Computación **Di**stribuida para **Se**guridad Informática
CoDiSe

Oscar Martín Bianchi
mbianchi@ejercito.mil.ar

Alejandro J. M. Repetto
arepetto@ejercito.mil.ar

CIDESO[1], DIGID[2] – Ejército Argentino
EST[3], IESE[4] – Ejército Argentino

RESUMEN

Tanto en el ámbito académico como en el comercial, la utilización de equipamiento de alto rendimiento para el cómputo de algoritmos complejos con alta demanda computacional es poco accesible dado su alto costo. Sin embargo, dada la dependencia tecnológica creciente, este tipo de cómputos es cada vez más necesaria.

En este panorama, 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 particular, cuando se experimenta en el ámbito de la seguridad informática poseer grandes capacidades de procesamiento disminuyen los tiempos de cómputo de manera oportuna incrementando así la posibilidad de realizar pruebas y validaciones. El tratar continuamente con grandes números, hace que el campo de la seguridad de la información en general, y la criptografía en particular, se encuentre bajo constante demanda de poder de cálculo.

En CoDiSe se plantea la utilización de un sistema distribuido, utilizando un *framework* del tipo Computación en Grilla de Escritorio (*Grid Desktop Computing – GDC*), que explotará la capacidad ociosa de la infraestructura informática disponible dentro de la organización, conformando una gran máquina de alto rendimiento fundamental en proyectos de investigación científica [1] en general, y en criptografía en particular.

Palabras Clave: *Computación Distribuida, Seguridad, Computación en Grilla de Escritorio, Factorización de Grandes Números.*

CONTEXTO

La arquitectura de GDC [2], es una forma particular de computación distribuida que ha atraído la atención tanto de los ejércitos alrededor del mundo como de los distintos institutos científicos debido al gran poder de cómputo que es posible extraer de ellos. Esta arquitectura es utilizada en áreas de interés tan críticas como las del criptoanálisis, toma de decisiones, análisis de imágenes por software, entre otros.

El Ejército Argentino, a través del Centro de Investigación y Desarrollo de Software del Ejército Argentino (CIDESO), implementó estos conceptos escalándolos y llevándolos a la practica en la forma de Sistemas de Simulación Distribuidos, en problemas para los cuales era necesario poseer gran poder de cálculo [3] [4].

Por otro lado, en el ámbito académico, la Escuela Superior Técnica del Ejército Argentino (EST), posee un laboratorio de investigación en temas de criptografía con amplia experiencia en el desarrollo científico y evaluación de algoritmos para la seguridad de la información, el CriptoLab.

Dentro de las investigaciones llevadas a cabo por el CriptoLab, la de identificar y evaluar algoritmos de factorización de grandes números - pilar fundamental de la seguridad de la información-, requería de gran poder de cálculo, disponible sólo a través de la utilización de computación paralela y/o distribuida.

Dada la estrecha vinculación entre el CIDESO y el CriptoLab, surgió la posibilidad

[1] CIDESO: Centro de Investigación y Desarrollo de Software
[2] DIGID: Dirección General de Investigación y Desarrollo
[3] EST: Escuela Superior Técnica – Facultad de Ingeniería del Ejército Argentino
[4] IESE: Instituto de Enseñanza Superior del Ejército – Universidad del Ejército Argentino

de realizar investigaciones en conjunto, aportando las experticias de cada uno, logrando una sinergia en la investigación.

Así, CoDiSe nace como un proyecto de investigación conjunto entre las dos organizaciones para poder ejecutar procesos que requieren gran poder de cómputo en el ámbito de la EST, logrando la sinergia entre ambas líneas de investigación.

