

84 >



ISSN 0186-4084

9

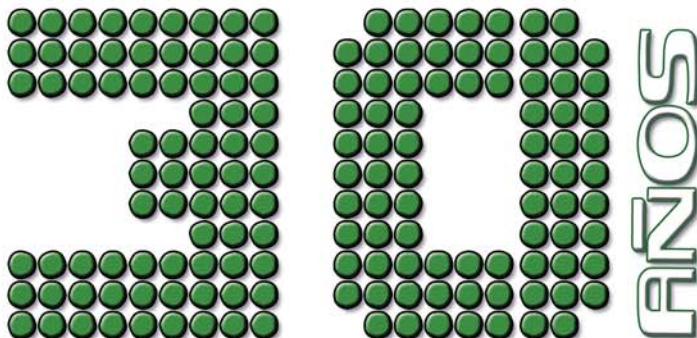
770186408006

ContactoS

ISSN: 0186-4084

REVISTA DE EDUCACIÓN EN CIENCIAS E INGENIERÍA

\$ 20.00 pesos



DEPARTAMENTO
DE INGENIERÍA ELÉCTRICA

*UAM/ en
sus inicios*

COPYLEFT

SUDOHU

P2P



Revista de las divisiones de CBI y CBS



UNIVERSIDAD AUTÓNOMA METROPOLITANA

Casa abierta al tiempo



Francisco Toledo

Francisco Toledo es un artista zapoteco conocido internacionalmente por su renovación en formas y técnicas antiguas, tales como la pintura, escultura y la cerámica. Comprometido con sus orígenes indígenas, difunde y promueve la cultura de su natal Oaxaca.

Un ser humano sencillo, observador, de ligera vestimenta, huaraches; cabello largo ondulado y ligeramente despeinado es lo que caracteriza a Toledo. La riqueza étnica y cultural de Oaxaca sumada a su visión de lo mítico, ha hecho que éste artista mexicano exprese con gran sentimiento, su sentido de la vida.

Obsesivo de las texturas y los materiales como la arena o el papel amate, así como la maestría con la que materializa su creación, consigue que su obra parezca vibrar como si la criatura híbrida de animal y hombre o cualquiera de sus seres tropicales pugnaran por cobrar vida.

El arte de Toledo refleja una gran apreciación de lo natural, particularmente la de animales que no son convencionalmente asociados con la belleza. Su visión moral afirma que el mundo de los humanos y el de los animales son uno con la naturaleza.



Mario Alberto Olmedo Juárez

Contenido

ContactoS No. 84, Abril–Junio 2012

Editorial	3	<i>Solución al acertijo del Sudoku con recocido simulado.</i>	31
<i>La Universidad Autónoma Metropolitana Iztapalapa en sus inicios.</i> Fausto Casco Sánchez.	5	Miguel Angel Gutiérrez Andrade, John Goddard Close y Sergio Gerardo de los Cobos Silva.	
<i>Influencia del desarrollo de las áreas de Ingeniería Biomédica de la UAM-Iztapalapa en los planes de estudio de la licenciatura en Ingeniería Biomédica.</i> Joaquín Azpiroz Leehan, Fabiola Martínez Licona y Edmundo Gerardo Urbina Medal.	9	<i>Comunicaciones en casos de misión crítica.</i>	40
<i>Centro Nacional de Investigación en Imagenología e Instrumentación Médica.</i> Andrés Morón Mendoza, Fernanda Maldonado Sandoval y Gabriela Beltrán López.	21	Enrique Rodríguez-Colina.	
<i>Reflexiones sobre las tecnologías de la información y la conciencia de lo “ecológicamente correcto”.</i> Gerardo A. Laguna Sánchez, Enrique Rodríguez de la Colina, Michael Pascoe Chalke y Víctor M. Ramos Ramos	25	<i>Configuración y Programación de un Clúster de GPUs.</i>	47
		Luis A. Regalado Orocio, Julio C. Boñaños Palacios, José L. Quiroz Fabián, Miguel A. Castro García y Manuel Aguilar Cornejo.	
		<i>De compartir contenido y encontrarlo usando P2P.</i>	56
		Geovanni Medrano Chávez y Elizabeth Pérez Cortés.	
		<i>Procesamiento de imágenes previo a la simulación de modelos biomecánicos.</i>	68
		Alejandro Guzmán De León y Hanspeter Frei.	
		<i>Internet es copyleft.</i>	81
		Alan Lazalde Cruz.	

cs

ContactoS en la WEB
Lea los artículos publicados en ContactoS en
<http://www.itz.uam.mx/contactos/>

Rector General

Dr. Enrique Fernández Fassnacht

Secretario General

Mtra. Iris Santacruz Fabila

UNIDAD IZTAPALAPA

Rector

Dr. Javier Velázquez Moctezuma

Secretario

Dr. Óscar Comas Rodríguez

Director de la División de Ciencias Básicas e Ingeniería

Dr. José Antonio de los Reyes Heredia

Director de la División de Ciencias Biológicas y de la Salud

Dr. Rubén Román Ramos.

ContactoS: Directora

M. en C. Alma Edith Martínez Licona.

Consejo Editorial: Dr. Javier Velázquez Moctezuma, Dr. Óscar Comas Rodríguez, Dr. José Antonio de los Reyes Heredia, Dr. Rubén Román Ramos., M. en C. Alma E. Martínez L., UAM-Iztapalapa

Editor responsable: M. en C. Alma Edith Martínez Licona.

Comité Editorial por CBI: Alberto Rojas, José Luis Córdova, Norberto Aquino; *por CBS:* Alejandra Quintanar, Margarita Salazar, Rocío Torres y Enrique Canchola.

Coordinación Editorial: Lourdes Barriga C.

Captura y Procesamiento: Oscar Ávila Mejía.

ContactoS. REVISTA DE EDUCACIÓN EN CIENCIAS E INGENIERÍA 3a. Época, No. 84, Abril–Junio 2012, es una publicación trimestral, editada, publicada y distribuida: UAM–Iztapalapa, Edificio E–317, Av. San Rafael Atlixco No. 186, Col. Vicentina Iztapalapa, C.P. 09340, México, D.F. Tel. 5804–4606, www.izt.uam.mx, cts@xanum.uam.mx, producida por la División de Ciencias Básicas e Ingeniería y la División de Ciencias Biológicas y de la Salud de la UAM–Iztapalapa. Editor responsable: M. en C. Alma Edith Martínez Licona. Reserva de derechos al uso exclusivo No. 04–2004–011510574000–102, ISSN 0186–4084. Licitud de título N o. 3769. Licitud de contenido No. 2546, ambos otorgados por la Comisión Calificadora de Publicaciones y Revistas Ilustradas del 7 de noviembre de 1986 de la Secretaría de Gobernación. Impresa por Galavisa y/o Roberto C. Galaviz Sánchez, Córdoba No. 1 A, Col. Roma, C.P. 06700.

Web: www.galavisa.com e-mail: ventas@galavisa.com
Este número se terminó de imprimir el 7 de Junio de 2012 con un tiraje de 1000 ejemplares

Las opiniones de los autores no necesariamente coinciden con las del Comité Editorial.

Se agradecerá la reproducción de los materiales citando la fuente.

Fecha de publicación: Abril–Junio de 2012.

Precio por ejemplar \$ 20.00

Los artículos publicados en **ContactoS** son sometidos a arbitraje; para ello se requiere enviar el original del trabajo en algún procesador de texto a doble espacio, dos copias claras del mismo y un CD con el archivo del artículo. Toda correspondencia deberá enviarse a:

Comité Editorial de la Revista **ContactoS**,
UAM–Iztapalapa, E–317, Telfax. 5804–4606
Av. San Rafael Atlixco No. 186, C. P. 09340, México, D. F.
Apartado postal 55-534
<http://www.izt.uam.mx/contactos>
e-mail cts@xanum.uam.mx

Para suscripciones enviar cheque o giro postal a nombre de: Universidad Autónoma Metropolitana, indicando claramente: nombre, dirección y números que comprende la suscripción. Suscripción anual (4 números) \$ 80.00 M.N. En el extranjero 60 Dls. U.S., a partir de 1995.

Editorial

En este número, nuestra Revista ContactoS aborda temas relacionados con el Departamento de Ingeniería Eléctrica de la División de Ciencias Básicas e Ingeniería (CBI), como uno de los eventos conmemorativos por el aniversario 30 de su creación. Representa una excelente oportunidad para celebrar logros significativos en el ejercicio de las funciones de docencia, investigación y difusión de la cultura a lo largo de su existencia. Desde la historia de la creación de sus grupos hasta la presentación de temas de avanzada en las Ciencias y Tecnologías de la Información, en este número tendremos la ocasión de leer aportes vinculados con este departamento. La División de CBI se congratula por los éxitos conseguidos y hace votos porque continúe su consolidación en beneficio de toda la comunidad de CBI.

ContactoS ha iniciado una serie de cambios en su imagen y presentación gráfica, que esperamos contribuyan a facilitar la difusión de la ciencia y el desarrollo tecnológico. Esta ha sido una labor que las Divisiones de Ciencias Básicas e Ingeniería y de Ciencias Biológicas y de la Salud realizan de manera conjunta y nos complace que evolucione favorablemente como un vehículo donde confluyan científicos no sólo de nuestra Universidad, sino de varias instituciones de México y del mundo de habla hispana. Esperamos que ContactoS llegue a un mayor número de lectores y de manera cabal cumpla con su cometido.

En épocas donde el conocimiento científico no se valora de manera adecuada, resulta primordial disponer de un vehículo para que se potencie su divulgación. Asimismo, habremos cumplido con una de las misiones de la Universidad pública al llegar cada vez a más generaciones de jóvenes mexicanos y de los países latinoamericanos.

Aprovechamos la ocasión para reiterar un reconocimiento al Dr. José Luis Córdoval Frunz, académico destacado del Departamento de Química de CBI, por su incansable labor durante varios años al frente de ContactoS. Apreciamos su contribución a que la revista se hubiera mantenido de manera significativa en el panorama universitario.

Finalmente, deseamos que la dirección de la Mtra. Alma E. Martínez Licona se conduzca por vías exitosas y que permitan reforzar las tareas de divulgación de la ciencia y los nuevos desarrollos tecnológicos. Consideramos que ContactoS seguirá siendo una ventana para que el quehacer científico de una comunidad destacada en la Unidad Iztapalapa siga llegando al público en general, igualmente para que académicos de otras latitudes puedan compartir los importantes trabajos científicos que realizan desde sus espacios de trabajo.

Dr. José Antonio de los Reyes Heredia
Director de la División de Ciencias Básicas e Ingeniería

Información para autores

ContactoS es una Revista de Educación en Ciencias e Ingeniería dirigida a profesores y a estudiantes de estas disciplinas.

Está registrada en Latindex, Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal.

Para publicar, los trabajos deberán ser originales y accesibles a un público amplio con formación media superior o universitaria pero no especializada; los temas deberán presentarse en forma amena. Cada colaboración debe incluir figuras, diagramas, ilustraciones, fotografías, etc., que hagan más accesible la presentación.

Las secciones que la constituyen son:

1. **Divulgación.** Artículos que presentan temas científicos con enfoques novedosos y accesibles (15 cuartillas).
2. **Educación Científica.** Enfoques originales en la enseñanza de temas particulares (15 cuartillas).
3. **Artículos Especializados.** Reportes breves de investigación, relacionados con una problemática concreta (15 cuartillas).
4. **Crónicas.** Historia y desarrollo de conceptos científicos, así como teorías alternativas (15 cuartillas).
5. **Divertimientos.** Juegos y acertijos intelectuales (5 cuartillas).
6. **Noticias breves.** Información de actualidad en el mundo de la ciencia (4 cuartillas).
7. **Los laureles de olivo.** Los absurdos de la vida cotidiana y académica (4 cuartillas).

En todos los casos se debe incluir los nombres completos de los autores con su adscripción, dirección, teléfono y dirección de correo electrónico.

Normas

Las colaboraciones a las secciones 1 a 4 deberán ajustarse a las siguientes normas:

1. Un resumen breve escrito en inglés.
2. 4 palabras clave en español e inglés.
3. Cuando se incluya una abreviatura debe explicarse por una sola vez en la forma siguiente: Organización de los Estados Americanos (OEA)...
4. Cuando se utilice un nombre técnico o una palabra característica de una disciplina científica deberá aclararse su significado de la manera más sencilla posible.

5. Las citas textuales deberán ir de acuerdo al siguiente ejemplo: En cuanto a la publicación del placebo se asevera que “el efecto placebo desapareció cuando los comportamientos se estudiaron en esta forma” (Núñez, 1982, p. 126).
6. Las referencias (no más de 10) se marcarán de acuerdo al siguiente ejemplo: Sin embargo, ese no es el punto de vista de la Escuela de Copenhague (Heisenberg, 1958), que insiste en...
7. Al final del artículo se citarán las referencias por orden alfabético de autores. Pueden añadirse lecturas recomendadas (no más de 5).
8. Cada referencia a un artículo debe ajustarse al siguiente formato: Szabadváry, F. y Oesper, E., Development of the pH concept, *J. Chem. Educ.*, 41[2], pp.105-107, 1964.
9. Cada referencia a un libro se ajustará al siguiente formato: Heisenberg, W., *Physics and Philosophy. The Revolution in Modern Science*, Harper Torchbooks, Nueva York, 1958, pp.44-58.
10. Los títulos de reportes, memorias, etcétera, deben ir subrayados o en itálicas.

Envío y características del artículo

El envío del artículo deberá ser en archivo electrónico, ya sea en WORD o TEX, tipo de letra Times New Roman, tamaño 12 con interlineado sencillo y uso de editor de ecuaciones.

En el caso de ilustraciones por computadora (BMP, JPG, TIFF, etc.) envíelos en archivos por separado.

El material es recibido en:

Revista Contactos, UAM-Iztapalapa, E-317,
cts@xanum.uam.mx, telfax 5804-4606,
S. Rafael Atlixco 186, C. P. 09340,
México, D. F. A. P. 55-534.

Arbitraje

El Comité Editorial utiliza un sistema de arbitraje anónimo que requiere un mes. Se entiende que los autores no han enviado su artículo a otra revista y que dispondrán de un plazo máximo de un mes para incorporar las observaciones de los árbitros.

La decisión final de publicar un artículo es responsabilidad exclusiva del Comité Editorial. Despues de la publicación del artículo los autores recibirán 5 ejemplares gratuitos.

cs



UAMI **sus inicios**

Fausto Casco Sánchez

Recibido: 13 de febrero de 2012.

Aceptado: 26 de abril de 2012.

La UAM abrió sus puertas en 1974. Debido a las necesidades de la nueva universidad se requirió contratar a la planta de profesores que atendiera las necesidades docentes y de investigación. Muchos profesores fueron contactados por sus colegas y también hubo campañas de reclutamiento en centros de estudio.

En sus inicios el entorno de trabajo en la UAM era novedoso. Las instalaciones eran nuevas, muchos de los profesores eran jóvenes que recién habían terminado sus estudios de posgrado. El modelo profesor-investigador en la universidad permitía que el profesor además de dar clases, perteneciera a un área de investigación. Así, el profesor podría continuar actualizándose y avanzando en su campo de conocimiento.

En la División de CBI de la Unidad Iztapalapa, existía el Departamento de Ingeniería con varias áreas de investigación, el jefe de este Departamento era el Dr. Diego Bricio Hernández Castaños. Como era de esperarse, con el paso del tiempo se comenzó a pensar en una partición de este departamento, ya que al ser demasiado grande, su administración se dificultaba.

Las correspondientes áreas de investigación del Departamento de Ingeniería en el año de 1974 eran Ing. Biomédica e Ing. Electrónica. En particular el área de Ing. Electrónica estaba formada por los profesores-investigadores: Daniel Barrios Zaragoza, Armando Maldonado Talamantes, Rafael Mantilla, Raúl Plascencia Amoroz, Luis Niño de Rivera y Oyarzabal, Emmanuel Moya Anica y Adriano de Luca Penachia, Guillermo Levine Gutiérrez, Daniel Salas Pedraza, Fausto Casco Sánchez, Ernesto Mercado Escutia y el técnico Elpidio Alcántara.

Primera etapa 1974

En esta primera etapa se atendió a la licenciatura en Ing. Electrónica, desde ese tiempo se fomentó la formación de profesores. Armando Maldonado Talamantes y Daniel Salas Pedraza participaron en esta labor, reincorporándose a inicios de 1982. Algunos directores de proyectos de investigación realizados en esta época fueron:

- Proyecto para captación de información (PRO-CAIN) Daniel Barrios Z., Raúl Plascencia A.
- Optimización de fortran estructurado (forest) para la HP 3000/ III. Guillermo Levine.G.

- Terminal de video alfanumérica. Emmanuel Moya A.

- Simulación de un Filtro Digital de Respuesta a Impulso Finita. Fausto Casco Sánchez.

Segunda etapa 1982

En el año de 1982 inicia una segunda etapa se dio la partición del departamento de Ingeniería dando lugar a los Departamentos de Ingeniería de Procesos e Hidráulica (IPH) e Ing. Eléctrica (IE). En paralelo con esto surgió la licenciatura en Computación, propuesta por la División. En ese entonces el jefe de División era el Dr. Eduardo Piña Garza. La idea original era que fuese atendida principalmente por profesores de Ing. Electrónica y profesores del Departamento de Matemáticas. Dado que la demanda para esta carrera era grande hubo necesidad de contratar a más profesores siendo la plantilla en aquel entonces:

Profesores titulares

Emmanuel Moya Anica
Armando Maldonado Talamantes
Yudiel Pérez Espejo
Adriano de Luca Pennachia
Sergio de los Cobos Silva
John Goddard Close

Profesores Asociados

Américo Vargas Villazón
Manuel Carrillo Ricalde
Daniel Salas Pedraza
Fausto Casco Sánchez
Omar Amin Abdel Rahaman
Ernesto Mercado Escutia
Jesús Barrios Romano
Juan Carlos Sánchez García
Raúl Plascencia Amoroz
Eduardo Acosta Cartas
Elizabeth Pérez Cortés
Graciela Román Alonso
Eduardo Rodríguez Flores
Pedro Flores Sánchez
Mariko Nakano
Luis Martín Rojas Cárdenas
René Mackinney Romero

Profesores Asistentes

Oscar Hernández Córdoba
Carlos Hahn Ramírez
Carlos Islas Pérez
Héctor Pérez Meana
Rafael Saavedra Barrera
Miguel Robles Cahero
Mauricio López Villaseñor
Alejandro Martínez González
Laura Ortíz Balbuena
César Jalpa Villanueva
Luis Castro Careaga
Jorge Lozano Moreno
Eleasid García Gómez
Gonzalo Duchén Sánchez
Miguel Ángel Guzmán López
Ricardo Marcelín Jiménez

Profesores de tiempo parcial

Guillermo Levine Gutiérrez
Cuauhtémoc Utrera
Othón Gандarilla Carrillo
Roberto Sotero Ortega
Javier Almanza Estrada
Lucía Pineda Cruz
Alejandro Ávila García
Claudia Gutiérrez Massotti

Algunos directores de proyectos desarrollados en el área para esta segunda etapa fueron:

- Predicción Lineal.
Daniel Salas Pedraza
- Sistema Automático para la Generación de Reportes para Proveeduría de la UAM-I.
Raúl Plascencia Amoroz.
- Sistema de Catalogación Bibliográfica y Control de Préstamo de Libros a Domicilio.
Américo Vargas Villazón
- Modulación delta de doble paso.
Fausto Casco Sánchez
- Adaptador de Línea para Transmisión de datos a distancias cortas.
Emmanuel Moya Anica
- Sistema de Programación para la solución de convolución y correlación para docencia.
Mauricio López Villaseñor
- Síntesis de Tonos Musicales.
Luis Niño de Rivera y Oyarzabal
Héctor Pérez Meana

Por otra parte, en esta segunda etapa varios profesores del área dejaron la Universidad, debido a la devaluación del salario. Por lo anterior se tuvieron años muy difíciles, que por fortuna fueron superados.

Tercera etapa 1992

Una tercera etapa inició en el año de 1992, con la división del Área de Ing. Eléctronica, dando lugar a las áreas de investigación:

- Computación y Sistemas
- Redes de Comunicación e Interconectividad
- Sistemas Digitales

Varios profesores de estas nuevas áreas participaron en un programa de formación de Doctores, por ejemplo en el Área de Redes de Comunicación e Interconectividad participaron los profesores:

Programa de formación de doctores

Fausto Casco Sánchez
Ricardo Marcelín Jiménez
César Jalpa Villanueva
Víctor M. Ramos Ramos
Miguel Ángel Ruiz Sánchez

La contratación de los profesores Víctor Ramos Ramos y Miguel Ángel Ruiz Sánchez tuvo lugar un poco antes de esta tercera etapa.

Poco a poco se fueron graduando los profesores que participaron en la formación de doctores. Precisamente, en el 2004 dos de ellos obtuvieron su grado, así se completaron cinco doctores en el área. Es en esta tercera etapa que se consiguieron metas académicamente importantes.

Cuarta etapa año 2006

En Junio de 2006 se formó el Área de Redes y Telecomunicaciones (anteriormente Redes de Comunicación e Interconectividad), la cual está conformada por once profesores que participan activamente en docencia, investigación y difusión de la cultura. Seis de ellos tienen el grado de doctor y tres son ingenieros con el 100 % de créditos de Maestría (Tabla 1).

Docencia

Sobre el Área de Redes y Telecomunicaciones recaía la responsabilidad de impartir la gran mayoría de las UEA del tronco básico profesional y especialidad de una de las licenciaturas con mayor demanda en la Unidad Iztapalapa: la licenciatura en In-

Tabla 1. Conformación del Área de Redes y Telecomunicaciones.

Titulares de tiempo completo	8
Asociados tiempo completo	3
Doctores	8
Maestros	0
Ingenieros	3

geniería Electrónica. Como consecuencia lógica de esta responsabilidad, en últimos años la coordinación de esta licenciatura ha estado a cargo de alguno de los miembros del Área. Al respecto, es de resaltar el trabajo del Área para lograr la acreditación de esta licenciatura ante las instancias correspondientes. En la Tabla 2 se presenta un resumen de la participación en docencia que incluye información sobre el número de cursos impartidos en ese tiempo.

Tabla 2. Participación en docencia durante 2001-2003.

Cursos impartidos en Ing. Electrónica	91
Proyectos terminales dirigidos	42
Proyectos de servicio social	17
Certificación de la Lic. en Ing. Electrónica	1
Aprobación por el Divisional de la Maestría en Ciencias y Tecnologías de la Información*	1

* en colaboración con el Área de Computación y Sistemas

La apertura de la Maestría en Ciencias y Tecnologías de la Información proporcionó un gran apoyo a la investigación que realizan los profesores-investigadores.

Investigación

Las líneas de investigación que se cultivan son: Multimedia Distribuida, Redes y Servicios de Telecomunicaciones y Sistemas de Comunicación Digital. Las líneas de investigación con sus respectivos integrantes son:

1. Línea de Investigación: Multimedia Distribuida

Integrantes

Dr. Alfonso Prieto Guerrero
Dr. Miguel Ángel Ruíz Sánchez
Dr. Víctor M. Ramos Ramos
Ing. Omar Amín Abdel Rahaman

2. Línea de Investigación: Redes y Servicios de Telecomunicaciones

Integrantes

Dr Ricardo Marcelín Jiménez
Dr. César Jalpa Villanueva
Dr. Víctor M. Ramos Ramos
Dr. Miguel Ángel Ruíz Sánchez
Dr. Miguel López Guerrero
Dra. R. Carolina Medina Ramírez

3. Línea de Investigación: Sistemas de Comunicación Digital

Integrantes

Dr. Fausto Casco Sánchez
Ing. Omar Amín Abdel Rahaman
Dr. Alfonso Prieto Guerrero
Ing. Mauricio López Villaseñor
Ing. Miguel Ángel Gutiérrez Galindo

Desde luego el entorno que se vive actualmente en la universidad es muy diferente al de hace 30 años, en la actualidad hay muchas áreas de investigación consolidadas y otras en formación. La productividad se ha incrementado notablemente, el número de doctores ha aumentado a través del tiempo, un aspecto importante es el número de alumnos que han egresado de las licenciaturas, los cuales se han incorporado al mercado laboral.

Respecto al modelo de la UAM que es el binomio docencia-investigación, pienso que su realización ha sido fructífera. Un factor que debe continuar en la universidad es el apoyo a los profesores-investigadores en la realización de proyectos académicos que les den la oportunidad de realizarse tanto en la docencia como en la investigación; sin perder de vista que el reto en este tiempo es mayor comparado con el de hace 30 años, la competitividad en el tiempo actual es mayor, de tal forma que los egresados deben estar preparados para emplearse en el mercado de trabajo. El alumno actual debe egresar con una mentalidad que considere los aspectos del ambiente, deben tener una actitud ante la vida muy positiva, con grandes anhelos de realizarse, de contribuir al desarrollo del país.

cs

Influencia del desarrollo de las áreas de Ingeniería Biomédica de la UAM-Iztapalapa en los planes de estudio de la licenciatura en Ingeniería Biomédica

Recibido: 26 de enero de 2012.

Aceptado: 01 de marzo de 2012.

Introducción

Se podría llevar a cabo una evaluación histórica acerca del desarrollo de las Áreas de Ingeniería Biomédica a través de los registros de los logros académicos de sus miembros. Sin embargo, dado que estos registros son escasos porque durante un largo período la preocupación principal del grupo de profesores dedicados a la Ingeniería Biomédica (IB) no fue la investigación sino la atención a los alumnos, el equipamiento (en buena medida autoequipamiento) de los laboratorios de docencia y su propia formación académica (ya que la mayoría de los profesores no contaban con estudios de posgrado en 1974), hemos decidido efectuar un recuento de la influencia que ha tenido el desarrollo histórico de las Áreas de Ingeniería Biomédica de la UAM-I en los planes de estudio que fueron promovidos e impartidos por sus profesores, ya que reflejan en buena medida las expectativas que se han tenido acerca de los alumnos, del entorno relacionado con la industria biomédica y del impacto que se deseaba tener sobre estas actividades en el país.

El origen de las reflexiones que llevaron a la escritura de este artículo se remonta a una serie de trabajos que se han publicado a lo largo de los años relacionados con la docencia y la historia de la IB, donde se efectúan análisis comparativos con planes de estudios de distintas instituciones nacionales e internacionales[1-6]. Sin embargo, su antecedente más importante es el Simposium sobre Diseño Curricular que se llevó a cabo del 11 al 14 de mayo de 2011 en la UAM y en San Miguel de Allende, Guanajuato. En este evento, donde participaron profesionales de la docencia en IB de 14 instituciones diferentes, nacionales e internacionales, se trataron temas relacionados con el análisis de los planes de estudio existentes y la forma que éstos deberán adoptar en el futuro.

Algunos de los temas importantes que se trattaron en ese evento fue la impartición de los cursos del primer año y la alta tasa de reprobación en éste. Parte de la causa del problema fue atribuida a la mala formación en matemáticas de los alumnos de primer ingreso, pero otra causa importante fue la impartición de asignaturas matemáticas por matemáticos que no han modificado sus cursos ni sus métodos de enseñanza para adecuarse a lo que se espera de un alumno de ingeniería: resolver proble-

mas en lugar de estudiar teoremas. Otro problema más fue el largo período de tiempo que separa a las asignaturas del primer año de las asignaturas propiamente de la licenciatura. Finalmente se mencionó que la gran influencia de las asignaturas originarias de un Departamento de Ingeniería Eléctrica podría ser una limitación para una evolución hacia subdisciplinas del futuro en IB como Proteómica, Genómica, Bioinformática, Micro y Nanotecnología.

En este trabajo se analizarán las causas y consecuencias de las distintas formas de impartir la licenciatura que se han implantado a lo largo del tiempo. Aunque ha habido cerca de ocho cambios en la licenciatura, agrupamos a éstos en cuatro grandes épocas para facilitar su estudio. Concluiremos planteando algunas perspectivas que orienten el desarrollo de los planes de estudio de IB en el futuro.

Primera etapa: 1974-1978

Aunque ya se ha expuesto con anterioridad el origen de la IB en la UAM, conviene volver a mencionar que el primer plan de estudios fue formulado por profesores que se habían formado en dos líneas: la medicina y la ingeniería eléctrica. En ese entonces sólo uno de ellos había obtenido recientemente su maestría en IB (F. Berdichevsky, en el Instituto Tecnológico de Israel-Technion) y dos tenían experiencia en el mantenimiento de equipos médicos (M. Lindig y M. Cadena). En esa época ya existían programas de formación en IB en otras partes del mundo, pero sólo a nivel de posgrado. Por otra parte muchos de los líderes de grupos de investigación de IB en el extranjero se habían formado primero como ingenieros eléctricos y electrónicos, mientras que algunos otros fueron fisiólogos-instrumentistas.

La licenciatura estaba inmersa en el Departamento de Ingeniería de la División de Ciencias Básicas e Ingeniería, donde éste coexistía con tres departamentos de ciencias básicas: Matemáticas, Física y Química. Esta circunstancia explica la estructura del plan de estudios, pues se requería un tronco común que sirviera de base a todas las licenciaturas de la División. Este tronco consistía en un año de trabajo en las disciplinas de Física, Química y Matemáticas, asumiendo que éste sería la base de todas las formaciones, tanto en ciencias como en ingeniería. El segundo año estaba dedicado a una formación común en ingeniería (tronco de ingeniería) que consistía en un año de cursos de Matemáticas aplicadas, Probabilidad, Estadística, Computación

y Métodos Numéricos. Al mismo tiempo se iniciaba la formación en asignaturas propias de la ingeniería eléctrica (Circuitos Eléctricos y cursos de Electrónica) y las ciencias médicas (Bioquímica, Fisiología y Farmacología).

En el tercer año se continuaba con la formación en ingeniería eléctrica (Control y Circuitos Digitales), matemáticas (Investigación de Operaciones y Diseño y Optimización), y se introducían asignaturas ya específicas de una licenciatura en ingeniería biomédica, como Bioelectricidad, Transductores e Introducción a la Medicina Asistencial. Por otra parte, al término de este año aparecían las asignaturas correspondientes a las áreas de concentración.

El cuarto año se dedicaba a asignaturas de las áreas de concentración y a otras de naturaleza optativa, que en realidad terminaron siendo obligatorias porque eran las únicas que se ofrecían en ese momento. Inicialmente hubieron dos áreas de concentración: Instrumentación Médica y Matemáticas aplicadas a la Medicina.

Para la primer área de concentración, entre las asignaturas obligatorias aparecían Ingeniería Aplicada a la Fisiología I, II y III, mientras que las que se ofrecían como optativas eran Cuidados Intensivos, Radiodiagnóstico y Radioterapia, Técnicas en Ultrasonido, Técnicas en Laboratorio Clínico, Técnicas en Audiología, Técnicas en Gineco-Obstetricia y Técnicas en Perinatología.

Es interesante notar que un número importante de estas asignaturas eran ofrecidas por médicos bajo un esquema similar al de una escuela de medicina. Esto obedece al hecho de que en la formación del plan de estudios intervinieron varios médicos de formación clínica (B. Muñoz, S. Carrasco, R. González). No pasó mucho tiempo para que asignaturas como Ciencias Fisiológicas II y III (Fisiología y Farmacología), Técnicas en Gineco-Obstetricia y Técnicas en Perinatología fueran criticadas en su época por tener demasiado contenido clínico y poco de ingeniería. Los cursos culminaban en dos seminarios de proyecto que sustituían al trabajo de tesis en licenciaturas clásicas.

Algunos aspectos relevantes tenían que ver con el hecho de que las asignaturas de Ingeniería Aplicada a la Fisiología (IAF) fueran asignaturas que integraran material de fisiología e ingeniería al mismo tiempo. Para ello éstas debieron ser impartidas

por dos profesores, un médico y un ingeniero, y buscaban conjuntar el conocimiento de los fenómenos fisiológicos y los de mediciones biomédicas. Estas asignaturas fueron innovadoras en su momento y marcaron la diferencia entre esta licenciatura y otras, como la ofrecida por la Universidad Iberoamericana y posteriormente por la UPIBI del IPN. En la práctica, la impartición de una asignatura entre un médico y un ingeniero se dio con dificultad. Por ejemplo, entre 1979 y 1980 la materia de IAF-I tuvo como contenido al sistema cardiovascular y sus mediciones, y fue impartida efectivamente por un profesor de cada disciplina (E. Hernández, que se integraba a su regreso de estudios de posgrado en Karlsruhe, Alemania y S. Carrasco). La asignatura IAF-II fue impartida exclusivamente por F. Prieto, quien se incorporaba después de obtener su doctorado en IB de Sussex, Inglaterra y quien era médico neumólogo de formación; el tema que se trató fue de equipamiento para terapia respiratoria. En tanto, IAF-III fue impartida por P. Huízar, doctor en fisiología, sin apoyo de personal de ingeniería.

La otra área de concentración, Matemáticas Aplicadas a la Medicina (MAM), tuvo poco éxito porque en esa época (antes de la existencia de las computadoras personales) había en las áreas de ingeniería biomédica pocos recursos humanos que las pudieran soportar, desapareciendo pocos años después. Dados los temas que en ella se trataban, como modelado y simulación de los sistemas cardiovascular y respiratorio, así como del tratamiento de la información en el sistema nervioso, con aplicaciones de la teoría de control moderna y del procesamiento de señales, hoy en día resulta una asignatura de gran interés para los alumnos de la disciplina.

La figura 1 muestra la distribución de créditos en las distintas asignaturas para el área de concentración de Instrumentación Médica. En ella se observa un énfasis muy importante en matemáticas debido a la gran cantidad de cursos en los primeros dos años (básicamente cálculo y matemáticas aplicadas a la ingeniería) y en instrumentación médica.

Sin embargo, si consideramos que algunas asignaturas optativas no se impartían propiamente como asignaturas de instrumentación sino más bien como asignaturas de medicina, y que asignaturas de electrónica analógica y digital, junto con los circuitos, se impartían como asignaturas propias de Ingeniería Eléctrica se observaría que se contaba con

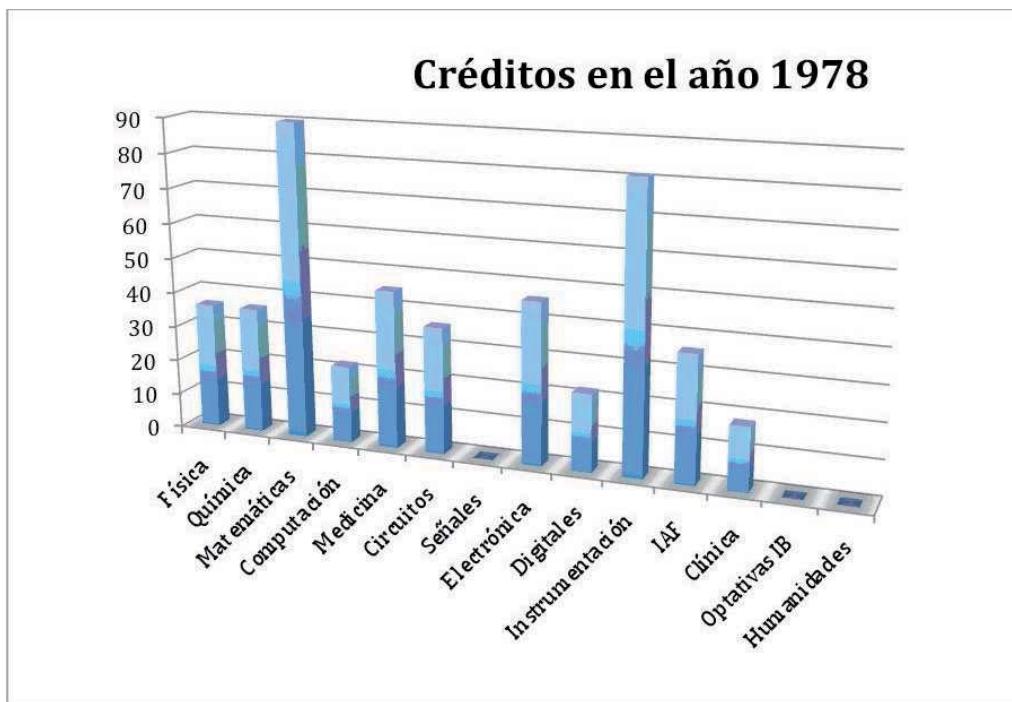


Figura 1. Distribución de créditos para el primer plan de estudios de IB de la UAM-I (1978). Se observa el énfasis en Matemáticas y en Instrumentación médica.

un fuerte componente en electrónica e instrumentación seguido por matemáticas, y medicina, mientras que asignaturas de señales y sistemas, optativas reales de ciencias biológicas o de otras disciplinas de ciencias básicas, ingeniería y humanidades no aparecen en la currícula.

En general se puede decir que este plan de estudios representaba bien al estado del arte de ese momento, cuando se pensaba que la política industrial debía orientarse a la sustitución de importaciones. Entre los logros de esa época están el desarrollo de algunos instrumentos para el autoequipamiento de laboratorios, el establecimiento de empresas por parte de profesores de la UAM, que construyeron equipos como electrocardiógrafos, termómetros digitales y medidores de pH entre otros. Desgraciadamente el mercado interno para estos productos no fue lo suficientemente grande como para permitir que estas empresas subsistieran a lo largo del tiempo.

Fue en este periodo cuando el trabajo en instrumentación para laboratorio clínico de algunos profesores se presentó por primera vez en conferencias internacionales de IB (M. Cadena, L. Hernández, M. Lindig) [7]. Al mismo tiempo, varios de los integrantes

del Área de IB comenzaron a dedicarse a la computación y el procesamiento de señales de manera incipiente. Prueba de ello es la aplicación que les dieron a sistemas de cálculo de filtros a partir de computadoras tipo PDP-8 y al uso de las primeras computadoras basadas en microprocesadores como las IMSAI y las Sol-20 anteriores a los años 80s.

Segunda etapa: 1982

La primera modificación importante al plan de estudios de la licenciatura en IB se dio en 1982, y obedeció a cambios importantes en la composición de los profesores. Por un lado el grupo de médicos que formaban parte del área de IB abandonó la División para integrarse a la División de Ciencias Biológicas y de la Salud, y aunque dos de ellos siguieron colaborando con la impartición de asignaturas, su influencia en la estructura del plan de estudios fue disminuyendo paulatinamente. Al mismo tiempo se integraron tres profesores con doctorados en biomecánica (C. López, J. L. Moreno, E. Buzo) y, como era de esperarse, influyeron en la conformación del nuevo plan de estudios. En el nuevo plan se abrieron tres áreas de concentración: Instrumentación Médica Electrónica, Instrumentación Médica Mecánica e Ingeniería Clínica.

El primer cambio importante se dio en el área de concentración en instrumentación médica electrónica, la que ahora tenía siete asignaturas de instrumentación médica electrónica, tres de instrumentación médica mecánica (sustituyendo las ingenierías aplicadas a la fisiología), dos de ingeniería clínica y cuatro de medicina.

El área de concentración en instrumentación médica mecánica era similar en estructura, donde se intercambiaban el número de asignaturas de instrumentación médica mecánica (siete) con las de instrumentación médica electrónica (tres), mientras que el área de ingeniería clínica eliminaba entre cinco y seis asignaturas, para sustituirlas con dos estancias hospitalarias de un trimestre de duración cada una. Es importante mencionar que con esta modificación al plan de estudios se inicia la filosofía que hay detrás de la conformación de éste y sus versiones posteriores (con excepción de la última) y que consiste en que las asignaturas y las áreas de concentración sean ofrecidas por expertos en esos campos que estén trabajando dentro de grupos de investigación, del Departamento de Ingeniería de aquel entonces y en la actualidad del Departamento de Ingeniería Eléctrica (el cual fue creado en 1982).

La incorporación de profesores con doctorado en biomecánica obedeció al crecimiento de las actividades del área de ingeniería biomédica. En ese entonces se pensaba en desarrollar líneas de investigación muy sólidas alrededor de este tema. Sin embargo, la gran crisis económica de los ochenta trabajo serias consecuencias, tanto para el área de IB como para el buen desarrollo de las actividades de formación a nivel de la licenciatura. Muchos profesores abandonaron la Universidad debido a los bajos salarios que percibían (entre 200 y 300 dólares al mes) y a la poca o nula inversión en infraestructura para poder desarrollar sus proyectos.

En muy pocos años los expertos en biomecánica fueron dejando la institución, y con ello una profunda crisis en la operación del plan de estudios, que se tradujo en la desaparición gradual del área de concentración en instrumentación médica mecánica, y en la sustitución de los contenidos de las asignaturas de mecánica y biomecánica por otros que no tenían nada que ver con esta área de conocimiento, debido principalmente a la falta de disponibilidad de profesores expertos en estos temas. Con el tiempo esta área de concentración terminó en la práctica siendo más parecida a la de instrumentación médi-

ca, y desapareció en el siguiente cambio de plan de estudios.

El caso del área de ingeniería clínica siempre fue controvertido ya que en algunos casos las estancias hospitalarias fueron bien estructuradas y muy enriquecedoras, pero en otros fueron estancias donde los alumnos sólo llevaban a cabo labores de técnico dentro de los servicios de Ingeniería Clínica (IC) dentro de los hospitales. Cabe mencionar que la experiencia en las instituciones de salud resultó ser trascendental para la incorporación de los egresados al ámbito laboral hospitalario; hubo un momento en el cual los ingenieros clínicos de la mayoría de los Institutos Nacionales así como de hospitales privados eran egresados de la UAM. En esta sentido sobresale el trabajo desarrollado en el Departamento de Ingeniería Biomédica del Instituto Nacional de Nutrición (ahora Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán) a cargo de T. Cadena, quien impartía las asignaturas de Ingeniería Clínica en aquel entonces.

Las figuras 2-4 muestran la distribución de créditos por temáticas para cada una de las Áreas de Concentración. En ellas se puede apreciar el efecto de una formación inicial común, dada por las asignaturas de ciencias básicas, para posteriormente ir variando el número de créditos para señales, electrónica, mecánica y clínica, de acuerdo al área de concentración elegida.

