

CardioGrid: una tecnología innovadora para el estudio de enfermedades cardiovasculares

D. Fernandez Slezak², P. Turjanski², J. Monetti¹, M. Risk²,
C. García Garino¹ y G. Marshall²

¹*LAPIC, Instituto Tecnológico Universitario, Universidad Nacional de Cuyo*

²*Laboratorio de Sistemas Complejos, Depto. de Computación, FCEyN, UBA.*

13 de agosto de 2008

Resumen

En los países desarrollados se evidencia una migración masiva del entorno académico e industrial hacia la utilización en investigación y desarrollo del modelo de eScience y Grid Computing. Basado en esta percepción surge en nuestro medio, CardioGrid, un proyecto original, que consiste en la construcción de un Portal Grid que permite introducir información sobre pacientes y sus respectivas fichas médicas, en particular electrocardiogramas (ECG), y el eventual procesamiento de los mismos mediante algoritmos especialmente desarrollados en un entorno de infraestructura distribuida (Grid) y de cálculo paralelo (MPI). De esta manera, Cardiogrid brinda a la comunidad médica un enorme poder de cómputo y almacenamiento distribuidos a lo largo de varios hospitales o centros de salud. En particular, Cardiogrid incorpora un original algoritmo de procesamiento paralelo de ECG's con el cual se obtienen resultados preliminares promisorios. Asimismo, Cardiogrid está diseñado para una integración al proyecto denominado Microsoft - Digital Inclusion Kit for Health and Higher Education, que resuelve el problema de la última milla, consistente en la captura de datos por medio de tecnología mobile. De esta forma el sistema integra todas las etapas del problema: captura de datos, centralización, distribución, procesamiento y acceso a resultados.

keywords: Grid Computing, Portal Grid, HealthGrid, ECG

1. Introducción

El término *HealthGrid* identifica un grupo de proyectos de investigación y desarrollo en el ámbito de la computación aplicada a la biología y medicina, que utilizan la tecnología Grid para resolver tareas relacionadas. En este contexto se encuentra CardioGrid, un original proyecto que se ha comenzado a desarrollar en el Laboratorio de Sistemas Complejos (LSC) en colaboración con el Laboratorio de Investigación y Desarrollo para la Producción Integrada por Computadora. (LAPIC) del Instituto Tecnológico Universitario (ITU) de la Universidad Nacional de Cuyo (UNCuyo) en Mendoza, que consiste en la construcción de un Portal Grid para que médicos e investigadores puedan acceder a una base de datos médica para introducir información sobre pacientes y sus respectivas fichas médicas, centrado específicamente en electrocardiogramas (ECG), permitiendo procesar estos últimos con un conjunto de algoritmos de tratamiento de señales biomédicas, pudiendo almacenar y visualizar en línea el resultado de este procesamiento [1].

El procesamiento de un ECG, por lo general, no consume grandes cantidades de poder de cómputo. Sin embargo, cuando la cantidad de pacientes alcanza un número elevado y se realizan estudios de tiempos muy largos, el tiempo de procesamiento computacional pasa a tener un rol fundamental para el buen funcionamiento de este portal. Por tal motivo es que se propone como solución la instalación de un portal basado en tecnología Grid Computing [2-5], que posee la escalabilidad necesaria para el proyecto. Por otro lado procesar tanto los datos como las señales con los mismos algoritmos a lo largo de una red de hospitales, permite homogeneizar los resultados y de esta forma posibilita una comparación correcta de los mismos.

Existen varios proyectos con características similares a CardioGrid orientados a otras áreas de la medicina, entre los cuales podemos mencionar a las iniciativas internacionales MammoGrid [6] y caGrid [7], entre otros. Ambas utilizan la tecnología Grid basada en Globus Toolkit, tecnología adoptada para CardioGrid.

2. Funcionalidades provistas

El LSC-LAPIC se encuentran desarrollando herramientas de software basadas en tecnologías de Grid Computing que permitirán a los grupos de investigación integrarse como nodos del nuevo desarrollo. Una vez conectados a la infraestructura Grid Computing, los investigadores podrán acceder y compartir datos experimentales, herramientas de análisis de datos y modelos computacionales relacionados al funcionamiento del corazón de individuos sanos y de aquéllos con problemas cardíacos. Para protección de la privacidad, ningún dato contendrá información del paciente del cual fue obtenido.

