Spring/Summer 2010

iIS Update

資訊科學簡訊

Newsletter of the Institute of Information Science
Academia Sinica, Taiwan

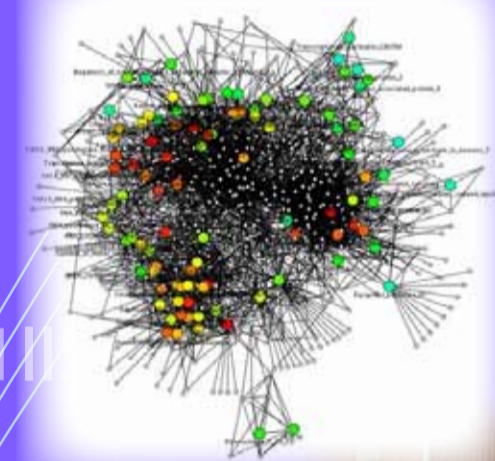## What's Inside:

Institute of Information Science

The Institute of Information Science (IIS) was established in 1982. We currently have 38 full time research faculty, 30 post-doctoral research fellows, and slightly more than 300 research associates and specialists. Our research is composed of eight specialized laboratories: Bioinformatics, Computer Systems, Information Processing and Discovery (iPAD), Multimedia Technology, Natural Language and Knowledge Processing, Network Systems and Services, Programming Languages and Formal Methods and Computation Theory and Algorithms.

IIS is not a degree-granting academic institution, with the important exception of the international graduate program in Bioinformatics, under the auspices of the Taiwan International Graduate Program (TIPG) in Academia Sinica. This PhD program was established in 2002, and has enrolled about 31 students over the last seven years.

Many of our research fellows hold joint faculty appointments at top universities in Taiwan. This allows our institution to play a very significant role in training and fostering advanced research talent in the IT industry and academia in Taiwan.

DIRECTOR:
Dr. Yew, Pen-Chung

DEPUTY DIRECTORS:
Dr. Chen, Meng Chang
Dr. Ko, Ming-Tat

GROUP COORDINATORS:
Dr. Sung, Ting-Yi
    Bioinformatics Lab
Director Yew, Pen-Chung
    Computer Systems Lab
Dr. Chen, Meng Chang
    Information Processing and Discovery (iPAD)
Dr. Liao, Hong-Yuang Mark
    Multimedia Technology Lab
Dr. Hsu, Wen-Lian
    Natural Language and Knowledge Processing Lab
Dr. Ho, Jan-Ming
    Network Systems and Services Lab
Dr. Chuang, Tyng-Ruey
    Programming Languages and Formal Methods Lab
Dr. Ko, Ming-Tat
    Computation Theory and Algorithms Lab

Message from the Director

The Institute of Information Science (IIS) at Academia Sinica is one of Taiwan's premier research institutes. It has been engaged in fundamental and applied research in information sciences and related technologies since its establishment in 1982. Currently, IIS has about 40 full time research fellows and slightly more than 300 research assistants and graduate students. They are conducting research on a wide variety of subject areas, ranging from bioinformatics, multimedia, and computer/network systems, to Chinese natural language and knowledge processing, theories, and algorithms. As a government-sponsored research center, the institute tries to distinguish itself from other research universities and industrial research laboratories in Taiwan by focusing more on interdisciplinary, team-based research projects that explore mid- to long-term fundamental and applied research issues.
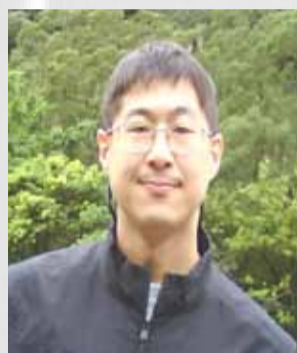
In this inaugural issue, and future biannual issues, of the IIS newsletter, we invite you to share with us the exciting events and research activities happening in the institute, find out about the people involved and their latest news, and explore some fun ideas on various things crossed our research horizons.

The purposes of the newsletter series are manifold. First of all, it provides a platform for the IT research community and the general public, particularly overseas readers, to find out about our research efforts and major results; the awards and recognition that IIS personnel have received through their dedicated work; as well as the activities and events that have occurred, or are being planned, in IIS. It also provides a means to reach out to a large number of IIS alumni, reconnect with them, and learn about their recent activities and news. Hence, we encourage them to write to us and share their latest news as well as their accomplishments. We also hope that, through the newsletters, we can encourage and recruit more talented researchers to join us and become part of the IIS team. Indeed, we hope that the newsletter will serve as a useful forum for all stakeholders to exchange ideas, news and information. We therefore welcome your inputs, comments and contributions.

