Supercarreteras de la Información

Andrés Albanese Raúl Rojas

International Computer Science Institute
Berkeley, California

1 Introducción

Mucho se ha hablado de las modernas redes de computadoras como columna vertebral del flujo de información que caracteriza a las sociedades postindustriales. Hasta hace pocos años solamente los Estados Unidos y algunos países europeos contaban con una infraestructura adecuada para el transporte masivo de información. Sin embargo hoy en día más y más países en vías de desarrollo adoptan esta tecnología y se incorporan a Internet, la red de redes de cómputo mundial [7]. Nosotros vemos este proceso como un signo positivo de integración de la economía mundial, pero sobre todo, como una gran oportunidad para países como México, Brasil or Argentina que pueden aprovechar este flujo de información haciendo uso de una inversión mínima en relación a las ventajas y oportunidades que dicha información proporciona. Es así, por ejemplo, que la comunidad académica puede accesar las bases de datos y la literatura científica producida en los mejores centros de investigación en el mundo sin retraso alguno. A través de la transmisión de imagenes de video un profesor en Venezuela puede tomar parte en un seminario remoto en la Universidad de Minnesota y puede incluso hacer preguntas. Internet es lo más cercano a la "cabaña mundial" pronosticada hace años por McLuhan.

En la primera parte de este artículo pasamos revista al desarrollo fenomenal del grado de conectividad mundial y a algunas de las aplicaciones que han surgido en el último par de años. Este tipo de aplicaciones demandan velocidades de transmisión varios ordenes de magnitud mayores de lo que hasta hace algunos años era considerado tecnología de punta. Sería utópico pensar que la red de cómputo mundial va a ofrecer el mismo ancho de banda en cada uno de sus segmentos y por eso discutimos en la segunda parte del artículo algunas alternativas que existen para resolver el problema de una red mundial *heterogénea*, es decir, compuesta de segmentos de diferente velocidad y grados de confiabilidad. Los protocolos de transmisión que hasta ahora existen no han abordado este problema, que sin duda se convertirá en una de las cuestiones centrales a resolver si Internet sigue creciendo al mismo ritmo que en los ultimos años.

2 Conectividad mundial

El mapa mostrado en la figura adjunta, producido por Larry Landweber y la *Internet Society*, es una representación gráfica del nivel de conectividad mundial alcanzado hasta julio de 1994. Los países en color negro cuentan con conecciones a Internet que permiten transmitir y recibir correo electrónico así como accesar computadoras y bases de datos remotas. Algunos países solo cuentan con correo electrónico (tonos grises) y algunos otros no participan en Internet, sobre todo países africanos. Como se puede deducir del mapa es solo cuestion de algunos años más para que Internet abarque prácticamente todo el mundo.

Más impresionante aún es la Tabla 1 que muestra el crecimiento anual del número de computadoras (nodos) conectados a Internet. Como se puede observar, hasta julio de 1994 mas de 3 millones de computadoras estaban enlazadas a través de la red de redes. Al ritmo de crecimiento actual, cada mes se integran a la red 128,440 nuevas máquinas o sea 4280 nuevos nodos *diariamente*. Una regla empírica utilizada por los administradores de redes consiste en contar con unos diez usuarios por cada nodo. Eso quiere decir que hasta julio de 1994 había aproximadamente 32 millones de usuarios de Internet en todo el mundo [4].