Dentro de las funcionalidades que ofrecerá CardioGrid podemos mencionar las siguientes:

- Administración de entradas y salidas de usuarios del sistema (médicos e investigadores) con estrictos métodos de autenticación y seguridad de datos.
- Almacenamiento de información del paciente y sus correspondientes ECG. Esta información actualmente se está generando a través del proyecto “RFP 2006- MSR Digital Inclusion in Health and Higher Education^[8]”, a cargo de otro equipo del LSC^[8]. Este proyecto consiste en un sistema de toma y almacenamiento de información de pacientes y sus respectivos electrocardiogramas utilizando equipos portátiles con tecnología mobile.
- Procesamiento en línea de las señales biomédicas mediante algún algoritmo seleccionado. El portal presenta algoritmos clásicos de la literatura, así como métodos originales desarrollados dentro del LSC^[9].
- Visualización de la información del paciente así como su ECG, tanto procesado (con cualquiera de los algoritmos) como aquel sin procesar.
- Descarga de información del paciente y su ECG.
- Homoginización de la estructura de datos, señales e imágenes, para facilitar el estudio posterior de minería de datos.

El acceso a la información y todas las funciones del proyecto estarán disponibles a través de un portal público, accesible desde Internet, permitiendo un uso global y masivo de esta

herramienta.

3. Arquitectura

El portal Grid está basado en el paquete Open Grid Computing Environments Portal (OGCE) [10]. OGCE es un conjunto de herramientas que facilita el desarrollo de portales grid web, basado en la tecnología Gridsphere [11]. Entre estas herramientas podemos mencionar el manejo de credenciales Grid, administración de archivos, ejecución remota utilizando protocolos de seguridad y monitoreo de la infraestructura grid (por ejemplo, uso de cpu o cola de espera).

Para el desarrollo de CardioGrid se desarrollaron complementos para OGCE (Portlets [12]), de uso específico que se encargan del manejo de datos y administración del portal. Los datos son guardados en una base de datos Postgres utilizando un servicio grid de Globus Toolkit 4 (GT4) [13] desarrollado especialmente para este sistema. El diseño de la base de datos fue pensado para ser compatible con la integración del proyecto “Digital Inclusion - Microsoft RFP”, y así poder incorporar la información que está siendo generada por dicho proyecto.

En la figura 1 se muestra un esquema de la arquitectura del sistema.

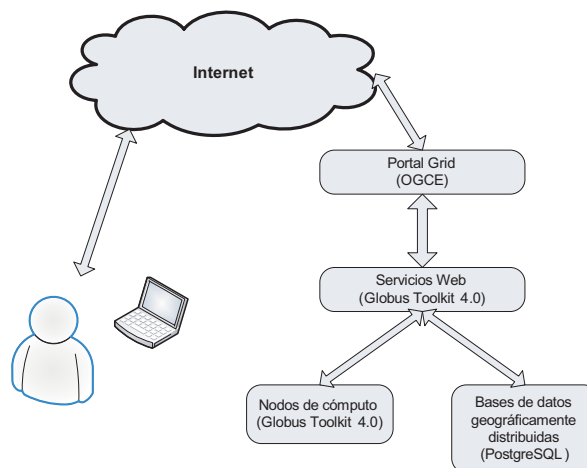


Figura 1: Esquema de alto nivel de la arquitectura del proyecto CardioGrid

El portal provee una interfaz para el desarrollo de nuevos algoritmos de procesamiento de electrocardiogramas (denominados plugins). Estos son ejecutados y procesados en las

unidades de cómputo distribuidas mediante GT4. Las unidades de cómputo cuentan con las bibliotecas para cálculo en paralelo, *Message Passing Interface*. Estas unidades de cómputo pueden provenir tanto de recursos dedicados como no dedicados.

Una vez obtenido el ECG procesado, la información es almacenada en la base de datos del sistema, a través de un servicio Web que permite mantener consistencia en múltiples bases de datos geográficamente distribuidas, con replica de información.

4. Resultados preliminares

Para realizar las primeras pruebas del sistema, se propuso utilizar el plugin presentado en [9] sobre procesamiento de series de tiempo RR. Este plugin fue desarrollado utilizando bibliotecas de cálculo paralelo MPI [14]. Las ejecuciones fueron realizadas en un cluster de 18 x Dual Pentium 4 (2.66 GHz), 2GB RAM, 2 x 250 GB HD. El archivo de prueba utilizado fue de aproximadamente 24 horas continuas de ECG obtenidas de un paciente anónimo. El speedup y eficiencia del procesamiento en paralelo se muestran en la figura 2. Como puede observarse, la utilización de procesamiento en paralelo disminuye el tiempo total de cómputo. Sin embargo, el speedup dista significativamente del ideal y la eficiencia disminuye con el número de nodos. Ambos pueden mejorarse aumentando la longitud de la serie temporal procesada en cada nodo, tarea que queda pendiente. Obviamente, para obtener escalabilidad, la eficiencia debe permanecer constante cuando se aumenta el número de nodos. De todas maneras y sin lugar a dudas, la naturaleza intrínseca del cálculo de la serie temporal, llevará a un algoritmo altamente escalable.

5. Conclusión

El proyecto CardioGrid es un desarrollo Grid original que integra múltiples herramientas de esta plataforma. Además acerca y pone a disposición de la comunidad de médicos e investigadores una herramienta de suma utilidad que les permitirá encontrar nuevas maneras para la prevención, detección y tratamiento de situaciones de riesgo de vida en enfermedades cardiovasculares en zonas geográficamente distribuidas, de una forma sencilla y amigable.