Honors and Awards

◆ **Dr. Jane W. S. Liu is awarded the Golden Penguin Award 2009**
◆ **Director Yew, Pen-Chung received the Information Science Honorary Medal 2009 from IICM (Institute of Information & Computing Machinery)**
◆ **Dr. Chen, Sheng-Wei received the K. T. Li Distinguished Young Scholar Award 2009 from ACM Taipei/Taiwan Chapter**
◆ **Dr. Chen, Ming-Syan received the Outstanding Research Award 2009 from National Science Council (NSC)**
◆ **Dr. Hsu, Wen-Lian received the Pan Wen Yuan Distinguished Research Award 2010**
◆ **Dr. Yang, De-Nian received the Pan Wen Yuan Research Exploration Award 2010**

◆ Dr. Yang, Bo-Yin and Dr. Yang, De-Nian received the 2010 Academia Sinica Career Development Award with the following projects respectively

Efficient Cryptographic Computing

This project aims to increase security and privacy for everyone by exploring how fast cryptographic primitives can be run and broken. A tenet of modern cryptography is: "any cryptosystem can be broken given enough money and time". Even though some proclaim that "security is more important than speed", there is a perpetual trade-off between security and efficiency, and we often must maximize encryption optimization, more so than for cryptanalysis. For example, today's web commerce depends on the security level of RSA (and other cryptosystems), which in turn depends on how fast we can implement RSA and detact attacking algorithms (e.g., Number Field Sieve). Conversely, more efficient encryption yields better security levels for the same price, or cheaper components for the same security. RSA, as a heavyweight, is often unusable in low-resource scenarios such as Ubiquitous Computing. NATO hence will switch to ECC (Elliptic Curve Cryptography) from RSA. This project will also explore Elliptic-Curves-based algorithms, as well as Post-Quantum ones (those expected to survive Quantum Computers). Dr. Yang and his team has worked with and optimized for many different algorithms and architectures (including PCs, micro-controllers, and video cards). In the past dozen half years, Dr. Yang's lab holds the best Taiwanese record at

Selective Network Coding for Internet Data Storage

This project proposes selective network coding for Internet data storage. Current Internet data storage servers store each individual data item. In contrast, this project enables more information to be stored in storage servers by mixing the data with the proposed selective network coding. Traditional network coding technology is designed for multicast/broadcast environments and applications. However, most Internet traffic is still unicast today. Moreover, traditional network coding technology is not compatible with current Internet standards and routers. Therefore, the impact of network coding technology for the Internet is limited. In contrast, this project proposes encoding Internet data storage according to user access patterns, an approach that is practical and fully compatible with current Internet standards and routers. This requires first addressing the fundamental problem of selecting the data items to be encoded in single cache storage so as to increase the cache hit ratio. Next, the project proposes implementing multiple data access, which enables each data item to serve more clients, compared to traditional techniques that serve only the clients who query the item. For on-line scenarios, the project proposes two novel cache management schemes, elastic access and adaptive adjustment, to increase the flexibility of cache allocation and reduce the number of cache misses. Finally, the project addresses the encoding of data items in distributed storage servers with various topologies.

◆ Dr. Liao, Hong-Yuan received the 2010 Academia Sinica Investigator Award with the following project
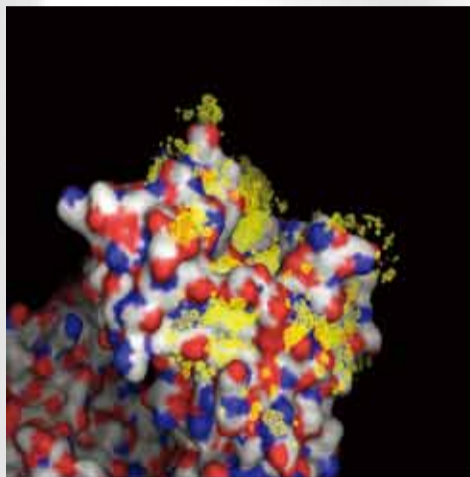
TELDAP-related Video Processing Techniques

