ReCIBE, Año 3 No.1, Enero 2014

Las simulaciones, una alternativa para el estudio de los protocolos P2P

Armando de Jesús Ruiz Calderón

Departamento de Sistemas y Computación
Instituto Tecnológico de Tlalnepantla
Av. Instituto Tecnológico s/n Col La comunidad
Tlalnepantla de Baz Edo. de México
CP. 54070
armandoruizmex@gmail.com

Abel González Cañas

Departamento de Sistemas y Computación
Instituto Tecnológico de Tlalnepantla
Av. Instituto Tecnológico s/n Col La comunidad
Tlalnepantla de Baz Edo. de México
CP. 54070
goncabel@yahoo.com.mx

Sofía Barrón Pérez

Departamento de Sistemas y Computación
Instituto Tecnológico de Tlalnepantla
Av. Instituto Tecnológico s/n Col La comunidad
Tlalnepantla de Baz Edo. de México
CP. 54070
sofia_barron@hotmail.com

Resumen: La arquitectura y funcionalidad de las redes P2P hacen que sean atractivas para ser utilizadas en ambientes distribuidos locales y aplicaciones de amplia distribución, el análisis de sus protocolos de ruteo bajo diferentes ataques como son los de negación de existencia y de servicio, así como su análisis estadístico, hacen que las simulaciones cobren gran importancia, y sean una alternativa adecuada para su estudio, pues existen varios protocolos

de esta categoría como Pastry o Chord, los cuales son de gran importancia dada su amplia utilización en diferentes aplicaciones para el envío y recuperación satisfactoria de información tanto en la nube como en aplicaciones distribuidas, razón por la cual su análisis es importante, este trabajo se centra en Pastry dado que es utilizado en la versión Azure de Microsoft Windows.

Palabras clave: Redes P2P, DoS, simulaciones, cómputo en la nube.

The simulations, an alternative to the study of P2P protocols

Abstract: Abstract The architecture and functionality of P2P networks makes them attractive for use in a local distributed environments and widespread applications, the analysis of routing protocols under different attacks such as denial of existence and service, and their statistical analysis make the simulations very important charge, and are a suitable alternative for study, because there are several protocols of this category as Pastry or Chord, which are of great importance because of its wide use in different applications to sending and recovery information, both in the cloud and distributed applications, which is why his analysis is important, this work focuses on Pastry as it is used in the Microsoft Windows Azure version.

Keywords: P2P networks,DoS attacks, cloud computing, simulations

1. Introducción

En la actualidad, las aplicaciones para redes de arquitectura Peer to Peer (P2P) han retomado popularidad de manera importante por su capacidad de compartir archivos entre ellas, y se considera que el 60% del tráfico de Internet proviene de aplicaciones de redes descentralizadas (Ankur Gupta, 2008). Sin embargo, el interés general de las redes P2P se centra en los aspectos técnicos, tales como: el control descentralizado, la auto organización, su adaptación y

escalabilidad, además de reconocer que esta arquitectura tiene variantes como un control centralizado, control descentralizado estructurado y control descentralizado no estructurado, (Yunhao Liu, 2007). Una red P2P se puede considerar como un sistema distribuido, el cual provee una base para la construcción de aplicaciones descentralizadas de gran tamaño, tales como almacenaje distribuido, grupos de comunicación, etc. (M. Castro P. Druschel, 2002). Estas características hacen que las redes de arquitectura P2P sean suceptibles a diferentes tipos de ataques y vulnerabilidades de seguridad (Marlin Engle, 2006), sin embargo, este tipo de redes tienen una alta capacidad de recuperación frente a un ataque ya que pueden enrutar mensajes de forma correcta aun cuando una fracción elevada de nodos fallen o la red esté particionada; pero esto no significa que sea segura, pues con una pequeña fracción de nodos maliciosos en falla, se genera una deficiencia en el envío y recuperación de mensajes al destino solicitado, esto también se conoce como "envío no correcto" (M. Castro P. Druschel, 2002); para este propósito, se requiere una topología que sea descentralizada, escalable, auto-organizable, y que se adapte a la llegada o salida de nodos así como a la existencia de fallas en los mismos. Su gran tolerancia a fallos durante el proceso de ruteo hace muy eficiente a este tipo de protocolos dentro de la nube; a lo largo de los años su utilización ha ido aumentado de manera considerable y este protocolo forma parte del conjunto de protocolos que se utilizan en "Windows Azure" (Y. Charlie Hu, 2002).