A pesar de este panorama se tuvieron avances importantes. Algunos de los logros en estas épocas fueron el inicio del programa de maestría en IB (1983), donde se inscribió un número importante de profesores de reciente contratación, y la conformación de grupos de trabajo dentro de laboratorios de investigación, contrario a la investigación individual que se llevaba a cabo antes. En esta época, bajo la propuesta de M. Cadena como Jefe del Área de Ingeniería Biomédica, se crearon primero los laboratorios de Procesamiento de Señales y Audiología, y posteriormente Instrumentación y Fenómenos Fisiológicos. El trabajo de estos laboratorios se orientaron cada vez más al procesamiento de señales y a la aplicación de los microprocesadores a la IB, donde el nivel de los proyectos era excepcionalmente alto.

En los años 84, 86 y 87 se reinició la presentación de resultados a nivel internacional, pero fue en 1988 cuando varios profesores (Aljama, Azpiroz, Cadena, Carrasco, Charleston, Medina, Yañez) presentaron un número importante de trabajos en el World

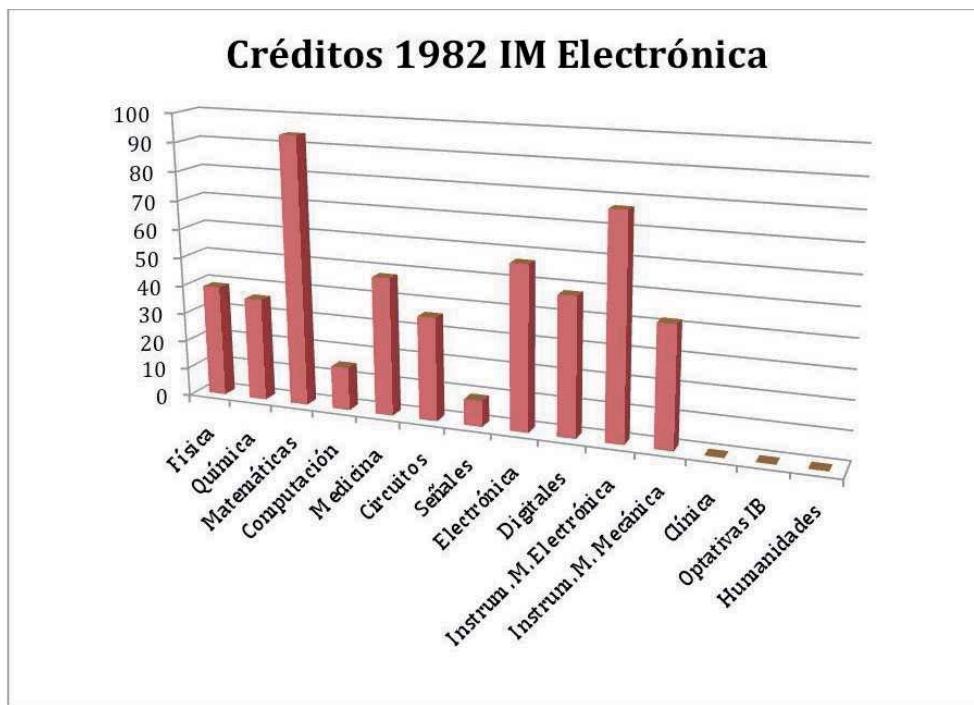


Figura 2. Plan de estudios para Instrumentación Médica Electrónica. Se observa el crecimiento de las asignaturas de Instrumentación, Mecánica, Sistemas digitales y Procesamiento de señales.

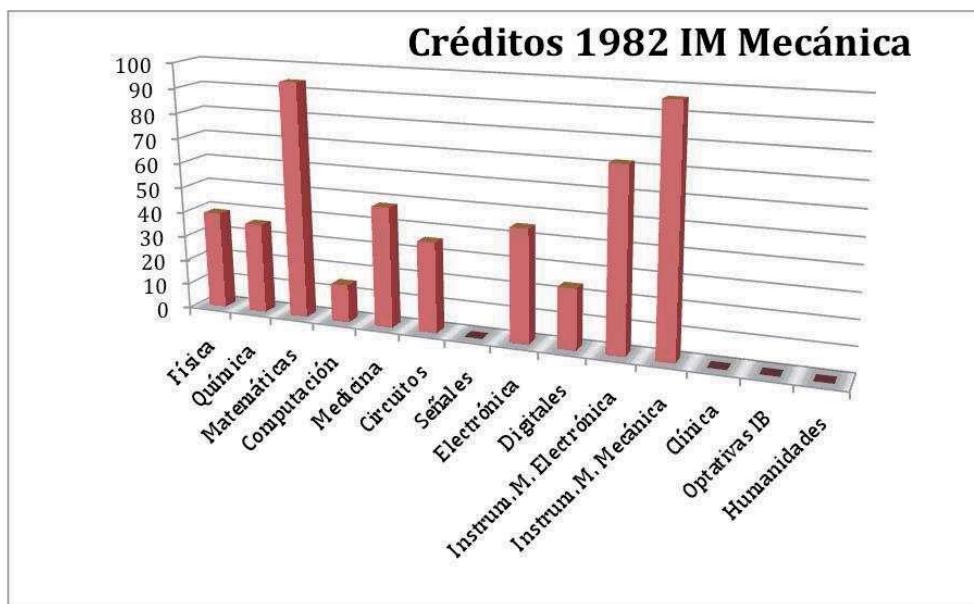


Figura 3. Plan de estudios 1982 para Instrumentación Médica Mecánica. Se observa el predominio de asignaturas de Mecánica y Biomecánica, la cancelación del curso de Procesamiento de señales y el descenso en el número de créditos para asignaturas de Electrónica.

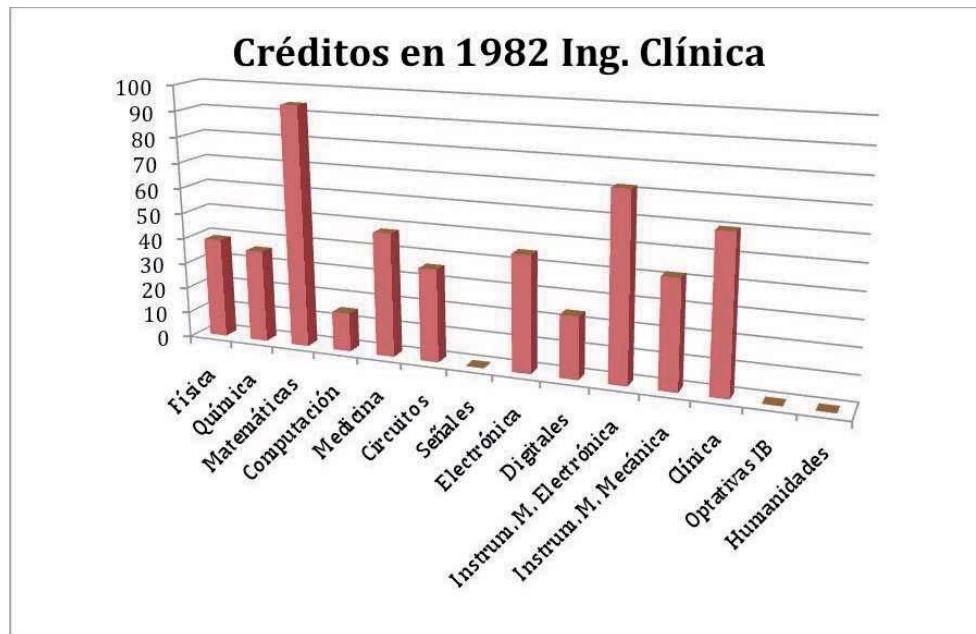


Figura 4. Plan de estudios 1982 para el área de concentración de Ingeniería Clínica. Se observa una carga comparativamente menor que se sustituye por dos trimestres de estancia en el hospital. En particular no se toman asignaturas de Aplicaciones de microprocesadores y Procesamiento digital de señales junto con una carga menor de las asignaturas de Electrónica.

Congress on Medical Physics and Biomedical Engineering en San Antonio, Texas[8-12]. Adicionalmente, hubo una notable asistencia de alumnos que aprovecharon la cercanía geográfica para asistir a este evento. Es en ese momento que se estableció definitivamente la costumbre de llevar a cabo proyectos de buen nivel, que se presentaran constantemente en eventos internacionales y promoviendo la participación de los alumnos de la licenciatura en ingeniería biomédica.

Además de estos resultados, es necesario resaltar el hecho de que a finales de esta década algunos profesores iniciaron sus estudios de doctorado en el extranjero, primero con recursos propios y con esfuerzos individuales y posteriormente ya con un programa de formación de profesores y apoyo consistente en becas por parte de la UAM. El regreso de estos profesores al inicio de la década siguiente terminó por consolidar a la investigación en IB. En ese momento el área había crecido tanto en número de integrantes como en líneas de investigación, que se decidió dividirla en dos: una de IB “clásica”, que continuaba con el trabajo en instrumentación y otra dedicada al procesamiento de señales e imágenes biomédicas.

Un hecho importante que permitió romper la “barreira de la publicación” fue la invitación hecha por el Prof. Max Valentinuzzi a colaborar en un número especial de la revista “*Medical Progress Through Technology*”, del cual fue editor invitado para conformar un número dedicado a América Latina. En ese número (Vol. 20, No. 1-2, 1994) se publicaron por primera vez en revistas internacionales indizadas trabajos de profesores de la UAM. De 14 trabajos en el volumen, cinco son de profesores de la UAM (Aljama, Azpiroz, Cadena, Cornejo, Granados, Jiménez, Medina, Muñoz)[13-17].

Si consideramos a esta como la etapa de fortalecimiento, hay que decir que aunque se tuvieron avances constantes en el desarrollo de las áreas, se necesitaron 20 años para llegar a tener un conjunto de profesores y laboratorios apropiados que pudieran desarrollar la docencia y la investigación al nivel que se espera en una institución como la UAM donde se enfatizan estas actividades.

Tercera etapa: 1996-2005

En estos años se hicieron varios cambios que llevaron a la desaparición del área de concentración de instrumentación médica mecánica. En cuanto al área de concentración de instrumentación médica

electrónica, ésta regresa al esquema del plan de estudios inicial, pero con algunas diferencias: se incrementó el número de asignaturas de computación a tres, las de ingeniería clínica a tres, cinco asignaturas de electrónica, tres de sistemas digitales, dos de procesamiento de señales, siete de instrumentación y por primera vez se agregaron dos asignaturas de humanidades.

En este momento se tenía una cantidad enorme de cursos muy relacionados con la ingeniería eléctrica, donde los alumnos llevaban más cursos de electrónica que aquellos de la licenciatura de ese nombre. Parece ser que el número tan grande de asignaturas está ligado a las características de los profesores, quienes al formarse mayoritariamente en las áreas afines al procesamiento digital de señales solían preferir a alumnos con mayores habilidades en computación y aplicaciones de microprocesadores, pero sin balancear la carga de las asignaturas clásicas de electrónica e instrumentación. Otros profesores interesados en impulsar al área de Ingeniería Clínica promovieron la impartición de tres cursos obligatorios. Las asignaturas de medicina se siguieron impartiendo de manera convencional con poca interrelación con temas de ingeniería; los temas que se trataban eran Bioquímica, Sistema Nervioso, Homeostasis y Fisiopatología.

La figura 5 muestra esta distribución de créditos donde se observa un crecimiento en las asignaturas de ingeniería clínica (para el área de concentración de instrumentación) y un nuevo grupo de asignaturas de humanidades. También se observó que el área de concentración de ingeniería clínica comenzó a ser preferida por alumnos que no deseaban llevar tantas asignaturas de electrónica y de procesamiento de señales; a cambio de esto pasaban una cantidad de horas equivalente en el hospital atendiendo las diversas actividades de la asignatura.

Desde el punto de vista del desarrollo de las áreas se puede decir que esta etapa (1996-2005) marca el fin del proceso de formación de profesores. Aunque algunos más se han graduado desde el año 2005, la mayoría de los profesores alcanzó sus grados máximos de estudios en esta época. Estos resultados se harán evidentes al finalizar esta etapa con un número importante de miembros de las áreas en el Sistema Nacional de Investigadores y en el reconocimiento de los grupos de trabajo como cuerpos académicos que se dará entre 2003 y 2010.

Cuarta etapa: 2006-2011

En esta etapa se redujo el número de créditos en asignaturas de matemáticas de 93 a 69, disminuyó el número de asignaturas de medicina a tres y de circuitos a una, al tiempo que se incrementó el número de asignaturas de procesamiento de señales de una a tres. Con respecto a electrónica el número de asignaturas pasó de cinco a dos mientras que en sistemas digitales éstas pasaron de tres a dos.

La parte más relevante de esta modificación es el hecho de haber eliminado las áreas de concentración y proponer 72 créditos de asignaturas optativas de IB o de la división de CBI (6 asignaturas de 12 créditos por lo menos). Esta situación promovió la flexibilización del plan de estudios dando la posibilidad de proponer un número importante de asignaturas optativas en temáticas como imagenología médica, procesamiento de señales, gestión tecnológica, instrumentación de laboratorio clínico, instrumentación de uso quirúrgico y terapéutico así como una asignatura denominada Temas Selectos de IB donde se presentan temas actuales de interés para el IB como biomateriales y nanotecnología. Adicionalmente esta versión del plan de estudios plantea un número de créditos a cursar en asignaturas de ciencias sociales y humanidades complementando la formación que fomenta el desarrollo de habilidades para discutir y relacionarse con profesionales de diversas disciplinas.

Esta modificación también trajo consigo una ruptura importante con respecto a la concepción y operación del plan de estudios. Por primera vez se pensaron contenidos de electrónica específicos para la formación de profesionistas de la ingeniería biomédica y se abrió la oportunidad para que en la formación de los alumnos interviniieran profesores de otros grupos de investigación diferentes a los que existen en el Departamento de Ingeniería Eléctrica, con lo que se amplían y diversifican las posibilidades de formación de los alumnos, más allá de los campos tradicionales cultivados por los miembros de las área de IB. Con respecto a la operación del plan de estudios, las modificaciones establecidas requirieron generar una serie de acciones para controlar los mecanismos de autorización para asignaturas optativas y para la parte final de los estudios (figura 6).

Discusión

Es claro que el plan de estudios de IB ha evolucionado a lo largo de los años y es incuestionable que la conformación de la planta de profesores ha

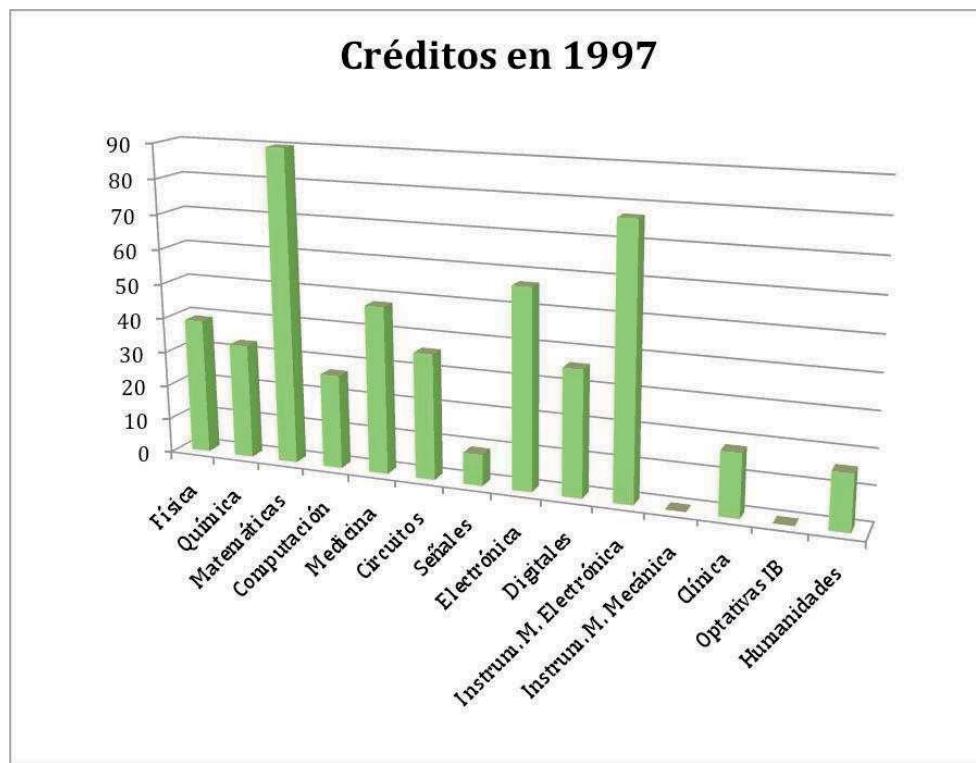


Figura 5. Distribución de créditos del plan de estudios de 1996-1998 donde se incluyen asignaturas de Ingeniería Clínica y Humanidades.

sido un factor fundamental para explicar los cambios de cada etapa, los cuales se pueden ver resumidamente en la figura 7. A fin de obtener algunas conclusiones, es conveniente hacer un balance de las modificaciones para apreciar su trascendencia e impacto.

El primer plan de estudios emergió con un espíritu más bien idealista ya que se conformó a partir de la revisión que hicieron profesores fundadores de la UAM, la cual se basó en la percepción de aquellos que tenían experiencia en mantenimiento de equipo médico, de docentes de otras instituciones como el CINVESTAV quienes en ese entonces recién estrenaban un posgrado en bioelectrónica, y en programas de posgrado en IB de universidades del extranjero como Case Western y Wisconsin. La mayor aportación a los estudios de IB fue la inclusión de asignaturas de ingeniería aplicada a la fisiología y asignaturas de técnicas en IB. Las primeras asignaturas (IAFs) eran un verdadero ejemplo de lo que se predicaba en la definición de la licenciatura: una colaboración entre médicos e ingenieros para la solución de problemas específicos en dis-

tintos órganos y sistemas del cuerpo humano, e incluían simulaciones y un poco de fisiología cuantitativa. Las asignaturas de "Técnicas" terminaron por ser el núcleo de lo que sería después el área de concentración de instrumentación médica, que sería una parte esencial de los programas en las siguientes modificaciones.

El segundo plan de estudios fue el resultado de la fortuita incorporación de profesores expertos en biomecánica que regresaban al término de sus estudios de posgrado en el extranjero. En términos prácticos, era una magnífica oportunidad para iniciar una nueva línea de investigación con profesores del mejor nivel. Sin embargo, su permanencia en la UAM no llegó a durar más de cinco años y por diversas razones abandonaron la institución. Un problema fue sin duda la falta de recursos, puesto que al pasar los años las ideas de fundar un laboratorio bien equipado de biomecánica se fue complicando por problemas financieros. El cambio trajo un abandono de la idea de las asignaturas de ingeniería y matemáticas aplicadas a la fisiología y la creación de un grupo nuevo de asignaturas de biomecánica. Se estable-

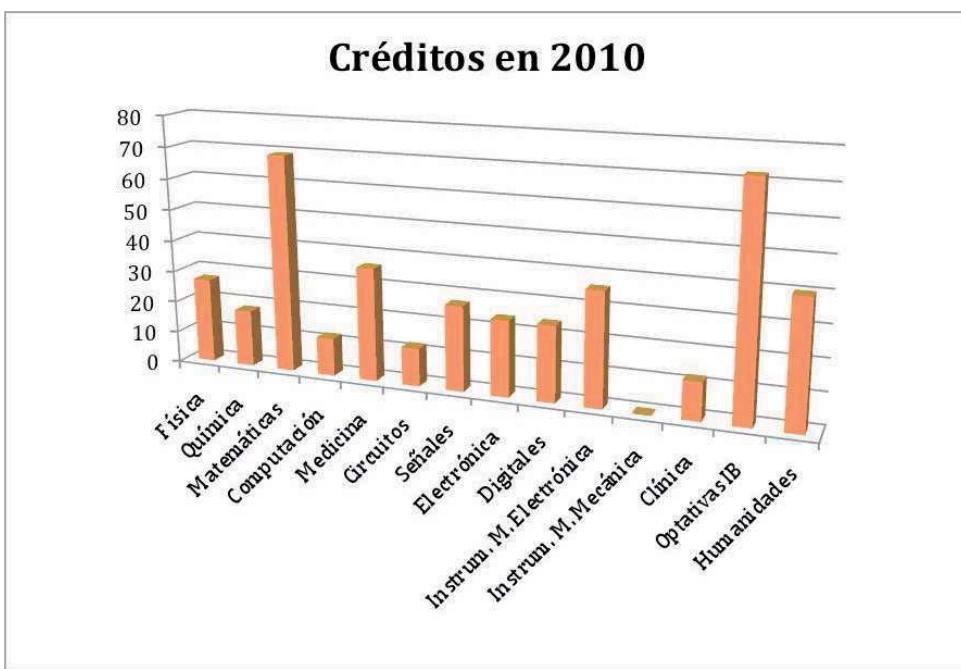


Figura 6. Se muestra la distribución de créditos para las temáticas consideradas para cada etapa del plan de estudios a fin de comparar los cambios descritos.

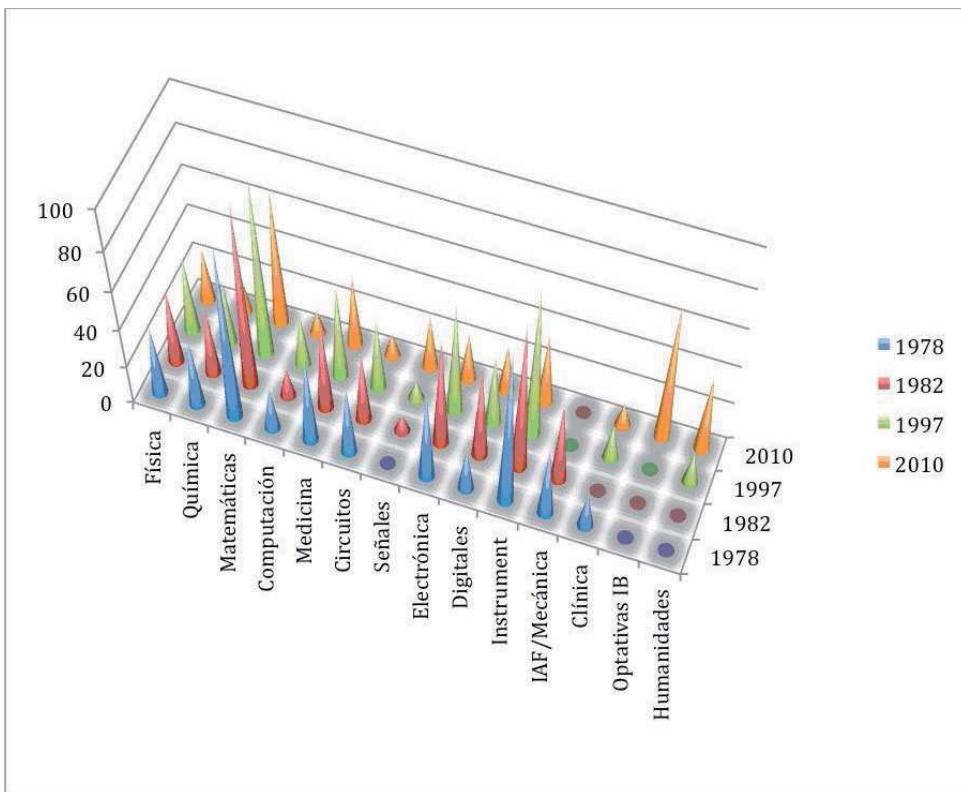


Figura 7. Distribución de créditos del plan de estudios en las cuatro etapas analizadas.

cieron tres áreas de concentración, de las cuales sólo persistieron la Instrumentación y la Ingeniería Clínica. Las asignaturas de biomecánica se tuvieron que dejar de impartir a los pocos años de aprobado este plan y el área de biomecánica dejó de existir antes de que terminara la década de los años ochenta. Como balance se puede afirmar que fue un experimento ambicioso, pero que su fracaso trajo no sólo la cancelación de esta área de concentración, sino que también se perdió la colaboración intensa entre médicos e ingenieros que se había propiciado en el plan de estudios anterior. Los médicos que estaban asociados al Área de IB pasaron a formar parte de la División de Ciencias Básicas y de la Salud.

El tercer plan de estudios fue producto de las ideas que aportó un grupo de profesores que regresó de sus estudios de posgrado al inicio de los años noventa. Se puede apreciar que hubo un incremento notable en dos tipos de cursos: procesamiento de señales y computación y sistemas digitales. Esta situación refleja directamente la influencia de las líneas de investigación que estos profesores desarrollaban, quienes se dedicaron fundamentalmente a proyectos de investigación en reconocimiento de patrones, procesamiento de señales e imágenes y aplicaciones de microprocesadores, que a su vez era un buen reflejo del estado del arte de la investigación en IB a nivel mundial en ese entonces. Como conclusión se observa que se siguió conservando la estructura del plan de estudios centrado en la instrumentación, de tal manera que hubo una expansión en la cantidad y complejidad de cursos que se impartieron en esta época. En lugar de reevaluar las líneas de interés para una licenciatura se amplió la diferencia entre lo esperado para un trabajo dentro de un hospital, en un área de concentración de Ingeniería Clínica, y en el área de Instrumentación Médica que tenía como objetivo graduar a expertos en instrumentación médica al mejor nivel posible.

Para el final de los años noventa se contaba con varios factores que generaban tensión en las áreas de investigación y en la docencia: un énfasis en añadir las experiencias novedosas de los profesores con formación de posgrado sin reducir los cursos de instrumentación tradicional, un estancamiento de los cursos de medicina que cada vez eran más convencionales y un área de concentración en Ingeniería Clínica que efectivamente evitaba las asignaturas más avanzadas del área de Instrumentación y las sustituía por trabajo en el hospital durante práctica-

mente el último año de la carrera. El plan de estudios propuesto a partir del año 2006 resolvió en parte esta problemática al cambiar radicalmente la estructura de los cursos durante el último año de la carrera.

El plan de estudios de la segunda mitad de la década del dos mil responde a la necesidad de ofrecer nuevos cursos que tratan temáticas novedosas dentro del campo de IB. Resuelve parcialmente la cuestión de acercar los temas tratados en las áreas de investigación con la docencia a nivel licenciatura e incorpora nuevos temas donde existe colaboración entre profesores e investigaciones de áreas tan diversas como economía de la salud y nanotecnología. Adicionalmente incorporó temas más específicos derivados de las investigaciones de los profesores como son las asignaturas de resonancia magnética, al tiempo que se le ofreció al alumno un conjunto de asignaturas optativas a cursar en su último año de estudios. Como resultado de los cambios se elimina la tendencia de los profesores del núcleo de docencia en IB a defender la idea de que las asignaturas que normalmente imparten son las más importantes de la licenciatura y que por ende deberían impartirse obligatoriamente. De esta manera desaparecen efectivamente las áreas de concentración pero se mantiene la posibilidad de que un alumno tome una serie de asignaturas relacionadas con su campo de interés o la línea de investigación dentro de la que se encuentra su trabajo de proyecto terminal.

Esta estructura también permite la incorporación o eliminación rápida de asignaturas cuando la disponibilidad de profesores cambie o en el caso de que se determinen líneas prioritarias de cursos especializados. En este sentido la flexibilidad del plan de estudios facilita su adaptación a los efectos de las tendencias futuras en el campo de la IB. Por otro lado la estructura tiene algunas deficiencias ya que mantiene una estructura muy arraigada en la ingeniería eléctrica, mientras que las nuevas tendencias de la IB se acercan mucho más a las líneas de micro y nanotecnología, bioinformática, genómica y otros temas biológicos.

Será importante en el futuro abordar seriamente tres de los temas que ya se han mencionado en este trabajo: La impartición de cursos orientados a la solución de problemas, en particular en los cursos básicos de matemáticas que deberán ser específicos para ingenieros; la neutralización de la influencia ex-

trema de material proveniente de las formaciones de profesores de ingeniería eléctrica y una integración eficiente de nuevas temáticas y metodologías que serán necesarias para una eficaz inserción de los estudiantes en el campo laboral dentro de los próximos 15 años. Los resultados derivados de esto permitirán que el plan de estudios ayude a que la formación de los egresados de la UAM siga distinguiéndose positivamente de aquellos de las otras instituciones que ahora pretenden formar ingenieros biomédicos.

Referencias

1. A. Velázquez, V. Gallegos, H. Suriano, J. Azpiroz, "Biomedical engineering education in México: 20 years of experience", Proc. World Congress on Medical Physics and Biomedical Engineering, 1984, pp 780.
2. E.G. Urbina, J. Azpiroz-Leehan, "Rethinking the biomedical engineering curriculum after an analysis of the previous 20 years", Proc. 25 Ann. EMBC, pp 3486-3409, 2003.
3. M. Cadena M, J. Azpiroz-Leehan, "Overview of the biomedical engineering history in Mexico: a personal point of view", Proc. 25 Ann. EMBC, pp 3450-3453, 2003.
4. J. Azpiroz-Leehan, G. Urbina, M.Cadena M., "New continuing education programs for hospital and clinical engineering", Proc. 26 Ann. EMBC, San Francisco, CA, Sept 2004, pp. 5205-5207.
5. Edmundo G. Urbina Medal, Joaquín Azpiroz Leeahan, "Análisis comparativo de planes de estudio de Ingeniería Biomédica: el caso de la Universidad Autónoma Metropolitana", *Revista Mexicana de Ingeniería Biomédica* vol. XXVIII, no. 2, pp121-129, 2007.
6. Fabiola Martínez Licona, Edmundo Gerardo Urbina, Joaquín Azpiroz-Leehan, "Updating the Biomedical Engineering Curriculum: Inclusion of Health Technology Assessment Subjects", Proc. 32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina, August 31–September 4, 2010, pp. 2967-2970.
7. Cadena M., Hernández L., "Design of an improved particle counter for use in hematology", *Proceedings of the San Diego Biomedical Symposium*. Vol. 17, pag. 143-148, San Diego, California. EUA, 1978.
8. V. Medina, J. Azpiroz, "Biomedical signal analysis and acquisition system", *Proceedings of the World congress on Medical Physics and Biomedical Engineering* Vol. 33, pp.277, 1988.
9. S. Charleston, T. Aljama, M. Cadena, "Personal computer based cardiac output impedance cardiograph", *Proceedings of the World congress on Medical Physics and Biomedical Engineering*, Vol. 33, pp.296, 1988.
10. O. Yañez V. Medina, S. Carrasco, "Determination of cardiac output by Fick's indirect method using a modular microcomputer system", *Proceedings of the World congress on Medical Physics and Biomedical Engineering*, Vol. 33, pp.297, 1988.
11. V. Carbajal, G. Urbina, S. Carrasco, "Mathematical modeling of measurement errors in gas analyzers during exercise", *Proceedings of the World congress on Medical Physics and Biomedical Engineering*, Vol. 33, pp.304, 1988.
12. M. Cadena, S. charleston, T. Aljama, S. Carrasco, "Data reduction and modeling of impedance cardiograms by linear prediction", *Proceedings of the World congress on Medical Physics and Biomedical Engineering*, Vol. 33, pp.384, 1988.
13. V. Medina, Y. Grall, C. Menguy, F. Langevin, Paul Gaillard, "Topographical characterization of normal visual evoked responses", *Medical Progress Through Technology*, Vol. 20, no. 1-2, Special Issue on Bioengineering in Latin America, pp. 15-22.
14. F. Hassainia, V. Medina, A. Donadey, F. Langevin, "Scalp potential and current density mapping with an enhanced spherical spline interpolation", *Medical Progress Through Technology*, Vol. 20, no. 1-2, Special Issue on Bioengineering in Latin America, pp. 23-30.
15. C. Muñoz-Gamboa, J. Jiménez-Cruz, "Human vestibular evoked responses", *Medical Progress Through Technology*, Vol. 20, no. 1-2, Special Issue on Bioengineering in Latin America, pp 31-36.
16. P. Granados, T. Aljama, M. Cadena, J. M. Cornejo, "Improved white noise method in the evaluation of linear hearing-aid characteristics", *Medical Progress Through Technology*, Vol. 20, no. 1-2, Special Issue on Bioengineering in Latin America, pp 37-42.
17. J. Azpiroz-Leehan, J-F- Lerallut, I. Magaña, "A multiprocessor architecture for medical image compression in a PACS environment", *Medical Progress Through Technology*, Vol. 20, no. 1-2, Special Issue on Bioengineering in Latin America, pp 101-110.



Centro Nacional de Investigación en Imagenología e Instrumentación Médica

Andrés Morón
Mendoza

Fernanda
Maldonado Sandoval

Gabriela Beltrán
López

Recibido: 28 de enero de 2012.

Aceptado: de 23 de marzo de 2012.

El Centro Nacional de Investigación en Imagenología e Instrumentación Médica (CI³M), se ubica dentro de las instalaciones de la Universidad Autónoma Metropolitana unidad Iztapalapa, (UAM-I), donde se desarrolla tecnología de punta en el área de la salud. El CI³M funciona gracias a la colaboración entre la UAM y el Consejo Nacional de Ciencia y Tecnología (Conacyt). Contando con equipos y áreas como la Resonancia Magnética de 3 Teslas, la Resonancia Magnética de 7 Teslas, el quirófano experimental y el laboratorio de instrumentación, el CI³M es un lugar único en Latinoamérica. Es importante que como miembros de la comunidad UAM, se conozcan las instalaciones y servicios de excelencia con los que se cuentan, ya que entre todos hacemos de esta Universidad una de las mejores en todo el país.

Historia

El CI³M es una iniciativa de la División de Ciencias Básicas e Ingeniería de la UAM-Iztapalapa que dio inicio en el año 2004 como parte de la convocatoria emitida por el Consejo Nacional de Ciencia y Tecnología: “Apoyos Complementarios para el Establecimiento de Laboratorios Nacionales de Infraestructura Científica o Desarrollo Tecnológico”; la UAM fue autorizada para recibir recursos e invertirlos en infraestructura del CI³M y fue nombrado Laboratorio Nacional por el Conacyt en el 2007.

Con la adquisición de dos equipos de resonancia magnética, uno de 7 Tesla para pequeños animales (VARIAN) y uno de 3 Tesla para humanos (PHILIPS), la UAM potencia sus actividades de investigación y formación de profesionales altamente capacitados, al contar con el laboratorio más importante en el campo de la Imagenología de América Latina y uno de los primeros doce del mundo.

El CI³M tiene la misión de albergar grandes instrumentos e infraestructura de primer nivel para fortalecer la investigación científica y tecnológica en el país, reforzar y complementar la docencia a nivel licenciatura y posgrado, propiciar la educación continua y capacitar personal técnico y profesional, crear recursos y promover transferencia tecnológica, generar vínculos academia-empresa-sector salud en torno a proyectos tecnológicos de alto impacto económico y social.

La creación del CI³M se realizó con el compromiso de que sus instalaciones e infraestructura estén

abiertas a cualquier investigador dentro y fuera de la UAM, así como a instituciones y empresas interesadas en realizar investigaciones, talleres, capacitaciones o consultas especializadas.



Figura 1. Centro Nacional de Investigación en Imagenología e Instrumentación Médica.

¿Qué es la Resonancia Magnética?

La Resonancia Magnética es un fenómeno físico que sucede como resultado de la interacción entre partículas atómicas que conforman algo, en este caso al cuerpo humano, sometidas a un gran campo magnético. No todas las partículas se comportan de la misma manera: la grasa no se comporta igual que el músculo u otros tejidos, es por eso que se pueden obtener imágenes del interior del organismo.

¿Qué es un bioriego?

El bioriego es una unidad especializada en producción y mantenimiento de animales de laboratorio, para ser utilizados en proyectos de investigación por la comunidad científica y académica, en este caso, de la UAM-I, la cual cuenta con un bioriego; el cual, además de las funciones ya mencionadas, fomenta la protección a la salud a los animales de laboratorio que ahí se producen y favorece el buen uso de los animales de laboratorio en investigación.

Es importante mencionar que se cumplen todas las normas de seguridad y ética para la experimentación en animales dentro del CI³M. Los experimentos e investigaciones que han requerido realizar pruebas clínicas en animales, se han realizado bajo una estricta relación de los responsables de cada proyecto con los especialistas y profesionales en el área de la salud veterinaria, en este caso, médicos veterinarios. Cuando se realizan cirugías como parte de protocolos de investigación, los animales son anestesiados.

dos y vigilados en todo momento para tener la seguridad de que se está haciendo buen uso de ellos, es decir, que no sufran. Igualmente por norma, para poder comercializar un equipo médico, se deben pasar por varias etapas antes de hacer pruebas en humanos, por eso es que algunas veces es necesario hacerlas primero en animales.

Servicios

El CI³M también se está desarrollando como un centro de investigación con aportes al área clínica, para este propósito brinda distintos servicios:

Resonancia Clínica

Utilizando el resonador magnético de cuerpo entero de 3 Teslas, se realizan estudios clínicos con los cuales se pueden detectar anomalías en los tejidos, visualizar fluidos o hallar deficiencias en el funcionamiento del organismo.

Se cuenta con un equipo de trabajo completo y profesional como en cualquier laboratorio de prestigio, incluyendo técnico y médico radiólogo.

Para los estudios se brinda la opción de aplicación de anestesia y medio de contraste en caso de requerirlo.

Resonancia Experimental

El servicio de estudios de investigación cuenta con un resonador magnético de 7 Teslas para animales pequeños. Para realizar protocolos de investigación se cotizan las horas de utilización del equipo, así como también se cuenta con bioriego y servicio de anestesia.

Cirugías experimentales con animales

El quirófano se encuentra equipado con mesa para cirugía (para animales) y lámpara de quirófano, máquina de anestesia, equipo para electrocirugía y endoscopio microscópico confocal. El quirófano también cuenta con servicios de bioriego y anestesiólogo.

Laboratorio de instrumentación

El laboratorio de instrumentación cuenta con el equipo y la tecnología necesarios para desarrollar pequeños dispositivos y prototipos para investigación e innovación tecnológica, incluidos equipos de medición. El laboratorio cuenta con espacio suficiente para albergar varios proyectos a la vez.

Clínica piloto de hemodiálisis

Se planea comenzar el servicio de Clínica de hemodiálisis en 2012. La clínica contará con todo el equipo

necesario para realizar hemodiálisis con el propósito de mejorar la calidad de vida de pacientes que padecen insuficiencia renal.

Dicho servicio será de excelente calidad y a precios accesibles con el fin de atender a pacientes y contribuir en proyectos de investigación dentro de esta área.



Figura 2. Equipo de Resonancia Magnética 3T de cuerpo entero.



Figura 3. Equipo de Resonancia Magnética 7T para animales pequeños.

Asesorías

Tomando en cuenta la capacidad de la Ingeniería Biomédica, el centro brinda a su vez el servicio de asesorías en el ámbito de la tecnología médica como gestión, análisis de la calidad o servicios de apoyo en el diseño de la infraestructura para hospitales o clínicas.

Cursos y capacitaciones

Estos cursos son especializados en cada una de las diferentes áreas o instalaciones con las que cuenta el



Figura 4. Quirófano experimental para pequeñas especies.

CI³M como resonancia magnética, quirófano y hemodiálisis. También se presentan nuevos proyectos de capacitaciones o cursos continuamente, como innovación y emprendimiento en el campo médico y de la salud.

El centro cuenta también con servicios de apoyo a la comunidad universitaria, como la apertura de realizar el servicio social dentro de las instalaciones, contando con rotaciones en las distintas áreas del centro.

Contacto

Para cualquiera de los servicios se realiza cotización gratuita comunicándose a:

Teléfono: 85-02-45-69

Página web: www.CI3M.mx

cs



Gerardo A. Laguna Sánchez
Enrique Rodríguez de la Colina
Michael Pascoe Chalke
Víctor M. Ramos Ramos

Reflexiones sobre las
tecnologías de la
información y la
conciencia de lo
“ecológicamente
correcto”

Recibido: 15 de enero de 2012.
Aceptado: 15 de febrero de 2012.

Abstract

During the last decades, we have been witnesses and protagonists of the evolution of science and engineering. In particular, today more than ever, the design of new information technologies must include some new criteria intended to reduce energy consumption and minimize their adverse environmental impact. The awareness of the “ecologically correct” must be encouraged not only by universities but also it must be included as part of the development strategies in our current society. In this article, we present an overview of the IT’s evolution in the last three decades giving emphasis to how significant it is for these ITs to include a sustainable design.

Keywords: Information Technology, carbon footprint, sustainable development and “ecologically correctness”.

Resumen

Durante las últimas décadas, hemos sido testigos y protagonistas de la evolución que han experimentado la ciencia y la ingeniería. En particular, hoy más que nunca, el diseño de las nuevas tecnologías de la información debe contemplar nuevos criterios destinados a reducir el consumo de energía, de tal manera que su impacto ambiental sea mínimo. La conciencia de lo “ecológicamente correcto” debe inculcarse no tan solo en las universidades, sino que también debe ser contemplada como parte de las estrategias de desarrollo en la sociedad actual. En este artículo, haremos un recorrido de la evolución que han tenido las tecnologías de la información en las tres últimas décadas con énfasis en la importancia de que éstas tengan un diseño sustentable.

Palabras clave: Tecnologías de la información, huella de carbono, desarrollo sustentable y ecológicamente correcto.

Introducción

Se cumplen 30 años del Departamento de Ingeniería Eléctrica de la UAM Iztapalapa, el cual alberga al área de Redes y Telecomunicaciones, afiliación de los autores de este artículo. Este aniversario también coincide con los 30 años de la popularización de las computadoras personales, lo cual motivó que la revista Time designa a la computadora personal como la máquina del año 1982 (*The Computer Moves*

In, 1983). Estas coincidencias nos invitan a la reflexión porque, durante estos 30 años, hemos sido testigos y protagonistas de la evolución tecnológica en este cambiante ámbito, tanto en la industria como en la academia, ya sea estudiando, aplicando o diseñando dispositivos de redes y telecomunicaciones o preparando profesionales en el área. Sobre todo, hemos visto como han ido evolucionando los enfoques de diseño en el área de las ahora denominadas como tecnologías de la información (TI). Antes, como ahora, se hacía énfasis en la eficiencia de los diseños, pero actualmente además se requiere que los diseños minimicen el consumo de energía y adopten mecanismos adaptables para que su impacto ambiental sea mínimo.

Las universidades tienen un papel preponderante en el proceso formativo de los individuos que componen a una sociedad. En particular, la formación de profesionales responsables con su entorno es una de las facetas que últimamente han adquirido más preponderancia. Las escuelas y facultades de ingeniería no son la excepción y en ellas empiezan a permear conceptos como lo “ecológicamente correcto” y el “desarrollo sustentable”.

