

[2004]九州大学情報基盤センター年報 : 2004年度

<https://doi.org/10.15017/4776948>

出版情報 : 九州大学情報基盤センター年報. 2004, 2005. 九州大学情報基盤センター
バージョン :
権利関係 :



第5章 イベント・プロジェクト等

5.1 韓国・アジアとの次世代インターネットに関する研究

岡村耕二 (九州大学情報基盤センター, Email: oka@cc.kyushu-u.ac.jp)

5.1.1 日韓拠点事業

九州大学情報基盤センターは、2003年度より日本学術振興会が支援する拠点大学交流のうち、インターネット分野で韓国韓国科学財団 (KOSEF) と実施している「次世代インターネット技術のための研究開発と実証実験」の日本側の拠点大学として、同じ韓国側の拠点大学である忠南大学とともに、本プロジェクトの拠点として、日韓の次世代インターネット技術に関する研究開発を進めています。このプロジェクトに参加している大学、研究機関は九州大学、忠南大学が拠点大学と呼ばれるのに対して協力大学と呼ばれます。日本側の協力大学としては、早稲田大学、九州産業大学、長崎大学、大分大学、佐賀大学、宮崎大学、京都大学、国立情報学研究所、東京大学、大阪大学、慶応大学、九州工業大学、広島大学、広島市立大学、情報通信研究機構、筑紫女学園大学、北海道大学、九州情報大学、長崎総合科学大学が参加しており、韓国側の協力大学としては、韓国科学技術大学校、ソウル大学校、高麗大学校、漢陽大学校、慶北大学校、慶熙大学校、全南大学校、韓国情報通信大学校、光州科学技術院、成均館大学校、梨花女子大学校、建国大学校、済州大学校、国立癌センター、韓国科学技術研究院、忠北大学校、順天郷大学校、慶南大学校、韓国先端網協会、釜慶大学校、釜山大学校、全北大学校、世宗大学校、慶尚大学校が参加しています。このように、本プロジェクトは日韓の主要な大学は全て参加しており、次世代インターネットの研究開発を推進しています。

本プロジェクトには現在以下に示す7つのテーマがあります。このプロジェクトのメインテーマは、次世代インターネット技術ですが、この次世代インターネットを用いた応用分野のテーマもたくさん盛り込んで、関係する研究者を広く取り込めるようにしています。

1. 次世代インターネット基盤技術の研究開発
2. インターネットにおける情報セキュリティ技術の研究開発
3. インターネットを介した仮想現実空間構築技術の研究開発
4. デジタルライブラリのための次世代インターネット基盤技術の研究開発
5. グリッドの基盤技術及びアプリケーションに関する研究
6. 高度マルチメディアデータ通信の研究開発と e-Learning・遠隔教育への応用
7. 次世代インターネットを用いた遠隔医療の開発と臨床応用

現在は、3年目ということもあり、それぞれの研究テーマでは韓国の研究者とともに順調に研究が進められています。

5.1.2 日韓高速インターネット

さて、九州大学が日韓拠点事業の拠点となっているのは、地理的に韓国に最も近い大学であるということもありますが、拠点プロジェクトが始まる前に日韓を高速に接続するインターネットの構築への貢献というの也能げることができます。この韓国とのインターネットに関する研究交流は、九経連が音頭をとって開始された KJCN という日韓光ファイバ (図 5.1) 敷設の計画が契機になり始まりました。

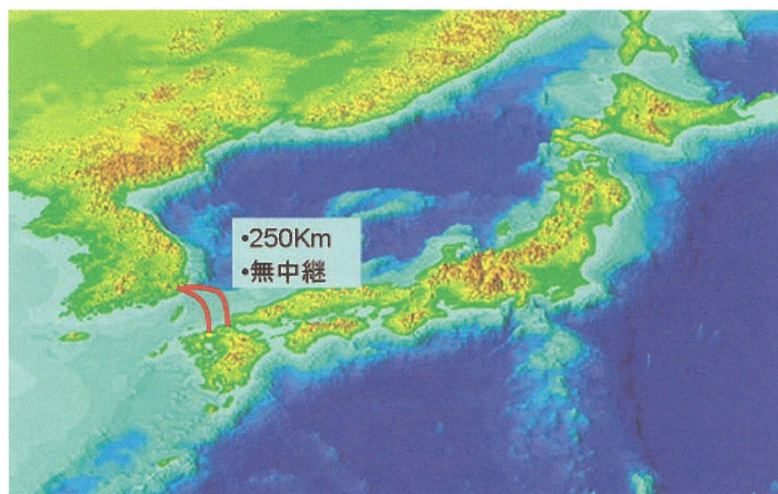


図 5.1: KJCN (Korea Japan Cable Network)