Descripción de "Pastry"

"Pastry" se presenta como una base adecuada para aplicaciones en redes P2P conectadas a Internet las cuales pueden ser potencialmente muy grandes. En "Pastry" cada nodo tiene un identificador único llamado "nodeId" este identificador se asigna aleatoriamente a partir de un espacio de 128 bits; el identificador de nodo al cual en lo futuro se referirá como "nodeId" está formado de tres partes:

- El “Leaf Set”
- La tabla de Ruteo
- El Conjunto de vecindario

1. El “Leaf Set” contiene la información correspondiente a los nodos geográficamente más cercanos al nodo actual, éste contiene a los *L/2* nodos numéricamente más grandes y más cercanos al nodo actual, y además contiene a los *L/2* nodos numéricamente más pequeños y más cercanos al mismo; siendo *L* típicamente *2b* , ó *(2 x 2b)*.
2. La Tabla de Ruteo está organizada en $log_{2^b}(N)$ filas y *2b* columnas (Zhang Rongmei, 2003), (M. Castro P. Druschel, 2002). Las *2b* entradas en la fila *n* de la tabla de ruteo contienen las direcciones IP de los “nodeId” que se encuentran en el “Leaf Set” (Antony Rowstron, 2001) y comparten los primeros *n* dígitos del “nodeId” del nodo actual.
3. El conjunto de vecindario *M* contiene a los “nodeId” y las direcciones *IP* de los *M* nodos más cercanos de acuerdo con la métrica de proximidad. El conjunto de vecindario normalmente no se utiliza durante el ruteo de mensajes, sin embargo es útil en el mantenimiento de las propiedades de la localidad sobre todo cuando hay movimientos en los nodos.

Ruteo en Pastry

De manera sucinta se puede decir que el ruteo se da cuando un nodo envía un mensaje a otro nodo; el nodo origen, revisa el “nodeId” que trae el mensaje a enrutar y lo verifica en el “leaf set”, en caso de no encontrarse aquí, entonces verifica en su tabla de ruteo, y envía al mensaje hacia otro nodo el cual comparte al menos un dígito o d dígitos de largo en el prefijo del “nodeId” del nodo. Si no hubiera un “nodeId” conocido, el mensaje se envía a un nodo que comparta en

su prefijo al menos la misma cantidad de dígitos que el nodo actual, y que numéricamente sea más cercano al destino.

El siguiente seudocódigo muestra el algoritmo de ruteo de “Pastry” (Antony Rowstron, Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems, 2001)

```
(1)    if ( L_{|L|/2} ≤ D ≤ L_{|L|/2}) {
(2)         // D está dentro del rango del "leaf set"
(3)         avanza a  Li, s.th. |D - Li| es el mínimo;
(4)   } else {
(5)         // Utilizar la tabla de ruteo
(6)         Let l = shl(D,A)
(7)         if (R_l^{Dl} ≠ null)    {
(8)               avanza a  R_l^{Dl};
(9)         }
(10)        else {
(11)              // caso raro
(12)              avanza a T ∈ L ∪ R ∪M, s.th.
(13)                    shl(T,D) ≥ l,
(14)                   |T - D| < |A - D|
(15)         }
(16) }
```

Nodos en Falla

Para el caso de las fallas se considera la utilización de las fallas bizantinas, las cuales son muy frecuentes en la actualidad, estas fallas pueden ser producidas por un ataque o por una falla propia del nodo (M. Castro, 1999); dado que el algoritmo funciona en sistemas asíncronos como Internet, es adecuado para la simulación. Se consideran dos tipos de fallas, estos tipos son:

1. Nodos coalicionados

Los coalicionados funcionan de la siguiente manera:

a) El primer nodo recibe un ataque y se altera la tabla de ruteo provocado por el atacante.
b) El segundo nodo recibe la información de la tabla de ruteo del primer nodo y actualice su información y envía su información de la tabla de ruteo hacia el siguiente nodo en coalición y así sucesivamente hasta que todos los nodos coalicionados hayan compartido su información.
c) Los nodos coalicionados, generan lo que se conoce como mal ruteo pero bajo el esquema de negación de existencia, esto es: Se enrutan los mensajes hacia un lugar distinto al que se pidió ir, diciendo que el sitio al que se pidio ir no existe.