El enfoque de los sistemas, profundamente empleado en el ámbito de la ingeniería, ha sufrido en los últimos treinta años significativos ajustes y reorientaciones en sus prioridades de diseño. Hoy, más que nunca, la minimización de los efectos ambientales de todo sistema es uno de los criterios de diseño que van volviendo prioritarios e, incluso, mandatarios. Estas nuevas tendencias de diseño también han alcanzado a las tecnologías que tradicionalmente se habían considerado como “limpias”, como lo es el caso de las TI, que incluye a los dispositivos que actualmente permiten la comunicación y la transmisión de información, ya sea mediante datos, audio, video o todo ello junto, además del procesamiento de la información por el mismo dispositivo, con una tecnología que ocupa el espacio de un teléfono celular.

A pesar de que en los actuales sistemas electrónicos (hardware), debido a la miniaturización de los componentes, han logrado una reducción significativa en el consumo de energía, el consecuente abaratamiento de estos productos ha propiciado que estos alcancen mercados de consumo con demanda masiva y alcance global, cuyo consumo energético neto es muy superior, por mucho, a los consumos de energía por equipo. Así, métricas como la “huella de carbono” han comenzado a aplicarse a las TI, pa-

ra disponer de una estimación de los efectos contaminantes, directos e indirectos, debidos al consumo de energía de estos dispositivos y asumiendo que en algún lugar esta energía se produce debido cierto consumo de combustibles fósiles.

Los signos de nuestro tiempo en las TI

En enero de 1983 en la revista Time, por primera vez en su historia, una máquina remplazó al “personaje del año”. Así, la máquina del año 1982 fue la computadora personal, después de comprobar el impacto en la sociedad de las computadoras personales Apple II e IBM PC. A partir de ese momento, se iban a gestar las historias de tres polémicos personajes que han influido, para bien o para mal, en las características de las actuales tecnologías de la información y sus aplicaciones: Steve Jobs, William Henry “Bill” Gates y Mark Zuckerberg. Todos ellos tienen en común, haber cofundado sus respectivas empresas apenas cumplidos los 20 años y habiendo iniciado alguna carrera relacionada con la electrónica y la programación de sistemas.

Steve Jobs (1955-2011), que en la década de los 70 desarrolló con su equipo la máquina personal Apple en un garaje, inició lo que a la postre se convertiría en la multimillonaria empresa de computación Apple. En 1986 Steve Jobs compró la división de cómputo gráfico de la empresa Lucasfilm Ltd. por \$10 millones de dólares y la estableció como una empresa independiente denominada como “Pixar”. Posteriormente, durante los años 90, Steve Jobs desarrolló a la subsidiaria adquirida y revolucionó la industria de animación con el lanzamiento de la película Toy Story (*Steve's Two Jobs*, 1999). En 2006 Jobs vendió Pixar a Disney, de la que había sido proveedora, a cambio de acciones por un valor de \$7,400 millones de dólares, convirtiendo a Jobs en el mayor accionista individual de Disney. Steve Jobs, recientemente fallecido a los 56 años, dejó como legado un conjunto de productos tecnológicos para consumo que se caracterizaron siempre por ser de diseño atractivo y novedoso: los sistemas operativos y entornos de ventanas, los teléfonos multimedia (iPhone) y las computadoras personales portátiles de pantalla de contacto (iPad). Su espíritu perfeccionista le valió estos logros, pero también detractores en el ámbito laboral.

Bill Gates (1955-) cofundó la empresa Microsoft (1976) y se insertó en el negocio de los sistemas operativos para computadoras personales debido a los

requerimientos de IBM (1981) y luego fue impulsado por la creciente demanda de sistemas operativos para computadoras como la IBM-PC y compatibles. Microsoft inició con el sistema operativo MS-DOS y, hasta la fecha, sus productos se han vuelto un estándar de facto para computadoras personales de plataforma abierta en todo el mundo. Su espíritu emprendedor y su visión tecnológica lo convertirían eventualmente en el hombre con la mayor fortuna personal del mundo, aunque también ha sido cuestionado por sus prácticas comerciales consideradas como monopólicas y desleales.

Mark Zuckerberg (1984-) es cofundador de la empresa Facebook (2004), el sitio con la red social más grande del mundo, con más de 800 millones de usuarios (hasta enero de 2012). Esta red social, desde su creación, ha tenido un crecimiento exponencial y, en la actualidad, caracteriza los hábitos y preferencias de muchos de los usuarios de Internet. Mark, a su corta edad (27 años), ya es multimillonario y es el arquetipo de una generación de tecnólogos de Internet que se han vuelto ricos, de la noche a la mañana, debido a sus habilidades técnicas y a los efectos propagantes de la red.

Hemos traído a cuenta estos personajes y sus empresas debido a su evidente influencia en las actuales TI de uso popular, tanto en la configuración de su diseño como en la definición de su valor agregado. En nuestros días es común que las personas puedan ingresar a Internet en forma inalámbrica, a través de computadoras portátiles miniatura, o mediante la red de telefonía celular, con sus teléfonos celulares con soporte de datos y multimedia (iPhone y Blackberry, entre otros), empleando alguno de los productos o servicios relacionados, directa o indirectamente, con alguna de las empresas de estos personajes. Tal demanda de estos servicios y productos es lo que ha hecho comercialmente exitosas a las empresas que los prestan o producen.

Nuevos criterios de diseño

En la década de los 80, igual que en estos momentos, tanto en el aula como en la industria se hacía énfasis en el papel del ingeniero como un profesional que resuelve problemas y realiza diseños eficientes. La eficiencia de un sistema puede ser vista como la maximización en el aprovechamiento de los recursos, es decir, la minimización del desperdicio o las pérdidas durante el proceso de transformación o realización de un trabajo. En la medida que el sistema minimiza las pérdidas de poten-

cia en el proceso, digamos por disipación térmica, entonces se vuelve más eficiente. De hecho, el sistema ideal no tiene pérdidas. En las máquinas que procesan información, se aplica el mismo principio, pero lo que se transforma es la información introducida. Aunque la información es intangible, su procesamiento consume energía y también es deseable minimizarla.

Sin embargo, actualmente, completada la primera década del siglo XXI, el énfasis en el diseño de los sistemas eficientes energéticamente, que por su puesto incluye a las TI, va adquiriendo un peculiar enfoque ambientalista y de apego a lo “ecológicamente correcto”. Hoy se habla del “desarrollo sustentable” y se da énfasis a los efectos ambientales de todo diseño o desarrollo. Esta conciencia de los efectos de la actividad humana en el calentamiento global, debido al consumo energético, cobró fuerza justamente a partir de la década de los 80, cuando los científicos tuvieron suficiente evidencia de ello. Aunque aún está en debate científico la afirmación de que el calentamiento global sea exclusivamente debido a la actividad humana, y si bien es cierto que en el calentamiento global intervienen factores cíclicos ajenos a la actividad humana (por ejemplo, las eras gélidas seguidas de eras cálidas), actualmente el efecto antropológico en el calentamiento global, debido a la obtención de energía a partir de combustibles fósiles, es uno de los más importantes. Debido a esto, ultimamente se viene proponiendo un uso más racional de la energía y se incentiva a las propuestas y los diseños que minimizan el consumo energético, además de emplear la energía de manera eficiente.

Concretamente en el área de las TI, existe una conciencia de la búsqueda de sistemas sustentables (ver Figura 1), que no sólo representen ahorro de energía y sean ecológicos, sino también que puedan ayudar a la creciente demanda de tráfico y seguridad de la información existentes en las nuevas generaciones de sistemas de comunicaciones.

Los nuevos conceptos de las TI deben incorporar los avances tecnológicos que ayuden a la reducción del uso de energía, aumentar la capacidad de los enlaces, la calidad de servicio, la seguridad y velocidad de transmisión y, para ello, considerar la gestión y control. Además estos nuevos sistemas deben reunir características que ayuden y contribuyan en el aspecto social, al medio ambiente y en la economía.



Figura 1. Desarrollo sustentable

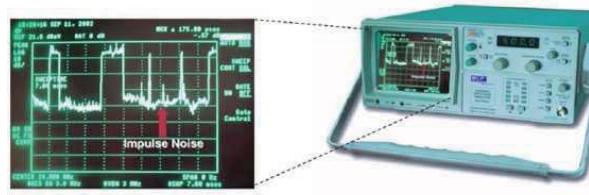


Figura 2. Monitoreo activo del espectro radio-eléctrico con un equipo analizador de espectros. En su pantalla se puede detectar la presencia de señales en cada frecuencia.

En cuanto al diseño de nuevos esquemas de comunicación, se habla de conceptos como “radio cognitivo” (Akyildiz, Won-Yeol, Vuran, y Mohanty, 2008), el radio que reconoce su entorno y se adapta, tanto para aprovechar canales de comunicación disponibles con un mínimo de interferencia como para minimizar el consumo energético. El ajuste de los parámetros de operación parte del monitoreo activo de varios factores externos e internos, tales como (Bolívar, Marzo, y Rodríguez-Colina, 2010; Rodríguez-Colina, Ramírez-Pérez, y Carrillo, 2011):

- el espectro de frecuencias de radio (ver Figura 2),
- la conducta del usuario,
- la posición,
- y el estado de la red de comunicación.

Más aún, incluso se habla del “radio autónomo” (Cardarilli, Re, Shuli, y Simone, 2011), el radio que es autosuficiente para proveerse de energía, configurarse, “defenderse” y adaptarse en un entorno dado. En general, en los nuevos diseños, se trata de usar en forma intensiva esquemas de modulación y de codificación de canal adaptables, a fin de

que un dispositivo de comunicación configure en forma automática los parámetros de operación que garanticen un mínimo en la calidad del servicio a la vez que minimice el consumo eléctrico. En cuanto a la conciencia de la sociedad sobre lo ecológicamente correcto, en los últimos 30 años también se han gestado cambios apreciables, aunque aún no son suficientes.

La huella de carbono de las TI

Hoy, a diferencia de hace 30 años, en la sociedad existe una influyente actividad de organizaciones, tanto civiles como gubernamentales, que tratan de concientizar a los usuarios de los beneficios, para el medio ambiente, de hacer un uso racional y justificado de los medios de comunicación electrónica. Así, surge el concepto de la “huella del carbono”, que es la totalidad de gases de efecto invernadero (GEI) emitidos por efecto directo o indirecto de un individuo, organización, evento o producto (Wiedmann y Minx, 2007). Entonces, resulta interesante hacer aquí una rápida revisión de los efectos sobre el medio ambiente de las tecnologías de la información más populares actualmente. Esperamos que ello contribuya a hacer un uso más racional de las mismas en nuestra sociedad.

Concediendo que las cifras publicadas en la red son aproximadamente correctas (Eden-project, 2011), con base en los datos proporcionados por el reporte de McAfee (McAfee y ICF, 2009) y el libro de Mike Berners-Lee (Berners-Lee, 2010), tenemos los siguientes efectos contaminantes de las, no tan limpias, tecnologías de la información (ver Figura 3). Para Internet tenemos:

- A un mensaje spam le corresponde una emisión de 0.3 g de CO_2 , que equivale a recorrer un metro en coche.
- A un mensaje personal de correo electrónico le corresponde una emisión de 4g de CO_2 , que equivale a recorrer 13 metros en coche.
- A un mensaje personal de correo electrónico, con un archivo grande adjunto, le corresponde una emisión de 50g de CO_2 , que equivale a recorrer 166 metros en coche.
- A todo Internet en un año le corresponde una emisión de 300 millones de toneladas de CO_2 , que



Figura 3. Huella de carbono de las TI más populares (tomado de www.edenproject.com)

equivale a recorrer 25 millones de veces la vuelta al mundo en coche, o el equivalente a todo el petróleo, carbón y gas quemado en Turquía y Polonia en un año.

Exclusivamente para la telefonía celular, tenemos:

- A un mensaje de texto le corresponde una emisión de 0.014g de CO_2 , que equivale a recorrer 5 centímetros en coche.
- A una llamada telefónica le corresponde una emisión de 190g de CO_2 , que equivale a recorrer 633 metros en coche.

Es de especial interés el caso de las redes sociales virtuales. Específicamente en Facebook, con más de 800 millones de usuarios en el mundo, de los que 350 millones ingresan mediante sus teléfonos móviles, suben un promedio de 250 millones de fotos al día (Facebook, 2011). Podemos estimar de manera aproximada y conservadora, suponiendo que las fotos se suben únicamente desde computadoras personales y que cada envío corresponde a un correo electrónico personal, que las fotos subidas diariamente a Facebook equivalen a una emisión de 1,000 toneladas de CO_2 , que equivale a recorrer 74,000 vueltas al mundo en coche.

Como referencia, tenemos la huella de carbono que resulta de enviar una carta por correo tradicional, que le corresponde una emisión de 200g de CO_2 y que equivale a recorrer 666 metros en coche. En apariencia, las TI son más limpias que los medios tradicionales, pero esto es cierto sólo hasta cierto punto, siempre que la intensidad en el uso de las TI no

supere cierta cantidad de información enviada. Por ejemplo, si un usuario envía 50 correos electrónicos, esto equivale justamente al efecto de enviar una sola carta por correo. El punto es que en el pasado, antes del abaratamiento y el uso masivo de las TI, la gente escribía cartas con moderación, pero en la actualidad es muy común que una persona en una semana envíe más de 50 correos electrónicos, y muchos más mensajes de texto, muchos de ellos con contenidos triviales.

Las emisiones de totales de CO_2 debidas a el uso mundial de Internet corresponden a cerca del 2% de las emisiones totales producidas por la quema de combustibles fósiles a nivel global y se espera que aumente a 6 % en el 2030. Todos estos porcentajes son significativos si consideramos que se trata de tecnologías que tradicionalmente se habían considerado como “limpias”.

Conclusión

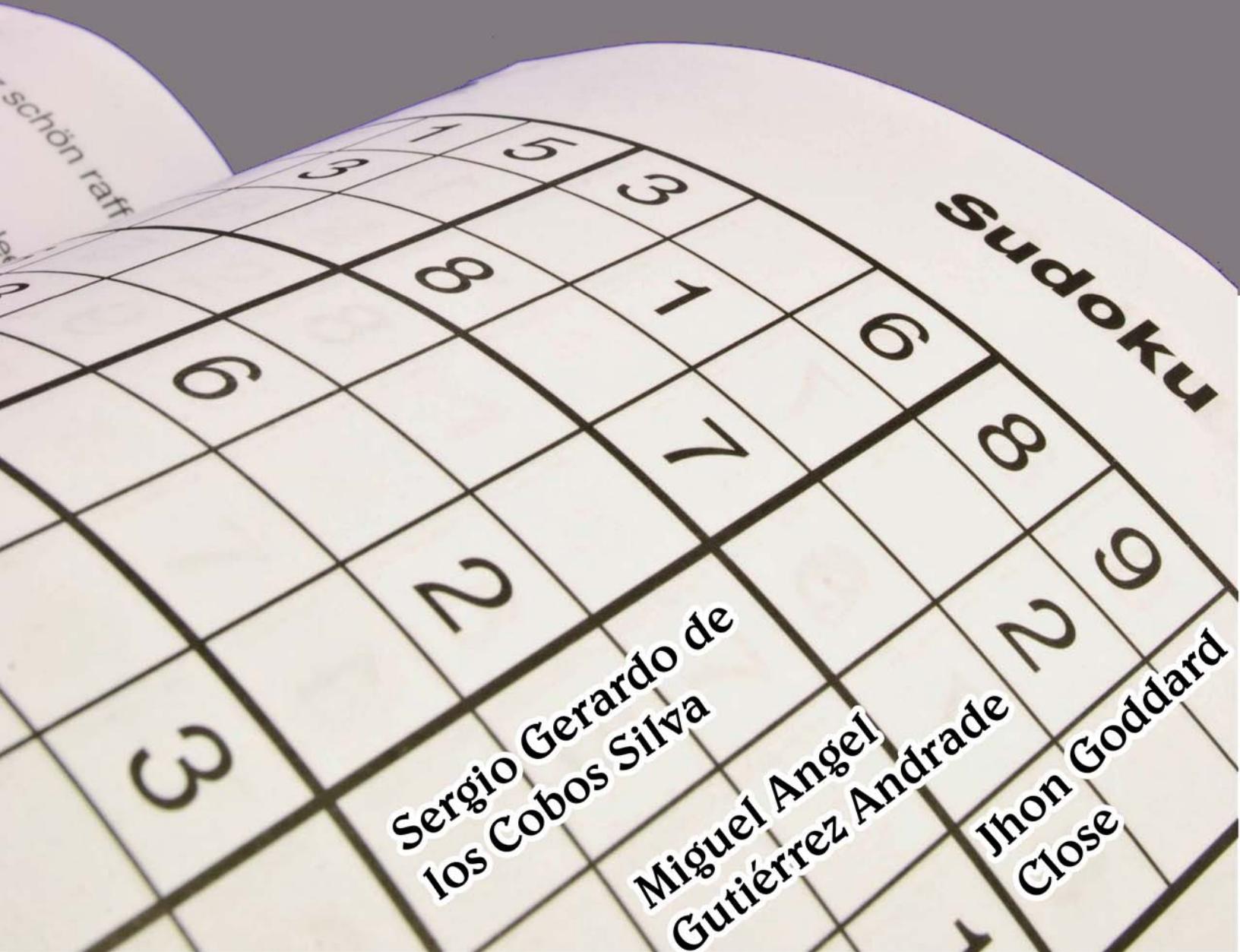
A 30 años del Departamento de Ingeniería Eléctrica, a 30 años de la designación de la computadora personal como el personaje del año y a 30 años de conciencia y reflexión sobre los efectos ambientales de nuestra actividad humana, tenemos tecnologías de comunicación e información relativamente limpias, si se comparan con las antiguas, pero también tenemos un abuso de ellas y, por el volumen implicado, resultan ser considerablemente más contaminantes cuando se toma en cuenta el efecto global.

Así, es importante que las universidades formen a ingenieros responsables, que consideren las repercusiones ambientales en sus nuevos diseños, y, aún más importante, que también contribuyan a concientizar a la población en general, como usuarios de estas tecnologías, para que hagan un uso racional de los mismos.

Referencias

1. Akyildiz, I., Won-Yeol, L., Vuran, M., y Mohanty, S. (2008). A survey on spectrum management in cognitive radio networks. *IEEE Communications Magazine*, 46 (4), 40-48.
2. Berners-Lee, M. (2010). *How bad are bananas?: The carbon footprint of everything*. Profile Books.
3. Bolivar, N., Marzo, J. L., y Rodriguez-Colina, E. (2010). Distributed control using cognitive pilot channels in a centralized cognitive radio network. En *Ieee computer society conference aict*.
4. Cardarilli, G., Re, M., Shuli, I., y Simone, L. (2011). Partial reconfiguration in the implementation of autonomous radio receivers for space. En *Workshop on reconfigurable communication-centric systems-on-chip (recosoc)* (pp. 1-6).
5. The computer moves in. (1983, 3 Jan). *Time Magazine*.
6. Eden-project. (2011). *What's your carbon footprint?* Eden project web page. (<http://www.edenproject.com/whats-it-all-about/climate-and-environment/show-me-the-carbon.php>)
7. Facebook. (2011). *Facebook: Estadísticas*. Facebook web page. (<http://www.facebook.com/press/info.php?statistics>, consultada en enero de 2012)
8. McAfee, y ICF. (2009). *Informe sobre la huella de carbono del spam por correo electrónico*. McAfee/ICF report. (<http://eu.vocuspr.com/ViewAttachment.aspx?EID=SVKCadtpAhOCUhahAzzOIhDjAmAaPIy2JF5OfEY8sAs%3d>)
9. Rodriguez-Colina, E., Ramirez-Perez, C., y Carrillo, C. E. (2011). Multiple attribute dynamic spectrum decision making for cognitive radio networks. En *Ieee international conference on wireless and optical communications networks (wocn)* (pp. 1-5).
10. Steve's two jobs. (1999, 18 Oct). *Time Magazine*.
11. Wiedmann, T., y Minx, J. (2007). *A definition of 'carbon footprint'*. ISA UK Research Report 07-01. (<http://www.censa.org.uk/docs/ISA-UKReport 07-01 carbon footprint.pdf>)

Solución al acertijo del SUDOKU con recocido simulado



Recibido: 30 de enero de 2012.

Aceptado: 22 de marzo de 2012.

Resumen

El Sudoku es un pasatiempo que en los últimos años se ha popularizado y para algunas personas se ha vuelto adictivo, quizá porque las reglas para resolverlo son muy simples. En este artículo se propone la solución a un Sudoku como un problema de optimización combinatoria y se resuelve utilizando el algoritmo de Recocido Simulado. Se presenta experiencia computacional en la solución de Sudokus que aparecen en periódicos de circulación nacional y algunas páginas de Internet con diferentes grados de dificultad que van desde fácil, media, difícil y muy difícil.

Palabras Clave: Sudoku, Recocido Simulado, metaheurísticas, heurísticas, optimización combinatoria.

Keywords: Sudoku, Simulated annealing, metaheuristics, heuristics, combinatorial optimization.

Introducción

El Sudoku es un pasatiempo cuyo objetivo es llenar una cuadrícula generalmente de 9×9 celdas (81 celdas en total) dividida en subcuadrículas de 3×3 con los números del 1 al 9, partiendo de algunas celdas previamente asignadas y siguiendo tres reglas muy sencillas. El pasatiempo aparece por primera vez alrededor de 1970 en Nueva York, Estados Unidos, en la revista de juegos y pasatiempos Dell con el nombre de “colocar el número”. Posteriormente, en 1984, aparece este pasatiempo en Japón con el nombre que actualmente se conoce de Sudoku, donde “Su” significa número y “Doku” significa individualmente, en japonés. A partir de ahí, se populariza hacia todo el mundo. El juego tiene sus orígenes en la idea de cuadrados latinos que introduce el matemático Euler en 1783. En México hay muchas revistas de pasatiempos y periódicos de circulación nacional tales como El Universal, Reforma, etc., que incluyen dentro de sus páginas Sudokus. Este artículo se desarrolla de la siguiente manera: en la sección 1 se define el problema del Sudoku y se dan algunos lineamientos para resolverlo. En la sección 2 se introduce brevemente la técnica de Recocido Simulado y su algoritmo asociado. La sección 3 se implementa el algoritmo de Recocido Simulado al Sudoku, se define una solución factible, se asigna un costo a cada solución factible y plantea el problema de resolver el Sudoku como un problema de optimización combinatoria. En la sección 4 se especifica el algoritmo de Recocido Simulado al Sudoku. La sec-

ción 5 se dan los resultados obtenidos por el algoritmo de Recocido Simulado al resolver Sudokus de grado de dificultad fácil, media, difícil y muy difícil. Finalmente se dan las conclusiones de este trabajo.

1. El Problema del Sudoku

El objetivo del Sudoku es llenar una cuadrícula de $n^2 \times n^2$ celdas, dividida en n^2 subcuadrículas de $n \times n$ en donde se deben colocar los enteros del 1 a n^2 de acuerdo a los siguientes tres reglas:

1. Las celdas de cada uno de los renglones deben contener los enteros de 1 a n^2 exactamente una vez.
2. Las celdas de cada una de las columnas deben contener los enteros de 1 a n^2 exactamente una vez.
3. Cada una de las n^2 subcuadrículas deben contener los enteros de 1 a n^2 exactamente una vez.

Estas reglas tan simples son quizás lo que ha hecho que el juego se popularice y se haya hecho adictivo para algunas personas. Al entero n se le llama el orden del Sudoku. En la Figura 1 se da un ejemplo de un Sudoku de orden 3, donde se puede observar que existen algunas celdas prellenadas con los enteros del 1 al 3^2 y el jugador debe completar la cuadrícula siguiendo las tres reglas descritas anteriormente. Cada Sudoku, bien planteado, tiene solución válida única y no requiere usar técnicas de ensayo y error o adivinar sino que se resuelve empleando las tres reglas y por medio del “razonamiento”.

El Sudoku de la Figura 1 tiene solución válida única y está catalogado entre los muy difíciles de resolver. Generalmente en los diarios de circulación nacional o internacional, así como, en revistas y otros medios impresos aparecen Sudokus de orden 3 y en algunos casos de orden 4 en donde aparece alguna clasificación que va desde fácil hasta muy difícil y tiene que ver con el número de celdas prellenadas y su colocación dentro de la cuadrícula. En el caso del Sudoku de la Figura 1, el número de celdas prellenadas es de 17, y aparentemente, es el número mínimo de celdas prellenadas que se pueden dar para Sudokus de orden 3, ya que con menos el Sudoku podría tener más de una solución válida.

Para ilustrar como se resuelve un Sudoku, considere la celda en donde se colocó el símbolo α , en la Figura 2. Las celdas donde aparecen los 1 (en negritas) hacen que las celdas marcadas con \times de la subcuadrícula donde aparece la letra α no puedan co-

1			7	3				
					2	5		
8								
			4					1
		5		6				
2								
4						3		
				2	4			
1		5						

Figura 1. Sudoku.

locarse el entero 1, ya que violarían las condiciones 1 y 2 anteriores, como las \times invalidan todas las celdas, excepto donde aparece la letra α , es claro que en esta celda debe colocarse el símbolo 1. Por otro lado, considere la celda en donde se colocó el símbolo β , en la Figura 3. Las celdas donde aparecen los símbolos 4 (en negritas) hacen que las celdas marcasadas con \times de la subcuadrícula donde aparece la letra β no pueda colocarse el símbolo 4, ya que nuevamente violarían las condiciones 1 y 2 anteriores, como las \times invalidan todas las celdas, excepto donde aparece la letra β , es claro que en esta celda debe colocarse el símbolo 4. Se sigue de la misma forma hasta completar todas las celdas vacías. En la Figura 4 se da la solución válida del Sudoku que aparece en la Figura 1.

1			7	3				
					2	5		
8								
\times	\times	\times		4				1
\times	\times	5		6				
\times	2	α						
4						3		
				2	4			
1		5						

Figura 2. Resolviendo el Sudoku.

Finalmente, se observa que el Sudoku de $n^2 \times n^2$ celdas, donde n se refiere al orden del problema, es un problema NP-Completo, como se ha mostrado en Yato y Seta (2002). La prueba usa una reducción simple para el problema de cuadrados latinos, el cual

ha sido probado que es NP-Completo por Colbourn (1984).

2. La Técnica de Recocido Simulado

El Recocido Simulado es un procedimiento introducido por Kirkpatrick, Gelatt y Vecchi (1982) y ha sido muy empleado para resolver problemas combinatorios. La idea surge del proceso físico conocido como recocido en el cual, se eleva la temperatura de un sólido hasta el punto que se vuelve líquido, a continuación la temperatura se disminuye de forma paulatina para obtener una estructura cristalina sin defectos y que puede considerarse como un estado de mínima energía. Cada descenso de temperatura debe ser lo suficientemente pequeño para que el sistema no adquiera una estructura cristalina con defectos, además el sistema debe permanecer un tiempo suficiente a una misma temperatura para permitir alcanzar un estado estacionario, en otras palabras, que las partículas vuelvan a reacomodarse.

1			7	3				
					2	5		
8								
			4					1
		5		6				
	2	1						
4			\times	\times	\times			3
			\times	\times	2	4		
1		5	\times	β				

Figura 3. Resolviendo el Sudoku (continuación).

1	5	2	7	3	8	9	6	4
7	3	9	4	6	1	2	5	8
8	6	4	2	5	9	1	7	3
6	8	7	9	4	5	3	2	1
3	4	5	1	2	6	8	9	7
9	2	1	8	7	3	6	4	5
4	9	8	6	1	7	5	3	2
5	7	6	3	8	2	4	1	9
2	1	3	5	9	4	7	8	6

Figura 4. Solución válida al Sudoku de la Figura 1.

$$P[\text{aceptar } x_j] = \begin{cases} 1 & \text{si } f(x_j) < f(x_i) \\ \exp\left(-\frac{f(x_i)-f(x_j)}{c_k}\right) & \text{si } f(x_j) \geq f(x_i) \end{cases} \quad (1)$$

El Recocido Simulado está catalogado como un procedimiento de Búsqueda Local Aleatoria: inicia con una solución factible x_i se selecciona a continuación una solución factible x_j dentro de una vecindad, enseguida se evalúa la calidad de la solución factible empleando una función de costo $f(x)$ asociada a cada solución factible x del problema.

Si la nueva solución factible x_j es mejor que la actual (de acuerdo al costo), se acepta, de lo contrario se selecciona de acuerdo a una probabilidad, para Recocido Simulado dicha probabilidad de selección está dada por la ecuación (1) que se conoce como el criterio de Metropolis; a c_k se le conoce como el parámetro de control (en el caso del sistema físico es la temperatura); este valor se inicia con un valor suficientemente grande para que cualquier solución tenga una probabilidad alta de ser seleccionada, a medida que c_k disminuye, la probabilidad de aceptar soluciones factibles de mala calidad disminuye (Kirkpatrick, Gelatt y Vecci (1982)). El enfriamiento se realiza empleando el sistema geométrico dado por la ecuación (2), donde α es un parámetro fijo, aunque a lo largo de los años se han desarrollado otros sistemas de enfriamiento.

$$c_{k+1} = \alpha c_k \quad k = 0, 1, 2, \dots \quad (2)$$

En la Figura 5 se da en pseudocódigo el algoritmo de Recocido Simulado. En las primeras líneas se definen los valores iniciales de los parámetros usados. Se puede observar que consta de dos ciclos uno exterior y el otro interior. El ciclo interior se ejecuta hasta que el número de *transiciones* se iguala a *equilibrio*. El ciclo exterior cambia el parámetro de control c_k de acuerdo a la ecuación (2) hasta llegar al punto de congelamiento.

3. Implementación de la Técnica de Recocido Simulado al Sudoku

En esta sección vamos a explicar cómo se implementa el algoritmo de Recocido Simulado al problema de resolver un Sudoku. Se pensará en un Sudoku de 9×9 celdas con 9 subcuadrículas de 3×3 , pero el procedimiento puede usarse para implementar

cualquier Sudoku de orden n . Se representará el Sudoku con un arreglo de 9×9 en donde existen celdas previamente llenadas. A las celdas previamente llenadas se les llamarán celdas fijas. Por ejemplo en la Figura 1, la celda (4,5) tiene el valor fijo igual a 4, o la celda (7,8) tiene el valor fijo de 3. A través del procedimiento, a las celdas fijas nunca se les cambiará el valor asignado.

3.1. Solución inicial

El objetivo del algoritmo que se desarrollará es obtener una solución válida al Sudoku, en donde solución válida significa un llenado del Sudoku donde se respeten los valores iniciales de las celdas fijas y se cumplan las tres reglas ya mencionadas. En lo que sigue, llamaremos una *solución factible* o simplemente *solución* de un Sudoku a cualquier relleno del Sudoku que cumpla con dos condiciones:

1. Respete los valores de las celdas fijas.
2. Respete la regla 3 de la sección 1.

Observe que esta definición de solución no implica que sea una solución válida del Sudoku porque no necesariamente cumple con las reglas 1 y 2.

La solución inicial se generará al azar, es decir, se genera una permutación aleatoria de los números del 1 al 9 no fijos en cada una de las subcuadrículas. Por ejemplo en la Figura 1, la primera subcuadrícula tiene dos celdas fijas la (1,1) y la (3,1) con los valores 1 y 8 respectivamente, por lo tanto, quedan libres los números 2, 3, 4, 5, 6, 7 y 9. Con estos números se genera una permutación aleatoria y se llenan las celdas faltantes. Por ejemplo, si la permutación aleatoria de los números anteriores fuera 9, 5, 7, 6, 2, 3, 4, se llenaría la primera subcuadrícula como en la Figura 6.

Procediendo de la misma forma con las siguientes 8 subcuadrículas obtendríamos el llenado dado en la Figura 7 (en donde aparecen numeradas las filas y las columnas). La solución inicial dada en la Figura 7 respeta las celdas fijas dadas originalmente (en negritas) y la regla 3 dada en la sección 1, pero no necesariamente respeta las reglas 1 y 2. Para esta solución inicial obtenida se puede observar, por ejemplo, que en la fila 1 no aparecen los valores 2 y 6 y se repiten los valores 5 y 7; por otro lado, en la columna 5, aparecen los valores 3 (tres veces), 4 y 8 (dos veces), y no aparecen los valores 1, 2, 5, 6. Lo mismo podemos observar del resto de las filas y columnas. Obviamente esta “solu-

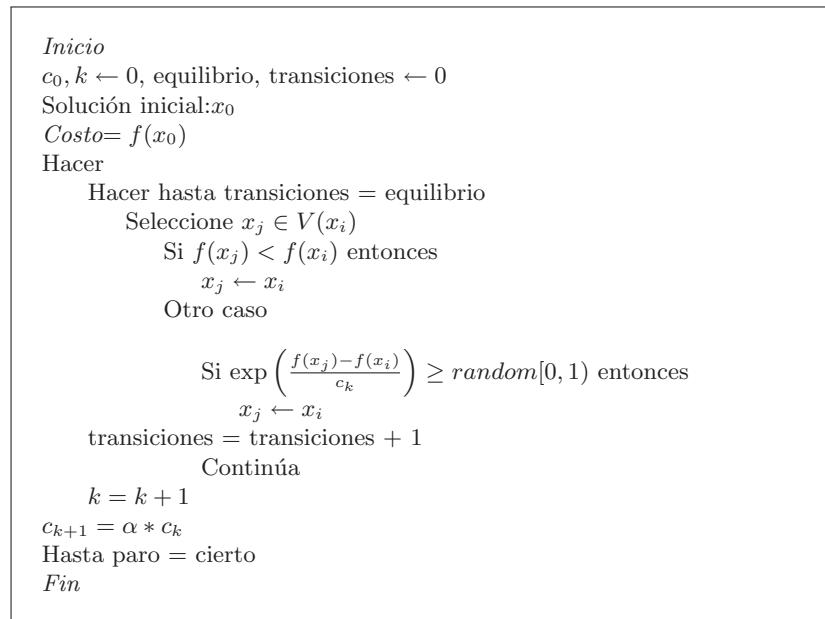


Figura 5. Algoritmo de Recocido Simulado.

ción inicial” no es una solución válida para el Sudoku ya que las reglas 1 y 2 no se cumplen.

3.2. Costo asociado a una solución

Vamos a asociar un costo (o una penalización) a cada solución factible que refleje que tan cerca o lejos esté de resolver el Sudoku. Entre más violaciones tengamos de las reglas 1 y 2, la penalización será mayor. Se asocia una penalización por cada una de las filas y columnas. Definiremos la penalización asociada a una fila como el número de enteros faltantes en la fila de entre los enteros 1, 2, ..., 9 y de la misma manera la penalización asociada a la columna es el número de enteros faltantes de entre los enteros 1, 2, ..., 9 en la columna. Por ejemplo para la solución factible de la Figura 7 la penalización de la fila 1 es igual a 2 ya que no aparecen el 2 y el 6. Y la penalización asociada a la columna 5 es 4 ya que no aparecen los enteros 1, 2, 5 y 6. Para la solución factible de la Figura 7 las penalizaciones asociadas se dan en la Tabla 1.

Tabla 1. Penalizaciones asociadas a la solución inicial.

Fila	1	2	3	4	5	6	7	8	9
Penalización	2	3	3	3	1	3	1	2	1

Columna	1	2	3	4	5	6	7	8	9
Penalización	2	1	2	3	4	3	2	2	3

El *costo total* de una solución factible la definiremos como la suma de las penalizaciones asociadas a las filas y columnas. Para la solución factible de la Figura 7 su costo total es de 41 unidades. Es fácil observar que si obtenemos una solución factible con costo total igual a cero, esta solución factible debe ser una solución válida para el Sudoku. Por lo tanto, el objetivo es ir mejorando las soluciones factibles hasta encontrar una solución cuyo costo total sea igual a cero.

1	9	5	7	3				
7	6	2			2	5		
8	3	4						
			4					1
			5		6			
			2					
4							3	
					2	4		
	1		5					

Figura 6. Llenado inicial de la primera subcuadrícula.

Observe que hemos definido un problema de optimización combinatoria donde se tiene un espacio de soluciones (las soluciones factibles dadas por las condi-

ciones 1 y 2 de la subsección 3.1) y cada solución tiene un costo asociado. La solución válida del Sudoku se obtiene cuando se descubre una solución con costo igual a cero. Si el Sudoku está bien planteado, debe de existir exactamente una solución con costo igual a cero.

	1	2	3	4	5	6	7	8	9
1	1	9	5	7	3	5	8	4	7
2	7	6	2	2	8	6	2	5	9
3	8	3	4	1	9	4	3	6	1
4	6	1	3	1	4	2	5	4	1
5	8	4	5	5	7	6	9	3	2
6	9	2	7	8	3	9	7	8	6
7	4	5	8	1	8	6	9	3	7
8	2	7	3	9	4	2	4	1	6
9	9	1	6	5	3	7	5	2	8

Figura 7. Solución inicial obtenida por el procedimiento.

3.3. Soluciones vecinas

Dada una solución factible actual, se desea generar otra solución factible, a partir de la primera, que cambie marginalmente. Para esto definimos las *soluciones vecinas* como todas aquellas que difieren de la actual al hacer un intercambio de los valores de dos celdas no fijas en la misma subcuadrícula. Observe que con esta definición se sigue cumpliendo que la nueva solución es factible ya que al hacer este cambio se cumplen las condiciones 1 y 2 de la subsección 3.1. Por ejemplo, si se seleccionan las celdas (5,4) y (6,5), de la solución inicial de la Figura 7 y se intercambian sus valores, se obtiene la solución vecina dada en la Figura 8.

	1	2	3	4	5	6	7	8	9
1	1	9	5	7	3	5	8	4	7
2	7	6	2	2	8	6	2	5	9
3	8	3	4	1	9	4	3	6	1
4	6	1	3	1	4	2	5	4	1
5	8	4	5	3	7	6	9	3	2
6	9	2	7	8	5	9	7	8	6
7	4	5	8	1	8	6	9	3	7
8	2	7	3	9	4	2	4	1	6
9	9	1	6	5	3	7	5	2	8

Figura 8. Solución vecina de la inicial.

3.4. Actualización de la función de costo

Si tenemos una solución factible y a partir de esta se genera una solución factible vecina, como se describe en la subsección 3.3, el costo total de esta nueva solución se puede modificar, pero no es necesario recalcular las penalizaciones en todas las filas y columnas para obtener el costo total de la nueva solución. Existen dos casos, tanto por fila como por columna. Para el cálculo de las penalizaciones por fila: si las celdas que intercambian valores aparecen en la misma fila, la penalización por fila queda exactamente igual ya que el intercambio afecta a una sola fila y este intercambio no agrega ni elimina ningún entero de esa fila, los enteros faltantes, si es que los hay, seguirían siendo los mismos. Por otro lado, si se intercambian valores en dos filas diferentes, entonces la actualización de la penalización en la fila se puede hacer en dos pasos: primero, si el entero que se elimina (se cambia por el otro) ya aparece en otra posición en la fila, quitarlo no hace que falte y por lo tanto no afecta a la penalización, en caso contrario, se aumenta en uno la penalización; segundo, si el entero que se agrega ya existe en otra posición en la fila, agregarlo no afecta la penalización de la fila, en caso contrario, la penalización disminuye en uno. La actualización por columna es exactamente igual que para la fila.

Por ejemplo para la actualización del costo de la solución factible vecina dada en la Figura 8. En la fila 5 primero se elimina el entero 5, pero como ya aparece este entero en la celda (5,3), no afecta su eliminación, por otro lado se agrega el entero 3 que ya aparece en la celda (5,8) y por lo tanto no afecta el valor de la penalización al agregarlo, en suma la penalización de la fila 5 en la nueva solución es igual a la de la solución inicial y su valor es 1. Para la fila 6, se hace el intercambio contrario, se elimina el valor 3 que no aparece en otro sitio de la fila y por lo tanto la penalización se aumenta en 1, por otro lado se agrega el entero 5 que no aparece en ningún otro lado de la fila y la penalización se disminuye en 1, finalmente la suma algebraica de ambas operaciones es $+1 - 1 = 0$ y la penalización de la fila 6 es igual a 3. Para el caso de las columnas: comenzemos con la columna 4, se elimina el entero 5, como también aparece en la celda (9,4), esta eliminación no afecta la penalización, por otro lado, se agrega el entero 3 que no aparece en el resto de la columna así que la penalización disminuye en 1 y por lo tanto la suma algebraica de ambas operaciones es $0 - 1 = -1$, así que, la penaliza-

ción de la columna 4 cambia a $3 - 1 = 2$. Para la columna 5, primero se elimina el entero 3 que aparece en las celdas (1,5) y (9,5) y por lo tanto la penalización queda igual al eliminar el entero 3, al agregar el entero 5 se observa que no existe en dicha columna y por lo tanto la penalización de la columna baja en 1, así que el efecto total es $0 - 1 = -1$ y por lo tanto la penalización de la columna 5, queda como $4 - 1 = 3$. Finalmente el costo total de la nueva solución vecina propuesta es 39. La nueva solución factible (de la Figura 8) se considera mejor que la inicial, dado que tiene un costo menor, es decir, tiene menos violaciones a las reglas 1 y 2.

3.5. Mecanismo de generación

El número de soluciones vecinas depende del número de celdas fijas, entre más celdas fijas haya se tendrán menos soluciones vecinas y viceversa. Se requiere crear un mecanismo de generación de una solución vecina a partir de la solución actual. La forma de hacerlo es: primero se selecciona al azar una celda no fija dentro de todas las celdas no fijas de la cuadrícula de 9×9 ; enseguida se selecciona al azar una celda no fija perteneciente a la subcuadrícula a la que pertenece la primera celda no fija seleccionada. Posteriormente se intercambian los valores enteros asignados a estas dos celdas seleccionadas.

Para crear una solución vecina de la solución actual dada en la Figura 7, se selecciona al azar una celda, es decir, seleccionamos al azar dos valores con distribución uniforme discreta entre 1 y 9, el primer valor se lo asignamos al número de fila y el segundo al número de columna. Por ejemplo, si obtenemos los valores 5 y 2 corresponden a la celda (5,2) que es una celda no fija (en caso que fuera una celda fija se intentaría nuevamente hasta obtener una celda no fija) esta celda corresponde a la subcuadrícula 4 por lo que la siguiente selección se haría seleccionando un número al azar con distribución uniforme entre 4, 5 y 6 y otro un número al azar con distribución uniforme entre 1, 2 y 3, nuevamente si está selección corresponde a una celda fija o a la celda ya seleccionada (la (5,2)), en este caso se seleccionarían otra pareja hasta obtener una celda no fija diferente a la (5,2); suponga que la celda seleccionada sea la (6,3), por lo que la solución vecina generada al azar sería aquella que resulta de intercambiar los enteros 4 y 7 en las celdas seleccionadas.