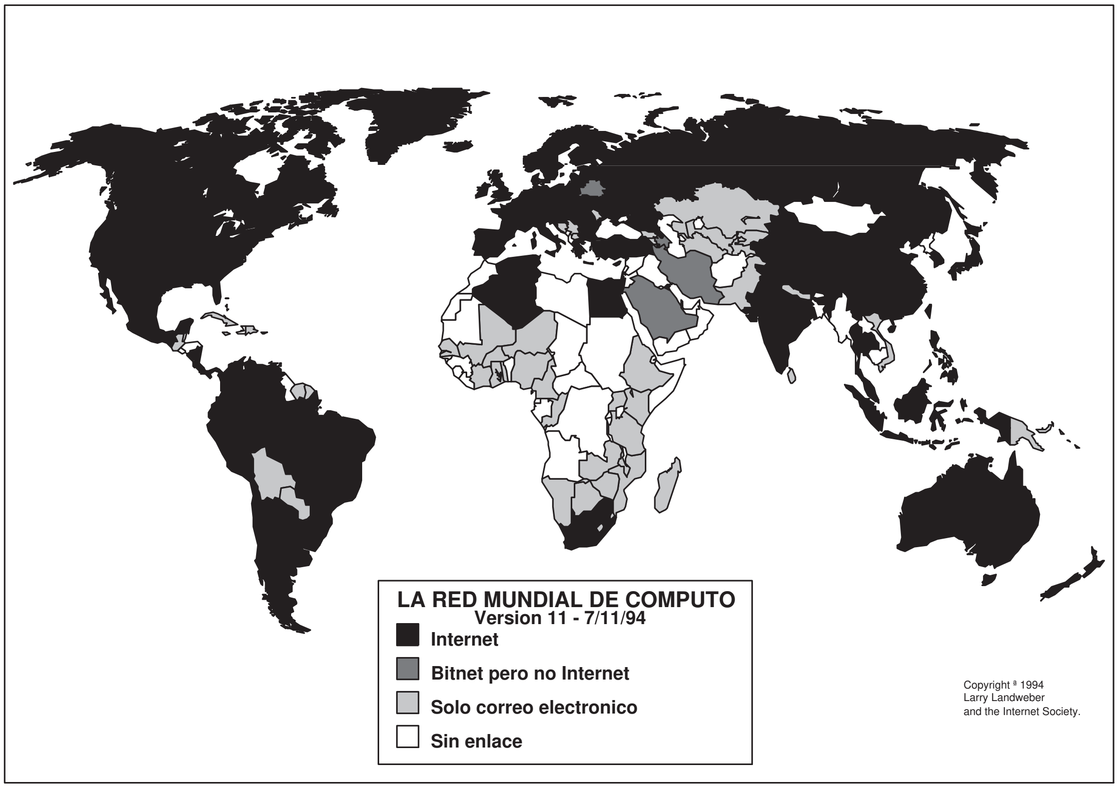

Figura 1: La red mundial de cómputo

Estadísticas recopiladas por los administradores de la red de la *National Science Foundation* de EU (NSFNET) muestran que el tráfico a traves de la red ha ido creciendo de manera vertiginosa desde 1991. La Tabla 2 muestra el número de gigabytes (miles de millones de bytes) transmitidos por NSFNET mes por mes desde marzo de 1991. Como se puede apreciar en la tabla, la cantidad de bytes transmitidos se duplica cada año y es ya mayor a los 13 *terabytes*. Los administradores de NSFNET han calculado que el tráfico total se divide en 36% correspondiente a intercambio de archivos, 18% a correo electrónico, 5% a comunicación interactiva y el resto a otros servicios

Tabla 1: Numero de nodos en Internet

Mes y año	Nodos
08/81	213
08/83	562
10/85	1,961
12/87	28,174
10/89	159,000
10/90	313,000
10/91	617,000
10/92	1,136,000
10/93	2,056,000
07/94	3,212,000

Fuente: Merit Network Inc.

(transmision de video, audio, etc.).

3 Bibliotecas virtuales

La gran oportunidad que Internet le ofrece a los países en vías de desarrollo es la de adquirir información de manera *eficiente, instantánea y a muy bajo costo*. Todos sabemos que hasta ahora las bibliotecas en los países latinoamericanos no han podido ofrecer la gama de servicios y el volumen de información que proporcionan sus equivalentes en países industrializados. Esta situación está cambiando de manera radical. Desde una terminal de computadora en Brasil o Chile se puede ya consultar los catálogos de las bibliotecas norteamericanas o europeas y, más importante aún, se puede transferir documentos a la estación local para imprimirlos posteriormente. Algunos de los servicios disponibles actualmente en Internet son los que a continuación se detallan.