The Multimedia Technologies group at the Institute of Information Science, Academia Sinica, recently focused its research into two main areas: multimedia signal processing and multimedia applications. The goal of research on multimedia applications is to develop useful systems that integrate existing technologies that were either invented by other groups or previously developed by the multimedia technology group. Over the past decade, the group has developed a number of novel techniques for image/video processing, 3-D graphics processing, machine learning, and multimedia security and protection. Some of these have been successfully transferred to related industries. At present, the group is engaged in two major projects: the Taiwan e-Learning and Digital Archives Program (TELDAP) and the Construction of Vision-based Intelligent Environment project (VBIE), which are sponsored, respectively, by the National Science Council (the second 5-year term, from 2007-2011) and the Ministry of Economic Affairs (the second 4-year term, from 2009-2012). TELDAP handles a large amount of digitized aged pictures and aged films. Since the collected pictures/films were taken a long time ago by analog devices, the quality of these pictures/videos is very poor. In the next five years, the group intends to address 4 major areas: (1) video inpainting; (2) video annotation and retrieval using heterogeneous features; (3) stabilization of digitized aged films; and (4) people counting in digitized aged films.
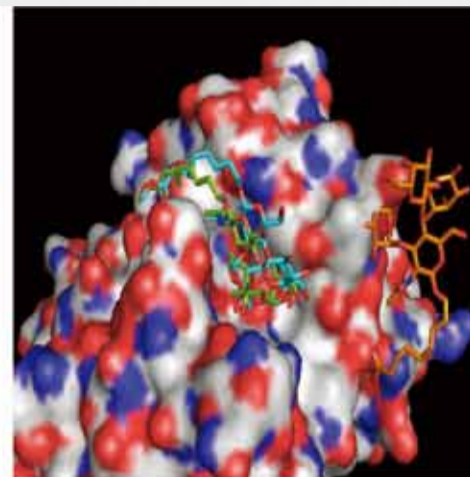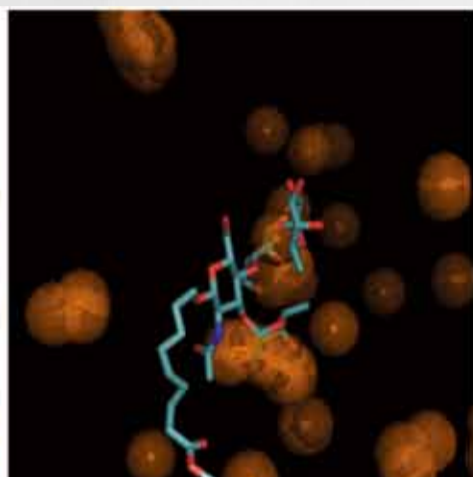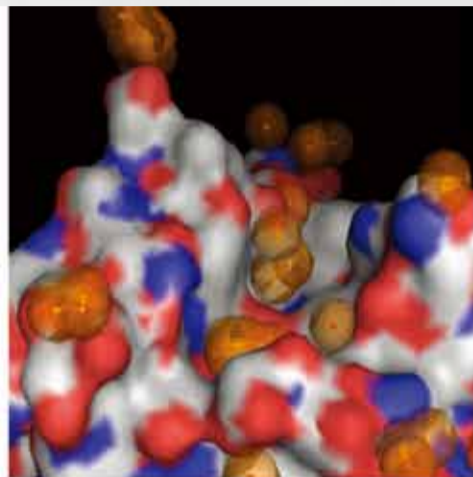
Lab Profile

Bioinformatics Laboratory

Protein-compound binding prediction.

The goal of the Bioinformatics Lab is to invent and develop new and urgently needed bioinformatics tools, and hopefully to facilitate new and significant biological discoveries.

Currently, their research is focused on the following areas:

1. Systems biology: (a) Regulatory mechanisms and networks: the study of the regulatory mechanisms in yeast and higher organisms (e.g., humans), including finding novel miRNAs, identifying transcription factor binding sites, and discussing the functionality of degenerate positions in TFBSs and the regulatory rule of adjacent genes. (b) Network biology: deciphering the cellular interactomes of viruses (e.g., herpes viruses and hepatitis C virus) and infected hosts, with the aim of identifying the protein complexes hijacked by pathogen proteins and finding ways to block the mechanism of infection. Furthermore, key proteins and motifs in the virus-host network will be identified and hopefully provide virologists with novel targets to study their involvement in viral pathogenesis. (c) Metagenomics: based on whole genome shotgun sequencing data, the lab will develop an integrated platform, including various databases, gene expression analysis, proteomic results and phylogenetic reconstruction to achieve a comprehensive view of microbials. (d) Transcriptome analysis: the integrative bioinformatic platforms for high-throughput sequencing of non-model species aims at supporting transcriptome studies of specific domestic species including A.hallerissp.

gemmifera, Formosan black bear, Formosan cypress, silvergrass, pteridophyte, and coral. Biologists will analyze the tran-scriptome of these non-model species, though their genomes have not yet been fully assembled. Overall, designing effective and efficient algorithms to deal with related problems is a challenge.
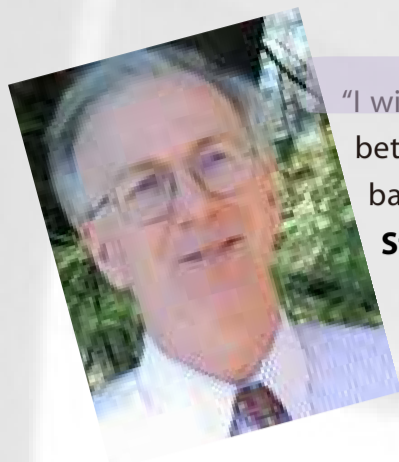
2. Proteomics: (a) Mass-spectrometry based proteomics: to facilitate biomarker discovery, the Bioinformatics Lab will develop methods and tools for large-scale protein identification and protein quantitation with post-translational modifications based on high-throughput mass spectrometry data, so that differentially expressed proteins between normal cells and tumor cells can be identified. (b) Protein prediction problems: using a machine learning approach, they will work on various protein prediction problems, including protein structure prediction. The lab is particularly interested in transmembrane proteins, since they count for about 50% of existing biomarkers, and subcellular locali-zation prediction, since the function of a protein depends on its localization site.