2. Falla de un nodo aislado

a) La falla de un nodo aislado, provoca la negación de servicio también conocido como “Denial of Service” (DoS).

Objetivo

Analizar el comportamiento del Protocolo “Pastry” bajo ataques de negación de servicio (DoS), a través de una simulación.

Contribución

El presente trabajo muestra la relación y la importancia que tienen las simulaciones para el análisis y desarrollo de los protocolos en redes P2P, estos protocolos se utilizan de manera segura en la nube siendo Microsoft Windows Azure, un producto comercial que será lanzado al mercado que utiliza Pastry como parte de su conjunto de protocolos, asi mismo el desarrollo de las simulaciones permite analizar el comportamiento de los protocolos bajo diversos escenarios, como el ataque de negación de servicio.

Metodología

Para el desarrollo de la simulación se consideran cuatro partes:

1. Generación de la red virtual
   a. Generación de los nodos
   b. Selección aleatoria de los nodos
   c. Generación de los identificadores de nodo “nodeId”
   d. Asignación aleatoria de los nodos
2. Ruteo normal
3. Generación de los ataques
4. Análisis de los resultados de la eficiencia del enrutamiento bajo diferentes niveles de ataques

Para la generación de la red virtual se proponen las siguientes variables y constantes:

1. Se considera a N como el número total de nodos en la red.
2. Se considera a ft como cantidad total de nodos en falla en la red y la cantidad se expresa en decimales.
3. Se considera a n_m como la cantidad de nodos maliciosos coalicionados que provocan mal ruteo.
4. Se considera a n_f como la cantidad de nodos en falla aislada.
5. Se considera a s como la probabilidad de ruteo exitoso, teniendo una cantidad n de nodos en falla.
6. Se considera a pf como la probabilidad que se tiene para que el ruteo de un mensaje falle cuando se tiene una cantidad ft de nodos en falla.
7. Se considera $b = 2$, siendo b un parámetro de configuración propio del protocolo el cual ayuda a determinar el tamaño de las tablas de ruteo.

Generación de la red virtual

La generación de la red virtual, es el inicio del proceso de simulación y de los ataques, ya que sin el ambiente, simplemente no se puede realizar la simulación.

- Se generan los “nodeId” estos se expresaron en base 2^b, y se guardan en una estructura de datos tipo celda (Antony Rowstron, 2001) de *1 x 1024*. Utilizando hashes según la descripción de Cormen (Thomas H. Cormen, 1990)
- Generación de los nodos.
  - Se generan las filas y las columnas
  - Se crean las tablas de ruteo
  - Se llenan las tablas de ruteo
- Utilizando la aleatoriedad, se asignarán los “nodeId” a cada nodo, tomando los valores de la estructura celda.

Ruteo

Una vez que se terminó de generar el ambiente, se realizó el enrutamiento correspondiente y se realizó de la siguiente manera:

1. De manera aleatoria se determina a que nodo se dirige el mensaje, considerando el siguiente procedimiento:

1. Aleatoriamente se selecciona el nodo por donde se va a iniciar la búsqueda.
2. Se revisan las tablas de ruteo para encontrar las coincidencias del prefijo del “nodeId” y saltar al siguiente nodo.
3. En caso de haber un nodo vacío, se revisa el renglón *i* de la tabla de ruteo.

Una vez terminada la implementación del ruteo, se pone a funcionar la red realizando envios y recepción de mensajes, considerando que en este punto el sistema funciona de manera satisfactoria.