Documentos

Las diversas comunidades de investigación que trabajan sobre un determinado tema se pueden enlazar a traves de correo electrónico común. Cuando un investigador termina un artículo lo pone a disposición de la comunidad,

Tabla 2: Historia del tràfico en NSFNET (en Gigabytes)

	1991	1992	1993	1994
Ene		2256	4782	8609
Feb		2371	5015	9303
Mar	1268	2761	6053	11226
Abr	1402	2848	6219	11587
May	1442	3061	5845	12187
Jun	1244	3274	6195	12466
Jul	1594	3373	6389	12764
Ago	1484	3200	6631	13385
Sep	1769	3315	7022	
Oct	1879	3903	8468	
Nov	1959	4651	8483	
Dic	1956	4372	8283	

Fuente: Merit Network Inc.

copiandolo a una computadora que funciona como depósito de información y notificando a sus colegas a través de la lista de suscriptores a un grupo de interés. De esa manera los resultados recientes de la investigación son accesibles a todos aquellos con connección a Internet. La mayor parte de los artículos publicados hoy en día en revistas especializadas pueden ser copiados de Internet semanas o incluso meses antes de que la revista aparezca impresa. Algunas revistas electrónicas se publican desde hace un par de años como es el caso del *Journal of Artificial Intelligence Research*. En 1995 se publicará en Internet el primer número de la revista *Journal of Universal Computer Science* dedicada a cubrir todos los aspectos de la informática.

Libros

Hasta ahora sólo pocos libros han estado disponibles a través de Internet. Esta situación cambiará de manera radical a partir de 1995 cuando la IEEE comience a almacenar el texto completo de los trabajos presentados en todas las conferencias internacionales auspiciadas por esta sociedad de ingenieros. No será necesario esperar a que aparezcan las memorias de los Congresos,

Tabla 3: Los 40 dominios con el mayor número de nodos de Internet

	Dominio	Nodos		Dominio	Nodos
1	USA (edu)	856234	21	Sudáfrica	15595
2	USA (com)	774735	22	Nueva Zelandia	14830
3	USA (gov)	169248	23	Corea	12109
4	GB	155706	24	Dinamarca	12107
5	Alemania	149193	25	Bélgica	12107
6	USA (mil)	130176	26	Taiwán	10314
7	Canadá	127516	27	Hong Kong	9141
8	Australia	127514	28	Israel	8464
9	Japón	72409	29	Polonia	7392
10	Francia	71899	30	Brasil	5896
11	USA (org)	66459	31	México	5164
12	Holanda	59729	32	Portugal	4518
13	Suecia	53294	33	Singapur	4014
14	Finlandia	49598	34	Chile	3703
15	Suiza	47401	35	Irlanda	3308
16	Noruega	38759	36	Islandia	3268
17	Italia	23616	37	Rusia (su,ru)	3467
18	España	21147	38	Grecia	2958
19	Austria	20130	39	Malasia	1322
20	USA (us)	16556	40	Turquía	1204

Fuente: Merit Network Inc.

muchas veces con meses de retraso. Cualquiera puede informarse de inmediato acerca de los trabajos presentados como si hubiera estado presente.

Por otra parte la biblioteca del Congreso de Estados Unidos inició hace poco un proyecto para capturar cada año el texto completo de miles de los libros que forman su colección. Se tiene proyectado gastar del orden de varios miles de millones de dólares en este proyecto para ir poniendo a disposición de los usuarios, si no todo, al menos una parte importante de su acervo bibliográfico.

Bases de datos

Otra de las areas en las que Internet esta posibilitando la cooperación internacional es en cuanto a bases científicas de datos. Un ejemplo notable son los proyectos que existen en todo el mundo para elaborar un "mapa" de los genes humanos. El carteo de pequeños segmentos de los genes es un proceso sumamente laborioso y que requiere de laboratorios muy bien equipados. Centros de investigación como los Laboratorios Nacionales de Livermoore y Los Alamos en EU han iniciado un proyecto para poner a disposición de los científicos interesados las sequencias descifradas de cromosomas humanos y de otras especies. Esta *Biblioteca Genética*, como se le ha llamado, será de gran importancia para futuros proyectos en los diversos campos de la medicina. Otro ejemplo notable es la base de datos de proteínas mantenida por el Instituto Nacional de la Salud de EU. Es posible accesarla desde cualquier computadora para examinar la sequencia de aminoácidos de alguna proteína conocida. Si se desea, se puede imprimir la imágen en colores de un modelo de barras y esferas de la macromolécula.

