

# SIMULACIÓN DE APLICACIONES PARALELAS Y MECANISMOS DE TOLERANCIA A FALLOS

Nilda M. Pérez Otero, Abigail R. N. Verazay, C. Marcelo Pérez Ibarra.

Grupo de Ingeniería de Software (GIS). Facultad de Ingeniería. Universidad Nacional de Jujuy

Ítalo Palanca 10. San Salvador de Jujuy. Jujuy. CP 4600

TE 54+388+4221591 CEL 54+388+154145307

[nilperez@fi.unju.edu.ar](mailto:nilperez@fi.unju.edu.ar), [abigailrn@gmail.com](mailto:abigailrn@gmail.com), [cmperezi@gmail.com](mailto:cmperezi@gmail.com)

## Resumen

Durante los últimos años, los sistemas de cómputo de altas prestaciones hicieron posible el tratamiento de grandes volúmenes de datos a altísimas velocidades de procesamiento para aplicaciones comerciales y científicas. La evolución de las arquitecturas paralelas y el desarrollo de los mecanismos de tolerancia a fallos permiten que la ejecución de las aplicaciones se realice de manera confiable. Teniendo en cuenta que la configuración de recursos, el tipo de aplicación y los mecanismos de tolerancia a fallos influye en el comportamiento de un sistema paralelo, su estudio se lleva a cabo a través de herramientas de simulación que permiten representar el sistema. En virtud de esto, el Grupo de Ingeniería de Software desarrolló un simulador de clusters que permite parametrizar la configuración física de un *cluster*, y las características de cómputo y comunicación de aplicaciones M/W y SPMD. En este trabajo se presentan la línea de investigación abordada.

## Contexto

La línea de investigación aquí presentada está inserta en los proyectos *Extensión de CluSim: Simulación de la arquitectura tolerante a fallos RADIC* (D/0120) y *Extensión Funcional de CluSim: Simulación de aplicaciones paralelas SPMD, Pipeline y Divide/Conquer* (D/0049). Ambos proyectos se encuentran acreditados y financiados por la Secretaria de Ciencia y Técnica y Estudios Regionales de la Universidad Nacional de Jujuy (SECTER-

UNJu), además, el proyecto (D/0120) se encuentra bajo el Programa de Incentivos.

## Introducción

En el mundo actual el procesamiento de gran cantidad de datos en el menor tiempo posible es una necesidad que se presenta tanto en el ámbito comercial como científico. Durante los últimos años, los sistemas de cómputo de altas prestaciones (HPC) contribuyeron gradualmente a suplir esta necesidad, no obstante los costos y complejidad asociados a ellos. Así fue que se desarrollaron técnicas y herramientas para la planificación y mantenimiento de tales sistemas a fin de disminuir el impacto negativo inherente a su implementación. Además, el avance tecnológico que acompañó a los sistemas de altas prestaciones permitió que las primeras arquitecturas, de enormes presupuestos, evolucionaran en alternativas tan accesibles como los *clusters* de computadoras [Catalán, 2003].

El uso de *clusters* se popularizó no sólo por su capacidad de procesamiento y menor costo sino por la confiabilidad y disponibilidad que alcanzaron mediante el empleo de mecanismos adicionales. La importancia de estos mecanismos se evidencia en los informes del Laboratorio Nacional de Los Álamos (LANL) que describen más de 20.000 eventos que causaron detenciones en la ejecución de aplicaciones considerando un período de 10 años y una veintena de *clusters* [Cappello, 2009]. Un estudio realizado por Schroeder y Gibson indica que algunos sistemas presentan más de mil fallos por año [Schroeder & Gibson, 2007] lo que

implica que si pueden ocurrir 3 fallas por día, las aplicaciones que utilizan todos los nodos de cómputo y demoran más de 8 horas, tienen pocas posibilidades de finalizar su ejecución correctamente.

Con el objetivo de mitigar los efectos de estos eventos, es importante contar con mecanismos que aseguren la ejecución exitosa de las aplicaciones en un entorno paralelo. Así, los llamados sistemas de tolerancia a fallos permiten la detección y recuperación de fallos, además de tareas de mantenimiento preventivo tales como el reemplazo de máquinas susceptibles a fallos sin interrupciones al sistema. [Santos *et al.*, 2008].