3.6. El parámetro de control

El valor inicial del parámetro de control c_0 se obtuvo realizando varias iteraciones con soluciones vecinas generadas a partir de la solución inicial aleatoria x_0 , (obtenida usando el procedimiento descrito en 3.1). Se hicieron $NOFIJA^2$ iteraciones, donde el valor $NOFIJA$ es el número de celdas no fijas del Sudoku que se va a resolver. Sea $f(x)$ el costo asociado a la solución x , calculado de acuerdo a la subsección 3.2. Para este conjunto de soluciones se calculó el costo promedio \bar{f} y la desviación estándar $\sigma_{\bar{f}}$. Se tomó el valor inicial del parámetro de control como: $c_0 = \sigma_{\bar{f}}$.

Los valores sucesivos del parámetro de control se obtuvieron mediante la ecuación (2) con un valor de $\alpha = 0.95, 0.97$ y 0.99 de acuerdo a la dificultad del Sudoku. Para cada valor fijo del parámetro de control se realizaron $NOFIJA^2$ iteraciones.

4. Algoritmo de Recocido Simulado para el Sudoku

El pseudocódigo del algoritmo de Recocido Simulado desarrollado para el juego del Sudoku se da en la Figura 9. En las líneas 2-4 se toman los valores iniciales del parámetro de control $c_0, k, equilibrio, transiciones$, el valor α que se toma igual a 0.95, 0.97 o 0.99 según sea el grado de dificultad del Sudoku y la solución actual x_0 . El ciclo interior (líneas 6-15) se ejecuta tantas veces como el número $NOFIJA^2$, donde se mantiene constante el parámetro de control. Si se alcanza una solución x_j que tenga un costo $f(x_j) = 0$, entonces se habrá obtenido la solución válida del Sudoku y el algoritmo termina (línea 8), en caso contrario se ejecuta totalmente el ciclo interior hasta obtener el equilibrio para el valor fijo c_k . El ciclo exterior (líneas 5-17) cambia el parámetro de control mediante la ecuación (2) en la línea 16 y controla el número de iteraciones hasta el punto de congelamiento (c_{final}) que se está definiendo como aquel en el que después de ejecutar un ciclo interior completo no hay cambios en la solución. Si el algoritmo termina con esta condición de congelamiento (línea 17) entonces no se habrá encontrado ninguna solución válida al Sudoku. En este caso se puede nuevamente “calentar” o recalentar, es decir, tomar nuevamente el valor inicial c_0 y la solución inicial igual a la solución actual y regresar a la línea 4 del algoritmo. Si después de seguir el procedimiento anterior, en reiteradas ocasiones, no se encuentra una solución válida al Sudoku se podría sospechar que el Sud-

```

1 Inicio
2  $x_i \leftarrow x_0$ 
3  $equilibrio \leftarrow NOFIJA^2, \alpha \leftarrow 0.95, 0.97, 0.99$ 
4  $transiciones \leftarrow 0, c_0 \leftarrow \sigma_{\bar{f}}, k \leftarrow 0$ 
5 Hacer
6     Hacer hasta  $transiciones = equilibrio$ 
7         Seleccione al azar  $x_j \in V(x_i)$ 
8             Si  $f(x_j) = 0$  entonces SALIR (solución válida del sudoku)
9             Si  $f(x_j) < f(x_i)$  entonces
10                 $x_j \leftarrow x_i$ 
11            Otro caso
12                Si  $\exp\left(\frac{f(x_j)-f(x_i)}{c_k}\right) \geq random[0, 1]$  entonces
13                     $x_j \leftarrow x_i$ 
14                 $transiciones = transiciones + 1$ 
15            Continúa
16             $c_{k+1} \leftarrow \alpha \times c_k, k \leftarrow k + 1, transiciones \leftarrow 0$ 
17 Hasta (No hay cambio de solución)
18 Fin

```

Figura 9. Algoritmo de Recocido Simulado para el Sudoku.

ku que se está tratando de resolver no tiene solución válida.

5. Resultados

Se tomaron Sudokus que aparecen en periódicos de circulación nacional tales como Reforma, El Universal, etc., y en portales de Internet como en <http://www.gamehouse.com/onlinegames/>, <http://www.sudoweb.com/>, <http://Sudoku.zeit.de/>, etc. en donde se catalogan en niveles de dificultad como fácil, medio, difícil y muy difícil. Se corrieron 100 Sudokus de cada nivel. Los resultados obtenidos aparecen en la Tabla 2 en donde se observa que el costo promedio para Sudokus de nivel fácil es mucho menor (en diez unidades) que Sudokus de nivel medio hacia arriba ya que el número de celdas fijas es mucho mayor que los demás y por lo tanto el número de errores al asignar valores aleatoriamente al resto de las celdas no crece mucho como en el caso de los otros niveles que no varía mucho el número de celdas fijas (en promedio entre 22 y 25). También se puede observar que la variación del número de celdas fijas entre los niveles medio, difícil y muy difícil no es mucho. Las desviaciones estándar promedio de los costos entre los diferentes niveles de dificultad no es muy significativa ya que va de alrededor de 9 a 10 unidades de manera creciente con respecto al gra-

do de dificultad. Para todas las corridas primero se calcula la desviación estándar de los costos de cada Sudoku y este valor se le asignó al valor inicial del parámetro de control. Los tiempos promedio de solución entre los diferentes niveles sí varía significativamente: de un tiempo promedio de 0.011 segundos, para el nivel fácil; se duplica en el nivel medio (.024 segundos), y se vuelve a duplicar el tiempo entre el nivel medio y difícil. La diferencia ya es más significativa entre el nivel difícil y muy difícil que se triplica. También se probaron varios valores para el parámetro de la ecuación (2) para los diferentes niveles obteniéndose valores de α idóneos para cada grado de dificultad. Para fácil y medio fue suficiente tomar el valor de $\alpha = 0.95$. Para el caso de nivel de dificultad difícil se tomó $\alpha = 0.97$ y para el nivel de dificultad muy difícil se tomó $\alpha = 0.99$. Con estos valores se garantizó la obtención de una solución válida a los Sudokus resueltos.

Conclusiones

En este artículo se da una descripción del Sudoku y se plantea como un problema de optimización combinatoria en donde se define un espacio de soluciones factibles y a cada solución se le asocia un costo. Se desarrolla un algoritmo de Recocido Simulado para resolver el Sudoku, la solución válida al Su-

Tabla 2. Estadísticas obtenidas en la solución de los Sudokus

Nivel	Costo Prom.	Desv. Estan.	Celdas Fijas Prom.	Tiempo Prom. (seg)	No. Iter. c_k fijo	α
Fácil	36.14	8.95	40.85	.01125	1612	0.95
Medio	45.78	9.51	25.05	.02465	3130	0.95
Difícil	46.55	9.70	22.65	.04445	3404	0.97
Muy Difícil	46.46	9.96	22.85	.13350	3381	0.99

doku es aquella cuyo valor objetivo es igual a cero. Se implementó el algoritmo y se corrieron instancias obtenidas de periódicos, revistas y algunas páginas de Internet cuyo grado de dificultad iba de fácil a muy difícil, obteniéndose siempre la solución válida en un tiempo muy pequeño.

Referencias

- Colbourn C. The complexity of completing partial latin squares. *Discrete Applied Mathematics*, (8):25-30, 1984.
- Kirkpatrick S., Gelatt C. D., Vecchi, M. P., Optimization by Simulated Annealing, *Science*, (220):671-680. May, 1983.
- Lynce I., Ouaknine J. Sudoku as a SAT problem, *Proceedings of the Ninth International Symposium on Artificial Intelligence and Mathematics*, Springer, 2006.
- Moraglio A., Togelius J., Geometric particle swarm optimization for the Sudoku puzzle, *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, 118-125, 2007.
- Yato T., Seta T. Complexity and completeness of finding another solution and its application to puzzles. In *Proceedings of the National Meeting of the Information Processing Society of Japan* (IPSJ), 2002.

cs

Comunicaciones en casos de misión crítica



Enrique Rodríguez de la Colina
Gerardo A. Laguna Sánchez

Recibido: 05 de octubre de 2011.

Aceptado: 03 de diciembre de 2011.

Abstract

The “mission critical communications” as main idea proposes to solve situations where communication is vital to save and protect people’s physical integrity or any type of life. In addition, this strategy of communication includes in its design, the transmission and reception care of the information that represents a significant application or value and where this could be irreplaceable. This article mention multidisciplinary concepts within a framework of the Information Technology (IT) and these are in particular to solve issues related to mission critical communications.

Keywords: Mission critical communications, communications during emergencies, communications after natural disasters.

Resumen

Las comunicaciones de misión crítica proponen resolver situaciones de comunicación en donde es fundamental salvar y proteger la integridad física de personas o la vida en general. Esta estrategia de comunicación incluye en su diseño un especial cuidado en la transmisión y recepción de la información de aplicaciones con un valor significativo que pudiera ser insustituible. En este artículo se mencionan algunos conceptos multidisciplinarios, pero dentro del marco de las Tecnologías de la Información (TI), que son de particular importancia para el diseño de esquemas de comunicación de misión crítica que ayuden a enfrentar mejor situaciones de emergencia.

Palabras clave: Comunicaciones de misión crítica, comunicaciones durante emergencias, comunicaciones y desastres naturales.

Introducción

Las comunicaciones de misión crítica, o *mission critical Communications* en idioma inglés, parten de la idea de resolver situaciones en donde es fundamental comunicarse para salvar y proteger la integridad física de personas o la vida en general, dando prioridad a la transmisión y recepción de la información que represente un valor significativo y que pudiera ser insustituible o crucial en una determinada crisis. Así, el estudio de este tema surge de la necesidad de contar con las técnicas y procedimientos adecuados para comunicarse en casos de situación crítica.

En otros países existen políticas de protección civil y militares que incluyen como parte fundamental el estudio de las comunicaciones robustas. Por ejemplo, en la Universidad de Cambridge, Reino Unido se organizan seminarios y talleres de comunicaciones de misión crítica dirigidos por asociaciones como los “Ingenieros sin Fronteras” (*Engineers Without Borders* o EWB). En este foro se tratan casos en los que es crucial contar con procedimientos, equipos y herramientas para hacer frente a problemas que pudieran presentarse debido a diversas causas como: desastres naturales, problemas civiles, conflictos militares, ataques a los sistemas de comunicación, cómputo y fallas en otros equipos.

Al desarrollar la idea de comunicación de misión crítica, se analizan posibles métodos para la anticipación a situaciones donde, considerando el desempeño humano en casos de emergencia o crisis y a pesar de contar con un plan de contingencia, al final puede presentarse la resolución de los problemas mediante cierta improvisación. Este comportamiento humano nos lleva a reflexionar sobre la necesidad de contar con equipos de trabajo, de múltiples habilidades y capacidades, que puedan afrontar los retos de las comunicaciones en casos “críticos”. En México existen organizaciones gubernamentales y civiles que se encargan de la protección civil en casos de desastre pero, en particular, consideramos que hace falta la creación de organizaciones que estudien, promuevan y desarrollen sistemas y procedimientos que se apoyen en las comunicaciones de misión crítica. Esto podría complementar de forma significativa los procedimientos existentes, que en su mayoría ven las situaciones de misión crítica desde una perspectiva general y no necesariamente desde el punto de vista de las telecomunicaciones.

Misiones críticas

Ya se dijo que las comunicaciones de misión crítica intentan ayudar a resolver situaciones en donde salvaguardar y proteger la integridad física, de personas o animales, pudiera ser la prioridad. Sin embargo, también se incluyen situaciones donde se necesita proteger a operaciones financieras, bursátiles, comerciales o proyectos de ingeniería donde los procedimientos y la infraestructura pudieran presentar fallas o ser destruidos. Otras comunicaciones de misión crítica contemplan la intervención en situaciones de crisis o emergencia tales como casos médicos, policíacos, de rescate, o bien, analizan asuntos en donde pudieran existir conflictos civiles y don-



Figura 1. Panorama general de las comunicaciones en casos de misión crítica.

de se requieren aplicaciones militares. Por otro lado, la protección y respaldo de la integridad de documentos importantes (quizás insustituibles en caso de pérdida) así como su difusión y control, tanto en electrónico o en papel, son objetivos analizados y resueltos por las redes en casos de misión crítica.

Entonces, las redes de comunicación de misión crítica son esenciales para garantizar la seguridad y la economía local, regional o inclusive mundial, en escenarios que se caracterizan por una gran incertidumbre generada por la incidencia de diversas variables externas que pueden afectar al sistema. Además, este tipo de redes resuelven una gama de problemas que pueden ir desde los más simples o controlables, como una transmisión de datos para verificar una transacción bursátil, hasta los más complejos e inesperados como, por ejemplo, la búsqueda rápida y efectiva de personas cuando ocurren desastres naturales. En la mayoría de los casos, las soluciones para comunicación de misión crítica deben contar con sistemas confiables y con un tiempo de respuesta lo más corto posible (*crítico*). Asimismo, deben contar con mecanismos y procedimientos adaptables a una amplia gama de equipos y sistemas, considerando la heterogeneidad de los sistemas de comunicación, tanto fijos como móviles.

La Figura 1 muestra los distintos enfoques de un entorno de comunicaciones de misión crítica donde se deben considerar diversos factores como el entorno geográfico (que puede ser clasificado como local, regional o global) y otros factores que tienen que ver con el hecho de que existen ambientes que requieren de seguridad informática, grandes volúmenes de información o comunicaciones de voz, en un tiempo relativamente corto, y donde existen políticas de privacidad de la información estrictas. La figura muestra que las comunicaciones en caso de misión crítica deben integrar sistemas capaces de afrontar los retos de situaciones de emergencia, como por ejemplo en desastres naturales, conflictos sociales y cuando se dan fallos inesperados en los sistemas de comunicación. Toda aplicación, al considerar los aspectos ilustrados en la Figura 1, debe garantizar confiabilidad en la transmisión de la información y tiempos de respuesta más cortos que los considerados en el diseño original del sistema.

Características y requerimientos

Es importante resaltar el hecho de que los sistemas de misión crítica requieren de una planeación y preparación adecuada en donde se deben considerar las diferentes etapas de operación y recuperación. Es decir, deben existir procedimientos en donde, al menos, se consideran etapas como son: la planeación (prevención), la operación (o puesta en ac-

ción) y la etapa de recuperación o restablecimiento de los sistemas.

En la etapa de planeación se deben formular preguntas como por ejemplo: ¿Existe en mi entorno de trabajo o familiar algún riesgo de emergencia que requiera que me comunique? o ¿Cómo se comunicaría (envío y recepción) información de gran importancia cuando los métodos tradicionales no están operando bien? ¿Existe algún procedimiento al respecto? Yo, o quien convive conmigo en el “día a día”, ¿Sabría cómo actuar en las situaciones donde fallan los sistemas de comunicación debido a alguna situación de emergencia? Pongamos por ejemplo, una inundación en donde el centro de servicio de datos pudiera resultar dañado.

Quizás estas preguntas no resuelven la situación en sí, pero nos alertan sobre la necesidad de contar con procedimientos que ayuden en una situación de comunicación de misión crítica. Incluso surgen más preguntas como: ¿Cómo puedo identificar una situación crítica? ¿Estoy preparado con lo básico? y ¿Qué puedo y debo hacer para estar preparado? Con estas preguntas, al menos de inicio, se plantea el problema de cómo abordar situaciones inusuales y nos concientiza sobre el hecho de que la solución depende de diversas variables (Krock, 2011; Oberg, Whitt y Mills, 2011). Lo cierto es que no existe un procedimiento estandarizado y general que pueda abarcar todos los casos, así que es labor de cada persona, grupo u organización, crear los procedimientos adecuados de acuerdo a la geografía, recursos económicos, herramientas y cantidades de personas o seres vivos en el entorno.

La Importancia de adecuar la tecnología a las aplicaciones

Analizando las comunicaciones de misión crítica, desde el punto de vista de la ingeniería electrónica, existen diversas soluciones para cada escenario, unas más avanzadas que otras, en cuanto a tecnología se refiere. Pero la creatividad, diseño y uso adecuado de cada tecnología pueden ayudar al desarrollo de sistemas eficientes de comunicación de misión crítica. Otra variable a considerar, en la planeación y puesta en operación, es la aplicación por sí misma, ya que define los ajustes a los procedimientos y sistemas diseñados para satisfacer las particularidades de cada escenario y geografía. Por ejemplo, en la República Mexicana tenemos un gran número de escenarios debido a la gran diversidad en la geografía (ver mapa de la Figura 2), aunque se pueden agru-



1. Sierra de Baja California
2. Llanura Costera del Pacífico
3. Sierra Madre Occidental
4. Altiplanicie Mexicana
5. Sierra Madre Oriental
6. Llanura Costera del Golfo
7. Sistema Volcánico Transversal
8. Depresión del Balsas
9. Sierra Madre del Sur
10. Sierra Madre Oriental
11. Sierras de Chiapas
12. Plataforma Yucateca

Figura 2. Mapa de la República Mexicana que ilustra diversidad en la geografía.

par como no habitados, rurales, metropolitanos y regionales.

La tecnología juega un papel fundamental en las comunicaciones para casos de misión crítica. Actualmente existen algunos sistemas avanzados de comunicación capaces de auxiliar a usuarios de forma colectiva o individual en situaciones críticas. Sin embargo, también pueden existir formas de comunicación sin el soporte de tecnología de punta, pero que, si existe la necesidad de improvisar en un momento dado, representan una alternativa para salir de un apuro. La improvisación no siempre se puede evitar, a pesar de que existan procedimientos y planeación adecuada ya que podemos enfrentarnos en una situación completamente nueva o desconocida, por ejemplo, si nos encontramos en un viaje o fuera de nuestro entorno conocido o simplemente donde no existe un plan de contingencia específico.

Aplicaciones y clasificación de las comunicaciones de misión crítica

En cuanto a las aplicaciones, éstas también abarcan una gama muy amplia. Un ejemplo, de aplicación en

misión crítica en los negocios, es aquella que garantiza el correcto funcionamiento de las comunicaciones de una organización o negocio donde una falla momentánea podría tener como consecuencia la salida permanente del mercado o la pérdida de la confianza de los usuarios (clientes). Si el negocio o empresa depende de clientes al mayoreo, una sola falla en una orden de compra puede ser de consecuencias graves. Por lo tanto, se busca minimizar las fallas en los equipos, en los paquetes y en los programas computacionales de los sistemas.

En general, las aplicaciones de misión crítica se pueden clasificar, de acuerdo a los daños que se pudieran ocasionar en caso de falla o cambios en el sistema u entorno, como:

- Pérdida de datos (información incompleta) por errores humanos o defectos del equipo.
- Caída del sistema.
- Desastres naturales como: incendios, inundaciones, terremotos entre otros. Los desastres naturales por lo general requieren de más cuidado en la planeación por ser prácticamente impredecibles.

La interrelación entre la necesidad de mantener la disponibilidad de un servicio y los factores de falla del sistema han provocado un enorme crecimiento de los sistemas de protección de datos (Viega y McGraw, 2002). Además existen programas y planes para recuperación de daños y rescate en casos de desastres naturales. Sin embargo, en cuanto a las comunicaciones se refiere, no se ha hecho suficiente énfasis. No sólo existen riesgos por desastres naturales, también existen por la interrupción de servicios debido a: problemas sociales, guerras o conflictos armados, que a su vez pudieran incluir ataques a los sistemas con sabotaje, e intrusiones de *non-ethical hackers*, conocidos como *crackers*.¹

Algunas prácticas y procedimientos para evitar la pérdida de información y comunicación contemplan la monitorización, mantenimiento, respaldo de los sistemas de información y la redundancia, todos los

cuales son temas relacionados con los sistemas distribuidos. Los presupuestos para crear estos procedimientos dependen del impacto que pudiera causar el fenómeno emergente y también dependen de los costos de la infraestructura y los bienes a proteger.

Seguridad de la información

Los sistemas de seguridad pueden definirse en función de la importancia de la información que protegen. Para este fin se pueden utilizar:

- Sistemas de respaldo de energía eléctrica con sistemas ininterrumpibles de energía.
- Respaldos de información en otros servidores o medios de almacenamiento como: CD, DVD, discos duros o memorias electrónicas de estado sólido.
- Duplicación de información en otros edificios, instalaciones o, inclusive, otras localidades nacionales e internacionales.
- Redundancia en los sistemas de comunicaciones. Para ello se implementa redundancia en transmisores, receptores, y en los medios de transmisión (fijos y móviles). Por ejemplo, este es el tipo de redundancia que se tiene, por lo general, en los bancos. En estos casos, con frecuencia se encuentran equipos de comunicación de distintos que además no tienen puntos de comunicación en común (Ramaswami, Sivarajan, Sasaki, 2010).

Para el caso de ataques ciberneticos, además de los sistemas de comunicación de respaldo para propósitos de prevención, deben existir respaldos de la información, copias de las aplicaciones y de los sistemas operativos. No basta con tener una buena política anti-intrusos, antivirus y de monitoreo de tráfico, sino que es necesario tener un plan de contingencia que contemple otras opciones de almacenamiento de la información.

Desde soluciones simples no-tecnológicas hasta soluciones con tecnología avanzada

Las soluciones no tecnológicas para comunicación de misión crítica pueden ayudar en casos de improvisación e, incluso, en los casos planeados, en donde lo esencial es comunicarse sea como sea. Estas soluciones no tecnológicas pueden ir desde: silbatos, sirenas,

¹No confundir a los *ethical hacker* (*hackers* éticos) con los *crackers* (no éticos). Los *hackers* tienen un gran conocimiento de la seguridad y ataque de sistemas y generalmente lo usan para un bien y no para irrumpir en los sistemas con fines personales. A diferencia de esto, los *crackers* que violan la seguridad de los sistemas con fines de uso personal y para hacer daño.

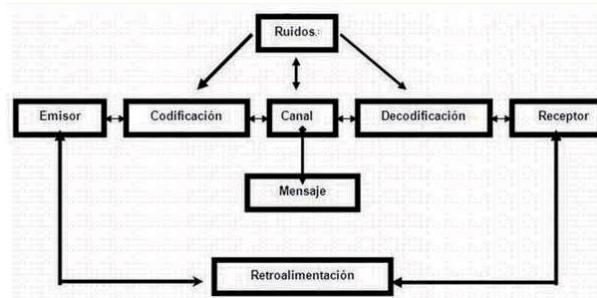


Figura 3. Sistema básico de comunicación.

alarmas, faros, fuentes de luz (como lámparas y fuego) y utilizar una forma de codificación,² tan simple como sea posible, de tal manera que en un caso crítico pudiera ser útil. Por ejemplo, un código internacionalmente conocido y empleado es el Código Morse. En la Figura 3 se muestra un sistema simple de comunicación electrónica con codificación. Se han realizado algunas pruebas en donde la comunicación con Código Morse a mostrado ser más rápida que un mensaje de texto enviado con un dispositivo móvil celular. A modo de ejemplo, se puede mencionar que en la UAM, en el marco del mes de la ciencia, se presentó recientemente un dispositivo que utiliza al Código Morse, como método de transporte de información, para comunicación en caso de siniestros tales como sismos o inundaciones (Meixueiro, 2011). Otro ejemplo puede ser la comunicación con banderas o símbolos, que es útil cuando los sitios a comunicar se puedan ver entre sí, es decir, que existe lo que se conoce en ingeniería de las comunicaciones como “línea de vista”.

En contraparte, ejemplos de soluciones que implican el uso de alta tecnología son los dispositivos inalámbricos móviles que permitan una comunicación eficiente entre lugares remotos y que, al operar prácticamente sin consumo de energía, permiten la operación por tiempos prolongados. Otros dispositivos de alta tecnología, son los que cuentan con sistemas avanzados de monitoreo y rastreo de personas y animales que incorporan los sistemas de localización conocidos como *Global Position Systems* o GPS. Otros sistemas más evolucionados pueden tener su base en sistemas inteligentes o autómatas diseñados para enviar información en medios hostiles

²Codificación se refiere al hecho de manipular la información con un método que permite convertir un carácter del alfabeto en un símbolo de otro sistema de representación, como puede ser un número o una secuencia de pulsos eléctricos en un sistema electrónico.

y con poca accesibilidad. Por ejemplo, aquellos lugares donde sólo un robot pudiera pasar.

Imaginemos el caso donde se emplea a un robot para auxiliar y acompañar a perros entrenados para la búsqueda de personas en un edificio derrumbado. Este dúo perro-robot permitiría la comunicación de personas atrapadas ya que, mientras que el perro puede regresar y llevar agua o alimentos, la persona podría comunicarse y describir su entorno a través de los dispositivos de comunicación del robot. Además el robot podría trazar la ruta para encontrar a la persona y mostrar por video el entorno. Otro ejemplo supone un robot con un detector de señales de celular para recorrer la zona afectada y detectar las señales de los celulares que estén encendidos, permitiendo el rastreo y la ubicación de las posibles víctimas que se encuentren inconscientes o que no se puedan comunicar por estar imposibilitados en ese momento.

El análisis de los sistemas de comunicación de misión crítica también debe considerar el desarrollo de aplicaciones que operen eficientemente y se adecuen al medio, a la vez que cumplan con los requerimientos de comunicación. Un ejemplo de esto, son los dispositivos que hacen uso a conveniencia del espectro radioeléctrico, es decir, que en emergencias son capaces de encontrar canales libres de comunicación en distintas bandas de frecuencia y así poder brindar los servicios requeridos en bandas que normalmente están ocupadas. Una tecnología adaptable que propone el uso a conveniencia del espectro, que podría ser de gran ayuda en las aplicaciones en casos críticos, es la conocida en inglés como *Cognitive Radio Networks*, que se puede traducir como redes de radios cognitivos (o inteligentes) (Mitola-III y Maguire, 1999). Estos radios cognitivos, con una tecnología que aún se está desarrollando, tienen como objetivo principal adaptar los parámetros de transmisión y recepción de los radios para cumplir de forma inteligente con los objetivos requeridos. Es decir, se trata de adaptar los parámetros de operación de los dispositivos de comunicación, de acuerdo a la aplicación, buscando bandas libres en el espectro radioeléctrico para poder transmitir y recibir señales.

Finalmente, dentro del enfoque centrado en las aplicaciones, es necesario mencionar que si la intención es comunicar datos, entonces los protocolos y estándares de comunicación tendrían que estar bien definidos para poder hacer frente a una contingencia.

cia en la comunicación. Esto sin ahondar en el mundo de los protocolos de comunicación, que por lo general se clasifican en niveles operativos que tienen que ver con estándares internacionales como los establecidos por la recomendación *Open Systems Interconnection* de la *International Standards Organization* (ISO/OSI).

Software y sistemas de información

Hoy en día las redes de datos representan un papel fundamental en las comunicaciones donde el *software* (constituido por programas y paquetes informáticos) es estudiado como parte esencial de las comunicaciones de misión crítica.

En cuanto a la seguridad del software, existe abundante bibliografía al respecto. La mayoría de los autores coinciden en que el diseño de un sistema debe considerar, la eficacia operativa de las instituciones para atender amenazas de seguridad y el tener una visión y entendimiento de cómo los recursos son compartidos entre las diferentes instancias de seguridad involucradas. Las soluciones deben ser independientes de marcas y tecnologías propietarias, además de que deben ser específicas para los sistemas requeridos, cumpliendo así con una operación independiente, coordinada, además de cooperativa (entre instancias) y segura en todos los aspectos. En el análisis de riesgos, se puede tener apoyo de un diseño de seguridad soportado por hardware, además de propuestas de sistemas novedosos y más efectivos. Después de la implementación, también es necesario realizar pruebas de robustez y confiabilidad de los sistemas instalados. Si no se realizan pruebas continuamente, los sistemas de emergencia no son confiables.

Por otra parte, la seguridad de los sistemas depende del valor de lo que se quiere resguardar. Así, con base en dicho valor, se crea el presupuesto del sistema a defender o proteger. Además se tiene que considerar qué tipo de atacante pudiera intentar vulnerar la seguridad y la integridad de la información. En general, existen normas y políticas de seguridad que ayudan en el diseño de los sistemas de comunicación y es una buena práctica incluirlos en el diseño del sistema propuesto.

Personas, actitudes y equipos de alto desempeño

A pesar de que las comunicaciones son un componente esencial en la respuesta ante una eventualidad o desastre, las personas y las organizaciones son

el punto clave para tener éxito en la operación del sistema. Recordemos que, en los casos críticos, es la tecnología o los sistemas los que se ven afectados directamente, por esta razón lo que se debe considerar fundamentalmente es el desempeño humano ante una situación desconocida. Es necesario contar con equipos especializados, capacitación y planes de acción donde el factor humano sea considerado como el elemento principal.

Los equipos deben trabajar de forma ordenada y con un liderazgo adaptable a las necesidades del momento. Es en los momentos de crisis donde el factor humano hace la diferencia para que una situación concluya felizmente o con consecuencias fatales. Por lo antes expuesto, continuamente deben ser fomentadas y practicadas habilidades de liderazgo y actitud ante un percance. Estos ejercicios deben tomarse con toda la seriedad y responsabilidad posibles y deben estar sustentados en un plan de contingencia, estudiado y meditado por todos los miembros del equipo o comunidad. Sin estas medidas de seguridad, difícilmente se puede llegar a una comunicación efectiva cuando es requerida.

Procesos y procedimientos

Generalizando los casos de comunicación de misión crítica, sobre todo si se analizan desde un enfoque sistemático, tenemos una etapa previa o de planeación, una etapa de ejecución en el momento de la misión crítica y una etapa de recuperación o reconstrucción. Entonces, se pueden establecer procedimientos que contemplen acciones para estas tres etapas, para que ayuden en la toma de decisiones y preparar herramientas en previsión de las situaciones consideradas como de riesgo. En particular, en la etapa de recuperación después de un desastre, se debe tener una estrategia clara donde se establezcan las prioridades para que la ejecución de las tareas a previstas fluya de la forma más natural posible. Este último punto se refiere al hecho de que, por lo general, se tiene un alto grado de estrés mientras se está luchando contra el tiempo y otras variables que son comúnmente desfavorables.

Existen diversos casos de estudio. Por citar algunos, de los más dramáticos y conocidos internacionalmente, tenemos el sismo de Wenchuan (Yang, 2011) con una escala de Richter de ocho, en mayo de 2008, que, en cuanto a su manejo de desastres y la velocidad de respuesta de los servicios de emergencia, puso a prueba a una gran región de China. Otro caso es el tsunami de Japón del 11 de marzo de 2011, debido a un te-

rrremoto de magnitud 9,0 MW que creó olas de mar remoto de hasta 10 metros. En este caso las telecomunicaciones fueron interrumpidas y algunas áreas sufrieron interrupción de los servicios de las Redes inalámbricas y de celular. Cabe destacar que algunos proveedores de servicio respondieron ofreciendo conexión gratis de larga distancia desde teléfonos fijos y móviles. También es ejemplar el terremoto sufrido en la ciudad de México el 19 de septiembre de 1985, donde los usuarios de servicios conocidos como “la banda civil” distribuyeron las primeras noticias de forma masiva.

Ante la experiencia de los casos anteriormente mencionados, podemos concluir que durante una catástrofe natural el acceso a los canales de comunicación se vuelve crítico, debido a que el personal de emergencia, los médicos, los voluntarios y la población en general intentan comunicarse a pesar de posibles irregularidades en los servicios de comunicación.

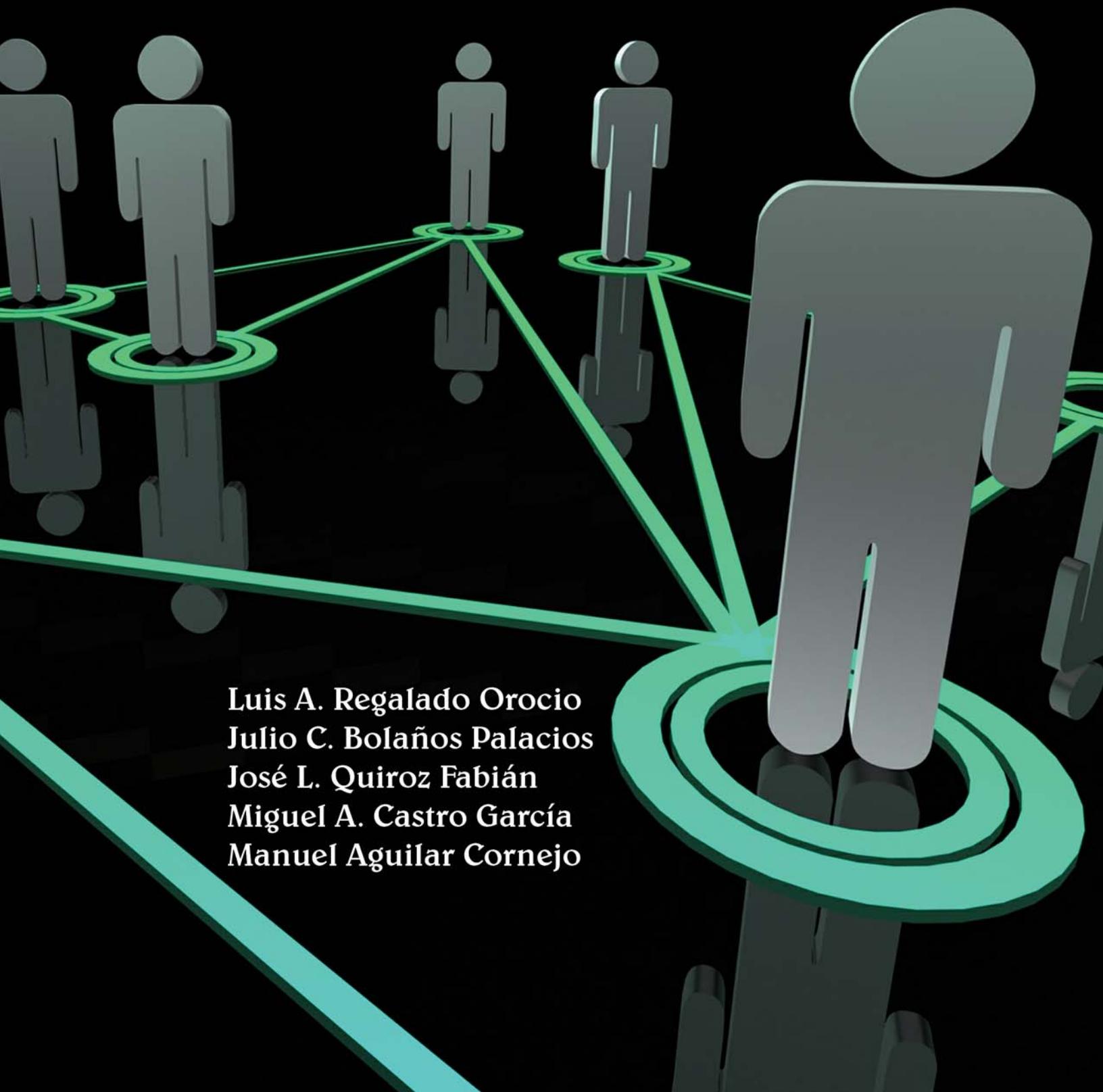
Conclusión

Este artículo muestra como los casos de comunicación de misión crítica pueden ser analizados y clasificados desde diversas perspectivas y ofrece un preámbulo para que exploremos este tema que nos ataña a todos, tanto a los especialistas en las comunicaciones como a la población en general. Debemos tener claro que las comunicaciones en casos de misión crítica pueden ayudar a los procedimientos creados por las organizaciones de protección civil, empresas y cuerpos de seguridad o policiales. Hacemos énfasis en el hecho de que es saludable contar con planes para prevenir, actuar y resolver problemas durante una situación de emergencia que pudiera afectar o incluso interrumpir los sistemas de comunicación.

Bibliografía

1. Krock, R. (2011, Enero). “Lack of emergency recovery planning is a disaster waiting to happen,” *IEEE Communications Magazine*, 49(1), pp. 48-51.
2. Meixueiro, E. (2011). “Vigencia del Morse en casos de emergencia,” *Publimetro educación*, 28-.
3. Mitola III, J., y Maguire, G. (1999, Agosto). Cognitive radio: Making software radios more personal, *IEEE Communications (Wireless Communications)*, 6(4), pp. 13-18.
4. Oberg, J., Whitt, A., y Mills, R. (2011, Enero). Disasters will happen - are you ready? *IEEE Communications Magazine*, 49(1), 36-42.
5. Ramaswami, R., Sivarajan, K., y Sasaki, G. (2010). “Optical Networks A Practical Perspective,” Elsevier.
6. Viega, J., y McGraw, G. (2002). *Building Secure Software* Addison-Wesley.
7. Yang, R. (2011, Enero). Considerations and suggestions on improvement of communication network disaster countermeasures after the wenchuan earthquake. *IEEE Communications Magazine* , 49(1), 44-47.

Configuración y Programación de un Clúster de GPUs.



Luis A. Regalado Orocio
Julio C. Bolaños Palacios
José L. Quiroz Fabián
Miguel A. Castro García
Manuel Aguilar Cornejo

Recibido: 31 de enero de 2012.

Aceptado: 03 de mayo de 2012.

Resumen

En los últimos años la tecnología de las unidades de procesamiento de gráficos (GPU-*Graphics Processing Unit*) ha sido utilizada para resolver problemas en diversas áreas de la ciencia, como son: criptografía, procesamiento de imágenes, predicción del clima, búsqueda en bases de datos, dinámica molecular, etc. Esta tecnología permite proponer soluciones que requieren una gran capacidad de procesamiento a un costo relativamente bajo, comparada con los procesadores convencionales.

En este artículo se presenta un panorama de los GPUs, así como una metodología para la construcción y puesta a punto de un cluster¹ de GPUs. Posteriormente se presentan algunos ejemplos de programación usando esta tecnología.

Introducción

Hoy en día debido a la reducción de costos del hardware y a la necesidad de mayor capacidad de cómputo surgen los procesadores multinúcleo (Multicore), los cuales están integrados por más de un núcleo (core) de procesamiento. Estos procesadores se encuentran desde simples computadoras portátiles (notebooks), hasta en estaciones de trabajo (workstations). El tener más de un núcleo permite realizar más de una tarea en forma simultánea por ejemplo, navegar por internet, reproducir música y descargar un archivo, todo al mismo tiempo.

Otra vertiente de los procesadores multicores son las GPUs. Estos dispositivos surgieron hace varias décadas como una solución al manejo de gráficos que requerían los sistemas operativos. Actualmente, las GPUs contienen cientos de núcleos, y son usados no sólo para apoyar al sistema operativo, sino para la gran demanda de gráficos que requieren aplicaciones como los videojuegos.

Buscando explotar las GPUs para aplicaciones en diferentes dominios, se ha introducido la filosofía GPGPU (General Purpose Computing on Graphics Processing Units), esto es, los GPU no solo son usados para el procesamiento de gráficos, sino para una gran diversidad de aplicaciones de propósito general. En la siguiente sección se presentan más generalidades de las GPUs.

¹Conjunto de computadoras interconectadas entre sí, con el fin de resolver un problema de forma conjunta, actuando como si se tratase de una sola computadora.

Unidad de Procesamiento de Gráficos (GPU)

Como se mencionó en la sección anterior, una GPU es un dispositivo que por sus características se encarga del procesamiento de gráficos en una computadora, aligerando en este aspecto la carga de trabajo de la CPU (Central Process Unit). Hoy en día se pueden encontrar GPUs en una gran variedad de dispositivos como: computadoras de escritorio, computadoras portátiles, consolas de videojuegos, videos proyectores, entre otros.

Las operaciones sobre gráficos que realizan las GPU, no son mas que operaciones sobre números reales. Siendo que además una GPU puede realizar varias de estas operaciones de forma simultánea, por esta razón se dice que una GPU trabaja de forma paralela.

Aunque una GPU y un CPU son dispositivos distintos en muchas formas, existen puntos de comparación entre ellos; uno de estos es la velocidad con la que pueden realizar una operación aritmética sobre un número real, el cual es representado en una computadora de forma eficiente y compacta usando la notación llamada: Punto Flotante². En la Figura 1 se muestra una comparación respecto a operaciones sobre números de punto flotante realizadas por CPUs Intel y GPUs NVIDIA. Como se observa en esta figura a principios del año 2003 tanto la CPU Intel como la GPU NV30 realizaban la misma cantidad de operaciones en un segundo, sin embargo conforme ha transcurrido el tiempo, la capacidad de calculó se ha incrementado en las tarjetas gráficas, a tal grado que en el año 2008 las capacidades de cómputo de la GPU modelo GT200 triplicó las capacidades del procesador Harpertown. Aunque no se muestra en la gráfica, esta capacidad de procesamiento sigue aumentando de forma considerable con la incorporación de tarjetas como la GTX 680 con 1536 núcleos y, recientemente con la GTX 690 la cual consiste del doble de núcleos que la GTX 680.

¿Cómo se programan las GPUs?

En un principio la programación en GPUs era muy difícil, ya que esto se hacía mediante interrupciones de hardware realizadas en el BIOS (*Basic Input/Output System*) de las computadoras donde estaban instaladas. Tiempo después, la programación se realizó utilizando el lenguaje ensamblador específico para cada modelo de GPU. Esta forma de

²La notación de punto flotante, es el método usual que emplean las computadoras para representar números reales.



Figura 1. Grafica comparativa respecto a las operaciones de punto flotante realizadas por CPUs Intel y GPUs NVIDIA del 2003 al 2008.

programación fue reemplazada al crearse un conjunto de interfaces de programación de aplicaciones, comúnmente llamadas APIs. Las APIs ayudaron a manipular las diferentes características de las GPUs, teniendo así un lenguaje de programación más homogéneo. Entre las APIs más importantes desarrolladas están OpenGL (*Open Graphics Language*) y DirectX.

Posterior al desarrollo de estas APIs, surgieron algunos lenguajes de programación orientados a gráficos; entre ellos podemos destacar a *OpenGL Shading Language* (GLSL), *C for Graphics* (Cg), y *High Level Shading Language* (HLSL); los cuales por sus características permiten definir secciones de código que se ejecutan directamente en la tarjeta gráfica.

