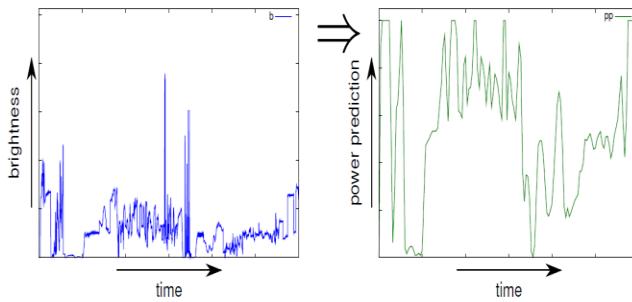


Figure 10. Power prediction is computed on frame brightness values

Experiments done with 653 content files allowed to state that detection of movies produced for cinema projectors was almost always a feasible task while many TV studio productions (e. g. talk shows, news) are difficult or impossible to identify when played as recorded content. Though it is still rather easy to determine which TV broadcast consumers are watching because just five minutes of consecutive playing of a movie is in many cases sufficient to identify the viewed content by analyzing the smart meter power consumption data along with the content of live stations besides Greveler et al. showed up that the content identification is also possible with CR tube and plasma TVs [14].

C. Security over NILM

As was shown in the subsection above, the need of security over this approach is really necessary, this is acknowledged for the scientific community, therefore there are currently two approaches of implementing privacy preserving smart meter data analysis.

The first approach relies on masking the meter readings, and the second approach relies on homomorphic encryption detailed information can be seen in [14].

VII. CONCLUSION

The advancements in the works of NILM and ILM are increasing diary as any other technology, it certainly will lead us to a broadly adoption of smart meters and self-sensing spaces once fine-grained energy monitoring is achieved, although there are public concerns about the use of this information considering the possibility of a misuse of this data by thirds parties chasing the electricity consumption of a customer is latent, because it can be used for marketing attack and even worst government or malicious eavesdropping. This opens the door to an inevitable and needed research of security over NILM and ILM.

Proper datasets is a prerequisite for enabling and advancing machine learning methods, therefore availability of those databases is transcendental to reach feasible machine learning techniques.

There are several machine learning methods used in for NILM and ILM for recognizing appliances, the most promised result are in the Neural Networks and Bayes approaches, although the works in variants of HMM, clustering and genetic algorithms focus on unsupervised learning are very important as it does not imply a training phase.

According to the information that has been pointed out in this paper, it is possible to infer that researches used some machine learning techniques to predict the power consumption from the brightness of a TV and in the same way it would be possible to made an inverse predicted function (i.e. from the power predict the brightness) thus allowing to identify the TV watching habits from a person with NILM or ILM approaches, another way to get this information from the customer would be having a database of the power metric movies (predicted from its brightness) and match with the power metrics done by the smart meter.

Making a projection of what can be accomplished for, after a fine-grained monitoring and control of appliance using ILM is a Smart House to provide healthcare, convenience, entertainment, energy efficiency and security being gentle with the environment at the same time. The same projection can be made by NILM approach that is and will be a key component of Smart Grid to improve the efficiency and sustainability of the production and distribution of electricity avoiding the wastage of electricity and thus improving the consumer economy and even most important helping to have a more health environment.

According to the information presented in this paper and the conclusions that have been made, it would be very important that most of the future work and research about NILM and ILM solutions focus on mainly two approaches:

1) *Security*: Methods of security over these approaches in order to find feasible solutions like encrypted data, which allow offering a reliable service and therefore the people approval for these solutions.

2) *Unsupervised learning*: The use of this machine learning technique, certainly will lead to a faster adoption of these approaches because as it does not need setup of instruments and human intervention for learning, it allows an efficient use of resources and hence a feasible solution from an economic point of view.

ACKNOWLEDGMENT

Thanks to all the scientific community that had been involve in the works and projects about NILM and ILM, all those researches indirectly helped to accomplish this work.

REFERENCES

- [1] Uteley, J.; Shorrock, L. Domestic Energy Fact File 2008; Technical Report for Building Research Establishment: Garston, UK, 2008.
- [2] Ahmed Zoha, Alexander Gluhak, Muhammad Ali Imra & Sutharshan Rajasegarar; Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing:A Survey, 2012.
- [3] Hart, G.W. Nonintrusive appliance load monitoring. IEEE Proc. 1992, 80, 1870–1891.

² <https://discovergy.com>

- [4] David C. Bergman, Dong Jin, Joshua P. Juen, Naoki Tanaka, Carl A. Gunter, Andrew K. Wright; Distributed Non-Intrusive Load Monitoring, 2011.
- [5] Sidhant Gupta, Matthew S. Reynolds, Shwetak N. Patel; ElectriSense: Single-Point Sensing Using EMI for Electrical Event Detection and Classification in the Home, 2010.
- [6] A.G. Ruzzelli, C. Nicolas†, A. Schoofs, and G.M.P. O'Hare; Real-Time Recognition and Profiling of Appliances through a Single Electricity Sensor.
- [7] Hicham Elzabadani, Abdelsalam (Sumi) Helal, Bessam Abdulrazak and Erwin Jansen; Self-Sensing Spaces: Smart Plugs For Smart Environments, 2009.
- [8] Jung Wook Park, Chao Chen, Hicham Elzabadani, Sumi Helal; SmartPlug: Creating Self-Sensing Spaces using Atlas Middleware, 2009.
- [9] Damien Zufferey , Christophe Gisler, Omar Abou Khaled, JeanHennebert; Machine Learning Approaches For Electric Appliance Classification, 2012.
- [10] Oliver Parson, Siddhartha Ghosh, Mark Weal, Alex Rogers; Non-intrusive Load Monitoring using Prior Models of General Appliance Types, 2012.
- [11] J. Zico Kolter, Matthew J. Johnson; REDD; A Public Data Set for Energy Disaggregation Research, 2011.
- [12] Shwetak N. Patel, Thomas Robertson, Julie A. Kientz, Matthew S. Reynolds, and Gregory D. Abowd; At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential, 2007.
- [13] Po-Hsiang Lai, Mark Trayer, Sudhir Ramakrishna, and Ying Li; Database Establishment for Machine Learning in NILM Power Line.
- [14] Ulrich Greveler, Peter Glosekotterz, Benjamin Justusy, Dennis Loehry; Multimedia Content Identification Through Smart Meter Power Usage Profiles, 2011.
- [15] Hugo Gonçalves, Adrian Ocneanu, Mario Bergés; Unsupervised disaggregation of appliances using aggregated consumption data, 2011.



Danilo Burbano Acuña received his B.S. degree in systems engineering from National Polytechnic School, Ecuador in 2006. From 2006 to 2013 he was a system analyst and software developer in three different enterprises. Starting in a banking software development company, then in a massive beverage production company and then in a government company which controls and administrates the taxes around Ecuador. He is currently pursuing the M.S. degree in computer science with a specialization in distributed systems in a Swiss joint master program at the University of Bern, Neuchâtel and Fribourg. He has been coauthor of papers regarding the internet of things and is currently working in his master thesis about indoor tracking. His research interest

Published by:

National Polytechnic School
Faculty of Systems Engineering
Department of Informatics and Computer Sciences
Ecuador

<http://lajc.epn.edu.ec/>
lajc@epn.edu.ec

May 2015

