Algoritmos + Criptografía + Estructura de Datos

CLCERT: CRIPTOGRAFÍA APLICADA Y SEGURIDAD

Departamento de Ciencias de la Computación, Universidad de Chile.

El CLCERT, grupo de criptografía aplicada y seguridad de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, focaliza su investigación en dos áreas. La primera es Criptografía Aplicada, incluyendo el diseño y estudio de protocolos orientados a mejorar la privacidad de los participantes; sistemas de comunicación anónima segura; votación electrónica verificable, y en general, computación distribuida segura. La segunda es Seguridad Aplicada, principalmente en el estudio de phishing y malware.

El grupo está compuesto por el profesor Marcos Kiwi (Departamento de Ingeniería Matemática, Universidad de Chile), el ingeniero Sergio Miranda (Universidad de Chile); los estudiantes de Doctorado Philippe Camacho, y Julio Quinteros; los estudiantes de Magíster Gaston L'Huillier, Patricio Seguel; y los actuales estudiantes memoristas Alonso González, Renata Faccilongo, Rodrigo Porras, Francisca Merino, y Felipe Troncoso. El director del grupo es el profesor Alejandro Hevia (Departamento de Ciencias de la Computación, Universidad de Chile).

En términos de investigación en las áreas mencionadas, nuestro grupo colabora con investigadores como Tamara Rezk (INRIA Sophia Antipolis, Francia) y Alfredo Viola (Universidad de la República, Uruguay), aunque a la lista de coautores se ha incorporado recientemente a Gilles Barthe (IMDEA Software Institute, Madrid), Bogdan Warinschi (University of Bristol, Inglaterra), Richard Weber y Sebastián A. Ríos (Departamento de Ingeniería Industrial, Universidad Chile), y Daniele Micciancio (Universidad de California San Diego, EE.UU.). Las publicaciones del grupo en conferencias y/o seminarios incluyen:

Sergio Miranda, Alejandro Hevia y Marcos Kiwi.

- IEEE Computer Security Foundations Symposium (CSF, IEEE Computer Society).
- Intl. Conference on Cryptology and Information Security (Latincrypt, LNCS Springer).
- IEEE International Conference on Intelligence and Security Informatics (ISI, IEEE Press).
- Intl. Conference on Information Security (ISC, LNCS Springer).
- Privacy Enhancing Technologies (PETS, LNCS Springer).
- Intl. Conference on Theory and Applications of Cryptographic Techniques (Eurocrypt, LNCS Springer).

En cuanto a publicaciones en revistas, éstas incluyen artículos en:

- Theoretical Computer Science (TCS, Elsevier).
- IEEE Selected Areas in Communication (JSAC, IEEE Press).
- Journal Computer and System Sciences (JCSS).

ACM Transactions on Information System Security (TISSEC, ACM Press).

Gonzalo Navarro:

ALGORÍTMICA Y TEORÍA DE LA INFORMACIÓN

Departamento de Ciencias de la Computación, Universidad de Chile.

La mayor parte de mi investigación reciente se enfoca en la intersección del área de Algoritmos y Estructuras de Datos, y la de Teoría de la Información. El objetivo es desarrollar estructuras de datos que utilicen poca memoria y alcancen una eficiencia comparable a la de las estructuras clásicas. Esto tiene interés por la diferencia creciente entre el desempeño de los distintos niveles de la jerarquía de memoria, y la disponibilidad de memorias mayores en todos esos niveles. El utilizar menos espacio hace que una

estructura de datos pueda residir en una memoria más pequeña y rápida, con lo cual puede resultar que, a pesar de requerir más operaciones que su versión clásica, la estructura compacta resulte ser más rápida. En particular, cuando la estructura compacta cabe en memoria RAM y la clásica necesita utilizar el disco, la diferencia de desempeño puede ser de varios órdenes de magnitud. Estas estructuras también son de interés en dispositivos de capacidad limitada, como celulares, sensores, routers, etc.