Es cada vez más frecuente realizar los estudios de las características de los sistemas HPC mediante modelos de simulación orientados al diseño y modelado de prestaciones [Denzel *et al.*, 2008], la exploración de arquitecturas o aplicaciones [Hammond *et al.*, 2009; Minkenberg & Rodriguez, 2009] o las herramientas de predicción de tráfico [Tikir *et al.*, 2009]. En la literatura existen muchos trabajos centrados en simular grandes redes y aplicaciones de HPC. Algunos de los trabajos orientados a la arquitectura de sistemas de altas prestaciones son:

- PARSEC: un entorno de simulación de eventos discretos que simula entidades y constructores de mensajes de comunicación entre entidades [Bagrodia *et al.*, 1998].
- SIMLAB: un entorno de simulación para SANs implementado en C++ que modela discos rígidos, nodos de enrutamiento e interfaces de redes. Se utiliza para el desarrollo y verificación de algoritmos distribuidos para la red PRES TO SAN [Berenbrink *et al.*, 2001].
- SIMCAN: un entorno de simulación para grandes redes complejas de almacenamiento que simula estas redes y sus subsistemas subyacentes (I/O, Networking, etc.) [Nuñez *et al.*, 2008].

Otros trabajos de simulación están centrados en las aplicaciones MPI. Entre ellos se encuentran:

- Un prototipo de un simulador híbrido que emula la ejecución de una aplicación paralela en modo stand-alone y la simulación de una red que se utiliza para el paso de mensajes de MPI. Este prototipo necesita que la aplicación simulada se ejecute en el hardware que se desea simular [Riesen, 2006].
- MPI-SIM: un simulador paralelo diseñado para predecir el rendimiento de aplicaciones MPI y MPI-IO ya existentes. MPI-SIM predice el rendimiento de estos programas en función de características de arquitectura (número de procesadores, latencias en la comunicación, algoritmos de caché, etc.) [Bagrodia *et al.*, 2008].

En algunos trabajos, ha sido útil la utilización de entornos de simulación de redes de propósito general, como OMNeT++ ([www.omnetpp.org](http://www.omnetpp.org)) y OPNET Modeler ([www.opnet.com](http://www.opnet.com)). Bajo estos entornos, es posible crear diferentes configuraciones de redes, con diferentes tipos de nodos, switches, topologías, protocolos, etc.

También es posible encontrar entornos de simulación de redes de propósito general que permiten crear diferentes configuraciones de redes, con diferentes tipos de nodos, switches, topologías, protocolos, etc. Ejemplos de éstos son OMNeT++ ([www.omnetpp.org](http://www.omnetpp.org)) y OPNET Modeler ([www.opnet.com](http://www.opnet.com)).

Considerando estos antecedentes, desde el año 2010, el Grupo de Ingeniería de Software (GIS) desarrolló CluSim, un simulador de clusters basado en OMNeT++ que permite parametrizar la configuración de recursos de un cluster, los patrones de cómputo y comunicación de la aplicación paralela y el tamaño del problema, de modo que sea posible evaluar y predecir el impacto en el rendimiento del sistema considerando diferentes configuraciones (número de nodos del

cluster). Este simulador, presentado en los trabajos [Pérez Ibarra *et al.*, 2010], [Valdiviezo *et al.*, 2010], [Lasserre *et al.*, 2011] y [García *et al.*, 2011], pretende servir como herramienta de soporte a la toma de decisiones para la selección de las configuraciones más adecuadas para un cluster que ejecuta un determinado tipo de aplicación paralela.

A fin de extender la funcionalidad de CluSim, el GIS definió las etapas de un proceso de modelado que permite la formulación de un modelo genérico de aplicaciones paralelas bajo distintos paradigmas de programación, proceso que ya fue utilizado para caracterizar aplicaciones Master/Worker y SPMD (Single Program Multiple Data) [Lasserre *et al.*, 2012; García *et al.*, 2012].

## Líneas de Investigación y Desarrollo

CluSim es un simulador de clusters homogéneos, desarrollado utilizando el framework de simulación OMNeT++, que permite parametrizar la configuración de recursos de un cluster, los patrones de cómputo y comunicación de la aplicación paralela y el tamaño del problema, de modo que sea posible evaluar y predecir el impacto de diferentes configuraciones en el rendimiento de un sistema. A partir de 2010 y a través de sucesivas extensiones se fueron incorporando nuevas funcionalidades, y a la fecha CluSim simula aplicaciones Master/Worker y SPMD.

En forma paralela a CluSim el GIS está desarrollando un modelo genérico para caracterización de aplicaciones paralelas que fue probado con los paradigmas Master/Worker y SPMD.