Presentaciones multimediales

Con la expansión del tipo de comunicación proporcionado por la interface gráfica conocida como la *World Wide Web* (WWW) no solamente se transmite texto a través de Internet. Imágenes, grabaciones de audio e incluso videos pueden ser accesados desde cualquier nodo de la red. Cualquier periodista puede copiar el ultimo discurso del Presidente Clinton del nodo mantenido por la Casa Blanca y escucharlo en su estación de trabajo. Es posible observar animaciones de experimentos o simulaciones realizadas por otros investigadores. Cuando hace algunos meses se produjo la colisión entre un cometa y Jupiter, la NASA se encargo de recopilar las imágenes capturadas por los observatorios de todo el mundo y de proporcionar simulaciones en video de los efectos esperados por los astrónomos.

4 Redes de alta velocidad

El crecimiento del tráfico de datos sobre las "carreteras" de la información está produciendo el mismo tipo de problemas que aflije a la carreteras reales: congestionamiento de las vías de comunicación hasta llegar a la parálisis total. Al ritmo de crecimiento actual de Internet (duplicación anual del tráfico) sería fácil llegar a una situación en la que la red fuera inutilizable, por lo menos en forma interactiva [2]. Es por eso que existen una serie de proyectos para aumentar las velocidades de transmisión y la capacidad de transporte de Internet. Este tipo de proyectos se conocen con el nombre genérico de *Redes Gigabit*. En Estados Unidos y Europa se han puesto en servicio redes piloto con capacidades de transmision que van desde los 100 a los 600 Mbit/s. En California, por ejemplo, BAGNet es una red piloto que enlaza a compañías y universidades ubicadas en el area de la Bahía de San Francisco. La velocidad de transmisión es tal que permite transmitir imágenes de video en tiempo real para teleconferencias. Está planeado aprovechar BAGNet para implementar laboratorios "virtuales", de tal manera que equipo de laboratorio pueda ser utilizado a distancia.

La red heterogénea

Sería ilusorio pensar que al ritmo actual de desarrollo de Internet todos los países van a poder contar con el mismo tipo de canales de comunicación y anchos de banda. La Tabla 3 muestra claramente que la distribución de équipos de cómputo es muy heterogénea entre los diversos países. Y dentro de cada país hay tambien diferencias entre una zona y otra así como entre los medios de comunicación accesibles para una institución y las demás. Algunas universidades podrán conectarse a través de fibra óptica, otras por medio de microondas y algunas más por línea telefónica. Lo que está surgiendo poco a poco es una red mundial de computo *heterogénea* y uno de los mayores desafíos para los investigadores trabajando en este campo es el de definir procolos de transmisión y algoritmos que permitan operar esta red heterogénea.

Un ejemplo que puede ayudar a comprender la importancia de resolver este problema es el de las transmisiones a través del MBone [3]. El MBone es una red virtual en la que se transmiten mensajes de un punto de la red a todos los demás. Si una universidad desea transmitir un seminario a través de Internet, las imágenes de video y el sonido se transmiten usando los protocolos adecuados y todos aquellos interesados y con acceso al MBone pueden "sintonizar" la transmisión. El flujo de paquetes proveniente de la institución que origina la transmisión no se interrumpe nunca. Si alguno de los nodos de la red pierde paquetes, ya sea porque la computadora no reacciona a tiempo o porque el congestionamiento de las líneas impide su transmisión, no es posible pedir una retransmisión. De alguna forma hay que reconstruir la información aún en el caso de que un cierto porcentaje de los paquetes se haya perdido.

Otro ejemplo sería que se tuviera que eliminar una parte de los paquetes al pasar de una red de cierta velocidad de transmisión a otra de menor velocidad. En el punto de enlace de una red de fibra óptica y otra cuyo medio de transmisión es línea teléfonica convencional, se tendría que eliminar una parte de los paquetes para poder transmitir un seminario en tiempo real. El problema es eliminar un cierto porcentaje de los paquetes sin degradar la señal hasta hacerla ininteligible. Se podría seguir transmitiendo imagenes de video, reduciendo el número de cuadros por segundo o la resolución de las imagenes de manera gradual.