Muchos de los desarrollos se concentran en el área de bases de datos de texto, donde se han conseguido avances espectaculares en la última década, tales como los llamados auto-índices. Estos representan un texto en un espacio cercano a su versión comprimida, pero dentro de ese espacio ofrecen búsqueda indexada, es decir de tiempo sublineal en el largo del texto. Las aplicaciones de este tipo de índices en áreas que necesitan manejar colecciones gigantescas de secuencias, como la bioinformática, recién están comenzando a explorarse. También hay varios resultados importantes para bases de datos de textos en lenguaje natural y en recuperación de información, donde la Web es un ejemplo obvio de la necesidad de utilizar el espacio en forma eficiente.

Como suele ocurrir, la investigación en esas áreas de aplicación ha llevado progresivamente a identificar problemas más básicos, donde se requieren estructuras de datos compactas para manejar árboles, secuencias, grillas, sumas parciales, grafos, relaciones binarias, permutaciones, y muchas otras. A su vez, los nuevos resultados en estructuras de datos básicas han dado lugar a resultados aplicados a problemas como indexación comprimida de colecciones XML, de objetos geográficos, de grafos Web y redes sociales, de índices invertidos, árboles de sufijos, y otras estructuras de relevancia para áreas como bioinformática, recuperación de información, sistemas de información geográficos, máquinas de búsqueda Web, etc.

Investigo también en otras áreas como Búsqueda por Similitud, Búsqueda Secuencial e Indexada en Texto, y Algoritmos y Estructuras de Datos en general. Con respecto a la primera área cabe destacar la creación en 2008 de la conferencia SISAP (Similarity Search and Applications), para focalizar la investigación que se realizaba en los aspectos algorítmicos de la búsqueda por similitud.

Software

Estoy convencido de la importancia de prestar atención tanto al componente teórico como al práctico en la investigación en computación. Si falta lo primero se cae en heurísticas sin fundamento teórico, que no se comprende por qué funcionan ni en qué contexto dejarán de funcionar, y se alejan de la ciencia, donde es fundamental comprender los porqués. Si falta el componente práctico, al menos en computación se cae fácilmente en investigación teórica sin ninguna conexión con la realidad ni probable utilidad.

En algoritmos, esto se traduce en que debe haber un componente de diseño, uno de análisis teórico y uno de experimentación. Intento además que los desarrollos experimentales se conviertan en prototipos públicamente disponibles, para que sean usados con fines de investigación, docencia, y a veces incluso comerciales. Como consecuencia, hay bastante software públicamente disponible en http://www.dcc.uchile.cl/gnavarro/software, y otros sitios mucho más sofisticados desarrollados por mí o por mis alumnos, tales como http://pizzachili.dcc.uchile.cl y http://www.recoded.cl. No creo en la utilidad de las patentes en computación, y en general estoy en contra de patentar ideas abstractas tales como algoritmos.

Colegas y alumnos

Trabajo mucho mejor con un buen partner que solo. Como resultado, casi todas mis publicaciones son con coautores y tengo una larga lista de colaboradores en todo el mundo. Un listado de muestra, considerando sólo los más recientes y recurrentes, y tomando

Gonzalo Navarro.

sólo los jefes de grupos de investigación, incluye a Jérémy Barbay (Universidad de Chile), Nieves Brisaboa (Universidad de la Coruña, España), Edgar Chávez (Universidad Michoacana, México), Paolo Ferragina (Università di Pisa, Italia), Johannes Fischer (KIT, Alemania), Veli Makinen (Universidad de Helsinki, Finlandia), Sebastián Maneth (NICTA, Australia), Simon Puglisi (RMIT, Australia), Luís Russo (Universidad Nova de Lisboa, Portugal), y Kunihiko Sadakane (Universidad de Tokyo, Japón).