A partir de estas dos líneas: simulador y modelo genérico, estos proyectos pretenden:

- Continuar agregando funcionalidad a CluSim, por un lado para que permita la simulación de aplicaciones Pipeline y Divide/Conquer y por otro, para que permita simular sistemas HPC con distintas configuraciones de tolerancia a fallos.

- Ampliar el modelo genérico hacia los dos paradigmas restantes, Pipeline y Divide/Conquer.

## Objetivos y Resultados

### Objetivos

Los proyectos aquí presentados pretenden extender la funcionalidad básica del simulador CluSim reformulando su modelo de simulación para incluir no sólo distintos tipos de aplicaciones paralelas sino también los aspectos más relevantes de un mecanismo de tolerancia a fallos. Con esto en mente, se establecieron los siguientes objetivos:

- Caracterizar aplicaciones SPMD, Pipeline y Divide/Conquer.
- Identificar los parámetros característicos de aplicaciones SPMD, Pipeline y Divide/Conquer.
- Formular modelos de simulación para aplicaciones SPMD, Pipeline y Divide/Conquer.
- Integrar los modelos de simulación formulados a CluSim.
- Formular un modelo de simulación para tolerancia a fallos.
- Integrar el modelo de simulación de tolerancia a fallos a CluSim.

### Resultados Obtenidos

En 2012 se obtuvieron los siguientes resultados:

- Definición de las etapas de un proceso de modelado
- Reformulación del modelo de simulación de CluSim para aplicaciones Master/Worker
- Caracterización de aplicaciones SPMD
- Incorporación de aplicaciones SPMD al modelo de simulación de CluSim.

### Resultados Esperados

Para 2013 se prevé obtener los siguientes resultados:

- Caracterización de aplicaciones Pipeline y Divide/Conquer

- Incorporación de aplicaciones Pipeline y Divide/Conquer al modelo de simulación de CluSim
- Incorporación de mecanismos de tolerancia a fallos al modelo de simulación de CluSim

## Formación de Recursos Humanos

En el marco de este proyecto trabajan 6 docentes y 8 alumnos pertenecientes a las carreras Informáticas de la Facultad de Ingeniería de la UNJu y del Departamento Académico San Salvador (DASS) de la Universidad Católica de Santiago del Estero. Entre los docentes se cuenta con un Magíster que está finalizando los estudios de Doctorado y 3 que están realizando carreras de posgrado.

En la actualidad, se desarrollan 5 tesinas de grado y una beca CIN afines a la temática del proyecto y los alumnos restantes, menos avanzados en la carrera, fortalecen su formación en investigación. Como resultado del trabajo realizado, se han finalizado satisfactoriamente una tesina de grado y una beca CIN.

## Referencias

- [Bagrodia *et al.*, 2008] Bagrodia, R., E. Deelman, and T. Phan. Parallel simulation of large-scale parallel applications. *Int. J. High Perform. Comput. Appl.*, 15(1):3–12, 2001.
- [Bagrodia *et al.*, 2008] Bagrodia, R., E. Deelman, and T. Phan. Parallel simulation of large-scale parallel applications. *Int. J. High Perform. Comput. Appl.*, 15(1):3–12, 2001.
- [Berenbrink *et al.*, 2001] Berenbrink, P., A. Brinkmann, and C. Scheideler. Simlab - a simulation environment for storage area networks. In *Workshop on Parallel and Distributed Processing (PDP)*, pages 227–234, 2001.
- [Cappello, 2009] Cappello, F. (2009). Fault Tolerance in Petascale/ Exascale Systems: Current Knowledge, Challenges and Research Opportunities. *International Journal of High Performance Computing Applications*, 23(3):212–226.
- [Catalán, 2003] Catalán i Coit, M. “Nuevo Modelado de Computación Paralela con Clusters Linux”. VI Congreso HIPALinux. miKeL a.k.a.mc2. Septiembre 2003.
- [Denzel *et al.*, 2008] Denzel, W. E.; Li, J.; Walker, P. & Jin, Y. A framework for end-to-end simulation of high-performance computing systems. In *Simutools '08: Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*, pp. 1--10, ICST, Brussels, Belgium, Belgium. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [García *et al.*, 2011] García, A.; Pérez Otero, N. M.; Pérez Ibarra, C. M. y C. M. Lasserre. Simulación de Clusters: Integración de INET a CluSim. XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011). ISBN 978-950-34-0756-1. Universidad Nacional de la Plata. Buenos Aires. pp. 367-373. Octubre 2011.
- [García *et al.*, 2012] García, A.; Lasserre, C. M.; Verazay, A. R. N.; Pérez Otero, N. M.; Pérez Ibarra, M.; Martínez, J. G. y S. A. Nolasco. Modelado de Aplicaciones SPMD. VIII Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA. Ed. Grupo Loza Impresiones S.R.L. ISSN 1853-7871 San Miguel de Tucumán. Septiembre 2012.
- [Hammond *et al.*, 2009] Hammond, S. D.; Mudalige, G. R.; Smith, J. A.; Jarvis, S. A.; Herdman, J. A. & Vadgama, A. Warpp: a toolkit for simulating highperformance parallel scientific codes. In *Simutools '09: Proceedings of the 2nd International Conference on Simulation Tools and Techniques*, pp. 1--10, ICST, Brussels, Belgium, Belgium. ICST (Institute for Computer Sciences, Social-