Los primeros hallazgos de experimentos realizados en redes de computadoras es que a grandes distancias y a velocidades altas de transmisión la diversidad en capacidad de procesamiento y de recepción aumenta y con ello las pérdidas. Entre los cientos de usuarios que pueden participar en una teleconferencia, donde todos pueden transmitir y recibir, hay aquellos que tienen la technología más reciente y envían una gran cantidad de información que otras estaciones más pequeñas no pueden recibir e interpretar en su totalidad. En las redes "gigabit" la informacion es segmentada en el transmisor en muchos paquetes de Internet para ser transmitida por la red global y ser ensamblada de nuevo en cada receptor final. Como un porcentaje aleatorio de estos paquetes se pierden o se retrasan y no llegan a su destino, se debe codificar la informacion antes de ser enviada para garantizar que cualquiera que sea la cantidad de paquetes recibidos uno siempre podrá recobrar la información en orden de importancia.

Consideraciones de este tipo son las que llevaron a investigadores del Instituto Internacional de Computación en Berkeley a diseñar un nuevo método de codificación de información adecuado para redes heterogéneas o con pérdidas aleatorias de paquetes.

Transmisión Codificada con Prioridades

El método desarrollado en Berkeley [1] se conoce como PET (*Priority Encoded Transmission*). Consiste en codificar la información con diferentes niveles de redundancia de acuerdo a su importancia. PET utiliza el concepto de dispersión de información propuesto por Shamir y Rabin [6]. La Figura 2 muestra el número de paquetes que se pierden durante una sesión de video transmitida de Palo Alto a Berkeley. El promedio de paquetes perdidos es en este caso de 4.5%. Imágenes de video transmitidas de forma digital utilizan el formato llamado MPEG, el cual se ha convertido en un estándar muy difundido. La pérdida de paquetes provoca que de cuando en cuando ciertas porciones de la imágen no aparezcan o que no se logre sincronizar la reconstrucción de cada cuadro.

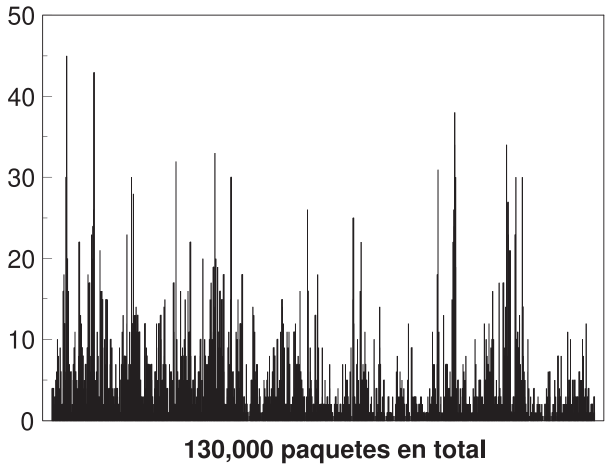

Figura 2: Porcentaje de pérdida de paquetes entre Palo Alto y Berkeley

La informacion de video es comprimida de acuerdo al estándar MPEG el cual elimina toda clase de redundancia entre un grupo de cuadros de video

[5]. Una escena de video consiste en mostrar 30 cuadros por segundo para dar la impresión de continuidad. Como se puede uno imaginar, los cuadros adyacentes son muy parecidos y las computadoras reducen la información que hay que almacenar y transmitir, calculando la diferencia entre los puntos cercanos de un cuadro y los puntos de cuadros contiguos. Esto produce un factor de compresión de casi 100 y hace más facil el almacenamiento, procesamiento y distribución del video. Sin embargo un video sin redundancia es muy susceptible a errores, lo que puede hacer imposible su descompresión, la cual consiste en un proceso de interpolación.

En MPEG se transmite un cuadro inicial (llamado cuadro I) que contiene toda la información necesaria para reconstruirlo y los cuadros siguientes se transmiten indicando solamente su *diferencia* con respecto al cuadro inicial. Los cuadros P (predecidos) se pueden generar teniendo el cuadro I y cierta información adicional. Los cuadros B indican como interpolar entre el último cuadro I y el próximo cuadro P para recuperar otro imagen completa. Ademas de esto un encabezado H (*header*) define la forma en que los cuadros siguientes se deben recuperar, de manera que si se pierde el encabezado todo el resto de la información ya no tiene sentido. Entonces es importante añadirle redundancia a la información comprimida de acuerdo a la importancia de cada cuadro, de manera que la información mas importante pueda tolerar mayores pérdidas que la información de menor importancia.

