

Construyendo un Sistema de Cómputo Distribuido Multipropósito

Oscar Martín Bianchi^(a,b)
mbianchi@ejercito.mil.ar

José Ignacio Ariznabarreta Fossati^(b)
ignacioariznabarreta@gmail.com

Alejandro J. M. Repetto^(a,b)
arepetto@ejercito.mil.ar

^(a)CIDESO⁰, DIGID¹ - Ejército Argentino

^(b)EST², IESE³ - Ejército Argentino

RESUMEN

El concepto de sistemas distribuidos, como la tecnología asociada a dichos desarrollos, es un concepto con una popularidad en constante crecimiento, así como un poderoso paradigma.

Dentro del ámbito académico en el que se enmarca este estudio, la computación distribuida prueba ser una herramienta valiosa para la experimentación tanto científica experimental como práctica, en cualquiera de las especialidades de ciencias donde se las desee enmarcar. Esto se debe en gran parte a la posibilidad de obtener grandes capacidades de cómputo con a bajo costo para quienes se plantean el uso del modelo, especialmente si se lo asocia con los sistemas de Computación de Escritorio de Grilla (Desktop Grid Computing).

En este entorno, la incorporación de tecnologías de computación de alto rendimiento a bajo costo, por ejemplo la computación voluntaria, surge como una opción interesante, especialmente por parte de organizaciones que cuentan con grandes cantidades de puestos de trabajo que presentan capacidad de cómputo ociosa.

En el presente trabajo se expondrán los avances realizados hacia la obtención de un sistema de computación distribuida multipropósito, que dé apoyo a los laboratorios de las distintas especialidades de ingeniería de la EST.

Palabras Clave: *Computación Distribuida, Multipropósito, Infraestructura como Servicio,*

Webservices, Computación en Grilla de Escritorio.

CONTEXTO

En las últimas dos décadas, las tecnologías orientadas a la Computación en Grilla (*Grid Computing*), experimentaron un avance exponencial de la mano de proyectos como Globus^a y BOINC^b, los cuales pusieron a disposición del público general la posibilidad de crear y gestionar proyectos de alta complejidad técnica.

Específicamente, la arquitectura de Computación en Grilla de Escritorio (*Grid Desktop Computing* – GDC)(1), es una forma particular de computación distribuida que implica la utilización de grandes grillas de computadores personales de forma organizada y sincronizada de manera que funcionen como un único computador, ofreciendo grandes cantidades de recursos con un costo relativamente bajo. Se pueden referenciar aplicaciones de este tipo de arquitectura en áreas críticas como la medicina (proyecto GENOMA), el estudio del espacio (con el proyecto SETI), el criptoanálisis, la toma de decisiones, el análisis de imágenes por software, entre otros.

En lo que hace al ámbito de aplicación dentro del sistema el Ejército Argentino, el Centro de Investigación y Desarrollo de Software del Ejército Argentino (CIDESO), implementó estos GDC en diversos proyectos, aplicándolo sobre todo a Sistemas de Simulación Constructiva Distribuidos. Estos sistemas

^a<http://www.globus.org/toolkit/>

^b<http://boinc.berkeley.edu/>

⁰ CIDESO: Centro de Investigación y Desarrollo de Software

¹ DIGID: Dirección General de Investigación y Desarrollo

² EST: Escuela Superior Técnica - Facultad de Ingeniería del Ejército Argentino

³ IESE: Instituto de Enseñanza Superior del Ejército - Universidad del Ejército Argentino

usualmente requieren un alto poder de cómputo y, por ser de simulación constructiva, una respuesta en cuasi tiempo real. La implementación de GDC al problema logró solucionar los problemas asociados, proveyendo una arquitectura adaptativa al problema del crecimiento del cómputo junto con el crecimiento en el número de participantes de los juegos (2)(3). Con relación a esto último, la mayoría de las fuerzas armadas y de seguridad alrededor del mundo están incorporando este tipo de tecnologías en asuntos críticos como las ejecuciones de modelos de propagación de epidemias, incendios, el uso de sensores distribuidos para facilitar la asistencia durante catástrofes o el control de zonas de alto riesgo, entre otras aplicaciones.