Ataques

Para realizar los ataques se considera que el ambiente descrito se encuentra funcionando de manera eficiente, esto es enviando y recibiendo mensajes sin pérdidas ni retrasos; en este momento se inicia con los ataques, los cuales se dan de la siguiente manera:

1. Se genera un número pseudo-aleatorio que se ubica entre *[0,N]* el cual se toma como *ft*
2. Una vez obtenido *ft* se calcula el valor de *s*
3. Ya calculado el valor de *s* se calcula *pf*
4. Teniendo el valor de *ft* se calcula la cantidad de n_m .
5. Ya que se calculó el número de nodos maliciosos, entonces la diferencia entre *ft* y n_m será *nf* que corresponde a los nodos que generan DoS.

Se considera a *s* como la probabilidad de enrutar satisfactoriamente un mensaje entre dos nodos en funcionamiento cuando existe una cantidad *ft* de nodos en falla, donde:

$$s = (1 - ft)^{h-1} \qquad (1)$$

Se considera a h como:

$$h = log_2{}^b(N) \qquad (2)$$

Se considera a *pf* como la probabilidad que se tiene para que el ruteo de un mensaje falle cuando se tiene una cantidad *ft* de nodos en falla, donde:

$$pf = (1 - s) \qquad (3)$$ (M. Castro P. Druschel, 2002)

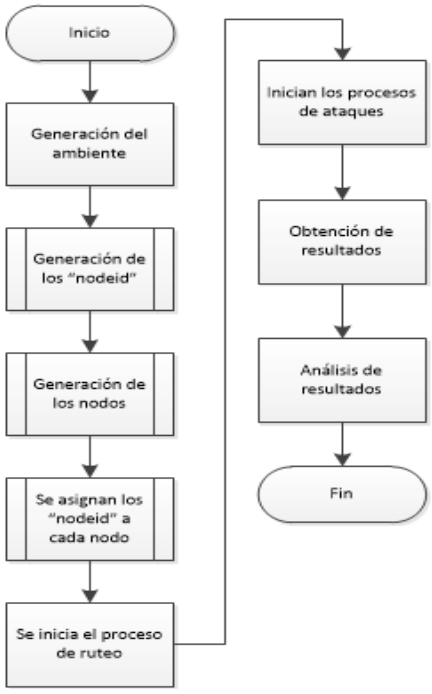


Figura 1. *Esquema simplificado del ciclo de la simulación*

Resultados

Para el desarrollo de este trabajo, se consideró una red de arquitectura tipo P2P con $N = 1024$ nodos, de manera inicial todos los nodos funcionan correctamente, esto es se realiza el enrutamiento de los paquetes de forma adecuada, considerando que se envían y se reciben los paquetes sin pérdidas ni retrasos, posteriormente se recibe un ataque y como consecuencia se detecta un número *ft* como la cantidad total de nodos que fallan. Se utilizó $b = 2$, ya que los recursos eran limitados y teniendo $b = 4$ provocaba desbordamiento de la memoria. Las fallas que se consideraron fueron de dos tipos:

1. Nodos coalicionados maliciosos que realizan mal ruteo
2. Nodos que fallan de forma aislada que provocan negación de servicio (DoS)

Durante los ataques se obtuvieron ambos valores, observados en las figuras 2 y 4 pero solo se utilizaron como datos estadísticos para el cálculo de las probabilidades de ruteo exitoso y probabilidad de falla en el ruteo.

La figura 2 muestra el comportamiento que tiene la probabilidad de que el ruteo de un mensaje sea exitoso teniendo una cantidad *ft* de nodos en falla, la prueba se ejecutó 1000 veces, y se puede observar que mientras más nodos se encuentren fallando es menor la probabilidad de enrutar exitosamente.

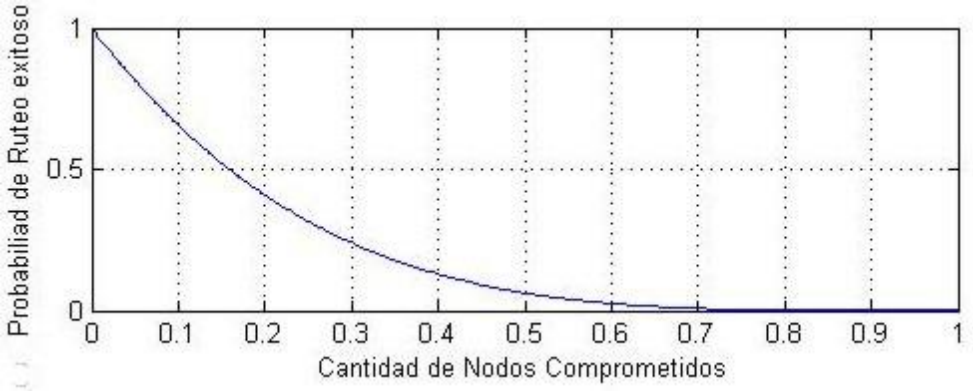


Figura 2. *Gráfica obtenida después 1000 ataques en la que se muestra la probabilidad de ruteo exitoso*