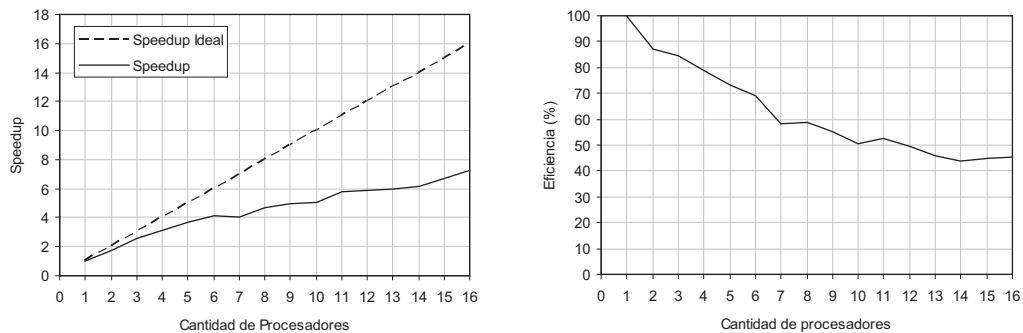


Figura 2: Speedup y eficiencia del procesamiento en paralelo de series de tiempo RR.

Cardiogrid fue desarrollado para su integración al proyecto Microsoft - Digital Inclusion Kit for Health and Higher Education, que resuelve el problema de la última milla, consistente en la captura de datos por medio de tecnología mobile para la prevención de la salud. De esta forma el sistema involucraría a todos los aspectos del problema: captura de datos, centralización, distribución, procesamiento y acceso a resultados.

Se pudo incorporar al portal CardioGrid un plugin de procesamiento de ECG ya publicado en la literatura, obteniendo resultados preliminares promisorios, mostrando que este tipo de herramientas puede ser de suma utilidad para el cómputo de grandes cantidades de datos.

Referencias

- [1] Laboratorio de sistemas complejos. Cardiogrid <http://www.lsc.dc.uba.ar/hpc-grid/cardiogrid>
- [2] Ian Foster. What is the grid? - a three point checklist. *GRIDtoday*, 1(6), July 2002.
- [3] Ian Foster and Carl Kesselman. *The Grid 2: Blueprint for a New Computing Infrastructure*. The Morgan Kaufmann Series in Computer Architecture and Design. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, November 2003.
- [4] I. Foster, C. Kesselman, J. Nick, and Tuecke S. The physiology of the grid: An open grid services architecture for distributed systems integration. In *Global Grid Forum*, 2002.

- [5] Ian Foster, Carl Kesselman, and Steven Tuecke. The anatomy of the grid: Enabling scalable virtual organizations. *Int. J. High Perform. Comput. Appl.*, 15(3):200–222, August 2001.
- [6] R Warren, T Solomonides, C del Frate, I Warsi, J Ding, M Odeh, R McClatchey, C Tromans, M Brady, R Highnam, M Cordell, F Estrella, and S R Amendolia. Mammogrid - a prototype distributed mammographic database for europe. *Clinical Radiology*, 62(11):1044–1051, june 2007.
- [7] Joel Saltz, Scott Oster, Shannon Hastings, Stephen Langella, Tahsin Kurc, William Sanchez, Manav Kher, Arumani Manisundaram, Krishnakant Shanbhag, and Peter Covitz. cagrid: design and implementation of the core architecture of the cancer biomedical informatics grid. *Bioinformatics*, 22(15):1910–1916, 2006.
- [8] Laboratorio de sistemas complejos. Digital inclusion kit <http://www.lsc.dc.uba.ar/digital-inlcusion>
- [9] MR Risk, DF Slezak, P. Turjanski, A. Panelli, RAM Taborda, and G. Marshall. Time Series Calculation of Heart Rate Using Multi Rate FIR Filters. In *Computers in Cardiology*, pages 541–544, 2007.
- [10] Jay Alameda, Marcus Christie, Geoffrey Fox, Joe Futrelle, Dennis Gannon, Mihael Hategan, Gopi Kandaswamy, Gregor von Laszewski, Mehmet A. Nacar, Marlon Pierce, Eric Roberts, Charles Severance, and Mary Thomas. The open grid computing environments collaboration: portlets and services for science gateways. *Concurrency and Computation: Practice and Experience*, 19(6):921–942, 2007.
- [11] Jason Novotny, Michael Russell, and Oliver Wehrens. Gridsphere: a portal framework for building collaborations: Research articles. *Concurr. Comput. : Pract. Exper.*, 16(5):503–513, 2004.
- [12] A. Abdelnur, E. Chien, and S. Hepper. Portlet specification 1.0, available from <http://www.jcp.org/en/jsr/detail?id=168>, 2003.