1. INTRODUCCIÓN

El concepto de computación distribuida es usado para solucionar problemas de carácter científico desde hace varios años. Sin embargo, cuando se hacer referencia a computación de alto rendimiento, surgen inmediatamente arquitecturas paralelas o distribuidas que acarrean altos costos, tanto en infraestructura como en la formación de recursos humanos. Aplicando el concepto de GDC, asociado a lo que se denomina computación voluntaria, utilizando computadores personales para generar clústeres de procesamiento que funcionan como una solo equipo, estos costes disminuyen rápidamente otorgando a las organizaciones una ventaja estratégica para procesar sus funciones de negocio [5]. Dentro de las arquitecturas de GDC, BOINC[a] es la referencia más conocida y utilizada a nivel mundial. Proyectos como SETI @Home o Einstein @Home [6] la avalan, demostrando su gran capacidad y estabilidad.

Como se ha señalado anteriormente, el CIDESO posee experiencia en las tecnologías de GDC [3] [4], la cual probó ser invaluable para la mejora de rendimiento en algoritmos de simulación que requerían cómputos intensivos dentro de sistema Batalla Virtual, logrando mejoras en la velocidad de cómputo de más del 400%.

Teniendo en cuenta este logro, y el hecho que organismos tales como compañías, centros de investigación y universidades tienen una participación fundamental en la validación de las técnicas y tecnologías existentes, así como en el desarrollo de nuevos y más seguros métodos, las tecnologías señaladas cobran una relevancia critica para estos mismos.

Por otro lado, el problema de la factorización de grandes números en el ámbito de las investigaciones criptográficas es uno de los más complejos desde el punto de vista del poder computacional asociado que se requiere. La criptografía actual se basa, justamente, en la complejidad computacional de dicho problema. Los números primos asociados a las claves criptográficas son pilar fundamental de la seguridad de la información a nivel mundial. Así, pues, un error, colisión o cualquier otra debilidad detectada en la generación o utilización de ellos representa un grave incidente de seguridad.

Poder evaluar la calidad de los números primo generados por las librerías de seguridad otorga una herramienta de alto valor agregado a nivel científico y operativo. Sin embargo, la evaluación de estos números requiere la ejecución de algoritmos con altas demandas de cómputo.

CoDiSe busca utilizar las tecnologías de GDC como una herramienta eficaz en el ámbito académico con la cual poder realizar estudios, experimentos y validaciones de las diferentes técnicas y tecnologías asociadas a la seguridad informática en general y a la criptografía en particular. Idea que tuvo origen conjunto entre los investigadores del equipo de Investigación Aplicada y Desarrollo Experimental del CIDESO y los investigadores del CriptoLab.

Durante el segundo semestre del 2011 se desarrolló un prototipo exploratorio con resultados positivos. Esta línea de investigación pretende profundizar los resultados alcanzados y mejorar su aplicación.

La experimentación durante el año 2011 consistió en la factorización del valor N, de los sistemas de clave RSA [7] generados por la librería OpenSSL[b] a través del uso de las claves públicas y privadas, obteniendo los primos generadores. Dicha factorización ha sido estudiada en varios trabajos de criptografía y, por más que se posea el par de claves

[a] http://boinc.berkeley.edu/

[b] http://www.openssl.org/

correspondientes al N dado, los algoritmos requieren gran poder de cálculo para lograr la obtención de lo primos.

El desenmascarar los primos generados por las librerías, otorga a los especialistas en criptografía una herramienta potente que permite evaluar potenciales colisiones que se traducen en debilidades de los sistemas criptográficos.

Para la factorización de N se evaluaron dos algoritmos, aprovechando la infraestructura para lograr un *benchmark* entre ambos. Los algoritmos evaluados fueron el de Menezes [8] y otro ideado por los investigadores del CriptoLab.

Los resultados preliminares mostraron una mejora de rendimiento cercana al 150% (ciento cincuenta por ciento) en los tiempos de obtención de claves utilizando una infraestructura de GDC en base a la arquitectura BOINC [9], con sólo cinco participantes en la grilla, contra su versión *standalone*. Adicionalmente, se probó de manera fehaciente una ventaja de rendimiento importante del algoritmo "CriptoLab" por sobre el Menezes, en una proporción mayor a 1:1000 para N de 512 bits.