Pero quizás el desarrollo más importante sobre GPUs se ha dado en los últimos años con la filosofía GPGPU. Por medio de esta filosofía, se busca aprovechar todas las capacidades de cómputo de las GPUs, con el fin de resolver problemas que por su naturaleza demandan una alta capacidad de cálculo. Para dicho propósito se han desarrollado diferentes lenguajes de programación como: BrookGPU, Sh, OpenCL (*Open Computing Language*), y CUDA (*Compute Unified Device Architecture*). Podríamos decir que este último es el más utilizado, ya que sus características lo hacen ideal para desarrollar aplicaciones altamente paralelas.

CUDA

CUDA es un lenguaje y conjunto de herramientas de programación desarrollado por la empresa NVIDIA para programar sus GPUs. Consta de un conjunto de aplicaciones y un compilador que permite a los programadores usar una extensión del lenguaje C y C++ para programar algoritmos que se ejecutarán sobre una GPU NVIDIA; además CUDA puede utilizarse también con los lenguajes Python, Fortran y Java.

Por medio de CUDA un bloque de código puede aislar como una función y compilarse para ser ejecutado en la tarjeta gráfica por uno o más hilos³

Instalación de CUDA SDK

Para la instalación de CUDA, se parte del supuesto que se tiene alguna distribución del sistema operativo GNU/Linux, posteriormente se siguen los pasos:

1. Instalar las librerías:
 - kernel-devel
 - gcc-c++
 - freeglut
 - freeglut-devel
 - libX11-devel
 - mesa-libGLU-devel
 - libXmu-devel
 - libXi-devel
2. Descargar el driver de la GPU desde la página de NVIDIA, seleccionando el controlador adecuado de acuerdo al modelo de la tarjeta y a la versión del sistema operativo.
3. Descargar las herramientas de desarrollo CUDA desde la página web de NVIDIA: CUDA Toolkit y GPU Computing SDK.
4. Instalar el controlador de la tarjeta con el modo gráfico del sistema operativo deshabilitado.
5. Instalar CUDA ToolKit y GPU Computing SDK. En este último paquete se encuentran algunos ejemplos.

Para comprobar que la instalación se realizó de forma exitosa se puede ejecutar el programa ejemplo DeviceQuery que se encuentra en GPU Computing SDK. Este programa detecta cada GPU instalada en el equipo y muestra en pantalla sus características.

En la Figura 2 se observa un ejemplo de la salida del programa DeviceQuery, mostrando las características de una GPU instalada.

Arquitectura de CUDA

Para entender el entorno CUDA es necesario tener claros algunos conceptos:

³Un hilo es la unidad de procesamiento mínima que puede ser planificado en un sistema

```
Found 1 CUDA Capable device(s)

Device 0: "GeForce GTX 460"
  CUDA Driver Version / Runtime Version      4.10 / 4.0
  CUDA Capability Major/Minor version number: 2.1
  Total amount of global memory:             1023 MBytes (1072889056 bytes)
    17 Multiprocessors x (48) CUDA Cores/MP: 336 CUDA Cores
    GPU Clock Speed:                         733 MHz
    Memory Clock rate:                      1850.00 Mhz
    Memory Bus width:                       256-bit
    L2 Cache Size:                          524288 bytes
    Max Texture Dimension Size (x,y,z):     1024 x 1024 x 64
    Max Layered Texture Size (dim) x layers: 10=(65536,65535), 20=(2048,2048,2048)
    Total amount of constant memory:          65536 bytes
    Total amount of shared memory per block:  49152 bytes
    Total number of registers available per block: 32768
    Warp size:                             32
    Maximum number of threads per block:     1024
    Maximum sizes of each dimension of a block: 1024 x 1024 x 64
    Maximum sizes of each dimension of a grid: 65535 x 65535 x 65535
    Maximum memory pitch:                   2147483647 bytes
    Texture alignment requirement:           128 bytes
    Concurrent copy and execution:          1 thread with 1 copy engine(s)
    Run time limit on kernels:              Yes
    Integrated GPU sharing Host Memory:     No
    Support host page-locked memory mapping: Yes
    Concurrent kernel execution:            Yes
    Alignment requirement for Surfaces:      Yes
    Device has ECC support enabled:         No
    Device is using TCC driver mode:        No
    Device supports Unified Addressing (UVA): Yes
    Device PCI Bus ID / PCI location ID:   6 / 8
    Compute Mode:
      < Default (multiple host threads can use ::cudaSetDevice() with device simultaneously) >

deviceQuery, CUDA Driver = CUDART, CUDA Driver Version = 4.10, CUDA Runtime Version = 4.0, NumDevs = 1
device = GeForce GTX 460
[deviceQuery] test results...
```

Figura 2. Ejecución del programa DeviceQuery incluido en el SDK de CUDA.

- **HOST:** Unidad Central de Procesamiento (CPU).
- **DEVICE:** Unidad de Procesamiento de Gráficos (GPU).
- **KERNEL:** Función que se ejecuta de forma paralela en la GPU.
- **THREAD o HILO:** Unidad mínima de ejecución en la GPU.
- **BLOCK o BLOQUE:** Conjunto de hilos agrupados, el número de threads dentro de un bloque es limitado y está definido por las características de la GPU. Cada Thread dentro de un bloque mantiene un identificador único.
- **GRID o MALLA:** Es a su vez un conjunto de Bloques, el tamaño de un GRID es también limitado, pero permite que una gran cantidad de threads sean ejecutados en una sola llamada a un Kernel. Un bloque dentro de un grid es también identificado mediante un ID.

En la Figura 3, se puede apreciar un esquema de organización de threads en la tecnología CUDA.

Los Threads durante su ejecución pueden acceder a distintos espacios de memoria:

1. La memoria local (*Local Memory*)
2. La memoria compartida (*Shared Memory*)
3. La memoria global (*Global Memory*)
4. La memoria constante (*Constant Memory*)
5. La memoria de textura (*Texture Memory*)

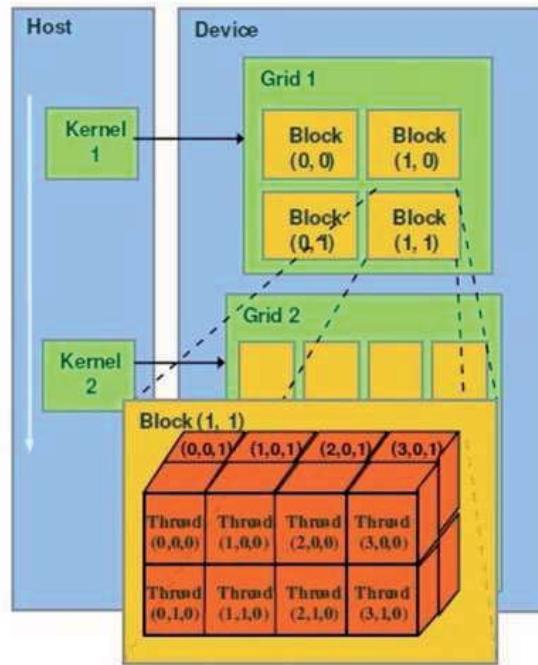


Figura 3. Organización de hilos en CUDA.

La *memoria local* es aquella que se le asigna a cada hilo. La *memoria compartida* permite comunicar (realizar operaciones de lectura y escritura) a todos los hilos de un bloque. La *memoria global* es aquella que sirve para comunicar a los hilos de todos los bloques y la CPU. La *memoria constante* y la *memoria de textura* son utilizadas para el procesamiento y visualización de imágenes y son solo de lectura para los hilos. Este par de memorias son especializadas, pues son del tipo memoria caché⁴. Toda esta jerarquía de memoria se puede apreciar en la Figura 4.

En CUDA se tienen una serie de calificadores que indican cierta particularidad para las funciones en CUDA; estas, se deben anteponer a la declaración de dicha función y definen las siguientes características:

- **`__device__`**: Indica que una función es invocada y ejecutada desde la GPU.
- **`__global__`**: Indica que dicha función será llamada desde el Host y ejecutada en el GPU; Se utiliza

⁴Memoria Caché: Es una memoria rápida y pequeña pero de alto costo, especialmente diseñada para contener información que se utiliza con frecuencia en un proceso.

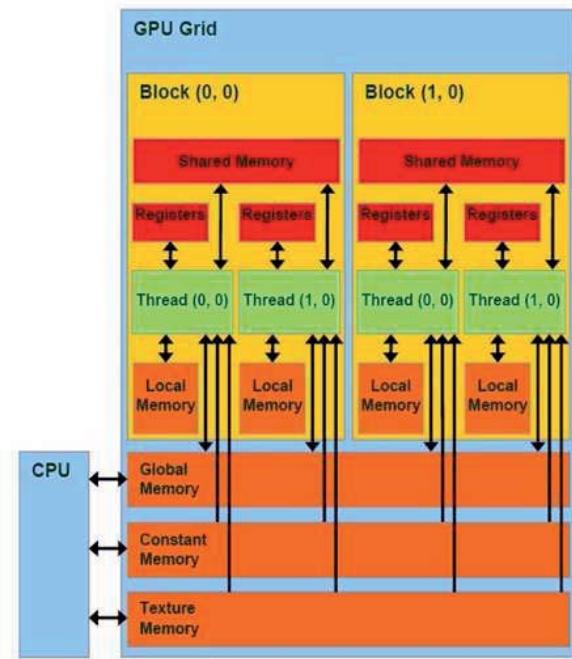


Figura 4. Organización de los diferentes tipos de memoria en una GPU NVIDIA, y sus diferentes niveles de acceso.

este calificador para caracterizar a las funciones Kernel en un programa CUDA.

- **`__host__`**: Indica que la función es invocada y ejecutada desde el Host.

Una parte importante en los programas realizados en CUDA, es la transferencia de información entre la memoria del CPU y la memoria global del GPU. Para ello se incluye en el lenguaje ciertas estructuras que realizan la copia de datos en diferentes flujos.

- `CudaMemcpyHostToHost`: Transfiere datos desde el Host al mismo Host.
- `CudaMemcpyDeviceToDevice`: Transfiere datos desde el Device al mismo Device.
- `CudaMemcpyHostToDevice`: Transfiere datos desde el Host hacia el Device.
- `CudaMemcpyDeviceToHost`: Transfiere datos desde el Device hacia el Host.

Con estos elementos se tiene las nociones básicas para escribir un programa en CUDA y ejecutarlo paralelamente en una GPU.

Primeros pasos en CUDA

Es común que cuando se aprende un nuevo lenguaje de programación, el primer programa que se realiza es el llamado: "Hola Mundo". Dicho programa muestra en pantalla la leyenda "Hola Mundo" y termina su ejecución. A continuación se muestran dos versiones de este programa, el primero es la versión en lenguaje C (Figura 5), y el segundo programa es la versión en el lenguaje CUDA (Figura 6).

```

1. //Hola Mundo
2. //Lenguaje C

3. #include<stdio.h>

4. int main(){
5.     printf("Hola Mundo\n");
6.     return 0;
7. }
```

Figura 5. Código en lenguaje C del programa Hola Mundo (archivo hola.c).

```

1. //Hola Mundo
2. //Lenguaje CUDA

3. #include<stdio.h>

4. __global__ void kernel(){}

5. int main(){
6.     kernel<<<1,2>>>();
7.     printf("Hola Mundo\n");
8.     return 0;
9. }
```

Figura 6. Código escrito en CUDA del programa Hola Mundo (archivo hola.cu).

En primer lugar, hay que resaltar que la extensión de los programas en CUDA es ".cu" mientras que en lenguaje C son ".c"; además observamos que el programa escrito en CUDA tiene una función llamada `void kernel` (Figura 6, línea 4). Esta función utiliza el calificador `__global__` que indica que el código que se escribe dentro de la misma se ejecutará en el Device (GPU). Cabe mencionar que en este programa la función no realiza ninguna operación, pues su cuerpo está vacío.

Otra diferencia es la invocación de la función `kernel` dentro de la función `main()` del programa `hola.cu` (Figura 6, línea 6), cabe destacar que dicha función puede llamarse de cualquier otra forma, respetando las reglas para definir funciones en el lenguaje.

La forma en que se invoca la función `kernel` antes mencionada es propia de lenguaje CUDA, donde `kernel` es el nombre de la función. El primer parámetro indica el número bloques que conformaran el Grid, el segundo parámetro indica la cantidad de hilos por bloque que ejecutarán de forma paralela la función `kernel`.

Para el ejemplo de la Figura 6, se ejecuta la función `kernel` con un bloque y dos hilos.

Para ejecutar el programa en CUDA se necesita la compilación del código, para ello se utiliza el compilador NVCC incluido en el SDK de CUDA.

```
$ > nvcc cod_fuente.cu -o cod_ejecutable
```

Otro ejemplo en CUDA: “Array_id.cu”

En la Figura 7 se muestra el código del programa `Array_id.cu`, en el cual se crea un arreglo de tamaño `N` donde cada una de sus entradas es inicializada por un hilo de la GPU.

Como se observa en el código, se crea un arreglo de tamaño `N` en la GPU usando la función `cudaMalloc` (línea 12), se define el total de hilos por bloque (línea 13) y el número de bloques que serán creados en la GPU (línea 14). Posteriormente se invoca la función `kernel` que será ejecutada en el GPU pasándole como parámetros el arreglo creado y el total de bloques e hilos por bloque (línea 16). Despues de terminar la ejecución de la función `kernel`, se transfiere el contenido del arreglo creado en la GPU a un arreglo en la CPU (línea 17) y se despliega su contenido (líneas 19-20). La implementación de la función `kernel` se muestra en las líneas 25-28. Cada hilo obtiene la entrada a inicializar en el arreglo en base a el identificador de su bloque (`blockIdx.x`), el total de hilos por bloque (`blockDim.x`) y su identificador en el bloque al que pertenece (`threadIdx.x`) (línea 26). Finalmente el valor de la entrada generado por cada hilo se asigna a un elemento del arreglo (línea 27).

- array_id.cu -

```
1. //Archivos de cabecera
2. #include<stdio.h>
```

```
3. //Definición de constantes
4. #define N 1000

5. __global__ void kernel(int *array);

6. //Función Principal
7. int main(void){

8.     int arreglo[N];
9.     int *dev_arreglo;
10.    int numBloques;
11.    int tamBloques;

12.    cudaMalloc ((void**) &dev_arreglo,
13.                N * sizeof(int) );
14.    tamBloques = 20;
15.    numBloques = N/tamBloques;

16.    //Invocamos al kernel.
17.    kernel<<<numBloques,
18.          tamBloques>>>(dev_arreglo);

19.    cudaMemcpy(&arreglo,dev_arreglo,
20.               N*sizeof(int) , cudaMemcpyDeviceToHost);

21.    printf("El Arreglo es: \n\n");
22.    for (int i=0 ; i<N ; i++)
23.        printf(" [ %i ] -> %d \n",i,
24.               arreglo[i]);
25.    printf("\n");

26.    cudaFree(dev_arreglo);
27.    return 0;
28. }
```

Figura 7. Código en lenguaje CUDA que crea un arreglo de tamaño `N` (archivo `array_id.cu`).

Trabajando con múltiples GPUs

Con lo mencionado hasta ahora se pueden implementar aplicaciones en CUDA que hagan uso de la GPU; pero ¿qué sucede si en una computadora se tiene más de una GPU?. Al aumentar el número de

GPUs, la capacidad de cálculo de la computadora aumenta; y ésta aumenta aún mas si la CPU es multicore.

Tradicionalmente cuando se tiene una CPU multicore en una computadora y, se desea aprovechar su capacidad de cómputo se hace uso de tecnologías como OpenMP, que es una API para la programación multihilo. OpenMP nos permite añadir concurrencia a programas escritos en lenguajes como Fortran, C, C++, y en nuestro caso permite trabajar con la plataforma CUDA. OpenMP trabaja con un modelo denominado FORK & JOIN (Figura 8); el cual permite definir una sección paralela dentro de nuestro programa. En esta sección paralela se crean varios hilos de ejecución, los cuales trabajan de forma conjunta para completar una tarea. Al finalizar la región paralela todo el trabajo realizado por los hilos se recolecta y se une en un solo resultado y de nueva cuenta se tiene un solo flujo de ejecución. Actualmente OpenMP es un estándar de la programación paralela.

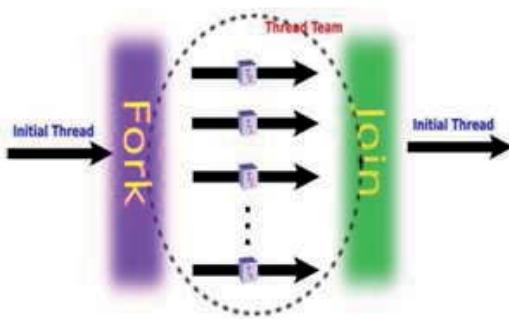


Figura 8. Modelo Fork & Join de OpenMP.

OpenMP + CUDA

Algunas funciones que necesitamos conocer de OpenMP son:

- `omp_set_num_threads()` : Fija el número de hilos.
- `omp_get_num_threads()` : Devuelve el número de hilos en ejecución.
- `omp_get_max_threads()` : Devuelve el número máximo de hilos que se ejecutarán en las zonas paralelas.
- `omp_get_thread_num()` : Devuelve el identificador del hilo (valor entre 0 y `omp_get_num_threads()`-1)

- `omp_get_num_procs()` : Devuelve el número de procesadores de la computadora o disponibles (para sistemas virtuales).

Una sección paralela en OpenMP se define como se muestra en la Figura 9.

```
1. #pragma omp parallel
2. {
3.     // Instrucciones que se ejecutarán
        // por los Hilos
4. }
```

Figura 9. Declaración de un bloque de código para ejecutarse en paralelo por un conjunto de hilos OpenMP.

Esquema básico de un programa: OpenMP + CUDA

A fin de usar OpenMP para trabajar con las GPUs y CPUs en una computadora se puede seguir la estructura de código que se muestra en la Figura 10. Como se observa, se incluye la librería `omp.h`, para poder tener acceso a todas las funciones de la biblioteca de OpenMP (línea 1). La función `Detección_GPUS()`⁵, se implementa usando algunas funciones que provee CUDA para detectar el número de GPUs instaladas en el sistema (línea 3). La función `Detección_CPU()`, se implementa usando la biblioteca de OpenMP (línea 4). `Init_data_cpu()` se encarga de inicializar los datos con los que se va a trabajar (línea 5).

Al inicializar la sección paralela, se crean los hilos CPU, algunos de ellos seleccionan una GPU para trabajar con ella (un hilo por GPU) (líneas 6 y 10). Cada hilo que controla una GPU, se encarga de copiar una porción de datos con la que va a trabajar (línea 11) y posteriormente invoca la función a ejecutarse en la GPU (línea 12). Los hilos que no controlan ninguna GPU, pueden realizar otras rutinas o incluso procesar parte de los datos (línea 16).

```
1. #include <omp.h>
2. void main(){
```

⁵Algunas funciones de detección de GPUs incluidas en el SDK CUDA, son: `cudaGetDeviceCount()`, `cudaSetDevice()`, `cudaGetDevice()` y `cudaGetDeviceProperties()`.

```

3. Deteccion_GPUS();
4. Deteccion_CPUS();
5. Init_data_cpu();

6. #pragma omp parallel
7. {
8. //Sección paralela de OpenMP
9. {
10. selec_GPU();
11. copy_data();
12. invlocar_kernel<<<>>>();
13. }
14. //Hilos Libres
15. {
16.   Otras_funciones();
17. }
18. }
19. }
```

Figura 10. Una posible estructura para un programa híbrido: OpenMP+CUDA.

Compilación OpenMP + CUDA

Para compilar un programa con directivas de OpenMP y funciones CUDA, es necesario informarle al compilador que el código fuente contiene funciones que no son propias de CUDA. Esto se realiza mediante la opción: Xcompiler de NVCC, de la siguiente forma:

```
$nvcc -Xcompile fopenmp origen.cu destino
```

Construyendo un cluster de GPUs

Hoy en día los clústers (grupos) de computadoras desempeñan un papel muy importante en diversas áreas de las ciencias de la computación, ya que al tener disponible la capacidad de cómputo de cada elemento del clúster, es posible atacar aquellos problemas que requieren una alta capacidad de cálculo. Este tipo de clúster son denominados: Clúster de Alto Rendimiento.

Por otro lado existen los Clúster de GPUs de alto rendimiento, que tiene la particularidad de estar formados exclusivamente por computadoras que a su vez tienen integrada una o varias GPUs. Dichas GPUs colaboran para resolver una problema computacional específico.

Para la implementación de un clúster no basta conectar varias computadoras a una misma red, si

no que se debe de proveer de un mecanismo de comunicación entre los procesos del clúster, este mecanismo comúnmente se denomina: paso de mensajes. En la actualidad una de las librerías más utilizadas para realizar paso de mensajes es MPI (*Message Passing Interface*). A continuación se describe como realizar un programa mediante CUDA y MPI para ejecutarse en un cluster de GPUs.

Paralelismo CUDA + MPI

Dentro de un cluster de GPUs podemos considerar a CUDA y MPI como entidades independientes, ya que el nivel de paralelización de CUDA es a nivel GPUs y el nivel de paralelización de MPI es a nivel nodo (máquina o computadora). Lo interesante de estas tecnologías es que pueden integrarse de manera sencilla usando un proceso MPI por GPU.

Para construir una aplicación que se ejecute en un cluster de GPUs son necesarios 3 archivos:

- *cuda.cu* (contiene código cuda).
- *mpi.c* (contiene código mpi).
- *machineFile* (contiene el nombre de cada nodo que conforma el clúster).

Se utiliza el compilador NVCC de CUDA para generar un archivo objeto del programa fuente .cu. Después se compila el archivo mpi.c haciendo uso del archivo objeto generado previamente. Además se agrega explícitamente la librería *cuda.h* propia del lenguaje CUDA.

Un aspecto importante a mencionar es que dentro del archivo *cuda.cu* se declara la palabra reservada *extern "C"* , que le indica al compilador que una función definida en el programa se implementa de manera externa. En la Figura 11 se ilustra lo anterior en la línea 4.

```

1. #include <stdio.h>
2. #include <unistd.h>
3. #include <cuda.h>

4. extern "C"
5. void run_kernel(int rank){

6. int num_gpus=0;
7. char host[20];
8. cudaGetDeviceCount(&num_gpus);
9. if(num_gpus>0)
10. {
```

```
11. .
12. .
13. .
```

Figura 11. Uso de la palabra reservada extern.

Para llevar a cabo la ejecución del programa se hace uso de `mpirun`, el cual es un programa que ejecuta cierto número de procesos en las máquinas indicadas en el archivo `machineFile`.

```
[gpu@pacifico5 simpleCUODMPI]$ mpirun --hostfile machineFile.txt -np 3 mpi
[pacifico5] ---[pacifico6]---[pacifico7]
[pacifico5] ---[pacifico6]---[pacifico7]

Proceso [ 0 ] : pacifico5 Tiene un GPU modelo: GeForce GTX 460. Leyendo dato = 10
Proceso [ 1 ] : pacifico6 Tiene un GPU modelo: GeForce 8600 GTS. Leyendo dato = 10
Proceso [ 2 ] : pacifico7 Tiene un GPU modelo: GeForce GTX 460. Leyendo dato = 10
```

Figura 12. Resultado de la ejecución de un programa simple en un clúster de GPUs.

La Figura 12 muestra una ejecución sobre un cluster con tres máquinas (pacifico5, pacifico6 y pacifico7), donde cada máquina tiene instalada una GPU con soporte a CUDA. Como se observa en la figura se crean tres procesos MPI, uno en cada máquina, los cuales toman el control de cada GPU.

Conclusiones y trabajo a futuro

En este artículo se presentó un panorama de la arquitectura y las aplicaciones que se pueden resolver haciendo uso de las GPUs. Se utilizó OpenMP para definir una arquitectura a fin de controlar más de una GPU y CPU en un programa en CUDA. Finalmente se definieron y describieron los elementos que se deben tomar en cuenta para implementar un cluster mediante MPI y CUDA.

Un trabajo a futuro sería buscar aplicaciones que requieran gran demanda de cómputo y puedan beneficiarse de tecnologías GPUs, por ejemplo aplicaciones en áreas como Física, Química, Biología, Procesamiento Digital de Imágenes, entre otras.

Referencias

1. Jason Sanders and Edward Kandrot. *CUDA by Example: An Introduction to General-Purpose GPU Programming*. Addison-Wesley Professional, 1 edition, July 2010.
2. F. Almeida, D. Giménez, J. M. Mantas, A. Vidal. *Introducción a la programación paralela*. Paraninfo Cengage Learning, 2008.
3. Rob Farber. *CUDA Application Design and Development*, Morgan Kaufmann Publishers Inc., 1 edition, San Francisco, CA, USA, 2011.
4. http://www.nvidia.es/object/what_is_cuda_new_es.html, Consultada Enero 2012
5. <http://graphics.stanford.edu/projects/brookgpu/>, Consultada Enero 2012
6. <http://www.opengl.org/>, Consultada Enero 2012
7. <http://directx.es/>, Consultada Enero 2012
8. <http://www.khronos.org/opencl/>, Consultada Enero 2012
9. <http://es.wikipedia.org/wiki/OpenMP>, Consultada Enero 2012
10. <http://www.open-mpi.org/>, Consultada Enero 2012
11. <https://www.wiki.ed.ac.uk/display/ecdfwiki/Use+multiple+GPU+devices+with+OpenMP+and+CUD>, Consultada Enero 2012
12. <http://www.open-mpi.org/faq/?Category=building#build-cuda>, Consultada Enero 2012

cs

De compartir contenido y encontrarlo usando



P2P

Geovanni Medrano Chávez
Elizabeth Pérez Cortés

Recibido: 23 de enero de 2012.

Aceptado: 22 de marzo de 2012.

1. Introducción

Actualmente existe una enorme cantidad de dispositivos de cómputo tales como computadoras personales, teléfonos inteligentes y tabletas electrónicas que almacenan distintas clases de contenido. Libros, música, video y software son algunos ejemplos de tal contenido. Concebir a estas entidades como dispositivos aislados no es correcto ya que generalmente están conectados a una red, como la Internet, que les permite comunicarse unos con otros. ¿Cuál es la mejor manera de organizar a estos dispositivos para intercambiar contenido?

El modelo Par-a-Par (P2P) intenta responder a la pregunta anterior. Bajo este modelo, los dispositivos que usan el sistema también aportan su contenido. Los sistemas P2P están diseñados para atender a cientos de miles o millones de usuarios. De hecho, su mayor fortaleza es su capacidad de crecimiento, pues mientras más dispositivos integren, más variado es el contenido y más usuarios pueden ser atendidos.

En la actualidad, los sistemas que emplean el modelo P2P gozan de gran popularidad. En febrero de 2001, Napster [7], una aplicación que permitía a sus usuarios compartir contenido multimedia, alcanzó los 26.4 millones de usuarios [1] mientras que en 2010, BitTorrent [3] tuvo 100 millones de usuarios en promedio al mes [2]. Hoy en día, se estima que más del 50 % de archivos descargados de la Internet fueron obtenidos mediante una aplicación P2P [5].

Considerando que los sistemas P2P atienden un gran número de usuarios y que el contenido se encuentra distribuido entre ellos, un reto importante es localizar un material en particular. Las técnicas empleadas para enfrentar este reto constituyen el típico central de este artículo. Sin embargo, antes de explicarlas, es necesario tener algunos antecedentes sobre los modelos utilizados en las aplicaciones de compartición de contenido. En la siguiente sección describiremos el funcionamiento, las ventajas y las desventajas de los modelos cliente-servidor (CS) y P2P.

2. Sobre cómo compartir el contenido

2.1. La solución centralizada: Modelo cliente-servidor

Una primera propuesta es concentrar el contenido en un equipo y dar a conocer la dirección a los consumidores potenciales. Este modelo se conoce como

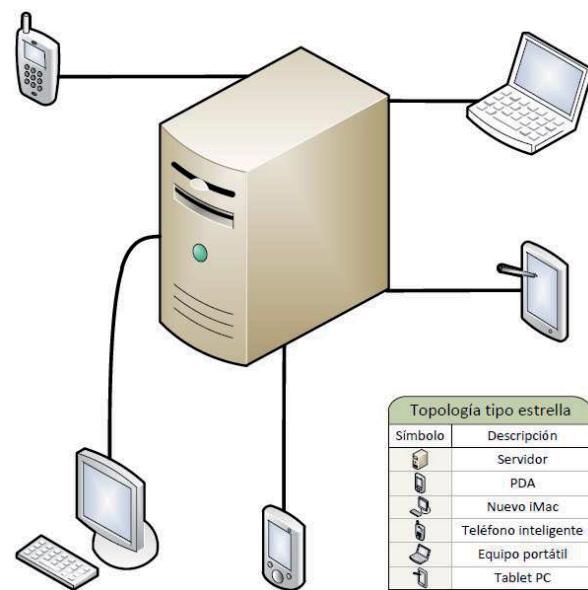


Figura 1. Modelo cliente-servidor.

cliente-servidor. El equipo que concentra el contenido y lo provee es denominado *servidor* y los consumidores son llamados *clientes*. En particular, un servidor es un equipo con una gran capacidad de recursos de cómputo (memoria, capacidad de almacenamiento, ancho de banda, poder de procesamiento), cuya función es proveer el contenido. Los servidores están disponibles la mayor parte del tiempo. Los clientes son máquinas, con una menor cantidad en recursos de cómputo, usadas por el público en general. Los clientes se conectan al servidor para conseguir algún contenido. La *web* es quizás el ejemplo más popular de aplicación basada en el modelo CS. La figura 1 ilustra una red tipo estrella representativa de este modelo.

El modelo CS posee dos importantes ventajas que lo hacen una solución atractiva. En primer lugar tenemos un control centralizado del contenido que permite su fácil actualización. En segundo lugar, la programación de aplicaciones y su mantenimiento son simples.

Sin embargo, a pesar de las ventajas anteriores, el modelo CS tiene dos inconvenientes mayores. Por una parte, el servidor tiene recursos finitos de cómputo, en consecuencia, la cantidad de contenido que almacena y el número de clientes que puede atender están acotados. Por otra parte, la falla del servidor implica la falla del sistema y la negación del acceso a todo el contenido.

Si se desea hacer frente a los inconvenientes anteriores mencionados es necesario invertir para mejorar el hardware del servidor. Cuando lo anterior ya no basta, se necesita adquirir más servidores que trabajen como uno solo. Sin embargo, el empleo de más servidores no implica que se supere la limitante en recursos, pero sí que la complejidad del modelo aumente.

2.2 La solución distribuida: Modelo P2P

La solución alternativa, llamada modelo P2P, consiste en conectar los dispositivos que emplean los clientes para crear una red que les permita intercambiar su contenido. En este modelo, el contenido se encuentra distribuido entre todos los participantes que conforman la red. Por tal razón, todos los dispositivos tienen el rol de servidor, cuando proporcionan contenido, y el rol de cliente, cuando lo descargan. Debido a la igualdad de roles, en el modelo P2P los participantes son conocidos como pares.

Los pares suelen ser dispositivos comunes de naturaleza diversa como computadoras personales, teléfonos inteligentes, tabletas electrónicas, etc. Cada uno de ellos es operado por un *usuario autónomo* que, de acuerdo a sus intereses particulares, determina el contenido que puede aportar. Los usuarios emplean una aplicación P2P con la que registran su identidad y el contenido que comparten. Además, la aplicación P2P les da acceso al contenido de los demás usuarios al conectarlos a alguna red P2P.

A diferencia del modelo CS, los sistemas P2P pueden atender a un número mayor de usuarios sin que su desempeño baje. Por lo anterior, resulta ser una gran ventaja que un sistema P2P tenga una membresía grande porque mientras más pares haya, más contenido estará disponible. Probablemente esta particularidad, llamada *escalabilidad*, es la de mayor importancia en los sistemas P2P.

Otra ventaja del modelo P2P con respecto al CS es que la ausencia de un par no implica que el sistema deje de funcionar. Los sistemas P2P están compuestos por una cantidad considerable de pares en los que está distribuido todo el contenido. Además, es muy probable que varios pares compartan el mismo contenido. Por tal motivo, solo se niega el acceso a un contenido cuando un único par lo comparte y éste se ausenta. Esta propiedad se conoce como *robustez*.

La escalabilidad y la robustez son dos ventajas atractivas del modelo P2P, sin embargo, localizar el conte-

nido no es tarea sencilla. Existen dos factores que dificultan esta tarea. Por un lado, el contenido se encuentra distribuido y, por otro lado, los pares son libres de entrar y salir del sistema. En consecuencia, no se puede asegurar la presencia de los pares ni la del contenido que comparten. Veamos qué es lo que le ocurre a un sistema P2P cuando muchos usuarios exhiben ese comportamiento.

3. Sobre los sistemas P2P y su demonio

Como mencionamos anteriormente, los pares poseen una característica especial: son autónomos. Por la autonomía, los pares pueden entrar al sistema a cualquier hora del día, permanecer el tiempo que consideren conveniente y luego irse, avisando o no de su partida. Los sistemas P2P no consideran a las entradas y salidas como fallas, sino como algo natural. Las entradas y salidas de los pares provocan que la topología¹ de la red P2P cambie constantemente. Por ejemplo, en la figura 2 mostramos la topología de una pequeña red P2P capturada en el tiempo 1. Después, en el tiempo 2, se aprecia que la red es diferente y que únicamente permanecieron los dispositivos marcados con estrellas.

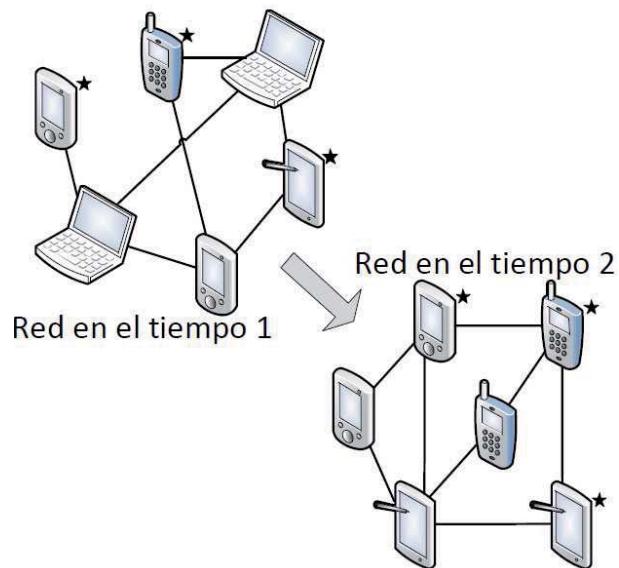


Figura 2. La transitoriedad en los sistemas P2P.

Debido a su robustez, los sistemas P2P pueden trabajar bajo tal ambiente dinámico. Sin embargo, su funcionalidad se ve comprometida cuando la membresía es modificada abruptamente porque un porcentaje considerable de nodos entró o salió del sis-

¹La topología es la forma de la red.

tema. A tal fenómeno se le conoce como *transitoriedad*. Los sistemas P2P tienen mecanismos para soportar cierto nivel de transitoriedad y seguir funcionando a pesar de un cambio drástico en la topología de la red. Sin embargo, si el nivel de transitoriedad es alto, entonces la forma de la red cambiará a una velocidad mayor y es probable que el sistema no lo tolere.

Como consecuencia de la transitoriedad, los usuarios pueden experimentar retrasos en la resolución de sus consultas o falsos negativos, es decir, no encontrar lo que desean a pesar de que sí está siendo compartido. En el peor de los casos, los usuarios notarán que la aplicación ha dejado de funcionar. Para entender mejor lo antes mencionado, es necesario analizar por separado el efecto de las entradas y el de las salidas de los pares sobre los sistemas P2P. Pensemos en una gran fiesta con invitados que llegan y que se van, tal como ocurre en los sistemas P2P, y supongamos los siguientes casos:

- Respecto a las entradas, supongamos que un gran grupo de invitados acaba de llegar. Juan y Paula ya estaban en la fiesta, pero se separaron y ahora él no la encuentra. Dado que el número de recién llegados es grande, Juan pregunta a uno de ellos si ha visto a Paula (dando una descripción). El recién llegado le responde que no. Aquí ocurrió un falso negativo, ya que Paula está en la fiesta. También se retrasó la consulta, porque Juan debe volver a preguntar la localización de Paula. A pesar de lo anterior, eventualmente Juan va a encontrar a Paula porque hay gente que ya lleva rato en la fiesta y porque los recién llegados podrán detectar a Paula por su descripción. Cuando la membresía de un sistema P2P crece abruptamente, los pares experimentan retrasos y falsos negativos, pero después de cierto tiempo, ya que los pares nuevos se hayan integrado, el sistema se estabiliza.
- Las salidas tienen diferentes implicaciones sobre los sistemas P2P y ocurren de dos maneras: informadas (1) y no informadas (2). Ahora supongamos que las siguientes situaciones se presentan colectivamente en varios grupos de la fiesta:
 1. De un grupo de diez amigos, ocho deciden abandonar la fiesta. Entonces, avisan a los otros dos de su retiro y les presentan a otros amigos con quien platicar. Los dos que permanecieron pueden seguir divirtiéndose porque no

quedaron aislados. En los sistemas P2P, las salidas informadas tienen el mismo efecto que las entradas masivas, por lo tanto, no son tan perjudiciales.

2. En otro grupo de diez, ocho amigos se van sin comentarlo con los otros dos. En tal situación, los que se quedaron buscarán a los otros sin éxito y después se quedarán aislados porque no conocen a otras personas. Para los dos que se quedaron, la fiesta dejó de funcionar, dejó de ser divertida. Si este comportamiento se presenta en muchos grupos de pares en un sistema P2P, puede ser que la robustez se pierda y en consecuencia, que el sistema falle totalmente.

La transitoriedad es un problema que baja el desempeño de los sistemas P2P, en particular, si su nivel es alto. Sin embargo, los sistemas P2P pueden localizar el contenido y mantener a los pares conectados en presencia de cierto nivel de transitoriedad. Enseguida describimos los mecanismos utilizados por diferentes sistemas P2P para estos fines.

4. Sobre cómo encontrar el contenido

Para permitir el intercambio de contenido, en las redes P2P se emplea un *servicio de localización*. A pesar de que el contenido se encuentre distribuido, el servicio de localización puede descubrir lo que un usuario solicite a través de su aplicación, si es que está. Además, el servicio de localización emplea distintos mecanismos que le permiten tolerar la transitoriedad. El tipo de servicio de localización empleado depende de la topología de la red P2P.

En este punto conviene aclarar que la topología de las redes P2P es lógica, es decir, las conexiones entre los pares no corresponden a las conexiones físicas reales. Regularmente, las conexiones físicas son muy complejas y muchas aplicaciones no necesitan tener conocimiento de tal complejidad. Para simplificar el entendimiento de las conexiones, las aplicaciones emplean un ingenioso concepto: el de las redes superpuestas. Éstas son redes lógicas que se colocan por encima de alguna red física o lógica y describen únicamente las conexiones que alguna aplicación emplea y hacen abstracción de los enlaces que se encuentran en las redes subyacentes. Por ejemplo, en la figura 3 se muestra una red lógica en forma de anillo superpuesta a una red física tipo malla. En este caso, la aplicación requiere de una comunicación en anillo, por lo que hace abstracción de los enlaces físicos de la malla.

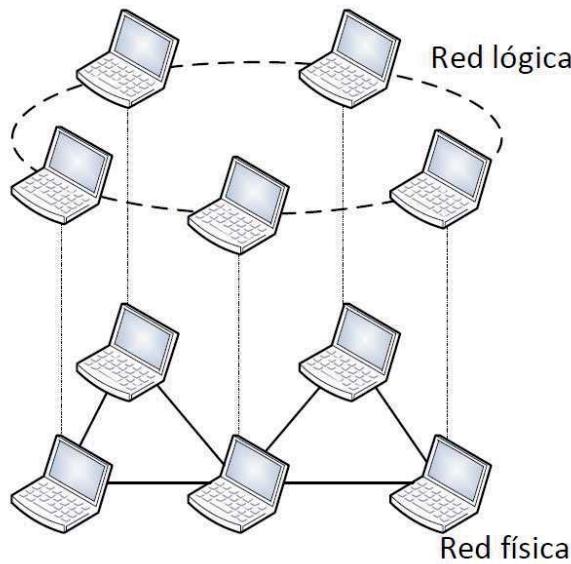


Figura 3. Redes superpuestas (malla subyacente a anillo).

El servicio de localización está pensado para descubrir el contenido que hay en la red lógica P2P y confía que las redes subyacentes son funcionales. En los siguientes apartados vamos a describir algunas de las topologías P2P junto con su correspondiente servicio de localización.

4.1 Sobre redes estrella y localización centralizada

Supongamos que vamos a un centro comercial tan grande que tomaría días recorrerlo en su totalidad y que buscamos una tienda en específico. ¿Qué podríamos hacer para encontrar el lugar que buscamos? La solución más ingenua es ir por todo el centro comercial hasta encontrar la tienda de nuestro interés. Una solución más inteligente consiste en consultar el directorio del centro comercial, que contiene un índice que detalla por categorías la dirección de todas las tiendas, e ir directamente al local indicado.

De manera similar se localiza el contenido en un sistema P2P centralizado. Se llama así porque emplea un servidor de localización que almacena un índice con información sobre todo el contenido disponible y los pares que lo poseen. La topología de la red superpuesta que describe esta clase de sistema es de tipo estrella y como se ilustra en la figura 4, todos los pares se conectan al servidor de localización.

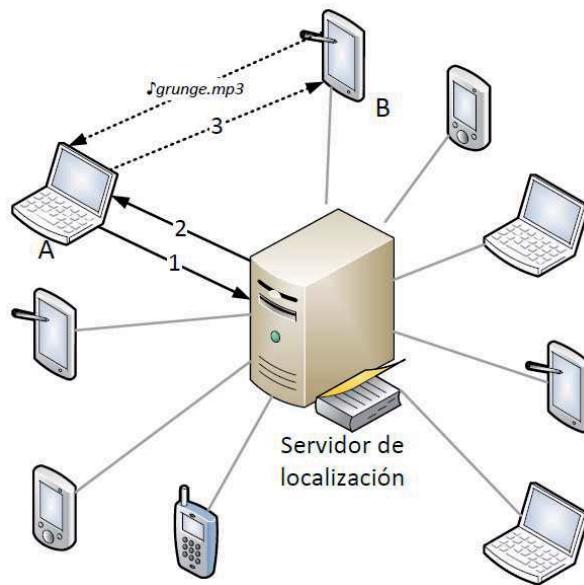


Figura 4. Red P2P centralizada.