Una sequencia típica de transmisión en MPEG sería por ejemplo la siguiente: H-I-B-B-B-P-B-B-B-I. Dado que para reconstruir los cuadros B y P se necesita contar con el último cuadro I, estos son los más importantes en MPEG. Si se pierde el cuadro I inicial, la información contenida en los siguientes 6 cuadros B y en el cuadro P no basta para recuperar la escena. La idea que surje en este caso es la de codificar cada cuadro con diferentes niveles de redundancia, por ejemplo:

- el encabezado H con factor de redundancia 10,
- los cuadros I con factor 1.67,
- los cuadros P con factor 1.33
- los cuadros B, los menos importantes, con factor 1.11.

La redundancia total, promediada sobre el total de cuadros, es 23%, es decir se transmiten 23% más bytes que sin utilizar el método PET. Pero aún cuando se llegaran a perder hasta el 90% de los paquetes transmitidos por la red sería posible recuperar el encabezado H, y los cuadros I cuando se llegara a perder hasta el 40% de los paquetes. El resultado es que videos que con este nivel de pérdida son prácticamente irreconstruibles, con PET se pueden recuperar en la computadora. Actualmente se esta trabajando en extender PET para utilizarlo para la codificación de sonido y para otro tipo de formatos multimediales.

¿Cómo funciona PET? Sin entrar en muchos detalles técnicos, basta decir que la idea básica es evitar transmitir primero paquetes conteniendo H, despues paquetes conteniendo el cuadro I, luego paquetes conteniendo un cuadro B, etc., sino *distribuir* la información sobre todos los paquetes transmitidos. Cada paquete debe contener información sobre H y los cuadros I, B y P. La Figura 3 muestra como se debe codificar la información. Supongamos que se quiere transmitir los números $a_1, b_1, \ldots, d_3$ mostrados en el paquete original. Los números a_1 y b_1 son más importantes que a_2, b_2 y c_2, y estos más importantes que el resto del paquete. Los paquetes PET reciben un número sequencial (de 0 a 4 en este ejemplo), de tal manera que se puede saber en que orden van llegando durante la recepción. Para transmitir los números binarios a_1 y b_1 se transmiten cinco números que corresponden a cinco evaluaciones del polinómio $a_1 + b_1x$ (módulo un cierto factor). Basta con recuperar dos de esos puntos para poder reconstruir los números a_1 y b_1, ya que dos puntos determinan de manera unívoca una línea recta. El factor de redundancia es en este caso 2.5 y hasta 60% de los paquetes se pueden perder sin afectar la reconstrucción de ambos números. Si queremos transmitir los números a_2, b_2 y c_2, se pueden transmitir cinco puntos del polinomio $a_2 + b_2x + c_2x^2$. Como basta recibir tres puntos de un polinómio cuadrático para poder reconstruirlo, eso quiere decir que 40% de los paquetes se pueden perder sin afectar la reconstrucción de a_2, b_2 y c_2. La idea es entonces utilizar polinómios de diferentes grados para proporcionar diferentes factores de redundancia y consiguientemente de protección. Como se ve la información de los cuadros MPEG se puede codificar haciendo uso de esta técnica.

PET ha sido implementado en ICSI totalmente en software. Es posible reconstruir videos en tiempo real sin utilizar hardware adicional. Una demostración del método aplicado a una transmisión de video está disponible en la página del World Wide Web de ICSI en Berkeley, California. El reporte técnico en que se describe el método PET está disponible ahí mismo. Este es un ejemplo del tipo de investigación que se está realizando en ICSI para

Paquete original

a_1	b_1	a_2	b_2	c_2	a_3	b_3	c_3	d_3

Polinomios para codificar

$$p_1(x) = a_1 + b_1 x \qquad (\mathrm{mod}\ 2^{16}+1)$$

$$p_2(x) = a_2 + b_2 x + c_2 x^2$$

$$p_3(x) = a_3 + b_3 x + c_3 x^2 + d_3 x^3$$

Paquetes PET

0	$p_1(0)$	$p_2(0)$	$p_3(0)$
1	$p_1(1)$	$p_2(1)$	$p_3(1)$
2	$p_1(2)$	$p_2(2)$	$p_3(2)$
3	$p_1(3)$	$p_2(3)$	$p_3(3)$
4	$p_1(4)$	$p_2(4)$	$p_3(4)$

Figura 3: El método de codificación PET

resolver los problemas planteados por la red heterogénea.