La figura 3 muestra la probabilidad de ruteo exitoso con una cantidad mayor de nodos que los que se utilizaron en esa simulación, observando el mismo comportamiento. Se puede observar una similitud en las gráficas, en donde se aprecia que la probabilidad de ruteo exitoso disminuye considerablemente, mientras aumenta la cantidad de nodos en falla.

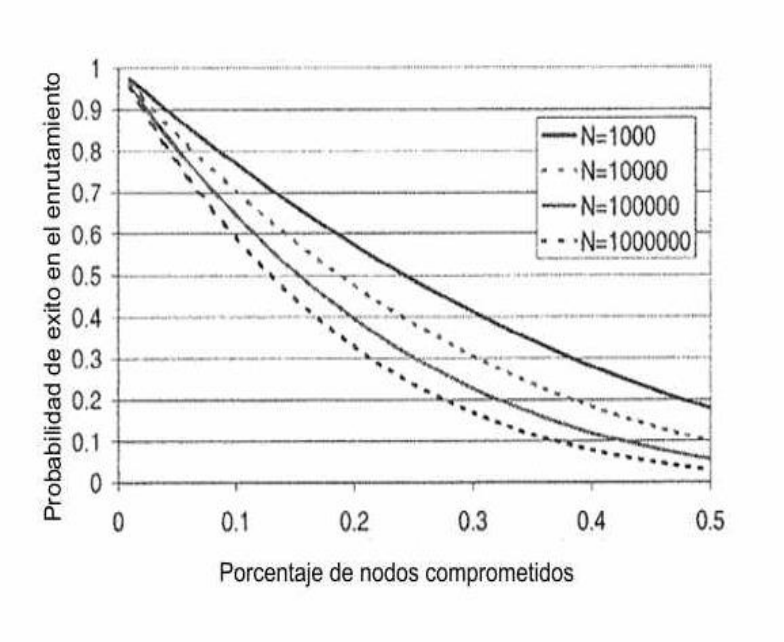


Figura 3. *Gráfica comparativa de probabilidad de ruteo exitoso tomada de (M. Castro P. Druschel, 2002)*

Discusión

Para el ruteo se consideró que todos los nodos funcionaban correctamente, y solo se revisa la tabla de ruteo, para esta simulación no se hace referencia al "leaf set". La figura 4 muestra la probabilidad de fracaso en el ruteo existiendo una cantidad *x* de nodos comprometidos, y se puede observar como la probabilidad de fracaso se eleva considerablemente. Se debe considerar, que en la simulación no existen falsos positivos, ni falsos negativos se puede observar que en estas dos figuras 4 y 5, la probabilidad de fracaso en el ruteo es similar cuando se tienen aproximadamente la mitad de los nodos comprometidos y aunque las gráficas no son muy parecidas, se puede observar que el comportamiento y los datos coinciden en ambas.

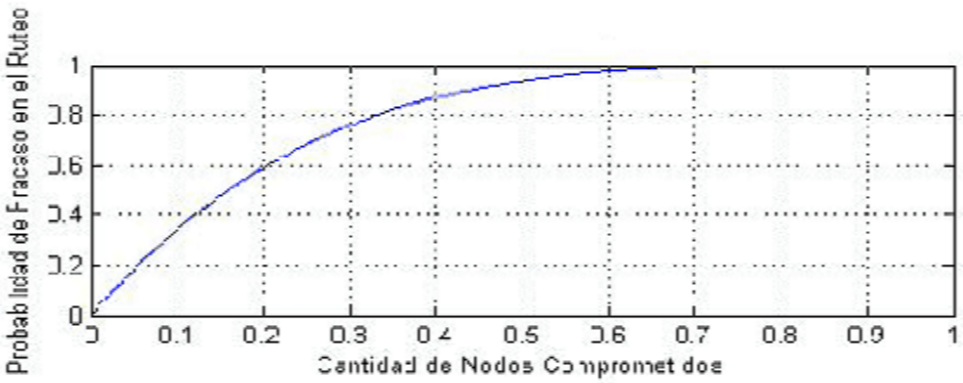


Figura 4. *Gráfica que muestra la probabilidad de falla en el ruteo considerando diferentes muestras*