The Bioinformatics Lab is currently partici-pating in various thematic projects, including the Bioethanol from Cellulosics project (NSC project hosted by Dr. Chi-Huey Wong, President of Academia Sinica), the C4 Rice Project (Aca-demia Sinica project hosted by Dr. Wen-Hsiung Li, Director of Biodiversity Research Center, Academia Sinica), Bioinformatics Core for Genomic Medicine and the Biotechnology Development project (National Research Pro-gram for Genomic Medicine (NRPGM) hosted by Dr. I-Shou Chang), and the Construction and Integration of Biological Resources project (NSC project, led by Director General Min-Liang Kuo, Department of Life Sciences, NSC, and Vice President Ching-Fong Chang, Nation Taiwan Ocean University).

They inaugurated the Bioinformatics Ph.D. Program in the Taiwan International Graduate Program, Academia Sinica. As of Fall 2009, two students received their Ph.D. degrees and 34 students enrolled, including local students and foreign students from America, Asia, and Europe, including Canada, Germany, India, Malaysia, the Philippines, Slovakia, the United States, and Vietnam.

Complex Network   Topological Analysis   Sub-network by hubs

Distinguished Lecture Series

"I will present, AlarmNet, a novel testbed system for health care that uses a two-way flow of data and analysis between front-end body networks, intermediate environmental sensing and communication networks, and back-end context aware protocols that are tailored to residents' individual living patterns." **Dr. John A. Stankovic. January 4th, 2010. Challenges in Mining Cyber-Physical Systems**

"Cyber-Physical Systems are integrations of computation, networking, and physical processes. Both static and dynamic data, in various types and formats, are generated and collected from the interconnection of cyber elements and physical elements." **Dr. Lionel M. Ni. March 8th, 2010. Challenges in Mining Cyber-Physical Systems**

"We will talk about how FTR pursues disruptive technologies, and also highlight several ways Intel Labs engages university and national lab researchers." **Dr. Andrew A. Chien. April 6th, 2010. Bold, Edgy Technologies for the Future of Computing**

" In this talk, we introduce the broadband network environment in Taiwan, and describe a Next Generation Network (NGN)/IP Multimedia Subsystem (IMS) deployed by Chunghwa Telecom. " **Dr. Yi-Bing (Jason) Lin. May 4th, 2010. Deploying Applications in Next Generation Networks**

Distinguished Lecture Series 2010

August	Dr. Chin-Hui Lee	Georgia Tech	Speech and Speaker Recognition
September	Dr. Jun'ichi Tsujii	Tokyo University	Natural Language
October	Dr. Jean Ponce	Ecole Normale Superieure	Computer Vision
November	Dr. Tom Mitchell	Carnegie Mellon University	Machine Learning
December	Dr. Bart Preneel	Katholieke Universiteit Leuven	Information Security

~Information is subject to change, please check our website for up-to-date schedule~

Activities

Visitors

December 2nd, 2009 visit by INRIA France. From left to right: Dr. Chuang, Tyng-Ruey; Director Yew; Dr. Dominique Scotteau; Dr. Emmanuelle Platzgummer; Dr. Arnaud Lalo; Vivien Lee

May 24th, 2010 visit by a group of students from the First-Girls Information Study Club (FGISC). Hosted by Dr. Chen, Sheng-Wei and Dr. Hsiao, Ching-Teng.

November 6th, 2009 visit by UIUC. From left: Director Yew; Head Andrea Cangellaris; Prof. Milton Fen; Ms. Beth Katsinas

Left: March 23rd, 2010 visit by students from Binus International School Serpong, Indonesia. Hosted by Dr. Allan Lo.

Below: December 8th, 2009 visit by Huazhong University of Science and Technology, China

Annual Field Trip

"And the destination is..."
(May 27-29, 2010)

After an intense competition between two destinations, a travel package including three days and two nights in Kenting was voted as this year's annual field trip - by 3 votes. Wow~what a vote!!! Congratulations to those who made it and to those who didn't...well, think on the bright side, you didn't get any sunburns.

Bon Voyage !!

Ladders to British Consulate.

Little Bay Beach in Kenting.

Tien-Ho Temple in Kaohsiung.

Taiwan International Graduate Program welcome party.

Taiwan International Graduate Program orientation.

IIS annual banquet 2009. Right: Open dance.
Below: Lucky Draw.

Formosan Summer School on Logic, Language, and Computation (FLOLAC '10)

Date: Monday, June 28th, 2010
Friday, July 9th, 2010
Time: Monday to Friday 9:00am to 5:00pm
Location: School of Professional and Continuing Studies, National Taiwan University, second floor, room 207
Phone: 02-23620502

FLOLAC provides a series of courses designed to aim at preparing the students with the knowledge needed to carry on research in foundational computing science. The courses during even-numbered years cover advanced topics in programming languages, while those on the odd-numbered years focus on model checking and program verification. The theme of this year is "Advanced Programming Languages: Semantics, Analyses, and Tools". FLOLAC has been held annually since 2007.

For more information, please visit their website at http://flolac.iis.sinica.edu.tw/flolac10/

December 14th to December 15, 2009. IEEE International Conference on Service-Oriented Computing and Applications SOCA 09
Left to right: Kwei-Jay Lin(Conference Chair), Raj Rajkumar(Keynote Speaker), Tharam S. Dillon (Keynote Speaker), Ming-Syan Chen (Director of CITI, Academia Sinica), and Chung-Ta King (Program Chair).