5 Conclusiones

Nosotros consideramos la difusión de redes de comunicación como Internet como una gran oportunidad para los países en vías de industrialización y al mismo tiempo como un reto. La comunidad académica y las compañías ubicadas en estos países cuentan ahora con acceso a un vasto acervo de información que pueden utilizar para sus propios fines. El reto es garantizar que a pesar de la heterogeneidad de la red mundial de cómputo, el nivel de acceso de cada país no se deteriore con cada nuevo salto tecnológico en los países industrializados. No hay que olvidar tampoco que dentro de cada país la red de transmisión es también heterogénea y que existe el peligro de dividir a la sociedad en dos sectores: uno con acceso a la biblioteca virtual mundial, otro sin acceso alguno. A los encargados de la política informática en cada país les corresponde garantizar el desarrollo a largo plazo de la infraestructura de comunicación. A los científicos trabajando en el área de redes les corresponde diseñar métodos para utilizar estos recursos de manera óptima y para garantizar que la heterogeneidad de la red no se convierta en un obstáculo insuperable. Es imposible pronosticar ahora todos los cambios que una difusión tan amplia de información va a provocar en el futuro, pero de algo estamos seguros: hay que participar. Y usted ¿ya está en Internet?

Referencias

[1] Albanese, A., J. Blömer, J. Edmonds, M. Luby (1994), "Priority Encoding Transmission", Reporte Técnico TR-94-039, *International Computer Science Institute*, Agosto.

[2] Claffy, K. C., H.-W. Braun, G. C. Polyzos, "Tracking Long-Term Growth of the NSFNET", *Communications of the ACM*, vol. 37, no. 8, pp. 34–45.

[3] Eriksson, H. (1994), "The Multicast Backbone", *Communications of the ACM*, vol. 37, no. 8, pp. 54–60.

[4] Goodman, S. E., L.I Press, A.M. Rutkowski (1994), "The Global Diffusion of the Internet: Patterns and Problems", *Communications of the ACM*, vol. 37, no. 8, pp. 27–31.

[5] Le Gall, D. (1991), "MPEG: A Video Compression Standard for Multimedia Applications", *Communications of the ACM*, vol. 34, no. 4, Abril, pp. 47–58.

[6] Rabin, M.O. (1989), "Efficient Dispersal of Information for Security, Load Balancing, and Fault Tolerance", *Journal of the ACM*, vol. 36, no. 2, Abril, pp. 335–348.

[7] Rojas, R., S. Castillo, H. Hernández (1990), "Redes de computadoras para la comunidad académica", *Ciencia y Desarrollo*, vol. 16, no. 7.

El Dr. Andrés Albanese es líder del grupo de Redes en el *International Computer Science Institute* en Berkeley. De origen italiano, realizó sus estudios de licenciatura en la Universidad Central de Venezuela, de maestría en la Universidad de Texas y su doctorado en ingeniería electrónica en Stanford. Es autor de muchos artículos en revistas especializadas y ha recibido 13 patentes por invenciones en las areas de redes de comunicación y lásers.
Correo electrónico: aa@icsi.berkeley.edu

El Dr. Raúl Rojas es profesor en la Facultad de Matemáticas de la Universidad Martín Lutero en Halle, Alemania. Estudió matemáticas en la Escuela Superior de Física y Matemáticas del Instituto Politécnico Nacional y obtuvo su doctorado y Habilitación en la Universidad Libre de Berlín. Es autor de diversos artículos en revistas especializadas y autor del libro "Teoría de las Redes Neurales", publicado en 1993 en alemán y por aparecer 1995 en inglés.
Correo electrónico: rojas@inf.fu-berlin.de