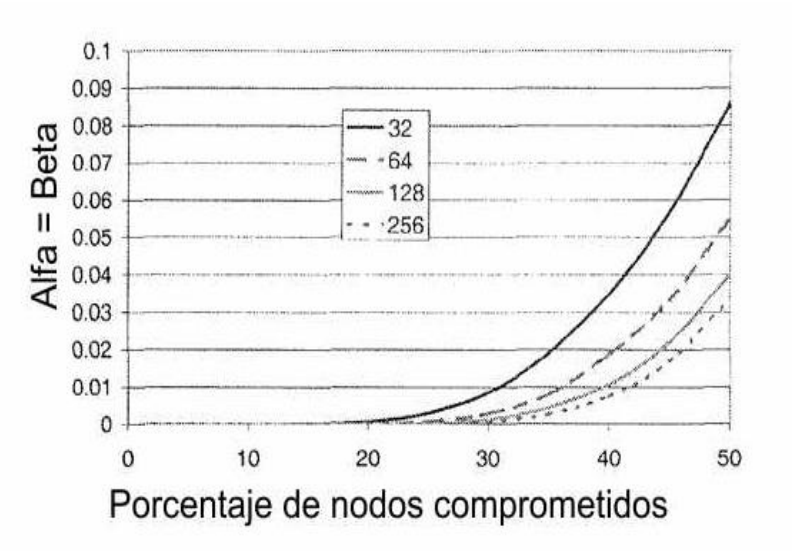


Figura 5. *Gráfica que muestra la probabilidad de falla en el ruteo con varias muestras tomada de (M. Castro P. Druschel, 2002)*

Conclusiones

- Las simulaciones desarrolladas para protocolos de redes P2P revisten gran importancia pues esta arquitectura al contrario de la arquitectura tradicional de redes es más robusta, mientras más nodos tenga mejor será su desempeño.
- De no existir las simulaciones, tratar de realizar un análisis de estos protocolos resultaría una tarea casi irrealizable.

- El análisis estadístico realizado, permite analizar el comportamiento que puede tener el protocolo, y por lo tanto, la red bajo una serie de diversas condiciones y escenarios.
- El protocolo “Pastry” es muy importante ya que fue desarrollado para ser utilizado como parte de la suite de protocolos que utilizará Microsoft en su versión de “Windows Azure” que trabajará en la nube, razón por la cual resulta importante su análisis.
- El protocolo presenta fallas en su diseño ya que no tiene fortaleza bajo ataques de negación de existencia, cuando existen varios nodos coalicionados.

Referencias

Ankur Gupta, D. M. (1 de 08 de 2008). : A Trust-based Scheme for Countering Distributed Denial-of-Service Attacks in P2P Networks NeighborTrust. ICON.

Antony Rowstron, P. D. (2001). Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. IFIP/ACM Middleware 2001, Heidelberg: ACM.

Antony Rowstron, P. D. (2001). Storage management and caching in PAST, a large scale, persistent peer to peer storage utility. Proc ACM SOS'P. Banff: ACM.

M. Castro P. Druschel, A. G. (2002). Secure Routing for Structured Peer-to-Peer Overlay Networks. Symposium on Operating Systems Design and Implementation. Boston: ACM.

M. Castro, B. L. (1999). Practical byzantine fault tolerance . Third Symposium on Operating Systems Design and Implementation (OSDI'99). New Orleans,: ACM.

Marlin Engle, J. I. (2006). Vulnerabilities of P2P Systems and a Critical Look at their Solutions,. Technical Report, Kent State University, Internet and Media Communications Research Laboratories, Kent.