En los sistemas P2P centralizados, el servicio de localización resuelve las consultas empleando el índice que posee el servidor. La información contenida en el índice es obtenida por medio de la aplicación P2P. Ésta pide al usuario que registre tanto su identidad como el contenido que va a ofrecer. Estos datos son enviados directamente al servidor para ser registrados en el índice.

Gracias al índice centralizado, la localización de contenido es relativamente simple. Para mostrar su funcionamiento, vamos a emplear el ejemplo ilustrado en la figura 4. En el paso (1), el usuario A hace una consulta al servidor y pregunta por el archivo 'grunge.mp3'. En el (2), el servidor busca en el índice global y le regresa a A una lista con la identidad y la dirección de los pares que poseen 'grunge.mp3' y que están en línea. En este caso el par B está en la lista. En el (3), A le solicita a B 'grunge.mp3'. Por último, en el (4) B le envía directamente el archivo a A.

No debemos confundir los sistemas CS con los sistemas P2P centralizados. En el primero, el contenido se encuentra alojado en un servidor y el intercambio de recursos se realiza entre el servidor y sus clientes. En el segundo, el servidor no almacena el contenido, sino las direcciones de los pares que lo poseen. Note que el intercambio de recursos se realiza directamente entre pares y el servidor no interviene en el proceso.

Los sistemas P2P centralizados heredan del modelo cliente-servidor ventajas y desventajas. En principio, como describimos en el ejemplo anterior, resulta simple encontrar el contenido ya que solo se necesita consultar el índice que posee la información global del sistema y que está alojado en el servidor. Sin embargo, la escalabilidad del sistema está comprometida ya que va a depender de cuántas direcciones puede almacenar el servidor y de cuántos pares puede atender. La funcionalidad del sistema también está comprometida porque si el servidor falla, el sistema también va a fallar.

Respecto a la transitoriedad, ésta repercute directamente en los resultados que entrega el servidor. Básicamente, el servidor debe tener conocimiento de quién está en línea y quién no lo está. El problema radica en que el servidor solo se percata de la ausencia de los pares cuando estos cierran su sesión. Por tal motivo, si el par falla o el usuario no cierra su sesión, existe la probabilidad de que el servidor entregue información imprecisa.

Una posible solución al problema anterior es que el servidor reciba mensajes de mantenimiento de los pares que están en línea cada cierto tiempo. Para que el índice esté actualizado se requiere que la velocidad del envío de mensajes de mantenimiento sea mayor o igual al nivel de transitoriedad del sistema. Por tal motivo, este mecanismo de actualización deteriora el desempeño del sistema porque añade carga extra que puede saturar al servidor.

4.2. Sobre grafos aleatorios y localización por inundación

La vulnerabilidad provocada por el uso de un servidor de localización centralizado hizo que se repensara la forma de localizar el contenido en los sistemas P2P. Al eliminar un único punto de conexión, surgen dos preguntas fundamentales ¿cómo conectar a cientos de miles de pares? y ¿cómo buscar el contenido una vez que se ha determinado la conexión entre los pares?

Para responder la primera pregunta pensemos en nuestras relaciones sociales. Como empíricamente sabemos, no conocemos a todas las personas del mundo, pero sí conocemos a cierta cantidad de personas. Según la teoría '*seis grados de separación*', cada quien conoce en promedio a cien personas, entre familiares, amigos y compañeros. Las personas que conocemos también conocen a otras cien personas más y de la misma manera esas personas conocen a otras cien y así sucesivamente. Gracias a es-

tas relaciones se garantiza que nosotros podamos tener contacto con cualquier persona a través de nuestros conocidos.

Del párrafo anterior se puede deducir que para evitar el servidor de localización de las redes P2P centralizadas basta conectar a cada par con otros pares y así sucesivamente. Comúnmente, cada par se conecta con sus pares vecinos y para determinar el vecindario de algún par se pueden emplear varios criterios. Por ejemplo, se puede establecer que un par es vecino de otro si sus direcciones son próximas. Otra forma es determinar la vecindad entre pares de acuerdo a su cercanía geográfica por medio de un sistema de posicionamiento global.

Conectar a cada par con algún grupo de pares genera una red P2P pura como la que se ilustra en la figura 5. Debido a que las conexiones entre los pares siguen la ambigua regla "conéctate con tus pares próximos", la topología que describe la red se conoce como *grafo aleatorio*. Se le da ese nombre a la topología porque la forma que tiene se puede producir por medio de un generador de grafos aleatorios. Un grafo aleatorio es un conjunto de vértices conectados por aristas, cuya conexión está en función de un proceso aleatorio.

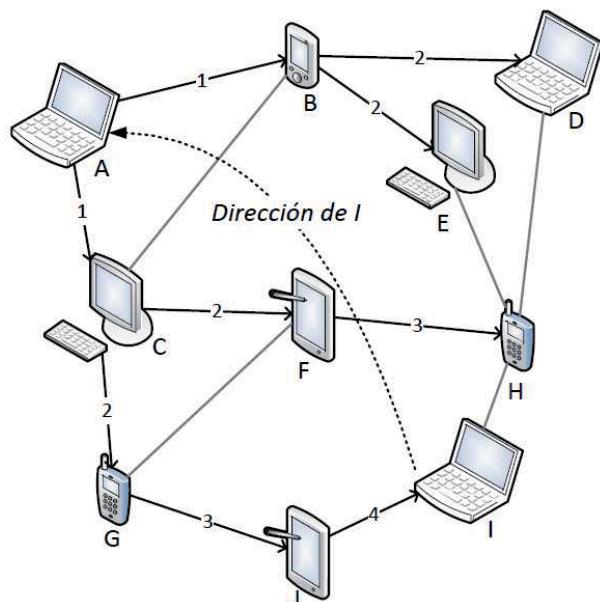


Figura 5. Red P2P pura.

Ahora es tiempo de responder la segunda pregunta. Ya tenemos una red lógica para comunicarnos con cualquier par que esté integrado a la red, sin embar-

go, no sabemos qué par comparte el contenido que buscamos y no existe un servidor que nos dé tal información. Aun así, cada par sabe qué comparte y además sabe quiénes son sus vecinos. Con eso es suficiente para construir una solución.

Consideremos el siguiente ejemplo, ilustrado en la figura 5. Suponga que el usuario *A* emplea una aplicación P2P pura. *A* hace una consulta por el archivo ‘*pop.mp3*’. Ya que *A* tiene como vecinos a *B* y a *C*, entonces les pregunta a cada uno de ellos si poseen ‘*pop.mp3*’ (paso 1). *B* y *C* reciben la consulta, buscan entre los archivos que comparten la petición de *A* y se dan cuenta de que no poseen el archivo. Entonces, *B* y *C* reexpiden la consulta de *A* a sus vecinos, a decir, *D*, *E*, *F* y *G* (paso 2). Del mismo modo, ocurre que los últimos tampoco tienen a ‘*pop.mp3*’, así que la consulta se reexpide a *H* y *J* (paso 3). Estos últimos tampoco encuentran el archivo así que la consulta se reenvía a *I*. Este último sí encuentra a ‘*pop.mp3*’ y envía su dirección a *A* para intercambiar el recurso. El mecanismo de localización anterior se conoce como *inundación*.

Propiamente, la inundación consiste en que cada par reenvíe una consulta a sus vecinos, excepto a aquel de quién la recibió. Si la consulta llega por segunda vez a un par, éste la descarta. El proceso se repite hasta que todos los pares de la red han recibido la consulta.

Una gran ventaja de las redes P2P no estructuradas es que son muy buenas resistiendo la transitoriedad. Las constantes entradas y salidas de los pares modifican el grafo aleatorio, pero como los mecanismos de localización basados en inundación no dependen de la forma de la red, entonces la probabilidad de tener éxito en la resolución de una consulta no desciende significativamente. Si el nivel de transitoriedad es elevado, entonces puede ocurrir que algunos pares con pocos vecinos queden aislados si estos abandonan el sistema abruptamente. Para que los sistemas P2P puros fallen, un porcentaje considerable de pares debe abandonar la aplicación al mismo tiempo para generar particiones de red, lo cual es poco probable en sistemas reales.

El principal problema de las redes P2P puras es el uso de mecanismos de inundación. La cantidad de tráfico que generan las consultas crece exponencialmente saturando el ancho de banda y restringiendo la escalabilidad del sistema. Por ejemplo, si ca-

da par tiene 10 vecinos, entonces en el paso n se habrá inundado la red con 10^n consultas.

No obstante, existen otros mecanismos que emplean menos ancho de banda que la inundación y que pueden ayudar a alcanzar una mayor escalabilidad de red. La inundación limitada, la inundación probabilística y la caminata aleatoria son ejemplos de mecanismos más económicos. En la inundación limitada las consultas están confinadas a ser reenviadas cierta cantidad de veces. En la inundación probabilística las consultas son reenviadas no por todo el vecindario, sino por un porcentaje de éste. Por último, la caminata aleatoria consiste en reenviar la consulta solo a un vecino, elegido aleatoriamente, en cada paso.

4.3 Sobre redes de grupos y localización híbrida

Las redes de grupos tienen el objetivo de aprovechar las ventajas que poseen los sistemas anteriores. Para ello combinan la topología estrella de las redes centralizadas con el grafo aleatorio de las redes puras mediante el siguiente criterio. Los pares son clasificados en dos tipos: pares hoja y superpares. Los primeros se conectan a un superpar formando una red centralizada. Los superpares se conectan entre sí, siguiendo el criterio de las redes puras, pero con la función añadida de ser servidores de localización. Para ilustrar la topología de esta categoría, en la figura 6 se muestra una red formada por 3 superpares, donde cada uno de estos administra un grupo de 5 pares hoja.

En las redes P2P centralizadas, los servidores de localización son máquinas muy potentes que son capaces de atender a un número grande de usuarios, entonces ¿por qué se puede crear una red centralizada empleando un dispositivo común? Considerando que la red puede estar compuesta por una gran variedad de dispositivos, es fácil observar que de entre éstos el hardware actual de muchas PC tiene capacidades que pueden ser usadas ventajosamente por los sistemas P2P. Por ejemplo, una PC de gama media puede realizar miles de millones de operaciones por segundo y tener acceso a una conexión a Internet de un ancho de banda de aproximadamente 10 Mbps². Por tal motivo, existe la posibilidad de que un par con tales características sea capaz de atender a un grupo de pares.

Con base en la observación anterior, para decidir qué pares son pares hoja y cuáles son superpares, tra-

²Megabaudios por segundo

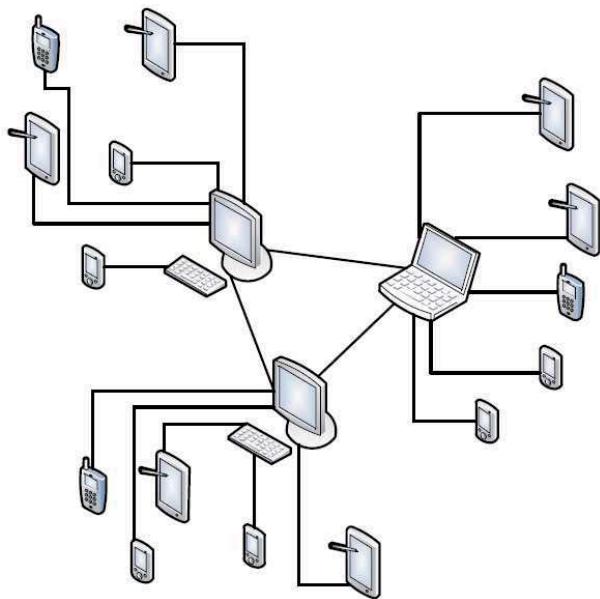


Figura 6. Red de grupos.

dicionalmente se decide que la máquina con hardware más poderoso es la que se convertirá en superpar. Cabe recalcar que la elección del superpar es dinámica ya que los superparas pueden ser reemplazados si es que se vuelven ausentes o si se conecta una máquina con mayor hardware al grupo. La información sobre el hardware de los pares puede ser solicitada a los usuarios o puede ser obtenida mediante el perfil del hardware que algunos sistemas operativos crean.

Respecto a la formación de grupos, existen diversos criterios que pueden ser empleados. Una propuesta común es agrupar a los pares hojas cuya dirección sea más próxima. Otra manera es hacer grupos de pares hojas cuya distancia geográfica sea mínima. Una propuesta más sofisticada es formar grupos con intereses similares, por ejemplo, grupos que comparten un interés literario. Este último tipo de agrupamiento crea una red P2P semántica.

En las redes P2P de grupos, el servicio de localización es un híbrido, ya que emplea en primera instancia localización centralizada y después inundación. Si un par hoja hace una consulta, entonces se dirige a su superpar para que éste se la resuelva. El superpar debe buscar en su índice lo solicitado y si no lo encuentra inundará la red de superparas. Una vez encontrado el contenido, se le envía directamente al par solicitante la dirección

del par que posee lo consultado. Cuando los superparas solicitan contenido, en primera instancia usan el índice que almacenan y si la consulta local no es suficiente, inundan con su solicitud la red de superparas.

Gracias a la diferencia jerárquica entre pares hoja y superparas, se obtienen las siguientes mejoras respecto a los sistemas P2P anteriores:

- La saturación de la red se reduce porque los mecanismos de localización por inundación se usan solo entre superparas, así el sistema puede atender a un mayor número de usuarios.
- La latencia en la resolución de las consultas es menor que en las redes P2P puras.
- La vulnerabilidad de las redes P2P centralizadas se minimiza ya que son varios los superparas que ofrecen el índice centralizado correspondiente a su grupo.
- Aumenta la probabilidad de encontrar contenido raro porque los superparas almacenan el índice del contenido que los pares hoja alojan.

A pesar de sus ventajas, las redes de grupos presentan ciertos inconvenientes. Ya que la red de superparas realiza la localización de contenido mediante inundación, es posible que se pueda saturar. Una solución a este problema es agregar una tercera jerarquía, los ultrapares (pares encargados de resolver las consultas de los superparas), pero se debe tener en cuenta que la complejidad de la red aumenta y con ello se dificulta su mantenimiento. Otra cuestión que debe de cuidarse es el tamaño de los grupos, ya que solo son administrados por un superpar. A pesar de que los superparas son nodos que tiene una mayor capacidad de cómputo respecto con los pares hoja, distan mucho de ser tan poderosos como un servidor. Es por eso que existe la probabilidad de que un superpar se congestione si tiene que administrar grupos grandes. Para que los grupos funcionen correctamente, se recomienda que no contengan más de 50 pares.

El ir y venir de los pares afecta de mayor manera a las redes P2P de grupos que a las puras, ya que el funcionamiento de los grupos depende fuertemente de los superparas y de su tiempo de sesión. Si los superparas tienden a abandonar abruptamente la red, entonces el desempeño del sistema se degradará porque se debe de elegir de nueva cuenta a

un superpar o porque los pares hojas quedaron aislados. Es por ello que en las redes P2P de grupos es necesario que la elección de superpar también tenga en cuenta, además de la capacidad de cómputo, el tiempo de sesión de los pares. Además, se tiene el mismo problema que en las redes P2P centralizadas porque los superpares tienen que actualizar frecuentemente su índice debido a la transitoriedad de los pares hojas.

4.4 Sobre redes estructuradas y localización por indexado distribuido

A comienzos de este siglo, la aplicación Napster elevó la popularidad de los sistemas P2P y con ello también aumentó el interés de la comunidad científica para mejorarlos. La propuesta que presentaron los investigadores fueron las redes P2P estructuradas. En éstas, el gran índice del contenido, que en los sistemas P2P centralizados está alojado en un servidor, es distribuido y administrado por todos los pares que integran el sistema. En contraste con las tres redes presentadas anteriormente, las redes estructuradas tienen una altísima escalabilidad, ya que evitan la inundación y son puras. Además, la topología de esta categoría está relacionada con el contenido que se comparte.

Las redes estructuradas emplean mecanismos basados en indirección por medio de *indexado distribuido*. Básicamente, el indexado distribuido utiliza una estructura de datos, por excelencia una Tabla Hash Distribuida (THD)³. Ésta es soportada por todos los integrantes del sistema y permite organizar y administrar a todos los pares y el contenido que comparten. Del uso de la estructura de datos distribuida viene el nombre de red P2P estructurada.

El funcionamiento de las THD es más fácil de entender si se comprende primero el funcionamiento de Tablas Hash Centralizadas (THC). Para ello usaremos la analogía del servicio de paquetería empleado en los centros comerciales. Básicamente, cuando queremos guardar un objeto se lo damos al empleado y él nos regresa una ficha que contiene el número de la casilla donde se guardará nuestro encargo. Cabe resaltar que no sabemos en qué casilla se guardará nuestro objeto y podemos suponer que su elección es aleatoria. Después, cuando pasamos a recoger nuestro objeto, solo le damos la ficha al empleado y él nos lo regresará directamente, sin buscar en todas las casillas del estante.

³Del inglés ‘*Distributed Hash Table*’ (DHT).

En la figura 7 mostramos una THC compuesta por la clave que identifica a un objeto y el objeto correspondiente. En las THC, la encargada de darnos una ficha aleatoria (que propiamente tiene nombre de clave) es la función *hash*. En nuestro ejemplo, a ‘rock.mp3’ le corresponde la clave hash ocho. Al igual que la paquetería, las THC son muy eficientes porque permiten encontrar directamente algún objeto sin buscar en toda la tabla, lo único que se necesita es calcular la clave *hash* de dicho objeto para saber en qué casilla de la tabla se encuentra.

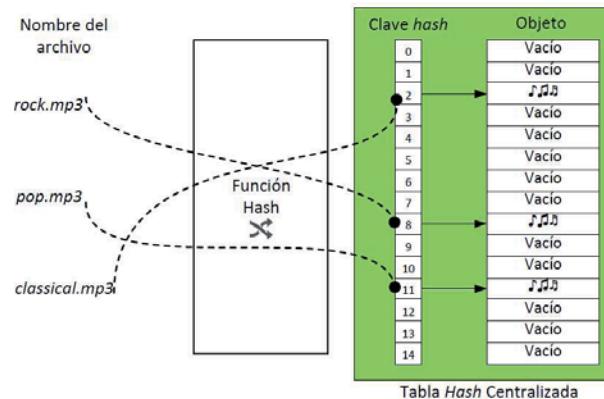


Figura 7. Tabla Hash Centralizada.

Las THD funcionan de manera similar, solo que la tabla es dividida en partes que son almacenadas y gestionadas por todos los pares que conforman la red. Además, las consultas por claves son resueltas por un protocolo de encaminamiento que dirige la consulta desde un par fuente hasta un par destino, a través de pares intermedios. Para exemplificar, en la figura 8 se muestra una THD en forma de anillo, compuesta por diez pares. Cada par almacena 1000 claves que están en el rango $[ID - 999, ID]$, donde ID es el identificador del par. Por simplicidad, elegimos múltiplos de mil para identificar a los pares. Las consultas por las claves se reexpiden en sentido horario, de par en par.

Apoyándonos en la figura 8, a continuación mostramos las características generales que poseen las THD.

- Indirección. En las THD no se almacena el contenido, sino la dirección del par que lo posee.
- Espacio de claves. Es el conjunto de claves que componen la THD. En la figura 8, el conjunto tiene 10000 claves. En sistemas reales, el espacio de claves suele tener varios millones de elementos.

- Uso de las claves. En las THD, las claves se utilizan para identificar tanto a los pares como al contenido y son obtenidas por medio de una función *hash* criptográfica.
- Secciones. En nuestro ejemplo, las secciones de la tabla son de tamaño fijo y cada una de ellas es de tamaño 1000. Sin embargo, en las THD las secciones son de tamaño variable, pudiendo tener en promedio N/P entradas, donde N es el número de claves y P el de pares.
- Responsabilidad por las claves. Regularmente, la asignación de las secciones de la tabla a los pares se basa en algún criterio de proximidad entre las claves de la sección y la clave del par.
- Topología. Existen muchas maneras de conectar las secciones de la THD. En nuestro ejemplo usamos una topología tipo anillo en el que las secciones se encuentran *conectadas* con su sucesor. Los investigadores han propuesto una amplia gama de topologías que tienen interesantes propiedades matemáticas y que sirven para poder resolver las consultas rápidamente.
- Resolución de las consultas. Las THD emplean un protocolo de encaminamiento que se encarga de dirigir las consultas. En la THD de la figura 8, si el par 1000 realiza una consulta por la clave 4200, ésta debe ser reenviada a los pares 2000, 3000 y 4000 hasta alcanzar a 5000. El mecanismo de encaminamiento que mostramos es ineficiente porque en el peor de los casos las consultas requieren N reenvíos, donde N es el número de pares en la red. Empero, las THD propuestas por la comunidad científica pueden resolver las consultas muy eficientemente, en aproximadamente $\log N$ reenvíos.

Las THD tienen propiedades interesantes, pero su funcionamiento depende de la topología de la red, por lo que se vuelven susceptibles a la transitoriedad. En nuestro ejemplo, la tabla está organizada en forma de anillo, cada par está conectado con su sucesor y cada par almacena y gestiona 1000 claves. Entonces, si un par sale del sistema el anillo se romperá y el protocolo de encaminamiento fallará en ciertos casos. Además se perderán todas las direcciones del contenido almacenado por el par que se retiró.

Para hacer frente a la transitoriedad, las THD emplean redundancia y un protocolo de autoorganiza-

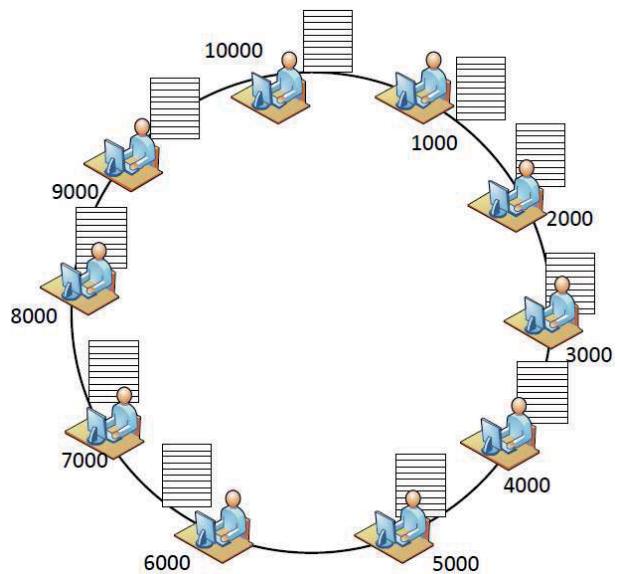


Figura 8. Tabla Hash Distribuida.

ción. La redundancia corresponde a que los pares almacenan posibles reemplazos para sustituir a pares que se han marchado. En nuestro ejemplo, los pares podrían almacenar los dos sucesores inmediatos de acuerdo a su identificador. Respecto al protocolo de autoorganización, éste se encarga de verificar las conexiones entre pares. El protocolo de autoorganización debe ajustar la topología de la red ante las entradas y las salidas de los pares. Para ello debe crear enlaces para los pares nuevos y debe eliminar y reemplazar (siempre que la redundancia lo permita) los enlaces de los pares que han salido del sistema. La redundancia y el protocolo de autoorganización tienen parámetros ajustables y su configuración debe estar en función del nivel de transitoriedad. No obstante, si estos mecanismos no se ajustan adecuadamente, es probable que la THD no soporte el nivel de transitoriedad y se rompa.

5. Sobre lo que son y serán de los sistemas P2P

A lo largo de este artículo conocimos qué son los sistemas P2P y cómo se encuentra el contenido en éstos. En primer lugar mencionamos que el contenido se puede obtener mediante dos modelos, el CS y el P2P. Después, mencionamos que los sistemas P2P están compuestos por entidades autónomas, llamadas pares, que solicitan y comparten toda clase de contenido. Además mencionamos que los sistemas P2P son escalables y robustos porque el contenido se encuentra distribuido entre todos los pa-

Tabla 1. Comparación entre sistemas P2P según el tipo de red.

Sistema P2P	Mecanismo de localización	Localización de contenido raro	Escalabilidad*	Robustez
centralizado	indexado centralizado	SI	alta (limitada por las capacidades del servidor)	baja (el sistema posee un punto único de falla)
no estructurado	inundación	poco probable y muy costoso	alta (limitada por el mecanismo de inundación)	alta
de grupos	híbrida	SI	muy alta	media (la transitoriedad de los superpares degradada el desempeño)
estructurado	indexado distribuido	SI	altísima	media (la THD tiene un límite en su resistencia a la transitoriedad)

*Escalabilidad comparada con el modelo CS.

res que conforma el sistema. Posteriormente mencionamos que existe un fenómeno que ocurre naturalmente en los sistemas P2P y se conoce como transitoriedad. El fenómeno de la transitoriedad es provocado por la autonomía de los pares y debido a éste se modifica la topología de la red, lo que implica que la localización del contenido no sea trivial. Después mostramos el funcionamiento de las diferentes categorías de los sistemas P2P, de acuerdo a la conexión entre los pares. Especialmente pusimos énfasis en cómo se localiza el contenido en los sistemas P2P centralizados, no estructurados, híbridos y estructurados, y de qué manera son afectados por la transitoriedad. En la tabla 1 resumimos las características de los sistemas analizados.

Respecto a la investigación relativa a los sistemas P2P, hay mucho trabajo por hacer. Es popularmente conocido que redes sociales como Facebook o Twitter tienen una gran cantidad de usuarios que pasan un tiempo considerable utilizándolas. Todos los recursos de los usuarios (memoria, ancho de banda, poder de procesamiento, etcétera) pueden servir de base para generar una red P2P que permita compartir contenido multimedia, comunicaciones interpersonales (llamadas o video-llamadas) o videojuegos masivos en línea (*Massively Multiplayer Online*, MMO) de forma más eficiente y económica que usando el modelo CS. Además, de la información de la red social se puede saber cómo se comportan los usuarios y con esta información aumentar el desempeño de los sistemas P2P [5].

Actualmente se encuentra en desarrollo el protocolo P2P-SIP⁴ (*P2P - Session Initiation Protocol, P2P-SIP*) que busca poder realizar llamadas telefónicas

sobre un sistema P2P. La idea de este protocolo es emplear la red P2P como infraestructura de comunicación. El protocolo está pensado para funcionar sobre una red P2P estructurada [6].

Otra idea en desarrollo, son las redes formadas únicamente por dispositivos inalámbricos como sensores, subportátiles⁵, tabletas y teléfonos inteligentes. Estas redes, llamadas redes móviles de iguales, son temporales, pueden carecer de infraestructura y pueden tener topología física altamente dinámica. Las ideas relativas a la localización en sistemas P2P pueden ser aplicadas en este tipo de redes inalámbricas, aunque no de manera directa [4].

Bibliografía

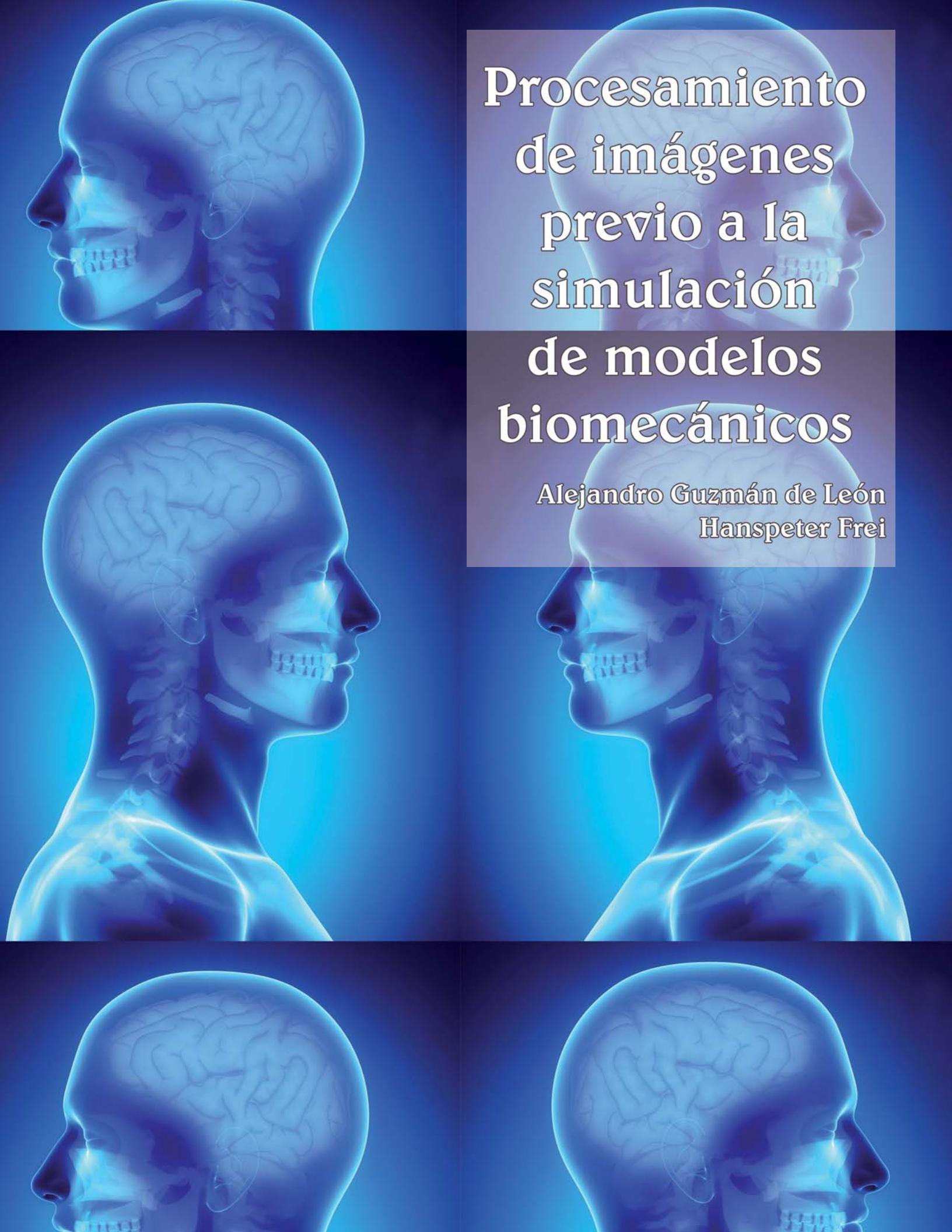
1. “Napster use slumps 65 percent.”, 2001. <http://news.bbc.co.uk/2/hi/business/1449127.stm>.
2. “uTorrent and BitTorrent Hit 100 Million Monthly Users.”, 2011. <http://goo.gl/c2o6L>.
3. Ahmed, Reaz, y Raouf Boutaba. “A Survey of Distributed Search Techniques in Large Scale Distributed Systems”. *IEEE Communications Surveys and Tutorials*, 13, no 2, (2011), 150–167.
4. Castro, Marcel C., Andreas J. Kassler, Carla-Fabiana Chiasserini, Claudio Caselli, y Ibrahim Korpeoglu. ((Peer-to-Peer Overlay in Mobile Ad-hoc Networks.)) En *Handbook of Peer-to-Peer Networking* (Xuemin Shen, Heather Yu, John Buffalo, y Mursalin Akon, eds.), Springer US, Boston, MA, 2010. 1045–1080.
5. Chen, Kang, Haiying Shen, y Haibo Zhang. “Leveraging Social Networks for P2P Content-Based File Sharing in Mobile Ad Hoc Networks.” *Mobi-*

⁴<http://www.p2psip.org/>

⁵Netbooks.

- le Ad-Hoc and Sensor Systems, IEEE International Conference on*, 0, (2011), 112–121.
6. Harjula, Erkki, Timo Koskela, y Mika Ylianti-lla. “Comparing the performance and efficiency of two popular DHTs in interpersonal communication.” En *WCNC. 2011*, 2173–2178.
 7. Yianilos, Peter N., y Sumeet Sobti. “The Evolving Field of Distributed Storage.” *IEEE Internet Computing*, 5, (2001), 35–39. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=957893.

cs



Procesamiento de imágenes previo a la simulación de modelos biomecánicos

Alejandro Guzmán de León
Hanspeter Frei

Recibido: 18 de enero de 2012.

Aceptado: 12 de marzo de 2012.

Abstract

The aim of this paper is to give readers an overview of the elements required for a patient specific biomechanical analysis. Special attention is given to the preprocessing stage of the analysis, which requires the use of volumetric digital image segmentation and meshing.

Initially general concepts are defined regarding the requirements of mathematical modeling used in mechanical engineering. The finite element method is a popular and powerful computational tool used to analyze complex mechanical structures. It requires to mesh the 3D structure into small finite elements.

These concepts are extended to the biomedical engineering field with the aid of medical imaging, digital image processing and three-dimensional visualization. Different approaches to incorporate these concepts for specific applications of interest to the authors are described and discussed. Here computed tomography as imaging modality is used to obtain the images. Then the deformable models as three-dimensional segmentation technique in their versions of the gradient vector flow (GVF) and the vector field convolution (VFC) are applied at the same time that a surface mesh with unstructured elements type triangle is obtained. These techniques are exemplified with two organs a femur and an aorta. The article concludes with a model based on micro-CT scan of human biopsies. The complex three-dimensional trabeculae structure was meshed with tetrahedron elements probing the feasibility of these concepts. The three examples provided structural information required for biomechanical finite element analysis.

Resumen

El objetivo del artículo es dar al lector una visión general de los elementos requeridos para un planteamiento de un análisis biomecánico. Se hace especial énfasis en la etapa de procesamiento de la estructura a analizar la cual requiere de la utilización de métodos de segmentación digital de imágenes volumétricas y mallado.

Inicialmente se definen conceptos generales sobre los modelos matemáticos usados en ingeniería mecánica. Dentro de las herramientas computacionales que ayudan a resolver el planteamiento dinámico de una estructura mallada se habla sobre el método numérico llamado elementos finitos.

La idea se extiende al campo de la ingeniería biomédica gracias a la inclusión de la imagenología médica, el procesamiento digital de imágenes y la visualización tridimensional. La presentación continúa al particularizar los conceptos en función de aplicaciones específicas del interés de los autores. Así se exponen: la tomografía computarizada como modalidad de imagenología utilizada y los modelos deformables como técnica de segmentación tridimensional en sus versiones flujo del vector gradiente (GVF) y convolución del campo vectorial (VFC) que incluyen un mallado superficial con elementos no estructurados tipo triángulo. Con esta última técnica se exemplifica la obtención del modelo biomecánico de un fémur y una aorta. El artículo termina con otro ejemplo que modela una muestra de trabécula con elementos tipo tetraedro en el mallado volumétrico. Los tres ejemplos aportan la información estructural necesaria para un análisis biomecánico con elementos finitos.

Palabras clave: Segmentación de imágenes volumétricas, Mallado, Modelos biomecánicos, Modelos deformables, Tomografía computarizada, Análisis por elementos finitos.

1. Introducción

Los modelos matemáticos y las simulaciones computacionales han sido muy exitosas en la ingeniería mecánica. Los modelos matemáticos son descritos con ecuaciones y fórmulas que son resueltas numéricamente con la computadora. Un método popular en ingeniería es el método por elementos finitos.

Actualmente estructuras biológicas son también modeladas matemáticamente usando la misma metodología para entender su comportamiento biomecánico.

Las modalidades de imagenología médica permiten a los ingenieros obtener las características geométricas de la estructura biológica estudiada, usando para ello técnicas de procesamiento digital de imágenes y visualización de imágenes tridimensional donde los modelos deformables son aplicados segmentando y mallando la región u órgano de interés.

Este artículo presenta los conceptos y la metodología requerida para aislar el órgano y mallarlo, etapa de suma importancia, previa al análisis biomecánico de la estructura y muestra algunos casos particulares

que fueron obtenidos usando la modalidad de tomografía computarizada.

Los tres primeros apartados: modelo matemático y simulación computacional (2.1), método de elementos finitos (2.2) y mallado (2.3). Son incluidos para poner en contexto al lector. Esos tres apartados junto con la visualización (2.6), son las etapas necesarias para cualquier modelado, diseño y análisis realizados en ingeniería mecánica. Para darle el enfoque biológico es necesario hablar de la imagenología médica (2.4), saber que hay diferentes equipos y que hay que saber escoger el adecuado en función la estructura biológica. A diferencia de los casos de ingeniería mecánica, en biomecánica se tiene que preprocessar la información que la imagen aporta mejorando su calidad, reconociendo y aislando la información de la estructura biológica a estudiar, para eso se habla del procesamiento digital de imágenes (2.5), en específico sobre la segmentación, que intenta automatizar este proceso que es tedioso y subjetivo para el usuario. La visualización volumétrica (2.6) se presenta al final de la sección de conceptos generales, considerando que cualquier análisis sobre los modelos mecánicos o biomecánicos son presentados gráficamente en una pantalla de computadora y que son objetos tridimensionales a los cuales se les debe de dar sombreado y perspectiva para que el usuario pueda tener una idea clara de lo sucedido en el proceso y visualizarlo desde diferentes ángulos al inicio y al final del análisis biomecánico.

En la sección 3 conceptos particulares, se habla sobre la tomografía computarizada (3.1) como modalidad de imagenología escogida para adquirir las imágenes de los ejemplos presentados al final del artículo y dos métodos de segmentación y mallado basados en modelos deformables 3D (3.2), flujo del vector gradiente (GVF) (3.3) y convolución del campo vectorial (VFC) (3.4) que presentan interesantes cualidades y son plataformas para futuros desarrollos.

Se presentan tres ejemplos de aplicaciones (4) que después de la segmentación y el mallado aportan los datos necesarios para el análisis biomecánico con elementos finitos. El artículo termina con una discusión (5) y la bibliografía (6).

2 Conceptos Generales

2.1 Modelos Matemáticos y Simulación Computacional

Un modelo es una simplificación de un objeto o fenómeno real, considera lo más relevante de él y puede simular o emular su forma o funcionamiento. Por ejemplo en Ingeniería, un modelo matemático representa mediante ecuaciones y fórmulas un comportamiento o una estructura, toma en cuenta los parámetros fundamentales o sea aquellos parámetros que lo definen, caracterizan o diferencian de otros modelos. Aprovecha el hecho de que estos parámetros pueden ser modificados con facilidad. Los cálculos son procesados por una computadora cuando se tienen muchas ecuaciones o se necesita de métodos numéricos para encontrar la solución. A esto se le llama simulación computacional y puede ser tan complicada para programar como la precisión que se quiera obtener, así, más precisión requiere de más parámetros o más ecuaciones que complican el cálculo de la solución y aumenta el tiempo de procesamiento dentro de la computadora. El simular mediante modelos permite analizar el fenómeno o estructura, reconociéndolo como señal o sistema. La simulación se puede repetir cuantas veces sea necesario y pueden cambiar las condiciones de uso mediante la modificación de los parámetros. La simulación computacional permite reducir costos de experimentación, hacer experimentos no destructivos o hacer experimentos que no sería posible realizar en la realidad permitiendo tener un entendimiento más particular y completo de la señal o sistema. Los resultados así obtenidos pueden ser generalizados para otros casos. Algunos ejemplos que permiten entender la utilidad de la simulación computacional con modelos matemáticos pueden ser el análisis de un reactor nuclear donde en la realidad no se podría estar experimentando con los parámetros por que un accidente sería muy costoso y los efectos serían peligrosos. Mediante la simulación computacional se pueden prever acontecimientos astronómicos donde en la realidad las distancias y los tiempos no son accesibles en la vida de un ser humano.

El planteamiento de las ecuaciones y fórmulas requieren de un entendimiento de la señal o el sistema que se analiza y mediante la experimentación se logra refinar el planteamiento. El planteamiento inicial puede provenir de teorías o de la experiencia y observación de la realidad. Las ecuaciones así planteadas generan una sencilla clasificación de modelos matemáticos, lineales y no lineales. Los pri-

meros al verlos como un sistema al cual le aplicamos una señal de entrada responde con una señal de salida, pero si a la señal de entrada se incrementa al doble o triple de su valor, la salida también se incrementará en el doble o triple de su valor. Los modelos matemáticos no lineales, no cumplen con esta propiedad y al incrementar su valor de entrada, la salida no será incrementada en la misma proporción.

2.2 Método de Elementos Finitos

Muchos sistemas que se analizan mediante modelos matemáticos pueden ser muy complicados en su aspecto estructural y las ecuaciones pueden no reflejar correctamente su comportamiento o son difíciles de resolver. Por otro lado, una excesiva simplificación del modelo puede no validar el caso real. Entonces, para analizar estos sistemas se ideó dividir la estructura en partes o elementos. De cada elemento se conoce su forma, fronteras, interacciones y propiedades físicas. Se pueden así plantear las ecuaciones para este elemento y conocer la respuesta a una excitación particular. Si esta respuesta es considerada como entrada del elemento vecino que posee su propio planteamiento, entonces se tiene un comportamiento encadenado de los elementos. Cuando se encadenan todos los elementos de la estructura inicial se obtiene el comportamiento global de la estructura y particular para cada elemento. Este método es conocido como método de elementos finitos FEM, de sus siglas en inglés *Finite Element Method*.

A pesar de que se tiene un gran número de ecuaciones, el modelo matemático así planteado es numéricamente resuelto dentro de la computadora con gran eficiencia por que las ecuaciones son más sencillas. La precisión del resultado tendrá ahora que ver con la validación del planteamiento de las ecuaciones para cada elemento y de la cantidad de elementos que se tenga representando a la estructura. Para estructuras cuyas propiedades sean homogéneas en todas direcciones (isotrópicas) las ecuaciones planteadas serán muy similares. En el caso ideal, la estructura estaría constituida por un número infinito de elementos, lo cual implica una infinidad de ecuaciones a procesar dentro de la computadora y por lo tanto irrealizable, entonces lo que se hace es; tener un número finito de elementos y buscar el resultado en una primera iteración, modificar el número de elementos y buscar de nuevo el resultado, seguir iterando de esta forma hasta que el resultado cumpla con un criterio de convergencia.

Normalmente, la solución numérica de un problema mediante elementos finitos, requiere de tres etapas, el preprocessamiento, el procesamiento y el posprocesamiento.