第十六屆 高性能計算及嵌入式編譯技術研討會
THE 16TH WORKSHOP ON COMPILER TECHNIQUES FOR HIGH-PERFORMANCE AND EMBEDDED COMPUTING
May 27th and 28th, 2010
Institute of Information Science, Academia Sinica

The 16th Worshop on Compiler Techniques for High Performance and Embedded Computing, hosted by the Institute of Information Science, Academia Sinica on May 27 and 28, 2010.

Project

Don't Feel Safe Hiking? Check Out the YushanNet !

The YushanNet project was initially funded by the Yushan National Park in 2008. The objective of this project is to develop a reliable and robust system for hiker tracking, searching, and rescuing in Yushan National Park. YushanNet is now funded by the National Science Council as a three-year deployment project with the goal of operationalizing the service within the national park system.

YushanNet is deployed on the Yushan Peak Trail, which is a 10.9km long trail with a 1302m altitudinal shift. In YushanNet, each tourist is asked to carry a small device called a `Black Box' (about 160 grams in weight) when he/she passes the entrance. The Black Box has a GPS receiver, Zigbee radio, and 10 KB of memory. When hikers encounter each other in the park, their black boxes will automatically exchange their IDs and locations, and store the received information in memory. When a hiker reaches one of the YushanNet base-stations, which are installed at various hot spot locations. The

station will upload all stored information to the Internet.

The YushanNet project has three working groups. The system deployment group is led by Professor Polly Huang (National Taiwan University), and it aims to 1) increase the scale of the deployment to 100 nodes by 2010 and 200 nodes by 2011; 2) deploy three base stations on the trail (i.e., at the entrance, West-Summit Pavillion, and Paiyun Lodge) in 2010 and two additional ones at the Yuan Lodge and Meng-Lu Pavilion in 2011; and 3) provide a pleasant web-based front-end for hikers and the general public to trace their hiking progress. The technology R&D group is led by Dr. Ling-Jyh Chen (Academia Sinica) and Professor Kun-Chan Lan (National Cheng Kung University), and its mission is to 1) tightly integrate the GPS receiver, accelerometers, and other necessary sensors on the WSN platform; 2) design proper duty cycle control algorithms specially tailored for mountaineering scenarios

Yushan landmark.

so as to better prolong the lifespan of the Black Box; and 3) improve the correctness, efficiency, stability, and reliability of the system. Finally, the innovative service group is led by Professor Hao-hua Chu (National Taiwan University) with a focus on 1) providing Yushan National Park with informative data analysis to develop advanced tourism services; and 2) working closely with the park to design brand-new e-tourism services.

Overall, this project aims to provide 1) a real-world and functional WSN system; 2) gain first-hand experience in the design, development, deployment, and maintenance of WSN systems; 3) make a positive impact on society and increase WSN visibility worldwide; and 4) stimulate interest in WSN among the research community and society as a whole.

The results of the YushanNet project may be extended to a variety of other applications, such as wildlife tracking, scientific monitoring, landslide and debris flow monitoring, and disaster response networks. For more detailed information about this project, please visit the project website at http://nrl.iis.sinica.edu.tw/YushanNet/, or contact YushenNet via e-mail at YushanNet@gmail.com.

Yushan Peak Trail.

Yushan project team members.

Building a Map of Wiring Diagrams for Olfactory Computation in the Drosophila (a type of fruit fly) Brain

**Collabration project between IIS
and National Tsing-Hua University
PI: Wen-Liang Hwang**

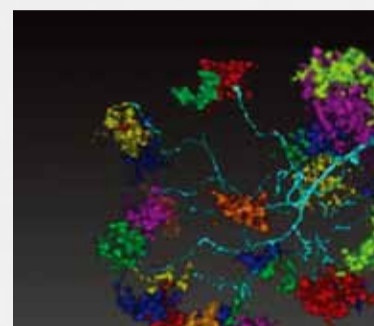## Specific Goals

The overall goal of this integrated project is (i) to survey the AL local circuits and (ii) to understand the basic computation mechanism in the AL. For this goal, the team proposes four sub-projects. Sub-projects 1 (PI: Prof. Yung-Chang Chen, NTHU) and 2 (PI: Prof. Henry Horng-Shing Lu, NCTU) focus on data analysis by using imaging processing and statistical methods. They aim to analyze high-resolution anatomic images, provided by sub-project 3 (PI: Ann-Shyn Chiang, NTHU & AS), to extract the information necessary for biologists to derive meaningful relationships in building a computational model, which is the main target of sub-project 3. The main goal of the sub-project 4 is to distribute, present, and protect the data from the model using a database format through the internet (PI: Wen-Liang Hwang, AS). The schematic diagram that demonstrates the integrations among the four sub-projects is shown here.