Thomas H. Cormen, C. E. (1990). Introduction to Algorithms. (M. E. Series., Ed.) MIT Press.

Y. Charlie Hu, A. R. (2002). Exploiting network proximity in peer to peer overlay networks. Microsoft Research.

Yunhao Liu, X. L. (2007). Defending P2Ps from Overlay Flooding-based DDoS. International Conference on Parallel Processing. Michigan: IEEE.

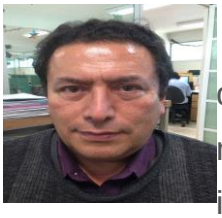
Zhang Rongmei, C. H. (2003). Borg: A Hybrid protocol for scalable application level multicast in peer to peer networks Monterey California. En ACM (Ed.), Nossdaw 03. Monterrey

.

Notas biográficas:

Armando de Jesús Ruiz Calderón Realizó estudios de Licenciatura en Biología en la UNAM y se graduó en el año 1993. En el año 2007 obtuvo el grado de Maestro en Ciencias en el Instituto Tecnológico de Estudios Superiores de Monterrey (ITESM), campus Monterrey, en el área de Ciencias en Tecnología Informática. Sus líneas de investigación están relacionadas con el análisis de ataques de negación de existencia y negación de servicio sobre el esquema de redes P2P, utilizando como base la arquitectura experimental de Pastry. De las investigaciones realizadas en este campo han resultado varios artículos publicados en diferentes eventos de carácter internacional, tales como: CIPITECH 2010, 2011; ANIEI 2010; Academia Journals 2013; Tekhne 2013. Trabajó en Becton Dickinson como administrador de redes y telecomunicaciones, en Ddemesis como administrador de sistemas, también ha colaborado en editoriales realizando la traducción y corrección de estilo en libros del área de medicina. Desde el año 1991 inició su trabajo docente en el Instituto Tecnológico de Tlalnepantla, ha colaborado como docente en los tecnológicos de Aguascalientes y Nuevo León y actualmente es Secretario del Consejo de Posgrado de la Maestría en Tecnologías de la Información en el Tecnológico de Tlalnepantla, y profesor investigador en el departamento se Sistemas y Computación; obtuvo el reconocimiento de profesor con perfil deseable en el año 2011 por parte de PROMEP.

Sofía Barrón Pérez Nació en Tlalnepantla Estado de México, el 16 de Diciembre de 1975, graduada de la carrera de Licenciatura en Informática en el Instituto Tecnológico de Tlalnepantla, ha ocupado cargos como Jefe de Sistemas en la empresa SKYDOM (Qro.) y como programadora en la misma empresa. actualmente docente en el Instituto Tecnológico de Tlalnepantla, así mismo ha acreditado diplomados en el CIIDET de Querétaro, como son Modelo Educativo Siglo XXI, Modelo Educativo por Competencias y Docencia Centrada en el Aprendizaje, ha participado como Facilitadora y Supervisora en el ESAD; actualmente esta concluyendo la Especialidad en Tecnologías de Información en el Centro Interdisciplinario de Investigación y Docencia en Educación Técnica, ha dirigido Tesis a Nivel Licenciatura, y participado como ponente en diferentes congresos de las asociaciones ANIEI (2010), CIIM(2011) y CIIM(2012), Congreso en el IT de Morelia (2013) así como en la Academia Journals en el 2012 y 2013.



Abel González Cañas Nació en Orizaba Veracruz en el 29 de octubre de 1957, se graduó como Ingeniero Eléctricista, el 6 de noviembre de 1987 en el Instituto Tecnológico de Cd Madero ingresó como docente en el Instituto Tecnológico de Tlalnepantla en el año de 1986.

Ha trabajado en empresas como Fabricaciones Ingeniería y Montaje, Cerveceria Moctezuma, y Comisión Federal de Electricidad,ha trabajado en institutos de investigación desarrollando partes robóticas; ha dirigido Tesis a Nivel Licenciatura, y participado como ponente en diferentes congresos de las asociaciones ANIEI (2010), CIIM(2011) y CIIM(2012), Congreso en el IT de Morelia (2013) así como en la Academia Journals en el 2012 y 2013, ha sido lider de proyecto de investigación, en el departamento de sistemas y computación.