Dados estos resultados, se procederá a la formalización de la arquitectura, extendiendo el uso de GDC a la mayor cantidad de equipos disponibles en el ámbito de la universidad, para ser utilizados tanto por el CriptoLab como por cualquier otro laboratorio que requiera procesar grandes niveles de información.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Lo que propone CoDiSe es la utilización de la capacidad de cómputo ociosa de la infraestructura disponible, para aportar:

- Gran capacidad de cómputo a bajo costo.
- Aprovechar la capacidad ociosa de la infraestructura informática a través del aporte voluntario, donde cada PC en el sistema ofrecerá los recursos que tenga disponibles en ese momento preciso.
- Utilizar lo antes mencionado para reducir drásticamente los tiempos empleados tanto en el análisis de robustez de claves como en el de las librerías de seguridad, tipo Kerberos o OpenSSL.

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

- Creación de un sistema GDC para cálculo científico-matemático distribuido multipropósito para uso académico.
- Expansión del actual proyecto de seguridad llevándolo a un sistema más complejo que se acomode a las necesidades del Ejercito Argentino en materia de seguridad informática.

Para lograr los dos objetivos planteados se debe profundizar en temas de GDC agregando cuestiones relativas a la seguridad desde el punto de vista de la información transmitida, procesada y utilizada por la grilla. También se deben incrementar las capacidades de chequeo cruzado que proponen las arquitecturas a través de los sistemas de votación para validación de resultados.

Por otro lado, se deben realizar cambios en el sistema nativo (BOINC) para permitir el proceso multipropósito, convirtiendo a CoDiSe en una facilidad genérica que exceda el uso direccionado del cómputo.

Por lo antedicho, se centrarán las líneas de investigación en computación en grilla de escritorio, seguridad informática en sistemas distribuidos y programación distribuida.

3. RESULTADOS ESPERADOS

Se espera en lo general para el año 2012 lograr el análisis y el diseño completo de la solución distribuida multipropósito así como también avanzar en el desarrollo del prototipo realizado durante 2011.

En lo particular se pretende afianzar el desarrollo direccionado hacia el ámbito de la seguridad informática, profundizando el circuito de prueba de librerías generadoras de claves. Para ello se debe consolidar el circuito generación de tuplas RSA - obtención de primos - análisis de resultados, habiéndose realizado a nivel prototipo sólo la obtención de primos (problema de factorización).

Además se extenderá el desarrollo para que el sistema acepte cualquier generador de cla-

ves y no sólo OpenSSL como en el prototipo experimental realizado.

Para tal fin se contará con apoyo de los dos laboratorios participantes así como también con el alumnado de la EST, en las materias relacionadas de Ingeniería en Informática e Ingeniería en Electrónica.

En resumen, al final del 2012, y como próximo paso de CoDiSe, se pretende:

- Dar el primer paso hacia la creación de un servidor de computación distribuida genérica, para ser usado en proyectos de distinta índole.
- Tener a disposición de los alumnos y profesores afectados al laboratorio de Criptografía de la EST, una herramienta robusta y potente, con un muy bajo costo de mantenimiento, que cumpla con los requisitos funcionales necesarios.
- Validar robustez de algoritmos de seguridad a nivel experimental.

Estos resultados, además, serán de gran interés para el CIDESO, que dispondrá de una herramienta fuerte para la evaluación de librerías de seguridad, permitiéndole seleccionar la mejor para el desarrollo de todos sus sistemas en general, y de su sistema de comando y control en particular, Sistema Táctico Integrado del Ejército Argentino (SITEA).

4. FORMACIÓN DE RECURSOS HUMANOS

CoDiSe tiene como característica principal haber nacido de la colaboración de dos centros de investigación dentro del Ejército Argentino. Y 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, el CIDESO tiene amplia 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.