Desde otro punto de vista, en lo que hace al ámbito académico particularmente, la Escuela Superior Técnica del Ejército Argentino (EST), posee experiencias previas en el uso de estos conceptos en sus laboratorios de investigación. La primera incursión en el uso de GDC estuvo relacionada a aplicaciones asociadas con la criptografía, orientada al desarrollo científico y evaluación de algoritmos para la seguridad de la información, dentro del CriptoLab.

Entre las investigaciones llevadas a cabo por el CriptoLab se puede destacar, por ejemplo, la de identificar y evaluar algoritmos de factorización de grandes números – esencial en la seguridad de la información actual –, la cual requería de gran poder de cálculo, difícil de obtener incluso con equipos de alto rendimiento pero disponible a través de la utilización de computación paralela y/o distribuida. Por eso, y gracias a un estrecho trabajo en conjunto entre el CIDESO y los laboratorios de la EST, surgió la posibilidad de realizar investigaciones en conjunto, aportando las experticias de cada uno, logrando una sinergia en la investigación.

Así nace el proyecto CODEC, de cuyos adelantos y propuestas futuras se ocupara el presente documento.

INTRODUCCIÓN

Luego de las experiencias obtenidas durante el año 2011 en el desarrollo del proyecto CoDiSe (4), se llevó adelante durante el año 2012 el proyecto CODEC (5), el cual, a diferencia del prototipo experimental, que solo estaba orientado a solucionar una problemática específica (validación de algoritmos asociados a la seguridad informática), está concebido para brindar apoyo a un abanico variado de necesidades de cómputo asociados a las distintas ramas de la ingeniería. Dichas necesidades varían enormemente tanto en su alcance como en su disciplina, yendo del contexto de la educacional civil, al de la aplicación militar, incluyendo a su vez proyectos mixtos entre las dos áreas de interés. Específicamente, se pretende que el proyecto CODEC aporte en los próximos años soluciones concretas a los laboratorios de las carreras de Ingeniería Química, Informática y Mecánica, realizando para cada una de ellas experimentos, pruebas y evaluaciones que de otra forma serían impracticables debido a su costo en tiempo y recursos.

Así mismo, en el marco de los organismos de Defensa y Seguridad, las aplicaciones de sistemas del tipo de CD y GDC sufrieron un importante incremento en los últimos años, específicamente en sus usos asociados en el apoyo a la comunidad y el control de crisis (a través de modelos de simulación predictivos de alta complejidad de cálculo). Además del uso exclusivo para un sector, podemos mencionar el caso de una utilización mixta, en donde estas tecnologías son utilizadas por laboratorios y universidades ajenas al ámbito castrense con el fin de realizar tareas de testing y certificaciones sobre los equipos y medios de comunicaciones utilizados por las distintas fuerzas; entre otros.

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

CODEC nos acerca en su propuesta la utilización de la infraestructura de cómputo ociosa disponible de la universidad, para aportar soluciones a problemáticas surgidas de la necesidad de realizar complejas experimentaciones, que requieren poderosos equipos de cómputo, para complementar el dictado de las clases, las prácticas profesionales, o las investigaciones de cualquier tipo.

Por lo mencionado, es que el sistema plantea dos grandes líneas de investigación, las cuales incluyen:

- Proveer a la universidad de una herramienta de cómputo multipropósito para sus investigaciones a un bajo coste.
- Utilización y aprovechamiento de la infraestructura informática ociosa disponible a través de las políticas incluidas en el VC.

De esta forma, las líneas de investigación que se llevaran a cabo tomando como base el prototipo realizado serán:

- Sistemas de cómputo distribuido multipropósito.
- Explotación de recursos computacionales ociosos en grandes organizaciones.
- Inclusión de BOINC en la nube (a través de webservices)
- Ejecución distribuida con foco multidisciplinario.