- Informatics and Telecommunications Engineering).
- [Lasserre *et al.*, 2012] Lasserre, C. M.; García, A.; Pérez Otero, N. M.; Verazay, A. R. N.; Pérez Ibarra, M.; Nolasco S. A. y J. G. Martínez. Hacia un Modelo Genérico de Aplicaciones Paralelas. XVIII Congreso Argentino de Ciencias de la Computación (CACIC 2012). pp. 296-305. ISBN 978-987-1648-34-4. Universidad Nacional del Sur. Bahía Blanca. Octubre 2012.
- [Lasserre *et al.*, 2011] Lasserre, C. M.; Pérez Ibarra, C. M.; Valdiviezo, L. M.; Verazay, A. R. N.; Quispe, G. L.; Nolasco, S. A.; Chosco, V. H. y N. M. Pérez Otero. Adaptación de CluSim a clusters heterogéneos. Investigaciones en Facultades de Ingeniería del NOA. Tomo 2 – 2011. ISSN 1853-7871. Editorial Científica Universitaria. San Fernando del Valle de Catamarca. pp 1053-1060. Octubre 2011.
- [Minkenberg & Rodriguez, 2009] Minkenberg, C. & Rodriguez, G. Tracedriven co-simulation of high-performance computing systems using OMNeT++. In Simutools '09: Proceedings of the 2nd International Conference on Simulation Tools and Techniques, pp. 1--8, ICST, Brussels, Belgium, Belgium. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [Nuñez *et al.*, 2008] Nuñez, A., J. Fernandez, J. D. Garcia, L. Prada, and J. Carretero. Simcan: a simulator framework for computer architectures and storage networks. In Simutools '08: Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops, pages 1–8, ICST, Brussels, Belgium, Belgium, 2008. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [Pérez Ibarra *et al.*, 2010] Pérez Ibarra, C. M., L. M. Valdiviezo, N. M. Pérez Otero, H. P. Liberatori, D. Rexachs, E. Luque y C. M. Lasserre. CLUSIM: Simulador de Clusters para Aplicaciones de Cómputo de Altas prestaciones basado en OMNeT++. XVI Congreso Argentino de Ciencias de la Computación. Buenos Aires (Morón). Octubre 2010.
- [Riesen, 2006] Riesen, Rolf. “Supercomputer Simulation Design Through Simulation”. Cray User Group (CUG). ISBN: 1-4244-0327-8. Barcelona. pp. 1–9. 2006.
- [Santos *et al.*, 2008] Santos, G.; Duarte, A.; Rexachs, D. & Luque, E. Providing non-stop service for message-passing based parallel applications with RADIC. In Luque, E.; Margalef, T. & Benitez, D., editores, Euro-Par, volume 5168 of Lecture Notes in Computer Science, pp. 58–67. Springer.
- [Schroeder & Gibson, 2007] Schroeder, B. & Gibson, G. A. Understanding failures in petascale computers. Journal of Physics: Conference Series, 78:012022 (11pp).
- [Tikir *et al.*, 2009] Tikir, M. M.; Laurenzano, M. A.; Carrington, L. & Snaveley, A. PSINS: An open source event tracer and execution simulator for MPI applications. In Euro-Par '09: Proceedings of the 15th International Euro-Par Conference on Parallel Processing, pp. 135--148, Berlin, Heidelberg. Springer-Verlag.
- [Valdiviezo *et al.*, 2010] Valdiviezo, L. M.; Pérez Otero, N. M.; Pérez Ibarra, C. M. y C. M. Lasserre. Caracterización de una aplicación paralela con distintas configuraciones en CluSim. Investigaciones en Facultades de Ingeniería del NOA – 2010. ISSN 3367-5072. Ed. EdiUNJu. S. S. de Jujuy. pp. 499-504. Noviembre 2010.