En el preprocessamiento se definen:

- La geometría de la estructura.
- Los grados de libertad.
- Las propiedades del material.
- El malla.

Este último consiste en la división de la estructura en elementos (número, tipo, tamaño y posición).

En el procesamiento se definen:

- Los parámetros de interés.
- Las ecuaciones y fórmulas del sistema.
- Las restricciones e interacciones con otros sistemas llamadas condiciones de frontera y condiciones de interface respectivamente.
- Las condiciones iniciales del sistema.
- El tipo de análisis estático o dinámico.
- Las restricciones para los parámetros y puntos de referencia o prueba.
- Se definen las señales de entrada, su localización, en qué elementos se quiere aplicar, su valor y dónde se quiere tener la salida particular o general.
- Se resuelve el problema y almacena el resultado.

En el posprocesamiento se pueden analizar los resultados, consiguiendo una cuantificación de los parámetros de interés en tablas. Su evolución temporal o espacial es visualizada en forma gráfica.

2.3 Malla

La Ingeniería Mecánica ha usado los elementos finitos para analizar las cargas, esfuerzos y deformaciones de estructuras mecánicas, así como las propiedades del material del cuál esta compuesto. Ha permitido analizar y diseñar partes de sistemas mecánicos de manera muy eficiente, no importando la complejidad geométrica de la estructura.

Para tener un buen resultado del análisis, se cuenta con una base del conocimiento y teorías bien establecidas que validan el planteamiento de los modelos matemáticos. Dependiendo de la aplicación se propone un tipo de malla. Una clasificación de tipos de malla alberga el malla estructurado y al no estructurado, el primero mapea la malla curvilínea del espacio físico a una malla en el espacio uniformemente cartesiano computacional (figura 1), esto permite identificar a los vecinos de un

nodo (arista de un elemento) y tener un fácil acceso a ellos.

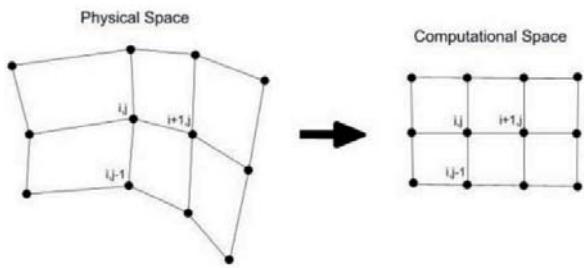


Figura 1. Mallado estructurado. Mapeo del dominio físico al dominio computacional.

La desventaja en este tipo de mallado se encuentra cuando se quiere refinar el mallado, no se puede hacer de forma local, tiene que ser global, por ejemplo si en la figura 1 se quisiera incrementar la resolución vertical del quinto elemento formado por los nodos (i, j) , $(i+1, j)$, $(i, j+1)$, $(i+1, j+1)$, se tendría que agregar y renombrar todo un renglón de nodos entre j y $j+1$, no solo un par de ellos en medio del elemento por que esto no tendría un mapeo discreto en la computadora.

El mallado no estructurado no posee este mapeo en la computadora, lo que lo hace más versátil en el remallado y la forma de los elementos. El costo computacional se refleja en un listado del vecindario y su necesidad de reajuste en cada refinamiento, suavizado o engrosamiento del mallado [innovative-cdf, 2011].

Entre las formas de elementos más usadas en superficies están los cuadriláteros (Quad) y los triángulos (Tri), para mallado volumétrico se tienen los hexaedros (Hex: seis caras cuadriláteras y ocho nodos), los tetraedros (Tet: cuatro caras triangulares y cuatro nodos) y prismas (cinco caras, tres cuadriláteras y dos triangulares con seis nodos) figura 2.

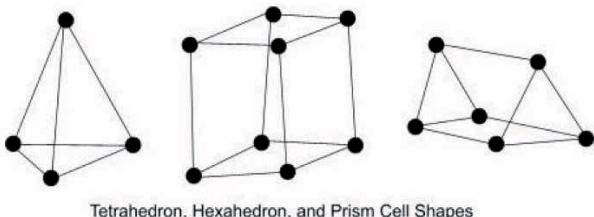


Figura 2. Formas de elementos volumétricos, Tet, Hex y prisma.

Cada arista de estos elementos tiene un nodo, la precisión del modelo puede ser incrementada al tener nodos intermedios en los flancos e incluso para el caso de volúmenes se pueden tener nodos centrales en las caras de la figura geométrica.

La elección del tipo del elemento dependerá de la aplicación y tiene que ser definido en los parámetros del preprocesamiento del método de elementos finitos. En algunos casos como el presentado por Longest *et al.* (Longest, 2007), se ha demostrado que los elementos estructurados, son mejores que los no estructurados por ser más precisos usando menos elementos y aunque su definición requiere de mayor esfuerzo y tiempo de implementación, el resultado en el tiempo de la ejecución del análisis completo se reduce en un 60 % respecto a los no estructurados.

2.4 Imagenología Médica

A diferencia de los casos de ingeniería mecánica donde el objeto de estudio se puede palpar o se tienen las especificaciones del diseño con todas las características del material y dimensiones de la estructura, como por ejemplo un puente, en el caso biomecánico las estructuras no son palpables y son solo observadas mediante la imagenología médica si se habla de casos *in vivo*, por ejemplo el fémur o la aorta.

La imagenología médica ha sido utilizada por los médicos para apoyar el diagnóstico ya que ésta permite visualizar estructuras anatómicas o comportamientos funcionales en un individuo, sin necesidad de invadirlo o intervenirlo quirúrgicamente. Algunas patologías requieren de la imagenología para dar seguimiento a la evolución del padecimiento bajo tratamiento o para hacer la planeación de una intervención quirúrgica o restauración de algún órgano con propósito no solo de salud sino también estético. De igual forma en la docencia clínica, la imagenología médica ha apoyado fuertemente en el análisis de casos, ayudando a los estudiantes a identificar diferentes patologías. En biomecánica la imagenología médica aporta la información de la posición, forma y dimensiones de la estructura a analizar y en algunos casos se logra correlacionar alguna propiedad del material con la intensidad de la imagen, por ejemplo la densidad ósea.

Las imágenes son formadas cuando algún tipo de energía interactúa con el cuerpo humano. Las energías más usadas son del tipo electromagnético, eléctrico, magnético, mecánico y nuclear. Las interacciones de estas energías con la ma-

teria pueden ser por ejemplo, transmisión, absorción, atenuación, reflexión, dispersión, relajación y aniquilación, entre otras.

Los diferentes equipos de imagenología médica llamados también modalidades de imagenología, poseen diferentes detectores de energías que miden alguna o algunas de estas interacciones dando así datos. Estos datos pueden ser del tipo directo o indirecto y ambos después de un procesamiento permiten crear una imagen.

Aquellas imágenes obtenidas de datos directos son llamadas imágenes proyectivas y aquellas de datos indirectos son llamadas imágenes reconstructivas.

Las modalidades de imágenes más comunes son los rayos X convencional o digital, la tomografía computarizada, la imagenología por resonancia magnética, el ultrasonido y la medicina nuclear. De estos hay derivados y combinaciones que generan una lista de equipos de imagenología médica.

¿Para qué hay entonces tantas modalidades?, la respuesta es porque dependiendo del padecimiento o el órgano a estudiar hay una modalidad de imagenología que se adecúa mejor y hace resaltar el problema o las características estructurales o funcionales.

2.5 Procesamiento Digital de Imágenes

Una vez escogida la modalidad de imagenología adecuada y hecho el estudio, las imágenes adquiridas tienen mucha información que no es relevante. Para analizar el órgano de interés, es necesario delinear y aislarlo del resto del cuerpo, método que en procesamiento digital de imágenes es conocido como segmentación y que filtra la información útil.

Existen muchos métodos de segmentación de imágenes y es un campo de investigación en el que las personas de la ingeniería, las matemáticas y las ciencias de la computación trabajan arduamente porque no existe un método de segmentación universal. El método de segmentación aplicado en imagenología médica depende de la modalidad de imagen usada, del órgano a segmentar e incluso de la patología. En general, cuando se habla de segmentación en el dominio espacial bidimensional y se considera el nivel de gris de la imagen como parámetro de control, los métodos de segmentación se pueden clasificar en dos grupos, segmentación por regiones o segmentación por contornos. Si se habla de segmentación tridimensional una clasificación similar es

usada aunque con otros nombres; estructural y estocástica. Si la segmentación es estructural, se segmenta mediante métodos como: medición de variaciones de intensidad de un órgano a otro, derivadas parciales de orden uno llamadas gradientes y de orden dos llamadas laplaciadas, combinados con otros criterios. Para el caso de segmentación estocástica la base es la intensidad del elemento mínimo volumétrico llamado voxel y su vecindario, con criterios de homogeneidad lo que lleva a crear regiones. Una tercera clase de métodos de segmentación es la combinación de las dos anteriores llamada híbrida. Una buena revisión de los diferentes métodos de segmentación volumétrica se puede encontrar en (Lakare, 2000) y (Hu, 2009).

2.6 Visualización de Imágenes Volumétricas

La visualización de imágenes volumétricas (3D) es una etapa muy importante en el diagnóstico médico, debido a la gran cantidad de imágenes bidimensionales que se obtienen en un solo estudio, no es difícil que el médico pase por alto algunos detalles. Por lo tanto el médico se apoya de una visión general y después pasa a los detalles de las subregiones o a los diferentes cortes en donde se puede analizar el problema.

Actualmente la forma de presentar en pantalla las imágenes obtenidas es mediante un apilamiento de varias imágenes bidimensionales (2D) que representan cortes corporales, formando un volumen en tres dimensiones (3D). Aplicando técnicas de procesamiento digital de imágenes se segmenta y aísla el órgano deseado, después con técnicas de visualización de imágenes se le da perspectiva, profundidad y sombreado proceso conocido como *rendering*, cuya presentación facilita la interpretación (figura 3).

En biomecánica también se acostumbra visualizar completa la estructura del órgano bajo estudio ya que es más intuitivo suponer el tipo de interacciones con los demás órganos en comparación con solo partes o secciones del órgano. El análisis mecánico podría dar el mismo resultado analizando la estructura completa o analizando alguna fracción de ella, una vez que se ha entendido el problema y a la fracción se le apliquen condiciones iniciales y de frontera junto con los esfuerzos atribuidos a las otras fracciones. Esos modelos de estructuras fraccionadas se usan conforme se va adquiriendo más experiencia y conocimiento del problema analizado.

Por estas razones clínicas y biomecánicas la visualización 3D es importante y se tienen dos diferentes

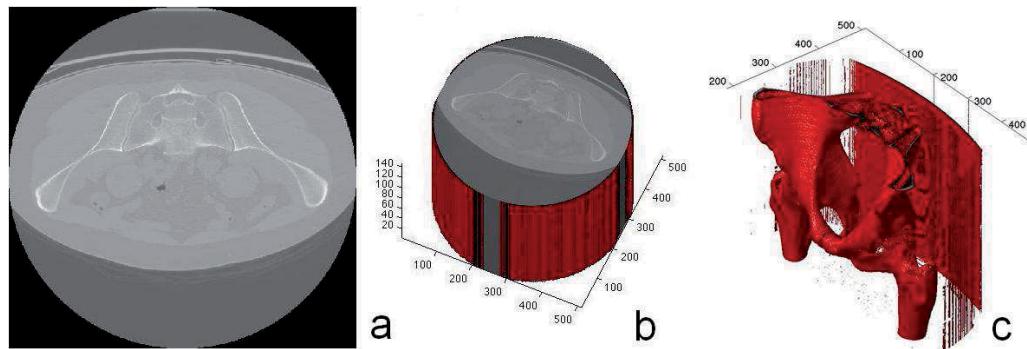


Figura 3. Estudio tomográfico de cadera. a) un corte axial 2D, b) volumen 3D, constituido de 147 cortes 2D apilados y c) segmentación y rendering 3D.

formas de visualización: el rendering y el volumétrico. En el rendering se tiene la visualización de una superficie del órgano, aislado previamente. En la visualización volumétrica se usan técnicas de transparencia que permiten visualizar al órgano deseado a través de los órganos circundantes dando una impresión de ser objetos transparentes, este último modo de visualización es el que se está desarrollando actualmente en la clínica.

Para el propósito de análisis biomecánico la visualización mediante rendering es la que se utiliza cotidianamente por aislar la estructura de interés, darle perspectiva, profundidad y sombreado. Tener acceso a rotaciones, translaciones y acercamientos de la estructura es parte del control de la visualización por rendering.

Dentro del rendering, podemos tener dos tipos de representación visual, un volumen, un sólido cuyo interior está lleno de elementos o una superficie, el cascarón del órgano. Estas representaciones visuales por rendering se relacionan con el tipo de elemento que se utiliza al hacer el mallado, usar tetraedros forman el volumen o usar triángulos forman la superficie, por ejemplo en la figura 4 se observa un cubo donde una cuarta parte de los elementos son visualizados con rendering por volumen o por superficie (Fang, 2009).

Para poder dar un sombreado se requiere de tres elementos una fuente luminosa, un objeto (órgano) que refleje la luz y el observador. Dentro de la visualización se emula la iluminación que será reflejada en cada cara de cada elemento que compone al mallado y esta reflexión será visualizada siguiendo la ley de reflexión especular.

3 Conceptos Particulares

3.1 Tomografía Computarizada

Una de las modalidades con más éxito en el análisis clínico es la tomografía computarizada. Utiliza la propiedad de atenuación de energía al interactuar con el cuerpo y presenta imágenes reconstructivas de alto contraste y resolución en dos y tres dimensiones como las mostradas en la figura 3. Para construir la imagen la tomografía maneja los datos de las proyecciones obtenidas a diferentes ángulos alrededor del sujeto quien es el eje de rotación del sistema tubo de radiación X y detectores de la atenuación radiológica. Estos datos indirectos son procesados en una computadora mediante algoritmos matemáticos, entre ellos la retroproyección filtrada, para reconstruir un corte axial corporal.

La energía que utiliza esta modalidad es la radiación electromagnética X, la cual es ionizante, es decir cuando la radiación X interactúa con cualquier tipo de materia, esta radiación libera electrones de los átomos que componen la materia. Al liberar estos electrones de átomos que por naturaleza son eléctricamente neutros se producen los iones, partícula con carga eléctrica, los iones negativos son los electrones liberados y el ion positivo el átomo del cual se desprendieron los electrones.

Cuando la radiación no es muy severa los átomos recuperan su neutralidad eléctrica en un corto tiempo, pero si la radiación es severa entonces se pueden producir distintos enlaces electroquímicos. En el caso de entes biológicos las células que componen al ente son dañadas por esta ionización severa. Por severa se debe entender como energía muy alta y/o mucho tiempo de exposición a la radiación ionizante.

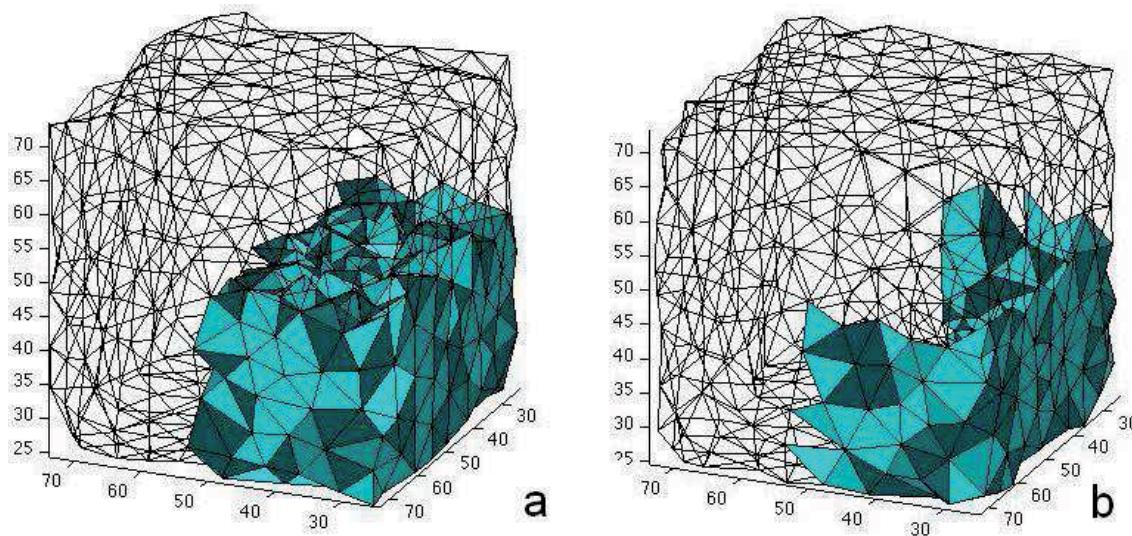


Figura 4. Mallado y rendering de un volumen cúbico. a) muestra una cuarta parte de los elementos tetraédricos, dando una representación por volumen y b) muestra una cuarta parte de los elementos triangulares, dando una representación por superficie.

Un médico es el responsable de evaluar la conveniencia de irradiar a una persona con este tipo de energías, tomando en cuenta el costo-beneficio que se tiene al tomar una tomografía y procurando siempre irradiar lo menos posible al paciente. El beneficio se valora con la rapidez y precisión con la que el médico puede hacer su diagnóstico a partir del estudio tomográfico. El costo y tiempo de un estudio tomográfico es intermedio, entre un estudio de rayos X convencional y una resonancia magnética. La elección de esta modalidad de imagenología por parte del médico radica en su inmejorable calidad de imagen para padecimientos del tejido óseo (huesos) y del tejido pulmonar.

3.2 Modelos Deformables

Dentro de los métodos de segmentación volumétricos clasificados como estructurales, encontramos los modelos deformables. Estos consisten en superficies volumétricas que son deformadas en la búsqueda de un equilibrio de fuerzas, la interna perteneciente a la superficie inicial y la externa perteneciente a la imagen tridimensional. La fuerza interna se ve representada por dos componentes; la fuerza de estiramiento y la fuerza de doblamiento, elasticidad y rigidez respectivamente de la estructura que constituye a la superficie deformable. La fuerza externa generalmente proviene del gradiente del contorno del órgano de interés, cuya función es atraer a la estructura deformable. Este tipo de modelado se funda-

menta en el conocimiento de la física; fuerzas, elasticidad, rigidez, esfuerzos, cargas y restricciones.

El problema se plantea mediante una función de costo. Esta representa energías del sistema y la idea es minimizarla. Las ecuaciones son planteadas aplicando los principios de la dinámica Lagrangiana los cuales simplifican la obtención de las ecuaciones de movimiento en sistemas complejos que incluyen varios grados de libertad (Kutz, 2009). El modelo deformable dinámico así construido da como resultado una interpretación a la forma y su deformación en el tiempo, o en el caso numérico para cada iteración, (McInerney, 2009) y (Hegadi, 2010) dan una excelente revisión de esta teoría.

En la etapa de segmentación, el mallado puede empezar con un modelo geométrico, por ejemplo una esfera, que se va ajustando dinámicamente a las fronteras del órgano, al final de esta etapa se puede extraer el órgano aislado y mallado. A partir de este punto se pasa a la etapa de análisis por elementos finitos que con el mallado proporciona los elementos para el análisis de esfuerzos dinámicos del modelo matemático en la simulación computacional. Este mallado sufrirá deformaciones si lo que se quiere ver es desplazamientos debido a cargas o buscar puntos de ruptura o de máximo esfuerzo y será visible gracias a la etapa de posprocesamiento donde la visualización es crucial en la interpretación de resultados.

Dos métodos de segmentación de imágenes basados en el principio de los modelos deformables son presentados a continuación. No son los únicos métodos ni los más potentes pero son métodos que cumplen con un compromiso entre el costo computacional y la calidad de la segmentación. Alrededor del primer método, llamado flujo del vector gradiente en inglés *Gradient Vector Flow* o simplemente GVF, se han generado variaciones y esa flexibilidad a sido uno de los principales intereses para usarlo. El segundo es una consecuencia del primero al intentar disminuir el tiempo de cálculo y ofrecer otras bondades, es conocido como convolución del campo vectorial o en inglés *Vector Field Convolution*, VFC.

3.3 GVF: Flujo del Vector Gradiente

Este método de segmentación basado en los modelos deformables (Xu, 1998) mejora a otros modelos deformables en dos aspectos importantes. El primero es que la superficie mallada es atraída a los bordes del órgano de interés desde distancias más lejanas y que esta distancia toma en cuenta la posición de los bordes vecinos lo que le da su segunda cualidad de segmentación de concavidades.

El método consiste en que se sintetiza una fuerza externa inicial proveniente del contorno de la imagen. Se genera un campo vectorial normalizado a partir del gradiente de este contorno. El campo vectorial se difundirá tan lejos como se desee generando así la fuerza de atracción que sentirá la superficie mallada del modelo geométrico inicial que puede ser por ejemplo una esfera. La superficie mallada inicial por su parte posee una fuerza interna constituida por dos términos, la tensión y la suavidad. La tensión protegerá a la malla de no romperse y la suavidad le permitirá doblarse. Así cuando las dos fuerzas, interna y externa, se equilibran se detiene el proceso de ajuste de la superficie mallada y normalmente esto ocurre justo en el borde del órgano de interés. Así al final se tiene la segmentación del órgano de interés junto con el mallado que se utilizará con los elementos finitos para su análisis biomecánico, (figura 5).

Aplicar el método de GVF implica difusión de un gradiente, y por difusión se puede entender como desparramar o extender el efecto. El algoritmo es iterativo es decir se repite varias veces según un criterio de parada que generalmente es un parámetro que se puede fijar con una constante. Un número común es 100 iteraciones. Entre más iteraciones más difundido el campo de vectores de gradiente, esto permi-

te tener más alcance en la atracción de la superficie mallada inicial que fue colocada en algún lugar que rodea o esta dentro del órgano de interés. El costo computacional aumenta conforme esta constante aumenta y es un precio que se tiene que pagar cuando la inicialización de la superficie mallada se encuentra muy lejos de los bordes del órgano de interés.

3.4 VFC: Convolución del Campo Vectorial

Investigaciones sobre el costo computacional del GVF llevaron a proponer una emulación de la creación del campo vectorial (Li, 2007). Considerando que el campo vectorial es una difusión del vector gradiente de los bordes deseados, lo que propusieron fue crear la difusión el campo vectorial no a partir de la difusión del gradiente si no a partir de una convolución de un kernel vectorial con el borde inicial del órgano de interés. La convolución es una operación matemática que involucra multiplicaciones, sumas y corrimientos. El kernel en este caso es un arreglo tridimensional cúbico dividido en celdas, en el cual se tienen vectores en cada celda y al ir barriendo este kernel dentro de todo el volumen en el cual esta inmerso el órgano de interés, se obtiene el campo vectorial.

Tres ventajas se obtienen a partir de este método de segmentación: primero que el tiempo de generación del campo vectorial se reduce sustancialmente por el tipo de operaciones realizadas; segunda que los vectores dentro del campo vectorial apuntan a los baricentros de la estructura mejorando la atracción a las concavidades o elongaciones y tercera que es un método menos sensible al ruido (figura 6). Este método es llamado convolución del campo vectorial o en inglés *Vector Field Convolution*, VFC.

4 Ejemplos de Aplicaciones

Para concretar lo dicho anteriormente, se muestran dos ejemplos de segmentación y mallado superficial con el método VFC en la cabeza de un fémur y en una aorta, imágenes que fueron obtenidas mediante tomografía computarizada. Un tercer ejemplo expone un mallado volumétrico aplicado sobre una muestra de trabécula.

4.1 Fémur

El fémur es un hueso que es visible perfectamente con rayos X usando la tomografía computarizada, generalmente se pueden utilizar métodos de segmentación básicos para segmentar en dos dimensiones pero si el propósito es segmentar con un mínimo de interacción con el usuario en tres dimensiones

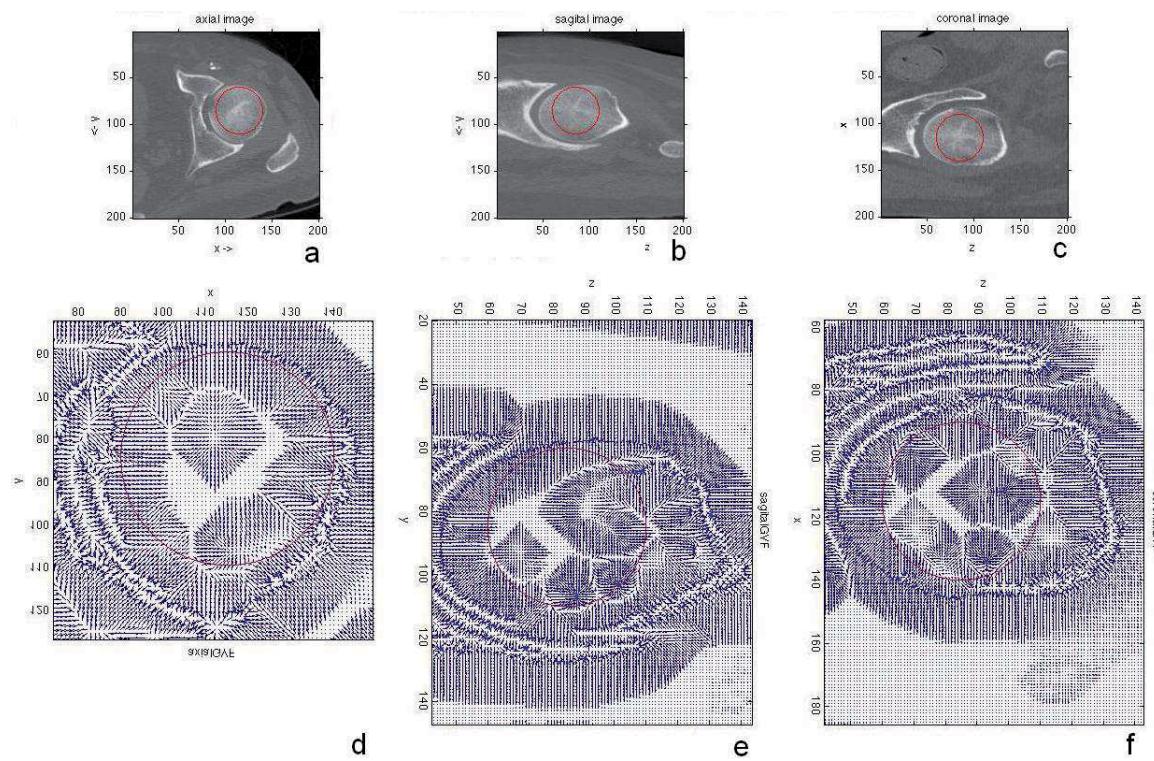


Figura 5. Cortes tomográficos en los planos a) axial, b) sagital y c) coronal de la cabeza de fémur y d-f) sus correspondientes campos vectoriales producidos por el método de GVF-3D. El círculo rojo representa la posición de la esfera inicial.

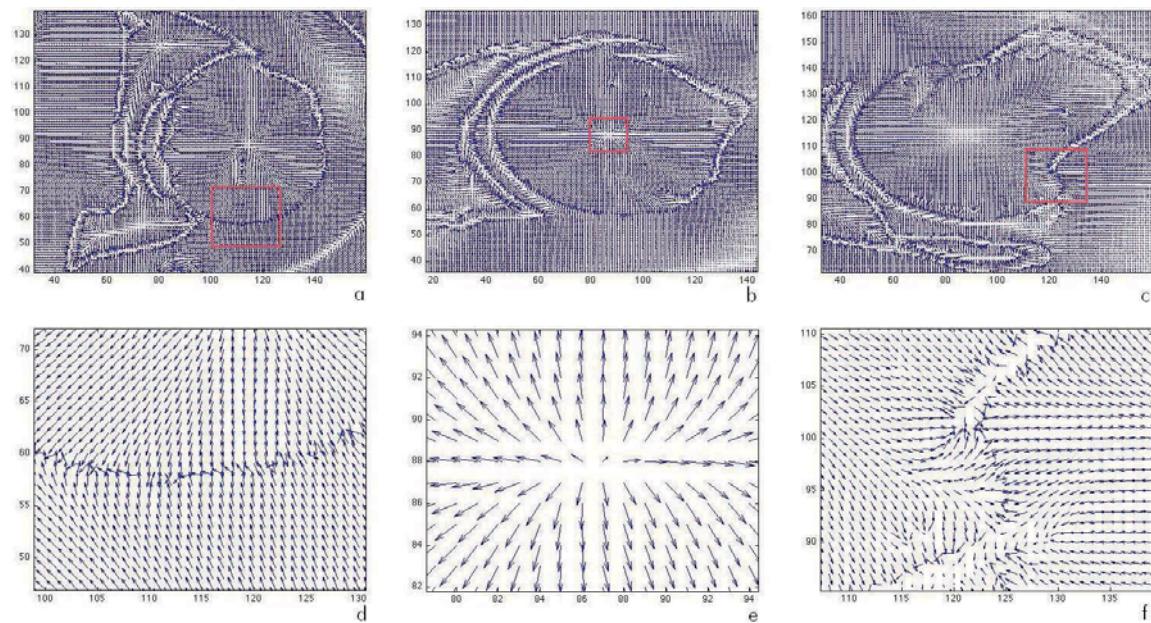


Figura 6. a-c) VFC-3D en los tres planos de la cabeza de fémur. d-f) Amplificaciones de las zonas enmarcadas en rojo de a-c) respectivamente, que muestran el campo vectorial de fuerzas externas.

el problema es algo más complicado por que la cabeza del fémur se encuentra muy cerca de la estructura ósea de la pelvis.

En este ejemplo se tiene un volumen de $201 \times 201 \times 201$ voxels que posee al fémur derecho y la mitad derecha de la pelvis. Los bordes iniciales se obtienen mediante un algoritmo de segmentación de Canny tridimensional del cual se estima la fuerza externa y de donde mediante la convolución con el kernel de vectores se genera el campo vectorial tridimensional. Se usa una esfera inicial que después de una inspección del volumen de bordes binario, se posiciona en el interior de la cabeza del fémur. La esfera tiene un mallado con un número constante de elementos que no varía durante el procesamiento (existen métodos que si permiten el remallado a medida que se va deformando el modelo). Se fijan las variables de tensión y suavidad que debe poseer este mallado y que van a constituir la fuerza interna. Se fija el número de iteraciones y se busca el equilibrio de fuerzas interna y externa. En la figura 7a se observa la esfera inicial que sufrirá la deformación (morphing) dentro del campo vectorial, en 7b el mallado que encuentra la convergencia dentro del campo vectorial y en la figura 7c se visualiza la segmentación de la cabeza del fémur la cual es la representación gráfica de un listado de caras y vértices de los elementos que constituyen su mallado. El listado es la entrada para el análisis por elementos finitos, al cual se le pueden aplicar cargas y saber como se deformará o si soportará una incrustación de un implante metálico por ejemplo.

4.2 Aorta

La arteria aorta esta pegada al corazón y soporta la alta presión del bombeo de sangre hacia el cuerpo. En este ejemplo se presenta un caso patológico de aneurisma, en el que en cierta región de la aorta, la pared se ha debilitado y sus propiedades mecánicas han cambiado comportándose como un globo. Al disminuir la rigidez de la pared existe un alto riesgo de romperse y desenlazar en la muerte del paciente. Es por eso que es importante hacer un análisis de esfuerzos mediante FEM y planear una intervención quirúrgica que solucione el problema.

El volumen es segmentado en la región donde se encuentra el aneurisma y mostrado en la figura 8.

4.3 Trabécula

Se presenta un ejemplo más, usando el método de segmentación por umbral seguido de un mallado volumétrico.

El fémur consta de dos tejidos óseos, cortical y trabecular. En la cabeza del fémur, la trabécula se encuentra en la parte interna radialmente hablando. La trabécula es una estructura porosa que mediante un tomógrafo normal no puede ser visualizada correctamente, por lo tanto se usa un micro-tomógrafo el cual consiste en un tomógrafo con fuente de radiación X y arreglo de detectores de dimensiones más finas, aumentando así la resolución espacial.

Conocer las propiedades mecánicas de la trabécula permitirá predecir en futuros estudios, posibles fracturas relacionadas con la osteoporosis. El análisis con elementos finitos permite ir encontrando los parámetros más importantes a considerar en la solidez de la estructura, esto se hace a la par de los esfuerzos físicos reales aplicados sobre la muestra en una mesa de compresión.

En la figura 9 vemos un corte axial de la muestra de trabécula, una imagen proyectiva del volumen constituido por 1016 cortes y el mallado volumétrico presentado mediante técnica de rendering. Las dimensiones reales de esta muestra son $1 \times 0.28 \times 0.25\text{cm}^3$.

La trabécula del fémur está lista para ser analizada mediante elementos finitos.

5 Discusión

En este artículo se han expuesto algunas bases teóricas que permiten incursionar en el campo de la biomecánica mediante la simulación computacional. Conceptos de procesamiento de imágenes tridimensionales y del análisis con elementos finitos que son herramientas matemáticas que están permitiendo investigar estructuras biológicas y apoyar al diseño de dispositivos biomecánicos que solucionen problemas estructurales. Dentro de la segmentación y mallado tridimensional se expusieron los métodos de GVF y VFC que son plataformas para nuevas propuestas. Estos métodos fueron programados mediante el lenguaje de programación *Matlab* dando flexibilidad al desarrollo de nuevas ideas. Los ejemplos mostrados de segmentación y mallado en fémur y aorta son preliminares ya que hace falta aún poder controlar el remallado, aumentando el número de elementos en regiones con muchos detalles y disminuirlo en regiones que varíen poco su estructura. También sería conveniente usar mallado del tipo estructurado. El caso de la trabécula ejemplifica un mallado bien terminado que está siendo ya analizado con elementos finitos para definir el modulo de elasticidad. Muchas más aplicaciones requieren de este tipo de análisis y este artículo, si cumple con su

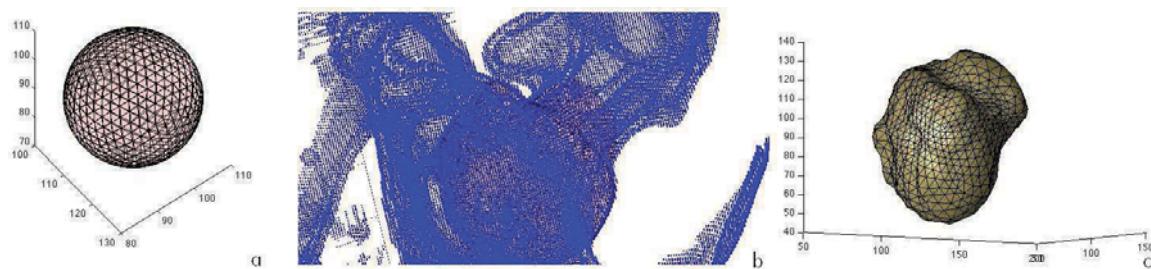


Figura 7. a) Esfera inicial, b) modelo deformable dentro del campo vectorial tridimensional, c) cabeza de fémur segmentada y mallada.



Figura 8. a) Un corte axial tomográfico. b) Campo vectorial tridimensional perfilando corazón, aorta y columna vertebral. c) Segmentación, mallado y rendering de la aorta con la región de aneurisma visible a la derecha del órgano.

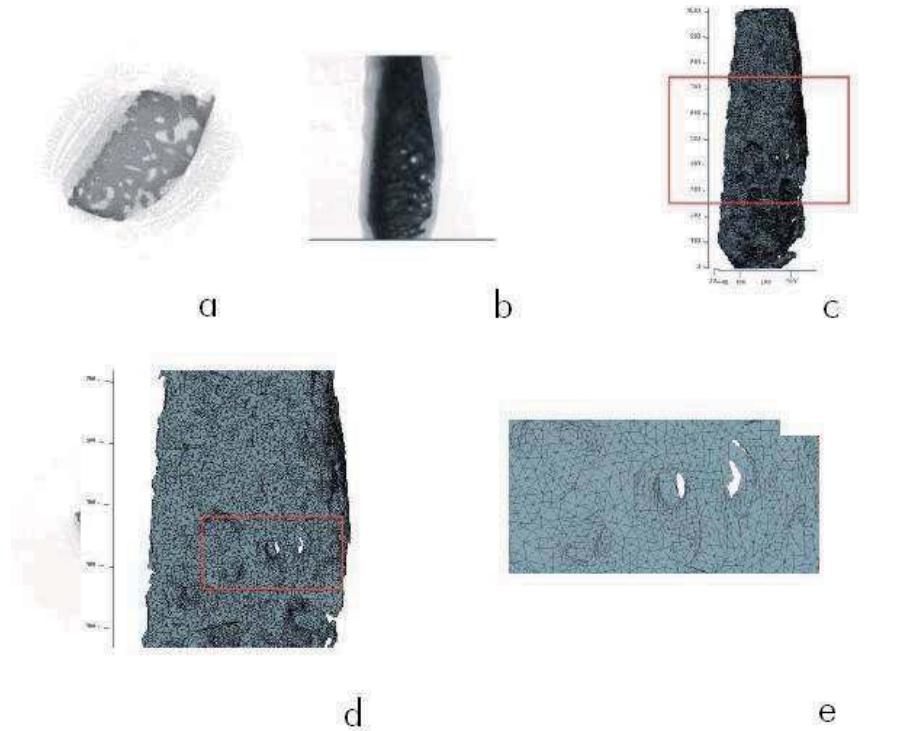


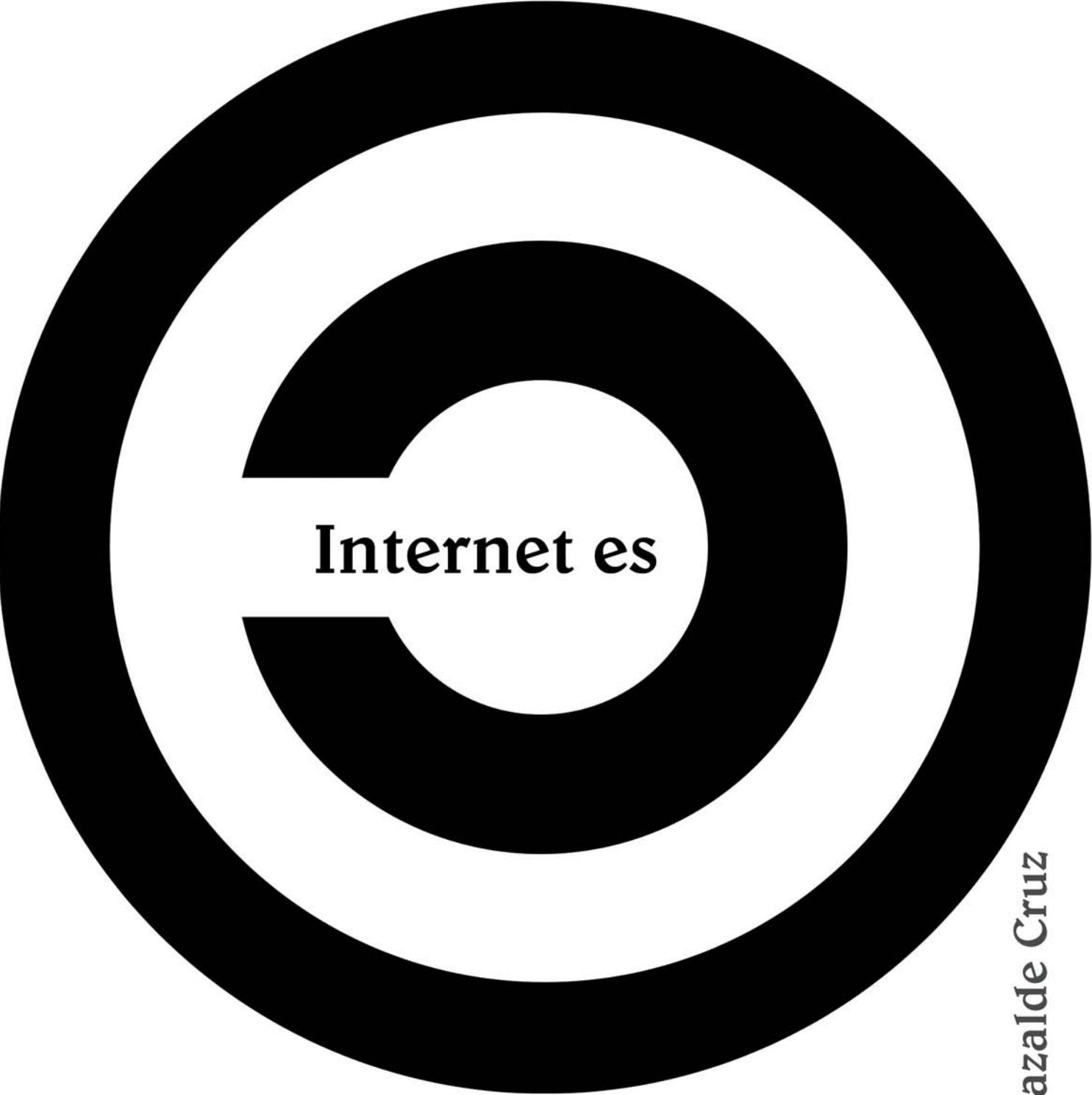
Figura 9. Trabécula: a) un corte axial, b) proyección sagital del volumen, c) mallado-rendering y d,e) detalles de la estructura porosa y Tets.

objetivo, debe permitir al lector reconocer el procedimiento para plantear problemas de biomecánica que podrán ser resueltos mediante simulación computacional.