Por el otro, el CriptoLab tiene experiencia en investigación básica en el terreno de la criptografía, asociado estrechamente al posgrado de Especialización en Criptografía y Seguridad Teleinformática que se dicta en la EST. En este laboratorio se dispone del material y los recursos humanos específicos para la investigación en esta área de la seguridad. Cuenta con expertos matemáticos y criptógrafos que dan cuerpo muchas veces a las investigaciones de lo trabajos finales de carrera que se realizan en el posgrado.

Ambos laboratorios, a través del dictado de materias de grado en Ingeniería Informática, aportan recursos humanos a la EST. Es así que gran cantidad investigadores de los laboratorios dan cátedras 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 CoDiSe 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 lo laboratorios y, potencialmente, becarios que se dediquen de modo formal (no sólo académico) a la profundización de lo modelos propuestos.

5- BIBLIOFRAFÍA

[1] A. B. Ajulo, Grid Computing An Advancement of E-Science To Computing And Beyond, Department of Computer Science, Federal University of Technology.

[2] V. Berstis, Fundamentals of Grid Computing, Redbooks Paper, IBM Corp, 2002.

[3] A. J. M. Repetto, «Hybrid Architecture for Constructive Interactive Simulation: Evaluation and Outcomes,» de */ITSEC'10, Interservice/Industry Training, Simulation and Education Conference*, Orlando, FL, 2010.

[4] A. J. M. Repetto, «Grid Desktop Computing for Constructive Battlefield Simulation,» de *XV Congreso Argentino de Ciencias de la Computación (CACIC 2009)*, San Salvador de Jujuy, 2009.

[5] Computerized Bussiness Solutions, «Centralized vs Distributed Computing: White Paper,» 2007.

[6] B. Javadi, D. Kondo, J.-M. Vincent y D. P. Anderson, Discovering Statistical Models of Availability in Large Distributed Systems - An Empirical Study of SETI@home, Berkeley: Space Sciences Laboratory, University of California.

[7] R. Rivest, A. Shamir y L. Adleman, «A Method for Obtaining Digital Signatures and Public-Key Cryptosystems,» 1978.

[8] A. Menezes, P. van Oorschot y S. Vanstone, Handbook of Applied Crytptography, CRC Press, 1997.

[9] K. M. Martin, R. Safavi-Naini, H. Wang y P. R. Wild, Distributing the Encryption and Decryption of a Block Cipher, London, England: Information Security Group, Royal Holloway, University of London.

[10] A. Ahmar, Grid Computing Technology: An Overview, Abbas, 2004.

[11] N.-Z. Constantinescu-Fulop, A Desktop Grid Computing Approach for Scientific Computing, Department of Computer and Information Science Faculty of Information Technology, Norwegian University of Science and Technology, 2008.

[12] E. T. O. Opiyo, E. Ayienga, K. Getao, B. Manderick, O. Odongo y A. Nowé, Computing Research Challenges and Opportunities with Grid Computing.

[13] D. P. Anderson, T. Estrada, M. Taufer y K. Reed, EmBOINC: An Emulator for Performance Analysis of BOINC Projects, University of Delaware, IBM, University of Berkeley.

[14] F. Berman, G. C. Fox y A. J. G. Hey, Grid Computing: Making the global infrastructure a reality, Wiley publishers, 2003.

[15] D. P. Anderson, E. Korpela y R. Walton, High-Performance Task Distribution for Volunteer Computing, Berkeley: Space Sciences Laboratory, University of California.

[16] R. Neves, N. Mestre, F. Machado y J. Lopez, Parallel and Distributed Computing BOINC Grid Implementation.

[17] J. Blythe, E. Deelman, Y. Gil y C. Kesselman, Transparent Grid Computing: a Knowledge-Based Approach, Marina Del Rey, CA: USC Information Sciences Institute.