Este último apartado merece una mención especial, ya que las aplicaciones específicas de las distintas disciplinas son índole muy variado y de gran importancia a nivel académico. Entre ellas podemos mencionar:

- Ingeniería Química: Para dar apoyo a experimentos asociados al cálculo de densidad de compuestos, basados en el modelo DFT (*Density Functiona lTheory*).
- Ingeniería Informática: Como se demostró con el proyecto CoDiSe, el uso de un sistema con las características del

planteado aplicado a la seguridad informática (tanto para el cálculo teórico, así como para la experimentación práctica) es de gran utilidad y provee un importante recurso.

- Ingeniería Mecánica: Cálculos como los asociados a la resistencia y estabilidad de los distintos componentes y materiales que forman parte de un motor o mecanismo sensible, así como la ejecución de modelos de simulación que permitan detectar y prevenir fallas desde las fases de desarrollo asociadas al diseño.

RESULTADOS ESPERADOS

Los objetivos propuestos para el próximo ciclo cubren desde la interacción del usuario para con el sistema hasta las cualidades que presente dicho sistema para con la arquitectura disponible.

Un objetivo es la consolidación de la interfaz proveedora de servicios del Sistema de Cómputo Distribuido con el fin de que la interacción usuario(no especializado)-computador pueda realizarse de forma más sencilla y que el objetivo principal de acercar una solución tratando de abstraer a los usuarios de la complejidad del sistema. El concepto subyacente es el de IaaS (*Infrastructure as a Service*). Esto se analizará mediante la estrategia de prototipos evolutivos, recurriendo a los docentes especialistas en cada rama de la ingeniería que queramos avanzar, con el fin de poder determinar el grado de aceptación que pueda tener dicha interfaz.

Para lograr el mayor nivel de usabilidad posible, se dividirá el problema en tres aspectos: acceso, envío de trabajos y respuesta de resultados. El primero se refiere al control de la seguridad básica del sistema en conjunto con un interface tipo panel de control. El envío de trabajos abarca la estandarización de las interfaces de BOINC reduciendo la complejidad a través de unmeta-lenguaje de *scripting*. Por último, la entrega de resultados también se abordará

como un esquema similar, estandarizando las interfaces de modo de unificar la visión del lado del usuario final.

Además de los aspectos amenos al usuario y la funcionalidad para con respecto a este, se propone un objetivo importante para garantizar el funcionamiento de los nodos (aquellos computadores cuya función será ceder los recursos) la generación de métricas de rendimiento del sistema, con el fin de poder evaluar el funcionamiento del sistema.

FORMACIÓN DE RECURSOS HUMANOS

Desde el punto de vista de su concepción, CODEC es fruto del trabajo conjunto entre dos centros de investigación enmarcados en el seno del Ejército Argentino, con la particularidad que uno de dichos centros es parte funcional de la EST. Esto pone al proyecto en un ámbito privilegiado para la formación de recursos humanos.

Por un lado la EST, específicamente la cátedra de Lenguajes de Programación I y el CriptoLab, de esa misma institución, tiene amplia experiencia tanto en la investigación básica en el terreno de computación distribuida como en el de la criptografía respectivamente. Así mismo, en los laboratorios de la institución se dispone del material y los recursos humanos específicos tanta para la investigación en estas áreas como para dar apoyo a los alumnos a cargo del proyecto.

Por otro lado, el CIDESO tiene experiencia en la formación de recursos humanos en el terreno de la investigación aplicada en sistemas de información de diversa índole, incluyendo sistemas de simulación para el adiestramiento, sistemas de información geográfica, sistemas de visualización, sistemas inteligentes, sistemas móviles, sistemas de comunicación de alta complejidad y sistemas de cómputo de alto rendimiento.

Ambos laboratorios, a través del dictado de materias de grado en Ingeniería Informática, aportan recursos humanos a la Escuela Superior Técnica del Ejército. Es así que

investigadores de los laboratorios son jefes de cátedra en la EST y, de manera análoga, alumnos de la escuela aportan sus análisis a los laboratorios a través de trabajos prácticos de laboratorio, prácticas profesionales supervisadas o tesis y tesinas de grado y posgrado.

En particular, el prototipo experimental del cual nace la idea de CODEC fue implementado por los alumnos de tercer año de Ingeniería Informática y Electrónica como trabajo práctico de laboratorio de la materia Lenguajes de Programación I, y tutorado por investigadores del CIDESO y el CriptoLab, que a la vez son docentes de dicha cátedra.