6 Bibliografía

1. Fang Q. and Boas D., “Tetrahedral mesh generation from volumetric binary and gray-scale images,” Proceedings of IEEE International Symposium on Biomedical Imaging 2009, pp. 1142-1145, 2009.
2. Hegadi R., Kop A. and Hangarge M., “A Survey on Deformable Model and its Applications to Medical Imaging”. IJCA, Special Issue on RTIPPR (2):64-75, Published By Foundation of Computer Science, 2010.
3. Hu, Grossberg and Mageras, “Survey of Recent Volumetric Medical Image Segmentation Techniques”, Biomedical Engineering, ISBN 978-953-307-013-1, Edited by: Carlos Alexandre Barros de Mello, Publisher: InTech, October 2009.
4. Innovative-cfd, Making Sense of CFD Grid Types. <http://www.innovative-cfd.com/cfd-grid.html>
5. Kutz, M., *Biomedical Engineering and Design Handbook, Second Edition, fundamentals, vol 1*, MacGraw Hill, 2009, Chapter 8.
6. Lakare S., “3D Segmentation Techniques for Medical Volumes”, Research Proficiency Exam, Center for visual computig, department of computer science, State University of New York at Stony Brook, December, 2000.
7. Li B. and Acton S.T., “Active Contour External Force Using Vector Field Convolution for Image Segmentation”, IEEE Transactions on Image Processing, VOL. 16, NO. 8, pp. 2096-2106, August 2007.
8. Longest PW., Vinchurkar S., “Effects of mesh style and grid convergence on particle deposition in bifurcating airway models with comparisons to experimental data”. *Medical Engineering & Physics*; 29(3): 350-66, 2007.
9. McInerney T., Terzopoulos D., “Deformable Models”, in Handbook of Medical Image Processing and Analysis (2nd Edition), I. Bankman (ed.), Academic Press, San Diego, 2008, Ch. 8, 145-166.
10. Xu C. And Prince J.L., “Snakes, shapes, and gradient vector flow”, IEEE Transactions on Image Processing, vol. 7, no. 3, pp. 359-369, 1998.

cs



Internet es

Alan Lazalde Cruz

Recibido: 13 de febrero de 2012.

Aceptado: 22 de marzo de 2012.

Resumen

Con este artículo hacemos una revisión histórica de Internet y las ideas libertarias que le sustentan. Intentaremos enfatizar cómo desde ARPANET hasta las redes P2P actuales, los padres fundadores de Internet y la redes que habitan encima, no han hecho sino trabajar para compartir y no restringir el conocimiento. Esta es la simple, no definitiva, y stallmaniana (hablaremos de Richard Stallman más adelante) razón por la que afirmamos que Internet es copyleft.

Arpanet: el primer Centro de Pensamiento

Hace medio siglo el psicólogo, físico y matemático J. C. R. Licklider escribió en su *Man-Computer Symbiosis*:

“Es razonable pensar, de aquí a 10 o 15 años, en un Centro de Pensamiento que incorporará las funciones de las bibliotecas de hoy en día junto con los avances en almacenamiento y obtención de información.

Esto implica una red de tales centros, conectados uno con otro en líneas de comunicación de banda ancha, y con usuarios a través de servicios cableados.”

El científico norteamericano reflexionó sobre la cooperación entre hombres y dispositivos de cómputo con la intención última de conformar un gran Centro de Pensamiento: una red humano-tecnológica para el conocimiento compartido, creada de manera colaborativa, para beneficio de la humanidad. Con esto, Licklider sentó las bases ideológicas sobre las que surgió la red que hoy conocemos como Internet.

Las ideas y liderazgo de Licklider llevaron a la implementación de ARPANET en 1969, la primera gran red de computadoras; un entramado de equipos de cómputo conectados sobre una infraestructura construida por universidades y la milicia de los Estados Unidos. Con esa red, grupos de trabajo técnicos y científicos fueron capaces de colaborar y compartir información de manera telemática.

Junto a Licklider, se sumaron hombres como Paul Baran —padre de las redes descentralizadas—, Bob Taylor —uno de los padres de ARPANET—, Leonard Kleinrock —padre de los ruteadores de Internet y la teoría detrás de ellos—, y muchos otros, quienes compartieron ideas y esfuerzos en torno a la implementación y perfeccionamiento de la ARPANET.

Aquí hay un hecho relevante: el principio de funcionamiento de la ARPANET diseñada por ellos era muy semejante al de las denostadas redes Par-a-Par actuales (P2P), en el sentido de que todos los miembros de esa red fueron concebidos equipotentes: todos con la misma jerarquía y responsabilidades, todos colaborando para compartir recursos e información entre sí.

Internet: conectividad para todos

El siguiente hito fue protagonizado en 1983 por Vinton Cerf y Bob Kahn, cuando terminaron de implementar la reglas que hoy dominan el diseño y construcción de redes computadoras en Internet: los llamados protocolos TCP/IP. El propósito primordial de Cerf y Kahn fue definir mecanismos robustos para la transmisión confiable de información a lo largo de esas redes.

Lo que sucedió con los protocolos TCP/IP, y con todas las ideas que hacen funcionar Internet, es que circularon entre la comunidad de ingenieros y científicos en documentos conocidos como RFC (*Request for Comments*). Fue gracias a esos documentos que se vio impulsado el debate intenso, entre pares, alrededor de cada una de las propuestas tecnológicas contenidas en ellos. Las discusiones llevaron a convertir muchas de esas ideas sobre cómo construir Internet en verdaderos estándares tecnológicos.

Por ejemplo, en el primer documento RFC, con fecha de 1969, están descritos con detalle las interacciones entre los componentes de una red de computadoras conectadas a través de la primera familia de dispositivos que en la actualidad conocemos como ruteadores. El documento RFC 1 simplemente describe los acuerdos que vieron nacer a la red ARPANET.

En general, un RFC es más que un documento, es una verdadera petición de comentarios, una invitación abierta a la participación en la discusión de las tecnologías que conforman a Internet. Por ejemplo, cuando un RFC es elevado a estándar, es el resultado del perfeccionamiento, o mejor dicho, la evolución de las ideas allí propuestas en pro de un cada vez mejor Internet. Una evolución guiada por los intereses de la comunidad, más que por individuos o grupos de poder.

En la actualidad, pocos saben que el bagaje intelectual de prácticamente todo Internet está disponible en alrededor de 5,000 documentos RFC, cuya

administración es llevada por un ejército de voluntarios llamado *The Internet Engineering Task Force* (IETF), con el único propósito de “hacer que Internet funcione mejor”.

Desde su fundación, con el auspicio de la *Internet Society*, la IETF trabaja bajo principios de participación voluntaria, en medio de procesos abiertos y consensuados. En la práctica, son los autores intelectuales de los protocolos de Internet quienes lideran a grupos de voluntarios en pro del desarrollo de sus ideas y el mantenimiento de una “propiedad responsable”.

Específicamente, el concepto de propiedad responsable tiene que ver con que la IETF, para cumplir con su misión, acepta la responsabilidad de difundir y mejorar esas ideas porque cree que “la existencia de Internet, y su influencia en la economía, la comunicación y la educación, [...] ayudará a construir una mejor sociedad humana.”

Web: Información para todos

En el año 1989, Tim Berners-Lee trabajaba para la Organización Europea para la Investigación Nuclear (CERN) cuando creó la tecnología que hoy en día hipervincula una ingente cantidad de información colocada sobre Internet, un tejido complejísimo que hoy llamamos *World Wide Web* (WWW), o simplemente, “la Web”.

Berners-Lee observó que si bien la Internet existente mejoraba las oportunidades de colaboración entre investigadores, estos no contaban con mecanismos adecuados para contener y presentar información. Fue bajo estas circunstancias que ideó e implementó los tres pilares que sostienen la Web:

- Un servidor de páginas web. Es decir, un software que entrega páginas web a sus solicitantes; navegadores web, en su mayoría.
- Un lenguaje informático llamado HTML (*HyperText Markup Language*) que utiliza etiquetas para componer documentos hipervinculados, los cuales son procesados por el navegador web para su presentación en pantalla.
- Y el protocolo HTTP para controlar la transmisión, entre el servidor y sus clientes a través de la red, de ese nuevo formato de representación de la información que son las páginas web.

Una vez que fueron plasmadas en documentos RFC para su discusión colectiva, las ideas de Berners-Lee

hicieron eco en miles de investigadores. Con esto, Internet fue catapultada rápidamente fuera del ámbito meramente científico. Y la Web ha servido desde entonces para conectar, a través de la información y la comunicación, la vida diaria de millones de personas: un beneficio con miras a ser universal, libre y gratuito.

¿Quién es dueño de la Red?

¿A quién le pertenecen Internet y la Web? Sabemos que los autores intelectuales de las redes que componen Internet, los protagonistas de la historia que algunos llamaron Revolución de la Información, son investigadores ampliamente reconocidos agrupados alrededor de la IETF.

Pero, ¿son ellos los dueños de Internet o las empresas o universidades para las que trabajaron? Más aún, ¿para quién trabaja la IETF? En su documento RFC 3935 se menciona:

“Queremos que Internet sea útil para las comunidades que comparten nuestro compromiso por la apertura y la imparcialidad. Adoptamos conceptos tales como control descentralizado, empoderamiento del usuario final y la distribución de recursos para ser compartidos...”

La IETF pone su énfasis en la competencia técnica, el consenso simple y la participación individual, que necesita ser abierta para recibir aportación desde cualquier fuente...”

Con esto, la IETF clama que su trabajo, “hacer que Internet funcione mejor”, es la suma del esfuerzo de muchos, en el entendido de que Internet es el bien público (procomún) digital más importante que posee la humanidad. Esto quiere decir que al cumplir con su trabajo, la IETF no sólo expande las fronteras de la investigación y desarrollo en el tema, sino que también le entrega a la humanidad una mejor plataforma para comunicarse.

El bien común que la IETF ofrece al mundo está en las ideas, no en su implementación. Aunque cada documento RFC procura describir ideas que ya han sido probadas en la práctica, la IETF sólo preserva los conceptos de fondo que luego pueden servir para su implementación por particulares. Así, por ejemplo, podemos encontrar que los protocolos TCP/IP tienen versiones específicas implementadas para cada sistema operativo en cada equipo que requiera conectarse a Internet, aún cuando las ideas detrás de cada una de esas versiones pueda ser la misma.

La mayoría de las ideas discutidas en documentos RFC son protocolos de comunicación. Donde un protocolo es una idea que resuelve un problema concreto, es decir, es un algoritmo, una descripción paso a paso para guiar, en este caso, la transmisión confiable y robusta de información a través de la red. Bien relacionados entre sí, el gran conjunto de protocolos usados en Internet han permitido la comunicación entre millones de equipos de cómputo durante más de tres décadas.

Los protocolos suelen ser implementados en software o hardware por empresas para su venta en productos específicos, o bien por universidades o particulares para su estudio, experimentación o mejora. Empresas como Cisco, IBM o Verizon realizan investigación y desarrollo de protocolos de comunicación, muchas veces de la mano de importantes centros de investigación en todas partes del mundo. Sin embargo, esto no los convierte de ninguna manera en los dueños de la red, gracias al esmerado trabajo de la IETF en la difusión y mantenimiento de tecnologías abiertas para todos, en vez de entregarlos al frenesí de los patentadores de ideas.

Código Fuente: El idioma hombre-máquina

Un algoritmo es un método que consta de un número finito de pasos a seguir para resolver un problema. Para que una computadora ejecute un algoritmo, éste se le debe presentar en un formato adecuado, normalmente en cadenas de ceros y unos (bits) inteligibles para ella, algo así como un lenguaje máquina.

Puesto que un ser humano no habla tal lenguaje máquina, ¿cómo hace para comunicarle órdenes a la computadora? ¿cómo logra que un dispositivo de cómputo ejecute un algoritmo que en principio no es sino una idea en la cabeza de su creador? En la práctica esto se logra con los llamados lenguajes de programación.

Un lenguaje de programación es un lenguaje a medio camino entre el natural, hablado por un ser humano, y el de la máquina, los bits. Una persona capaz de aprender un algún lenguaje de programación puede especificar, siguiendo una sintaxis precisa, el algoritmo que desea implementar en una computadora.

La especificación de una algoritmo escrita en cierto lenguaje de programación se conoce como código fuente (almacenado normalmente en un archivo de texto). La transformación de ese código al

lenguaje de la máquina ocurre en varios pasos que son guiados por software llamado compilador. La transformación final ocurre con la creación de un archivo listo para ser ejecutado por el equipo de cómputo; ese archivo es el programa de software. Dicho esto, el compilador no es sino un software que transforma código fuente en su versión de software ejecutable.

La computadora ENIAC, la primera construida para uso de propósito general —según la versión más aceptada—, encendida a partir de 1947, era programada mediante cables e interruptores. En este sentido, puede decirse que los programadores debían conocer muy de cerca el funcionamiento de la máquina ENIAC para introducir en ella los algoritmos de su interés.

En la actualidad los programadores requieren conocer cada vez menos detalles de las máquinas donde implementarán sus algoritmos y en cambio, se concentran todavía más en los problemas que buscan resolver con ellas. Es probable que en el futuro los lenguajes máquina y natural converjan a través de un mismo código fuente.

Software y Conocimiento

Ahora bien, si los algoritmos son ideas que resuelven problemas concretos y su implementación para ser ejecutados por una computadora está descrita mediante código fuente escrito con algún lenguaje de programación, ¿a quién le pertenece esa obra intelectual que es el software?

Y es que si el software es la suma de ideas (los algoritmos) y su implementación (el código fuente), ¿quién sino el programador (o quien le contrata) es el propietario legítimo?

Es un ejercicio común que la propiedad intelectual del software pertenezca legalmente a quien registre ese derecho ante la oficina correspondiente. Por ejemplo, en los Estados Unidos es posible patentar algoritmos y su implementación en software. Esto facilita que el modelo de negocios de una empresa de desarrollo de software esté basado en la protección de su software, con el fin de mantenerlo en secreto y alejado de sus posibles competidores.

Las licencias de software, por su parte, son permisos que suelen estar muy bien especificados con el fin de advertir al usuario final de las maneras legales en las que puede utilizar el software en cuestión sin caer en prácticas que vayan en contra de los intereses de su propietario.

Sin embargo, hay quienes piensan que las patentes y licencias tradicionales de software son en realidad herramientas que limitan las libertades fundamentales de las personas para acceder al conocimiento. Una de esas personas es Richard Stallman, quien en 1983 se dio a la tarea de iniciar un movimiento técnico e ideológico —y también religioso, en cierto sentido— que llevó a muchos a reconsiderar profundamente el asunto de la propiedad intelectual del software.

La lucha de Stallman comenzó sustentada en los viejos principios con los que él, un hacker informático, creció; esto es, colaborar para compartir el conocimiento usando esa lingua franca llamada código fuente.

La lógica detrás de los esfuerzos Stallman puede resumirse como sigue:

Las ideas son conocimiento y el conocimiento le pertenece a todos. Así que si el software es un conjunto de ideas implementadas desde código fuente, entonces el software es un repositorio de conocimiento. Ergo, todos somos dueños del software.

Software libre: Conocimiento para todos

Con el convencimiento de que el software es conocimiento, Richard Stallman creó el Proyecto GNU (1983) y luego la Fundación por el Software Libre (FSF, 1985); las organizaciones encargadas de promover los aspectos técnicos y filosófico-legales, respectivamente, de sus ideas. Con el abrigo de las dos emergió una nueva generación de software calificado como libre en términos de su valor económico y, sobre todo, social. Y es que el Software Libre, aduce la FSF, no sólo es gratuito, además es un bien común tecnológico sobre el que es posible obtener y generar conocimiento para todos.

Recordemos que en inglés Software Libre es Free Software, con la doble connotación del término free, que equivale a “gratis”, pero también a “libre”. Pero ¿qué hace que un software sea libre en los términos que sugieren Stallman y la FSF? La respuesta está en la aplicación de una licencia de software expresamente diseñada para preservar los principios que hagan del software un artefacto tecnológico libre. Esa licencia es la GPL (la Licencia Pública General del Proyecto GNU), publicada en 1989, y de la que a la fecha existen varias versiones.

Quien distribuya software con licencia GPL, ofrece estas cuatro libertades, numeradas a partir del cero, como suele ocurrir con cierto código fuente:

- La libertad para ejecutar el software para cualquier propósito.
- La libertad para estudiar el software accediendo a los algoritmos y su implementación.
- La libertad para distribuir (i.e., copiar y compartir) el software.
- La libertad para mejorar el software y publicar versiones modificadas del mismo.

Cada una de esas libertades está limitada únicamente por el hecho de que deben de preservarse intactas cada una de ellas para todos los usuarios. Esto es de particular importancia para el caso cuando el software sea modificado con la intención de redistribuirlo.

Es importante observar que todo software libre con licencia GPL produce también software libre GPL a partir de él. De modo tal que con esta licencia no existe vuelta atrás una vez publicada junto al software; razón por la que algunos dicen que la GPL es una licencia con efecto viral. En cambio, el software privado puede llegar a liberarse, en etapas o en un solo paso, hasta convertirse en GPL, o de dominio público incluso.

La licencia GPL es también conocida por ser una licencia copyleft (quizá la primera), un concepto antagónico del copyright. En este sentido, la FSF clama que los

“...desarrolladores de software privativo usan copyright para quitarle la libertad a los usuarios; nosotros usamos los derechos reservados para garantizarles su libertad. Por eso es que invertimos el nombre, convirtiendo los derechos reservados (“copyright”) en “copyleft”.”

Los valores libertarios detrás del copyleft ofrecido por la GPL pueden rastrearse históricamente a principios de los años 70 con la Principia Discordia de Greg Hill y Kerry Thornley, donde promovían el uso del término “Kopyleft” y la frase “All Rites Reversed”, como parodias del copyright, al tiempo que animaban a sus lectores a “reimprimir aquello que les guste”. Todo esto enmarcado por el gran ánimo contracultural típico de aquella época; una herencia ideológica bien recibida por Stallman.

Es interesante observar que la FSF no aboga por hacer software de dominio público. De hacerlo así el software peligra con caer en manos egoístas que lo

convirtan en software privado. Además, esa privatización iría contra el deseo del programador original de compartirlo para todos. En este sentido, podemos decir que la GPL es una suerte de licencia “distribuida” que busca el beneficio de todos y elimina los exclusivamente personales, como sucede con las licencias “centralizadas” fundadas en el copyright.

A la licencia GPL le acompañan otras de espíritu semejante, e.g., Apache, BSD, Artistic, MIT. Cada una de ellas representa un punto de vista distinto (filosófico y técnico) sobre lo que significa hacer software libre. Por ejemplo, la licencia BSD incluye una cláusula que solicita el reconocimiento de los autores originales del software, por lo demás éste puede considerarse de “dominio público”. El problema con esas licencias, argumenta Stallman, es que no evitan el abuso del software por parte de particulares.

Linux: Arquetipo del software social

Desde Internet como plataforma, en pocos años el Proyecto GNU fue liberando software libre con licencia GPL con el objetivo de obtener un sistema operativo (el software que administra memoria, discos duros, procesador y demás recursos de un equipo de cómputo) completamente libre. Sin embargo, a pesar de haber creado prácticamente todas las partes periféricas a un sistema operativo, aún restaba ofrecer su componente modular, algo conocido como kernel o núcleo.

En 1991 un joven estudiante finlandés, Linus Torvalds, decidió que prefería crear su propio sistema operativo antes que conformarse con el que le proveía su nueva computadora. Lo que Torvalds proyectó en su momento como un mero pasatiempo, llegó a convertirse en el kernel de un sistema operativo en toda forma, y la pieza faltante del anhelado Proyecto GNU.

Se sabe que Torvalds bosquejó dicho kernel en pocos meses, al tiempo que lo fue vistiendo con software libre distribuido por el Proyecto GNU, hasta obtener un pequeño sistema operativo. Con ese prototipo apenas funcionando, fue que escribió estas palabras a un grupo de programadores en Internet:

“Hola a todos... Estoy haciendo un sistema operativo (libre)... (es sólo un pasatiempo)... Toda sugerencia es bienvenida, pero no prometo implementarlas :-)”

Sin más, esa sencilla invitación a la participación atrajo el esfuerzo y talento de decenas de programadores entusiastas alrededor del mundo. No faltó mucho para que el sistema operativo libre emprendido por Linus Torvalds comenzara a ser llamado Linux dentro de la creciente comunidad de usuarios; también una franca alusión al viejo sistema Unix.

Hoy podemos decir que Linux es consecuencia del trabajo de una organización integrada y defendida por legiones de voluntarios, participando todos a través de una jerarquía prácticamente horizontal y sólo sometida a la “dictadura benévolas” de Torvalds. Esa benevolencia viene del hecho de que la comunidad respeta la palabra de Torvalds como definitiva en caso de controversias sobre la dirección del proyecto, aún cuando todas las aportaciones deben permanecer en comunión con las libertades que la licencia GPL (incluida con Linux) ofrece.

Alrededor de Linux podemos encontrar literalmente cientos de proyectos de todo tipo, que van desde los que adaptan Linux para administrar teléfonos móviles hasta quienes lo optimizan para ofrecer servicios de nivel crítico en empresas y gobierno, pasando por los que instalan Linux en cúmulos o clusters de consolas de videojuegos para construir supercomputadoras.

¿A qué se debe la flexibilidad de Linux? En buena medida a su arquitectura de software, así como en las bondades inherentes a su licenciamiento. Empresas como Google, por ejemplo, deben buena parte de su infraestructura de bajo costo al uso y apoyo sistemático de todo tipo de software libre. Esto incluye la instalación de Linux en decenas de miles de equipos de cómputo de su propiedad; algo difícilmente sostenible con esquemas de licenciamiento tradicionales.

La participación incentivada desde la GPL, así como el trabajo colectivo guiado por objetivos simples y de corto plazo (una de las principales aportaciones de Torvalds), han convertido cada línea del código fuente de Linux en un ejemplo de evolución y refinamiento técnico sin precedentes. Hoy en día, Linux es software libre en toda su extensión y uno de los principales estandartes del movimiento engendrado por Stallman.

Buena parte del éxito de Linux también recae en la protección social y legal que la GPL le confiere. Y es que las aportaciones que Linux recibe por parte de programadores en todo el globo no están únicamente

atraídas por la flexibilidad y creciente popularidad del sistema, sino también al hecho de que la GPL evita que su trabajo sea enajenado por los intereses de unos pocos. De esta manera, el proyecto Linux queda únicamente bajo la dirección y amparo de su conglomerado de usuarios.

Redes P2P: Redes de colaboración

Para finales de la década de los años 90 Internet ya era una tecnología consolidada con millones de usuarios en todas partes del mundo. Muchas actividades cotidianas tenían su símil virtual en Internet. De la complejidad de tal red de comunicación también emergieron esfuerzos colaborativos, fuertemente influenciados por los grupos de desarrollo de software libre, para resolver asuntos de importancia individual y colectiva. Sin duda un asunto vital para los usuarios de Internet sigue siendo desde entonces el relacionado con el libre acceso a la información.

¿Y si la información es música? En 1999 Shawn Fanning, joven programador norteamericano, desarrolló una pequeña aplicación para compartir archivos de música en formato MP3 entre sus amigos; la llamó Napster en alusión a su apodo en la red. Bastó un par de años para que Napster se convirtiera en punto de encuentro obligado para decenas de millones de personas. Con esto, el sencillo software escrito por Fanning abrió las puertas a un fenómeno social que traspasó las fronteras de Internet, uno que hizo usar (y abusar) de la red en niveles sin precedentes. Y es que luego de Napster, quedó claro que todos quisieron conectarse a Internet para obtener y compartir información de todo tipo.

¿Dónde radica el éxito de Napster? Es muy probable que en su modelo de participación y colaboración masiva y entre iguales. Es así que Napster también dio pie a una nueva generación de aplicaciones para Internet, las llamadas aplicaciones Par-a-Par (P2P, por sus siglas en inglés).

Las aplicaciones de tipo P2P crean redes de colaboración con tres características básicas:

1. Todos los miembros de la red P2P son iguales en sus responsabilidades. En este caso, todos reciben y todos entregan información. En otros modelos de organización en red, ciertos miembros tienen una jerarquía centralizada, encima de los demás.
2. La red puede incrementar el número de sus miembros sin afectar negativamente los servicios que

ofrece; es decir, escalar. En el caso de las aplicaciones P2P para distribución de contenido, poseer más miembros es mejor porque todos se apoyan compartiendo recursos e información.

3. Las redes P2P presentan cierto funcionamiento orgánico. Por un lado, los miembros de la red son autónomos al regular su propio ciclo de vida en la red P2P (al entrar, salir y participar en la red a voluntad). Por otro, la totalidad de la red se auto organiza para mantenerse operando y seguir ofreciendo sus servicios.

Las redes de colaboración creadas con aplicaciones P2P también pueden emplearse para resolver problemas. El proyecto SETI@Home, por ejemplo, utiliza el modelo P2P para procesar colectivamente la información generada por sus telescopios con el fin de hallar vida extraterrestre. Otros proyectos similares analizan moléculas, encuentran nuevos números primos o ayudarán al Gran Colisionador de Hadrones recién instalado en el CERN en el procesamiento de una cantidad inmensa de datos.

Sin embargo, la denostación mediática de las redes P2P viene de su relación con la distribución de pornografía, virus informáticos, actividades anónimas que pueden afectar asuntos de seguridad nacional, entre otros. Pero sin duda el tema más polémico relacionado con el uso de las redes P2P es la distribución colectiva de material protegido por derechos de autor.

En su momento Napster fue demandado por integrantes de la banda de rock Metallica y finalmente obligado a retirar toda clase de contenido considerado ilegal, hasta el punto de hacer desaparecer el servicio. No obstante, a raíz de tal demanda la popularidad de Napster alcanzó alturas insospechadas. Napster desapareció al poco tiempo de todas maneras, pero le sucedieron toda una serie de nuevas redes P2P, tales como Kazaa, Gnutella, BitTorrent y Freenet; cada una de ellas con servicios cada vez más eficientes, colectivos, anónimos y especializados.

La historia de cada red P2P es casi una copia fiel de la historia de Napster; esto es:

- nacimiento tímido,
- desarrollo exponencial,
- polémicas querellas durante meses,
- éxito masivo,

- resolución judicial con gran alcance mediático,
- clausura abrupta del servicio,
- y surgimiento de sistemas derivados más sofisticados que el original.

Así las redes P2P son una suerte de organismo vivo que nace, se desarrolla y adapta a su entorno tecnopolítico hasta morir a causa de vacunas legales y luego renacer con mutaciones que lo fortalecen ante nuevas amenazas.

También los demandantes han sofisticado sus tácticas de defensa y ataque. Las grandes empresas de producción de contenidos provenientes de las industrias del cine, la música, así como las instituciones civiles y gubernamentales que abogan en pro de la propiedad intelectual, todos ellos han establecido un compacto grupo para derribar cada intento por el uso de sus potestades en Internet.

Para alcanzar su propósito, lo llamados lobbies de la Propiedad Intelectual han decidido ir más lejos: no sólo a los proveedores de servicios P2P, sino también al interior de Internet para regular el flujo de información entre sus usuarios.

Eso quiere decir que con la implantación por la vía legal de tecnologías de detección de información protegida por copyright, las empresas serán capaces de demandar directamente a los usuarios que supuestamente dispongan de información protegida.

Lo anterior polariza las opiniones sobre el tema. Por un lado están los defensores de la propiedad intelectual. Y por el otro los que alzan la voz ante lo que consideran ataques flagrantes a sus derechos fundamentales como ciudadanos en la red; es decir, privacidad, equidad y libertad. Uno y otro bando son los protagonistas de una batalla en curso, con enfrentamientos en tribunales y en las entrañas de Internet. Se trata de una batalla que definirá, según sus resultados y alcance, cómo accederemos a buena parte de nuestros bienes culturales, aún fuera de la red.

Comentarios finales

Podemos resumir las condiciones que hicieron patente una nueva era tecnológica y social basada en la participación y la información.

- La proliferación de dispositivos de almacenamiento y transmisión de información de mayor calidad en sus prestaciones.

■ La creación de aplicaciones especialmente diseñadas, como las P2P, para cooperar masivamente en la distribución de la información. Muchas de ellas dirigidas por notables grupos de investigación.

- La emergencia de redes sociales virtuales por todas partes, conformados por una escalada de miembros ávidos de información, con niveles de conectividad y comunicación nunca vistos.

En particular, vista de manera simple la información en Internet consiste de un tren de bits que según su disposición puede representar texto, audio, vídeo o cualquier otro tipo de archivo cuando sea requerido. Pero, ¿dónde está la información en Internet?

- En sus extremos, donde están los usuarios con sus equipos de cómputo la información habita en dispositivos de almacenamiento (e.g., discos duros, memorias Flash).
- Y en el interior, donde la información reside temporalmente en los dispositivos (routers) que los conducen a su destino de extremo a extremo y a través de la red.

Por otra parte, sin importar el contenido (libros, películas, etc.), una vez digitalizada la información mantiene propiedades independientes del medio de transmisión en la red. Ya sea que la información viaje a través de una red P2P ilegal, o legalmente vía iTunes luego de una compra electrónica, sufre las mismas transformaciones inherentes a los protocolos de comunicación de Internet.

Así pues, ¿en qué momento la información original deja de ser ella misma si cuando viaja por la red debe literalmente despedazarse para llegar a su destino? ¿Acaso cada pieza de esa información también debe protegerse con copyright?

Es decir, ¿el software de los protocolos de comunicación en Internet violan alguna ley de derechos de autor al copiar, transmitir y modificar información que puede estar protegida?

De manera similar, ¿qué hay de la propiedad intelectual de la información cuando en el fondo consiste de sólo bits y debe manipularse para convertirse en otro tipo de información cuando se utiliza criptografía (como sucede durante las compras en línea)?

¿Cómo detener esa máquina de copiar información llamada Internet?

Cuando la información crece en Internet con todo tipo de nuevos formatos y medios, ¿cuál es la última palabra sobre domino público, copyleft y copyright? ¿Hasta qué grado puede estirarse el copyright para que alcance a cada nuevo integrante de esa selva de información que es Internet? ¿Cómo regular blogs, foros, redes sociales y la descomunal energía de millones interactuando en torno a la información compartida? Estas son preguntas que instancias técnicas, culturales, empresariales o sociales, responden (o no) sólo desde perspectivas parciales, muchas veces sin mirarse a la cara.

Ya sea para la creación de software libre o para la difusión de nuevas metodologías de acceso a la información (alejadas del concepto tradicional de propiedad intelectual), podemos decir que un mundo alterno se está produciendo alrededor en las llamadas Culturas Libres. Desde los grupos de trabajo con filosofía Hazlo tú mismo (DIY, por sus siglas en inglés), hasta los Partidos Piratas (nacidos en Suecia en 2006, pero multiplicados en docenas de países), pasando por los grupos hacktivistas (informáticos ocupados en pro del libre acceso a la información y los derechos civiles en la red, generalmente reunidos en hacklabs), las Culturas Libres buscan convertirse poco a poco y desde distintos frentes en oportunas alternativas de acceso a la cultura, el conocimiento y la educación, ajenas todas a las formas de control centralizado.

Lawrence Lessig escribía en su *Free Culture* de 2003, visto por muchos como una de las piedras angulares de la Cultura Libre:

“En toda nuestra historia nunca ha habido un momento como hoy en que una parte tan grande de nuestra “cultura” fuera “posesión” de alguien.”

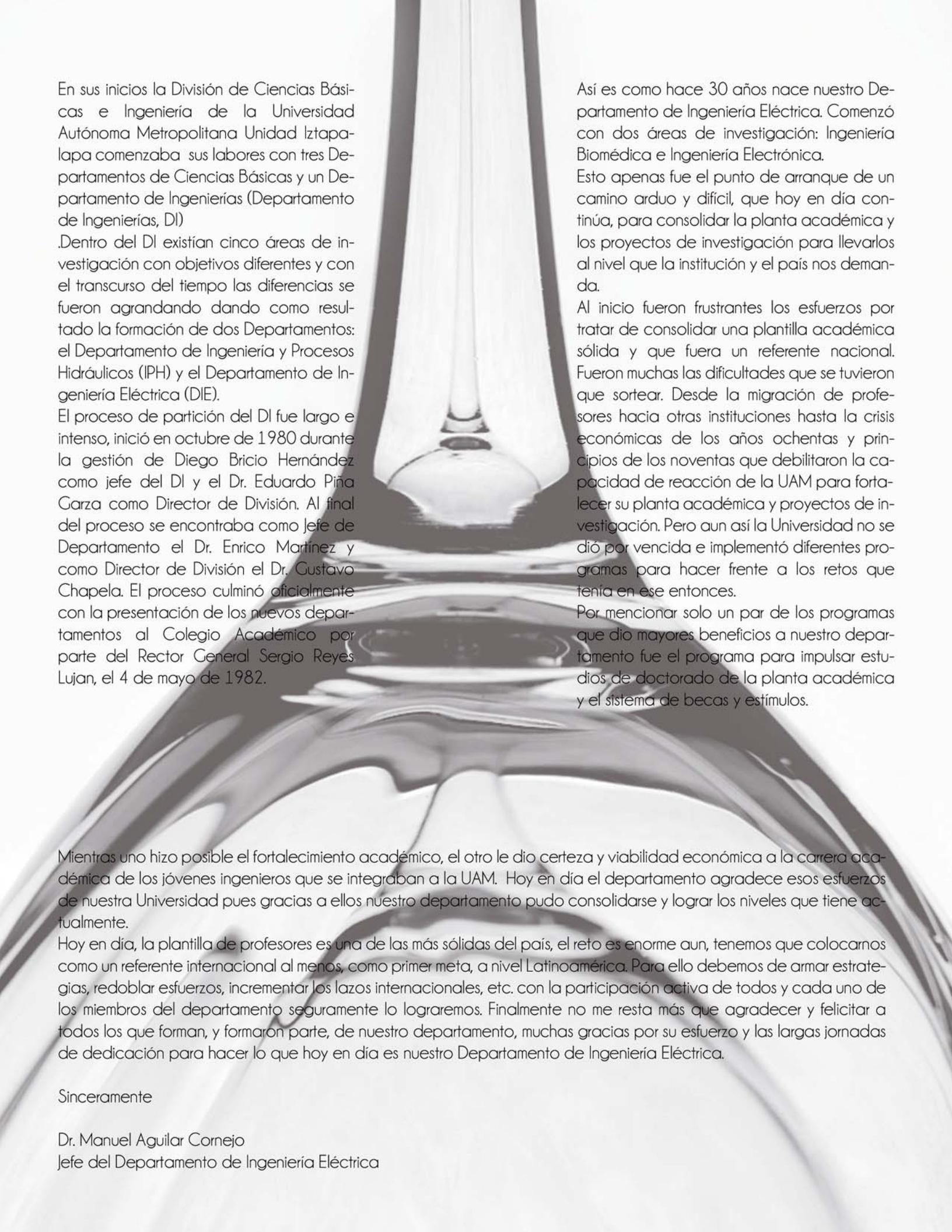
Las licencias Creative Commons, diseñadas bajo el liderazgo de Lessig, conforman una de las respuestas más importantes a un mundo gobernado por modelos económicos basados en el copyright. Hoy en día, esas licencias le muestran al mundo cómo compartir información en vez de pedir permiso por ella y de conformidad con un espíritu llamado Share-alike.

Las Creative Commons, junto con las GPL y otras licencias libres, integran una trinchera ideológica, técnica y legal, que con mucha seguridad irán perfeccionándose a favor de devolver la cultura sus verdaderos dueños: todos nosotros.

Al día de hoy, sin saberlo, sin pensarlo casi, hemos convertido Internet en el más grande enemigo del concepto tradicional/industrial de propiedad intelectual. Con Internet hemos llevado esa idea a los límites de lo absurdo en cada proceso de copia y descarga de información. Y así hemos establecido Internet como la más poderosa herramienta a favor de una cultura copyleft.

Referencias

1. Lessig, L., *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. CreateSpace. Paramount, California, 2008.
2. Stallman, R., *Free Software, Free Society. Selected Essays of Richard M. Stallman*. 2a. Edición. GNU Press. Boston, Massachusetts, 2010.
3. Zittrain, J.L., *The Future of the Internet*. Yale University Press. New Haven, Connecticut, 2008.



En sus inicios la División de Ciencias Básicas e Ingeniería de la Universidad Autónoma Metropolitana Unidad Iztapalapa comenzaba sus labores con tres Departamentos de Ciencias Básicas y un Departamento de Ingenierías (Departamento de Ingenierías, DI).

Dentro del DI existían cinco áreas de investigación con objetivos diferentes y con el transcurso del tiempo las diferencias se fueron agrandando dando como resultado la formación de dos Departamentos: el Departamento de Ingeniería y Procesos Hidráulicos (IPH) y el Departamento de Ingeniería Eléctrica (DIE).

El proceso de partición del DI fue largo e intenso, inició en octubre de 1980 durante la gestión de Diego Bricio Hernández como jefe del DI y el Dr. Eduardo Piña Garza como Director de División. Al final del proceso se encontraba como Jefe de Departamento el Dr. Enrico Martínez y como Director de División el Dr. Gustavo Chapela. El proceso culminó oficialmente con la presentación de los nuevos departamentos al Colegio Académico por parte del Rector General Sergio Reyes Luján, el 4 de mayo de 1982.

Así es como hace 30 años nace nuestro Departamento de Ingeniería Eléctrica. Comenzó con dos áreas de investigación: Ingeniería Biomédica e Ingeniería Electrónica.

Esto apenas fue el punto de arranque de un camino arduo y difícil, que hoy en día continúa, para consolidar la planta académica y los proyectos de investigación para llevarlos al nivel que la institución y el país nos demanda.

Al inicio fueron frustrantes los esfuerzos por tratar de consolidar una plantilla académica sólida y que fuera un referente nacional. Fueron muchas las dificultades que se tuvieron que sortear. Desde la migración de profesores hacia otras instituciones hasta la crisis económicas de los años ochentas y principios de los noventas que debilitaron la capacidad de reacción de la UAM para fortalecer su planta académica y proyectos de investigación. Pero aun así la Universidad no se dió por vencida e implementó diferentes programas para hacer frente a los retos que tenía en ese entonces.

Por mencionar solo un par de los programas que dio mayores beneficios a nuestro departamento fue el programa para impulsar estudios de doctorado de la planta académica y el sistema de becas y estímulos.

Mientras uno hizo posible el fortalecimiento académico, el otro le dio certeza y viabilidad económica a la carrera académica de los jóvenes ingenieros que se integraban a la UAM. Hoy en día el departamento agradece esos esfuerzos de nuestra Universidad pues gracias a ellos nuestro departamento pudo consolidarse y lograr los niveles que tiene actualmente.

Hoy en día, la plantilla de profesores es una de las más sólidas del país, el reto es enorme aun, tenemos que colocarnos como un referente internacional al menos, como primer meta, a nivel Latinoamérica. Para ello debemos de armar estrategias, redoblar esfuerzos, incrementar los lazos internacionales, etc. con la participación activa de todos y cada uno de los miembros del departamento seguramente lo lograremos. Finalmente no me resta más que agradecer y felicitar a todos los que forman, y formaron parte, de nuestro departamento, muchas gracias por su esfuerzo y las largas jornadas de dedicación para hacer lo que hoy en día es nuestro Departamento de Ingeniería Eléctrica.

Sinceramente

Dr. Manuel Aguilar Cornejo
Jefe del Departamento de Ingeniería Eléctrica

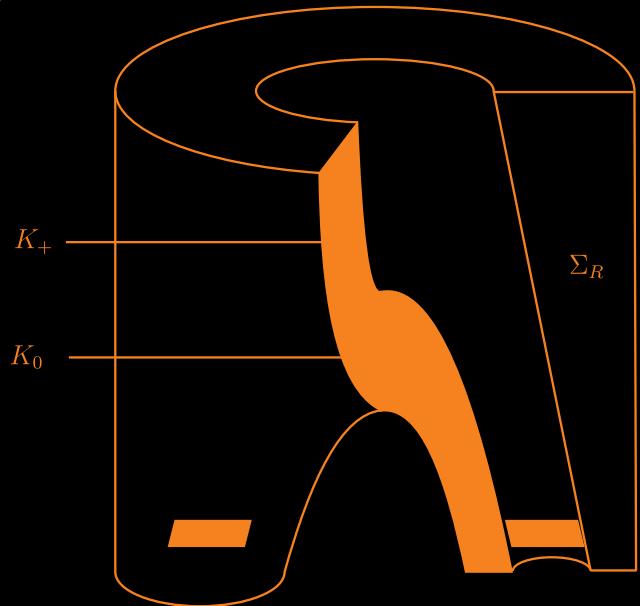


Departamento de Matemáticas
UAM Iztapalapa

Maestría en Matemáticas Aplicadas e Industriales

Maestría y Doctorado en Matemáticas

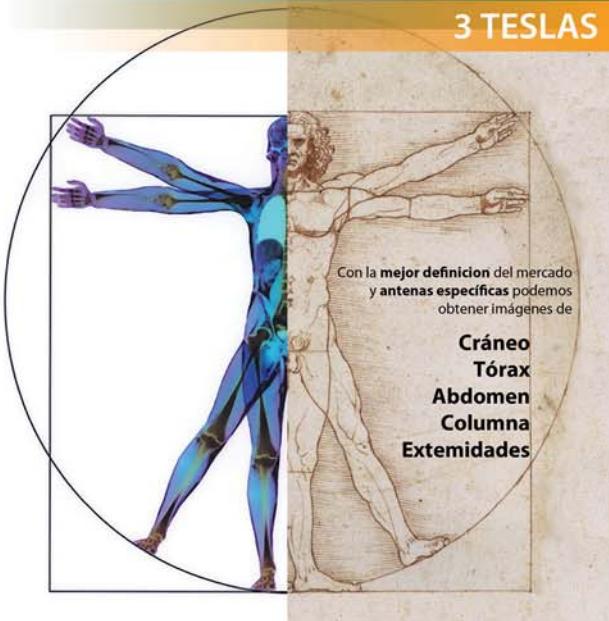
<http://pmat.itz.uam.mx>
<http://mcmai.itz.uam.mx>



El Centro Nacional de Investigación en Imagenología e Instrumentación Médica te ofrece los servicios de

RESONANCIA MAGNÉTICA

3 TESLAS



WWW.CI3M.MX

5970 48 70 · 5804 49 03 · 5970 48 69
contacto@ci3m.mx

POSGRADO EN INGENIERÍA BIOMÉDICA

- Maestría y Doctorado
- Programa de Tiempo Completo
- Becas disponibles a todos los alumnos aceptados
- Posgrado consolidado en el PNPC
- Admisión trimestral

Informes:

<http://itzamna.itz.uam.mx/pib/>
pdiv@xanum.uam.mx



Lineas de Investigación

Audiología
Fisiología perinatal
Instrumentación Médica
Órganos artificiales
Resonancia magnética
Fisiología cardiopulmonar
Potenciales evocados y EEG
Procesamiento de Imágenes
Procesamiento de señales
Inteligencia artificial
Biofísica
Biomateriales
Ingeniería de tejidos
Nanotecnología

UNIVERSIDAD AUTONOMA METROPOLITANA – UNIDAD IZTAPALAPA
DIVISION DE CIENCIAS BASICAS E INGENIERIA