A esta lista debo agregar a mis ex-alumnos y ex-posdocs, con la mayoría de los cuales mantengo una relación de colaboración: Joaquín Adiego (PhD, U. Valladolid, España), Diego Arroyuelo (PhD, Yahoo! Research Chile), Benjamin Bustos (MSc, U. de Chile), Rodrigo Cánovas (MSc, U. de Chile), Francisco Claude (MSc, doctorando U. Waterloo, Canadá), Antonio Fariña (PhD, U. Coruña, España), Karina Figueroa (PhD, U. Michoacana, México), Travis Gagie (posdoc, U. Aalto, Finlandia), Rodrigo González (PhD, Index Technologies, U. de Chile), Gilberto Gutiérrez (PhD, U. Bío-Bío), Rodrigo Paredes (MSc y PhD, U. de Talca), y Diego Seco (posdoc, U. Coruña, España).

Finalmente, la mayoría de mis alumnos y posdocs actuales también son importantes colaboradores para la investigación: Carlos Bedregal (PhD), Ana Cerdeira (PhD, U. Coruña, España), Violeta Chang (PhD),

Cecilia Hernández (PhD), Norma Herrera (PhD, U. San Luis, Argentina), Sebastián Kreft (MSc), Fernando Krell (MSc), Susana Ladra (PhD, U. Coruña, España), Miguel Ángel Martínez (posdoc, U. Valladolid, España), Eliana Providel (MSc), Nora Reyes (PhD, U. San Luis, Argentina), Carina Ruano (MSc, U. San Luis, Argentina), y Daniel Valenzuela (MSc).

Busco completar la formación de mis alumnos enviándolos a congresos internacionales, en particular para que presenten los artículos en los que participan, e incluso a veces como oyentes si se lo han ganado con la calidad de su trabajo. Organizo siempre que puedo los "Miércoles de Algoritmos", reuniones donde se exponen nuevas ideas, se analizan artículos de la literatura, se practican charlas, etc.

Publicaciones

Una lista de las conferencias relevantes donde he publicado más recientemente (últimos cinco años) incluye: ACM-SIAM SODA, STACS, ESA, ICDE, ACM SIGIR, CPM, ISAAC, DCC, LATIN, SPIRE, SEA, ACM-SIAM ALENEX, ACM RECOMB, y MFCS.

Una lista similar de revistas incluye: ACM Trans. Alg. (TALG), ACM Trans. Inf. Sys. (TOIS), ACM Trans. Web (TWEB), ACM Comp. Surv. (CSur), ACM J. of Exp. Alg. (JEA), Algorithmica, Theor. Comp. Sci. (TCS), Softw. Pract. Exp. (SPE), J. Comp. Biol. (JCB), Inf. & Comp. (IC), e Inf. Retr. (IR).

Desde 2006 he publicado 35 artículos en revistas internacionales y 62 en conferencias internacionales.

Otras distinciones

He sido Chair del Comité de Programa de siete congresos internacionales, estoy en el Comité Editorial de dos revistas internacionales (ACM JEA e IR), y he dado charlas plenarias en ocho congresos internacionales, entre otras muchas de menor importancia. En 2008 recibí el Premio Scopus, de Elsevier y Conicyt, en el área de Matemáticas, Computación e Ingeniería, a los autores más prolíficos de Chile.

Jérémy Barbay junto a alumnos del curso Alice.

Jérémy Barbay:

ANÁLISIS ADAPTATIVO: MÁS PRECISO, RÁPIDO Y PEQUEÑO

Departamento de Ciencias de la Computación, Universidad de Chile.

Mi nombre es Jérémy Barbay. Nací y estudié en Francia, trabajé seis años en Canadá, y llevo tres años en Chile. Soy matemático por formación, teórico en Ciencias de la Computación por vocación, y usuario de computadores por hobby. Mi tema principal de investigación se relaciona con refinar las técnicas de análisis de "rendimiento de los algoritmos" y del "espacio de las estructuras de datos". Esto, dentro de otros intereses como el mejoramiento de las técnicas de docencia, la teoría de la evolución y el diseño de mecanismos sociales en la red.