Their interdisciplinary approach will greatly enhance their chance to be ahead of their international competitors and have great impact for understanding how the brain computes and turns reception into perception.
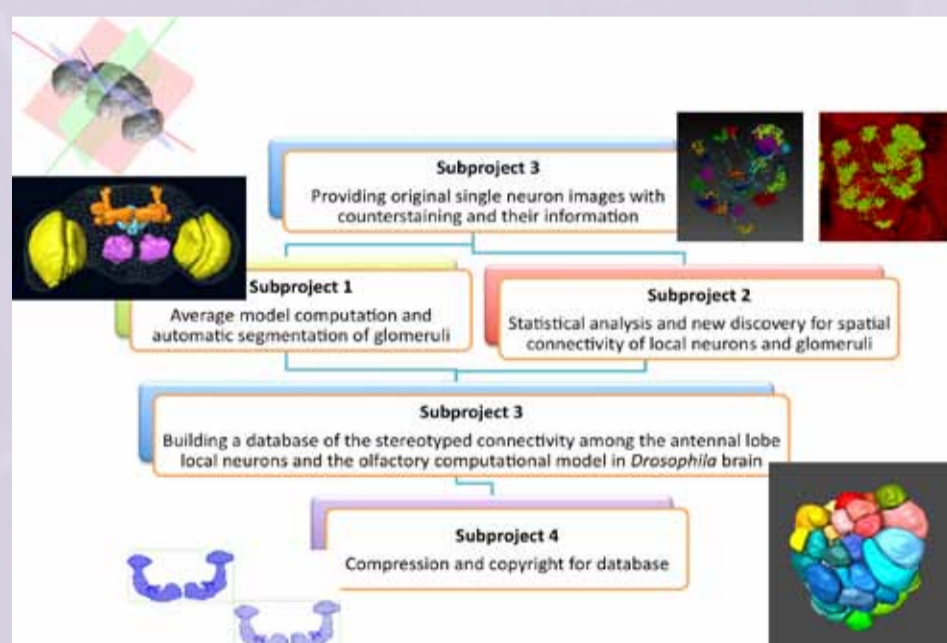
Innovation and Competitive Edge

The competitive edge of this team is that they have the highest resolution data of the Drosophila brain in the world. One of their PIs and his team has developed a set of novel bioimaging technologies for high-resolution 3D reconstruction of biological structures, a significant accomplishment for any other research groups in the world (Liu and Chiang, 2003). This unique technology breakthrough leads to unprecedented high-resolution 3D imaging of the brain circuits at single neuron resolution. This sets an entry barrier for other neuroscientists in the same field. Initial application of the developed imaging technologies has allowed Chiang's group to discover stereotyped neural connections for olfactory computation and new structures involved in long-term memory formation in the Drosophila brain (Lin et al., 2007; Wu et al., 2007).

Anticipated Achievements

The goal of this project is to build a model of olfactory computations in the AL of the Drosophila brain; in other words, a hardwired map including all the connections of all players within the AL, on which one can deduce real olfactory computational ways to fit the functional outputs: either the activities of neural circuits or the behaviors of the individual flies. To the best of our knowledge, if successful, this map will become the first example of brain hardwiring with known functions to satisfy the dream of artificial intelligence.

To achieve the above goal, Sub-project 1 will segment the images of glomeruli in the AL and build a standard model with image processing techniques. Meanwhile, Sub-project 2 will statistically analyze the parameters extracted from the segmented single neuron images and develop a probabilistic connectivity map. Afterwards, sub-projects 1 and 2 will work together to build the connectivity map of glomeruli and neuron in the AL. Sub-project 3 will analyze from the glomeruli and local neuron connectivity information and provide a preliminary model for the olfactory computational circuitry of the Drosophila brain. This rare and precious information of olfactory computation will be represented, protected, and distributed on a database by the collaboration between sub-project 3 and sub-project 4.

Subproject 3 — Providing original single neuron images with counterstaining and their information

Subproject 1 — Average model computation and automatic segmentation of glomeruli

Subproject 2 — Statistical analysis and new discovery for spatial connectivity of local neurons and glomeruli

Subproject 3 — Building a database of the stereotyped connectivity among the antennal lobe local neurons and the olfactory computational model in *Drosophila* brain

Subproject 4 — Compression and copyright for database

Efficient Cryptographic Computing (story continued from page 2)

conferences sponsored by the IACR (International Association for Cryptologic Research), the sole appearances at CHES (Cryptographic Hardware and Embedded Systems) and FSE (Fast Software Encryption). There is no other author from Taiwan at the flagship conferences other than their esteemed colleague Dr. Lu. Dr. Yang's team latest research is running algebraic attacks on video cards, accepted to CHES 2010, partly thanks to the career development award. They hope to continue as respected top speed demons in the near future.

Spotlight

Dr. Sheng-Wei Chen (also known as Kuan-Ta Chen) is an assistant research fellow at the Institute of Information Science and the Research Center for Information Technology Innovation (joint appointment) of Academia Sinica. He received his Ph.D. in Electrical Engineering from National Taiwan University in 2006, and his B.S. and M.S. in Computer Science from National Tsing Hua University in 1998 and 2000, respectively. Prior to taking his academic path, he was known as a programmer specializing in Windows and system programming, a technical writer of four books, a technical lecturer of various programming courses, and a shareware developer.