Para el próximo paso, se pretende continuar con esta interacción fluida entre los centros de investigación y el alumnado, formado profesionales con conocimientos de campo en el terreno de la computación de alto rendimiento y un conocimiento acabado sobre temas criptográficos.

Además, al expandirse el sistema para ser aplicado en cualquier problema que requiera altos niveles de cómputo, se pretende incorporar alumnos y docentes de otras cátedras, de cualquiera de las ingenierías que se dictan en la Facultad.

Por otro lado, se continuará con la formación de profesionales en investigación utilizando estas nuevas tecnologías en los dos laboratorios participantes a través de la utilización de la GDC.

Así, pues, se formarán recursos humanos de todos los niveles, grado, posgrado o investigadores activos, incorporando más alumnos a los laboratorios y, potencialmente, becarios que se dediquen de modo formal (no sólo académico) a la profundización de los modelos propuestos.

Referencias

1. **Berstis, Viktors.** *Fundamentals of Grid Computing*. s.l. : Redbooks Paper, IBM Corp, 2002.
2. *Hybrid Architecture for Constructive Interactive Simulation: Evaluation and Outcomes.* **Repetto, Alejandro J. M.**

Orlando, FL : s.n., 2010. /ITSEC'10, Interservice/Industry Training, Simulation and Education Conference.

3. *Grid Desktop Computing for Constructive Battlefield Simulation*. **Repetto, Alejandro J. M.** San Salvador de Jujuy : Red UNSI, 2009. XV Congreso Argentino de Ciencias de la Computación (CACIC 2009). 978-897-24068-4-1.

4. **Schroeder, Ralph.** *The Usability of Collaborative Virtual Environments and Methods for the Analysis of Interaction*. [ed.] University of Oxford Oxford Internet Institute. Oxford : Massachusetts Institute of Technology, 2006.

5. **Craig, Donald.** *Advantages of Simulation*. St. John's : Memorial University, Faculty of Science.

6. **Moriello, Sergio A.** *Inteligencias Sintéticas*. Buenos Aires, Argentina : Alsina, 2001.

7. *Interoperability-Ready, Training-Focused Architecture for Command and Control Systems*. **Repetto, Alejandro Juan Manuel**. Orlando, FL : s.n., 2011. Interservice/Industry Training, Simulation and Education Conference.

8. *Framework de Interoperabilidad para Sistemas de Comando y Control*. **Repetto, Alejandro Juan Manuel**. El Calafate, Santa Cruz : WICC'10, 2010. ISBN 978-950-34-0652-6.

9. **Ajulo, Adeshina B.** *Grid Computing An Advancement of E-Science To Computing And Beyond*. s.l. : Department of Computer Science, Federal University of Technology.

10. *Computación Distribuida para Cálculo Científico*. **Bianchi, Oscar Martín and Repetto, Alejandro**. Posadas : XIV Workshop de Investigadores en Ciencias de la Computación, 2012.

11. **Bianchi, Oscar Martín and Repetto, Alejandro.** *Sistema de Computacion Distribuida para Seguridad Informatica*. Posadas, Misiones : Workshop de Investigadores en Ciencias de la Computación 2012, 2012.

12. **Computerized Bussiness Solutions.** *Centralized vs Distributed Computing: White Paper*. 2007.

13. **Berman F, G. C. Fox and A. J. G. Hey.** *Grid Computing: Making the global infrastructure a reality*. s.l. : Wiley publishers, 2003.

14. **David P. Anderson, Eric Korpela, Rom Walton.** *High-Performance Task Distribution for Volunteer Computing*. Berkeley : Space Sciences Laboratory, University of California.

15. **Rodrigo Neves, Nuno Mestre, Francisco Machado, Joao Lopez.** *Parallel and Distributed Computing BOINC Grid Implementation*.

16. *The Challenge of Volunteer Computing*. **Christensen, Carl, Aina, Tolu and Stainforth, David**. Oxford, United Kingdom : Clarendon Laboratory, University of Oxford, 2009.