Mi ejemplo favorito de un problema práctico que requiere un análisis más fino de complejidad que el tradicional es la "intersección de arreglos ordenados", que ocurren, por ejemplo, cuando los motores de búsqueda como Google tratan de resolver las consultas de sus usuarios. Dadas tres palabras u,v,w, correspondientes a tres arreglos ordenados U,V,W con las referencias a las páginas asociadas con u,v,w, respectivamente, se pide responder la consulta "u,v,w"; es decir, se requiere buscar las referencias que los tres arreglos U,V,W tienen en común.

Un análisis tradicional agrupa las instancias por tamaño, y, además, analiza el comportamiento de los algoritmos en términos de su comportamiento "en el peor caso". Este análisis intenta encontrar cuál es la peor clase de consulta que podría ser dada como entrada al algoritmo. Por ejemplo, en el caso del problema mencionado antes, una instancia de entrada que es muy compleja de tratar por cualquier algoritmo "razonable" es la siguiente: U={1,2, 4,5, 7,8 }, V={1, 3,4, 6,7, 9} y W={ 2,3, 5,6, 8,9}. De hecho, para validar su resultado sobre esta instancia, cualquier algoritmo de intersección tiene que indicar nueve comparaciones, esto es, el tamaño del conjunto.

Sin embargo, tales instancias son muy artificiales, y, además, bastante diferentes a las que ocurren realmente en la práctica. El problema con el análisis de complejidad "en el peor caso" es, por tanto, que no logra diferenciar los distintos tipos de algoritmos con respecto a su "performance" sobre instancias que efectivamente ocurren en la práctica. Esto se debe a que las instancias prácticas son más fáciles y el análisis

tradicional es demasiado pesimista. Por otro lado, un análisis más fino de complejidad identificaría un parámetro adicional como, por ejemplo, la cantidad de comparaciones que se necesita por demostrar el resultado. Este medida es muy alta para el ejemplo que mostramos arriba, pero es mucho más baja para una instancia como: U={1,2,3, 7,8,9}, V={ 4,5,6,7,8,9} y W={1,2,3,4,5,6 }, que ocurre más usualmente en la práctica. Es por eso que un análisis fino puede ser un mejor predictor de la performance práctica de un algoritmo.

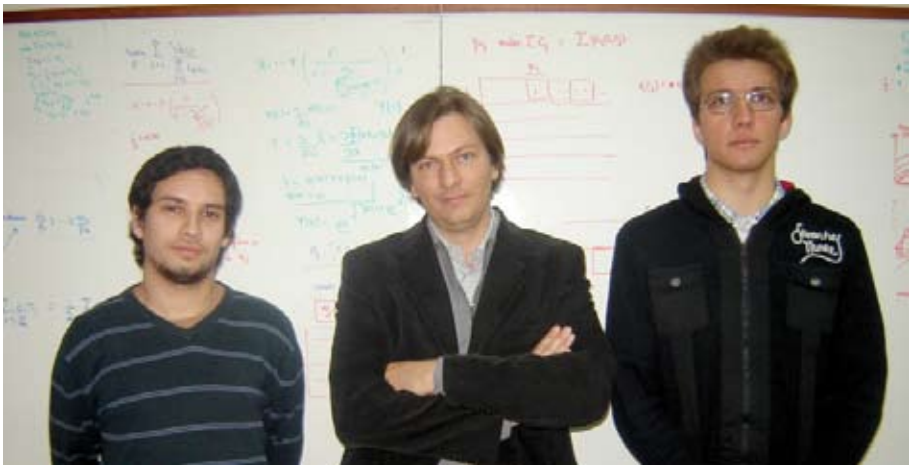
El problema de intersección es solamente un ejemplo, ya que también trabajo en variantes del problema de intersección, en algoritmos de ordenamiento, en problemas de geometría computacional, y en el análisis y diseño de estructuras de datos comprimidas, en colaboración con Gonzalo Navarro y Carlos Bedregal. Estos resultados fueron presentados en las conferencias de más alto nivel del campo, como SODA y FOCS, a razón de una o dos al año, y publiqué algunos en revistas prestigiosas como "ACM Transaction of Algorithms" y "Algorithmica". Enseño estas técnicas en cursos avanzados y también en cursos básicos de Ciencia de la Computación: el objetivo es enseñar a los alumnos este tipo de análisis lo más temprano posible.