His research interests include Internet and Multimedia Quality of Experience (QoE) management, Internet measurement, VoIP, network security, and online games. Many of his recent works focused on the human factors in network systems, including QoE measurement, user perception and behavior modeling, and QoE-aware system design. He received a Best Paper Award (with Ieng-Fat Lam and Ling-Jyh Chen) in IWSEC 2008 and K. T. Li Distinguished Young Scholar Award from ACM Taipei/Taiwan Chapter in 2009.

Dr. Chen's general research area is computer networking, and one of his major interests is improving users' network experiences and perceptions of network quality. Computer networks are almost ubiquitous, and the concept of "anytime-anywhere" communication is rapidly becoming a reality for millions of people. However, people must still invest time and effort learning about and

Someone once asked Prof. H.J. Eysenck (a famous psychologist): "when a research brings you satisfying results, how do you feel?"

His answer is: "The feeling is just like (the joyfulness of) a cat acquires the cream, which is hard to describe. I think it's a substantial and wonderful feeling, and you will feel everything is so delightful. I am often somewhat surprised when what I expected finally realized, as I always think it's almost impossible.

However, it is the truth now, it merits to celebrate. Thus, you will be happy for yourself, and feel happy about the nature, for the whole world. You will feel the life is wonderful, and worth to continue to live in the world."

adapting to this relatively new phenomenon of networking. His goal is to make it a seamless part of everyday life.

He focuses on developing new Internet services and making the medium more user-friendly. His current research interests span the following areas:

1) Internet Quality of Service: He hopes that network systems will become smarter and more responsive to people's needs. For example, if people find the network quality is unacceptable when attending a video conference or playing an online game, the systems will "perceive" this situation and know how to adjust the parameters and re-balance the network resources to mitigate the problem. Designing such smart systems is challenging. Thus, he is investigating how to estimate and quantify human perceptions and reactions, and then use these measures in the design of computer systems.

Dr. Chen hopes
that network systems
will become smarter
and more responsive to
people's needs.

---

2) Network Security: Dr. Chen hopes that, in the not too distant future, people will not have to constantly update antivirus tools or

"Sheng-Wei loves cats." The above picture is his cats, Mao-Mao and Ding-Ding.

Would you like to know the feeling of "a cat that got the cream?"

This is an open invitation from Dr. Chen to anyone who intends to devote to academic research.

worry about trojan horses, computer viruses, and frauds when they surf on the Internet. He envisages a time when users' network experiences will not be ruined by threats to privacy and acts of piracy; for example, the deletion of documents by malicious software, the theft of credit card details and online game accounts, and the unauthorized publication of sensitive information, such as how often a user links to dating websites. Thus, how to detect and prevent malicious/fraudulent activities on the Internet is also one of his research areas.

3) Online Gaming: Although online gaming has become one of the major Internet activities, when playing online games, users are often bugged by serious network lags, or they must tolerate seeing the characters of cheaters wearing sparkling magical armor fighting mighty dragons while their own characters can only wear secondhand scale armor and defeat wild dogs, or they may experience being shot by bot-controlled characters with all-the-time perfect accuracy while playing first-person shooting games. Moreover, some players may find that their characters' valuable virtual goods have been ripped off by crackers when they log into a game. Dr. Chen has teamed up with online game companies in order to solve these problems related to performance, fairness, and security from the aspect of game design mechanisms and human behavior.

You can find more details about him at http://www.iis.sinica.edu.tw/~swc or just google "Sheng-Wei Chen" or "Kuan-Ta Chen" on the web.

Great Ideas

A Survey of Binary Search

By Shin-Cheng Mu

Hoare logic and Dijkstra's weakest pre-condition appear to be surprisingly little known among students. Therefore I'd like to demonstrate their usefulness by showing something about binary search that you might not know.

Given a sorted array of N numbers and a key, the task is to either locate, in $O(\log N)$ time, the position where the key resides in the array, or report its absence. You would not expect it to be a hard programming task. Yet J. Bentley noted in Programing Pearls that surprisingly few professional programmers managed to implement the algorithm without bugs at their first attempt. You have to try it to believe!

"If you think you know everything about binary search but have not read N. van Gasteren and W. Feijen's note on The Binary Search Revisited, you should."

Van Gasteren - Feijen Approach

Van Gasteren and Feijen pointed out a surprising fact: binary search does not apply only to sorted lists! The association with sorted lists, they believe, is in fact a major educational blunder.

They considered a more general problem: let M and N be integral numbers with $M < N$, and let Φ be a predicate such that $\Phi(M, N)$ holds, with some additional constraints to be given later. The task is to find l such that

$$M \le l < N \wedge \Phi(l, l+1)$$

This is the program:

```
{M < N ∧ Φ(M, N)}
l, r := M, N
    {Inv: M ≤ l < r ≤ N ∧ Φ(l, r),  Bound: r - l}
; do l + 1 ≠ r →
        {l + 2 ≤ r}
    m := (l + r)/2
    ; if Φ(m, r) → l := m
    [] Φ(l, m) → r := m
```

```
    fi
    od
{M ≤ l < N ∧ Φ(l, l + 1) }
```

To verify the correctness of the algorithm, the reader is invited to check that:

● the loop invariant Inv: $M \le l < r \le N \wedge \Phi(l, r)$ is established by $l, r := M, N$;
● if Inv holds before entering the loop, it still holds after one iteration of the loop;
● the bound $r - l$ gets strictly smaller after each iteration of the loop. When it equals one, the loop has to terminate;
● and if the loop terminates, $l + 1 = r$ and Inv together imply $M \le l < N \wedge \Phi(l, l + 1)$, the postcondition we want!

We need another side condition, however: at least one of the guards of the if statement must be always satisfied, that is:

$$\Phi(l, r) \wedge l < m < r \Rightarrow \Phi(l, m) \vee \Phi(m, r) \quad (*)$$

What predicates satisfy (*)? Examples given by van Gasteren and Feijen include:

● $\Phi(i, j) = a[i] \ne a[j]$ for an array a. They suggested using this as the example when introducing binary search.
● $\Phi(i, j) = a[i] < a[j]$,
● $\Phi(i, j) = a[i] \times a[j] \le 0$,
● $\Phi(i, j) = a[i] \wedge a[j]$, *etc.*

Searching in a Sorted Array

To search for a key K in an ascending-sorted array a, it seems that we could just pick this Φ:

$$\Phi(i, j) = a[i] \le K < a[j]$$

and check whether $a[i] = K$ after the loop. However, we can not establish the precondition $a[0] \le K < a[N]$!

Van Gasteren and Feijen's solution is to add two imaginary elements to the array: for $a[0..N]$, let $a[-1] = -\infty$ and $a[N] = \infty$, which is equivalent to using this Φ:

$$\Phi(i, j) = (i = -1 \vee a[i] \le K) \wedge (K < a[j] \vee j = N)$$

which still satisfies (*) if a is sorted.

Do not worry about the idea "adding" elements to a. Inv implies that $-1 < m < N$, thus $a[-1]$ and $a[N]$ are never accessed, and a needs not be actually altered. They are just there to justify the correctness of the program. It also enables us to handle possibly empty arrays.

Van Gasteren and Feijen pointed out a surprising fact ...

Bentley's Program

Bentley's program for binary search in Programming Pearls can be rephrased as below:

```
l, r := 0, N - 1
; do l ≤ r →
    m := (l + r) / 2
    ; if a[m] < K → l := m + 1
    [] a[m] = K → found := true; break
    [] K < a[m] → r := m - 1
    fi
    od
; found := false
```

I would like to derive this program in class, since it appears to be more popular. However, to relate the test $a[m] < K$ to $l := m + 1$ I will have to bring in the fact that a is sorted in an earlier stage of the development. Thus it is harder to put it in a more general picture.

For several reasons I used to believe that Bentley's program could be preferred, for example, it seems to shrink the range more effectively, assigning l and r to $m + 1$ and $m - 1$, rather than m. On a second thought I realised that it might not be true. Variable l can be assigned $m + 1$ because the possibility of $a[m] = K$ is covered in another case with an early exit, and r is assigned $m - 1$ because this algorithm represents an array segment with an inclusive right bound.

The story is not finished yet. Bentey's algorithm exits the loop using a *break* once K is found. Is it really more efficient? Why did a Google engineer recently claimed that almost all binary search implementations still contain a bug? For more info., check my blog! http://www.iis.sinica.edu.tw/~scm/

The Institute of Information Science (IIS) at Academia Sinica, Taiwan, R.O.C. invites all qualified candidates to apply for the positions of junior and senior research fellows of all ranks (equivalent to the ranks of tenure track assistant, associate and full professors in a regular academic department without teaching responsibility) in all areas of Computer Science. In particular, candidates in the areas of computer systems, machine learning and natural language processing are strongly encouraged to apply.

Academia Sinica is a national academic research institution in Taiwan that conducts research on a broad spectrum of subjects in science and humanities. IIS is committed to high quality research in computer and information science and engineering. In addition to research funding supported by Academia Sinica, external funding through government agencies and industry-sponsored institutions is also available.

Full-time research fellows are free to set their own research directions. IIS currently has about 40 full-time research fellows and close to 300 full-time post doctoral fellows and research assistants. The areas of their current research include Systems Technology, Bioinformatics, Multimedia, Natural Language and Knowledge Processing, Network and Theoretical Computer Science and Parallel Processing.

All Candidates should have a Ph.D. degree in computer science or closely related fields with a strong research and publication record. Senior candidates must demonstrate strong leadership, and have an international reputation evidenced by publications, patents, industrial experiences, or other academic and scholarly achievements. Salary is commensurate with qualifications.

All candidates should send detailed curriculum vitae, and at least three letters of recommendation to

Dr. Pen-Chung Yew, Director
Institute of Information Science
Academia Sinica
Nankang 115, Taipei, Taiwan
Tel: 886-2-2788-3799 ext.1729
Fax: 886-2-2782-4814
E-mail: recruit@iis.sinica.edu.tw

Fluency in Chinese is an advantage, but not required. For additional information about IIS, please visit http://www.iis.sinica.edu.tw.