José Rafael Correa:

ENTRE INVESTIGACIÓN DE OPERACIONES Y TEORÍA DE LA COMPUTACIÓN

Departamento de Ingeniería Industrial, Universidad de Chile.

Tras estudiar Ingeniería Matemática en la Universidad de Chile, partí a hacer un Doctorado en Investigación de Operaciones al MIT. En este marco me dediqué a trabajar en el diseño y análisis de algoritmos de aproximación para problemas de optimización combinatorial NP-difíciles. Lo que se busca en este contexto son algoritmos eficientes (a tiempo polinomial) que entreguen soluciones con una garantía de aproximación. Así pues, la pregunta fundamental que está detrás es: ¿Qué podemos hacer con un problema si restringimos el tiempo de ejecución a ser polinomial?



Omar Larré, José Rafael Correa y Charles Thraves.

De regreso en Chile desarrollé también un interés por el estudio de los algoritmos "en línea" donde la idea es diseñar algoritmos que obtengan buenas soluciones, a pesar de que el input no es conocido de antemano sino que se revela en el tiempo. La pregunta en este caso es: ¿Qué podemos hacer con un problema si restringimos la información disponible?

Finalmente, en los últimos años me he interesado cada vez más en la Teoría Algorítmica de Juegos. Esta área, que ha visto muchos desarrollos en la última década, estudia sistemas distribuidos, en que diversos agentes toman decisiones en forma simultánea. Un tema que me ha interesado particularmente en este ámbito es el estudio del llamado "Precio de la Anarquía", el cual cuantifica la pérdida de optimalidad de un sistema descentralizado respecto de una solución coordinada. En este contexto, el paradigma de la eficiencia computacional, en el caso de los algoritmos de aproximación, o el paradigma de información de los algoritmos en línea se reemplazan por el paradigma de la coordinación entre los distintos agentes del sistema.

En este último tema trabajo actualmente con dos alumnos del Magíster de Gestión de Operaciones del Departamento de Ingeniería Industrial de la Universidad de Chile: Omar Larré y Charles Thraves. Omar estudia el problema de ruteo en un grafo donde múltiples agentes buscan llegar a su destino en el menor tiempo posible. La dificultad es que los links sufren congestión, por lo que cuando muchos agentes quieren usar un determinado arco del grafo, los tiempos de todos los usuarios de ese arco se ven afectados en forma negativa. Por otra parte, Charles trabaja en un problema de políticas de precio con consumidores estratégicos. En este contexto, una firma quiere desarrollar un plan para vender un determinado producto, en que el precio irá decreciendo en el tiempo (liquidación). Los consumidores, entonces, se enfrentan a la disyuntiva de comprar hoy a precio normal o esperar a que éste baje. Pero esperar puede significar que el producto ya no esté disponible.

Como se desprende de lo anterior, mi área de investigación está en la frontera entre Investigación de Operaciones y Teoría de la Computación. Así pues, con frecuencia participo en reuniones de ambas comunidades. En febrero de 2010, por ejemplo, asistí a un workshop de computación en Dagstuhl, Alemania, mientras que en septiembre di una charla en el Departamento de Gestión de Operaciones de la Escuela de Negocios de la Universidad de Nueva York.

Lo mismo ocurre con las publicaciones. He publicado en conferencias de computación teórica como ICALP, IPCO, SODA, STOC y WINE, así como en revistas de Investigación de Operaciones como Operations Research y Mathematics of Operations Research.