日韓光ファイバは、福岡県と釜山を無中継で接続するもので、間でアンプをメンテする必要がないので、比較的安価な運用が可能です。そのため提供価格も安い、ユーザである我々から見れば国際回線を安価で利用できることが期待できます。これは福岡と韓国の距離の近さを典型的に示している例だと思います。それで、2001 年春にこの日韓光ファイバを研究ネットワークとして利用しないかという九州電力の方のありがたいお言葉から 2 年に及ぶ茨の道が始まったのでした。それで、この茨の道の話はまた今度の機会、もしくは別の方にお任せるとして、2003 年 1 月 (正式には 4 月) から日韓の研究ネットワークのインフラとしてこの日韓光ファイバを利用することが可能になりました。帯域はギガビットでまさに次世代インターネットのインフラが日韓の間にもできあがりしました。この日韓の研究用のインターネットは色々な呼ばれ方があるのですが、本稿では「玄海インターネット」と呼ぶことにします。説明するまでもないと思いますが、玄海灘を通るインターネットというわけです。ちなみに、玄海に相当する英語はなく、韓国語では Hyunhae (ヒョネ) といいます。よって、この玄海インターネットを面倒をみるプロジェクトとして、玄海プロジェクトというのを非公式に作っていますが、これはその後 Genkai/Hyunhae Project になりました。

さて、日韓光ファイバを研究ネットワークとして利用する、つまり、玄海インターネットを実現するために必要なものはたくさんありますが、その中で何が一番重要かというと、それは当然予算です。その予算をなんらかの方法で獲得する必要があったのですが、結局、日本、韓国の政府から頂くという方針でそれを実現しました。その過程において、「日韓にギガビット級のインターネットがあれば何ができるか」についてが最も重要なテーマで、それについて日韓の研究者の間で議論が続けられました。ある程度まとまったら、それを日韓の両政府に持って行って説明するというのを繰り返すのです。その議論は先にも書いた通り約 2 年間続いたわけですが、結局、これが現在の日韓のネットワーク研究者交流の基盤となりました。ところで、福岡空港から韓国の仁川空港への飛行時間は 50

分で、羽田空港よりも近いのです。また、インターネットで、九大からソウル大学まで ping コマンドによるパケットの往復時間は 20 ミリ秒未満なのに対して、東京大学までは 20m 秒以上かかります。いままであまり気がつかなかったのですが、こういう客観的な数字をみてもその近さを感じます。

拠点大学の話は急にやってきました。玄海インターネットが稼働をはじめた 2003 年 1 月の終りに早稲田大学の後藤茂樹先生から九州大学でインターネット分野の拠点大学になれないか電子メールが来たのでした。しかし、学振が日韓でインターネットに関する拠点大学交流を始めようとしたきっかけは、学振が我々の日韓の深いインターネットに関する交流を知っていたからではなく、学振の拠点大学交流プロジェクトでインターネットに関するものがないということからでした。いずれにしてもこれはいい話でしたので、できる限り努力しますというお返事をして、玄海インターネットの茨の道がやっと終わったすぐあとにまた茨の道がはじまりました。もっとも、後藤先生が九州大学にこれを依頼されたのは玄海インターネットの実現に深く関わっていたというのを評価されてということだそうで、まさに茨の道がまた茨の道と呼んだことになったのでした。

5.1.3 実証実験の紹介

本拠点プロジェクトでは遠隔講義、日韓民間テレビ局 (TNC,PBC) の生放送コンテンツの共有実験をはじめ、すでにいくつかの実用的な実証実験を行なっていますが、ここでは 2005 年 5 月 14 日に韓国で開催された、韓国医学会 全国大会でのデモンストレーションを紹介したいと思います。韓国医学会 全国大会は、日本のそれを同じく、何年かに 1 度開催される、医学会を代表する非常に大きな重要なイベントの一つです。この全国大会中に、会場のホテルと、ソウル大学病院、梨花大学病院、九州大学病院を超高速インターネットで接続し、韓国国内外の医師によるパネルディスカッションのようなものが行なわれました。この時には画像としては無圧縮の DV (Digital Video, デジタルビデオ) が用いられ、低遅延で本当に高品質な映像が用いられました。また、ソウル大学病院から会場のホテルに HDV (High Definition Video, 超高解像度映像) を用いた、手術映像の通信のデモも行なわれました。

この韓国医学会全国大会でのデモは結局、大成功に終り、韓国 KBS (日本の NHK に相当) でも、大々的に取り上げられ放映されました。図 5.2 に KBS テレビで放映された様子を示しますが、4 地点を結んだパネルの様子で九州大学 大学病院の清水先生も見えます。なお、テレビのアナウンサー、なお、テレビのアナウンサーのコメントの内容を要約すると、次のようになります。

「外国にいる有名医者の助言を聞きながら実際の手術を国で受けるのが可能でしょうか? 夢だけで思われたこんな遠隔手術の道が遂に開かれました。執刀医と国内医療陣、日本九州大学病院医療陣までリアルタイムで繋がれて手術画面を見ながら熱っぽい討論をすることができることは超高速インターネットが送る高画質立体画面のおかげです。この技術を利用すれば国内医療陣が海外有数の医療陣と自問を取り交わす遠隔手術ができるようになるはずです。」

さて、この実証実験では、4 地点の DV 通信をまとめる処理などは、九州大学情報基盤センターで行なわれていました。また、本プロジェクトテーマの一つである「次世代インターネットを用いた遠隔医療の開発と臨床応用」は、月に 1 回以上のペースで遠隔医療教育を日韓で開催しています。もちろん、技術的には、次世代インターネット基盤技術に取り組んでいる先生が協力しています。このような実績にいる技術の積み重ねと体制が整っているからこそ、このような一般の人にも認めもらえる大規模なデモを成功させることができました。いずれにしても、このような日韓での実証実験が行なえるようになっていることは、本プロジェクトが順調に進行していることを証明していると思います。



図 5.2: KBS テレビでの放映の様子

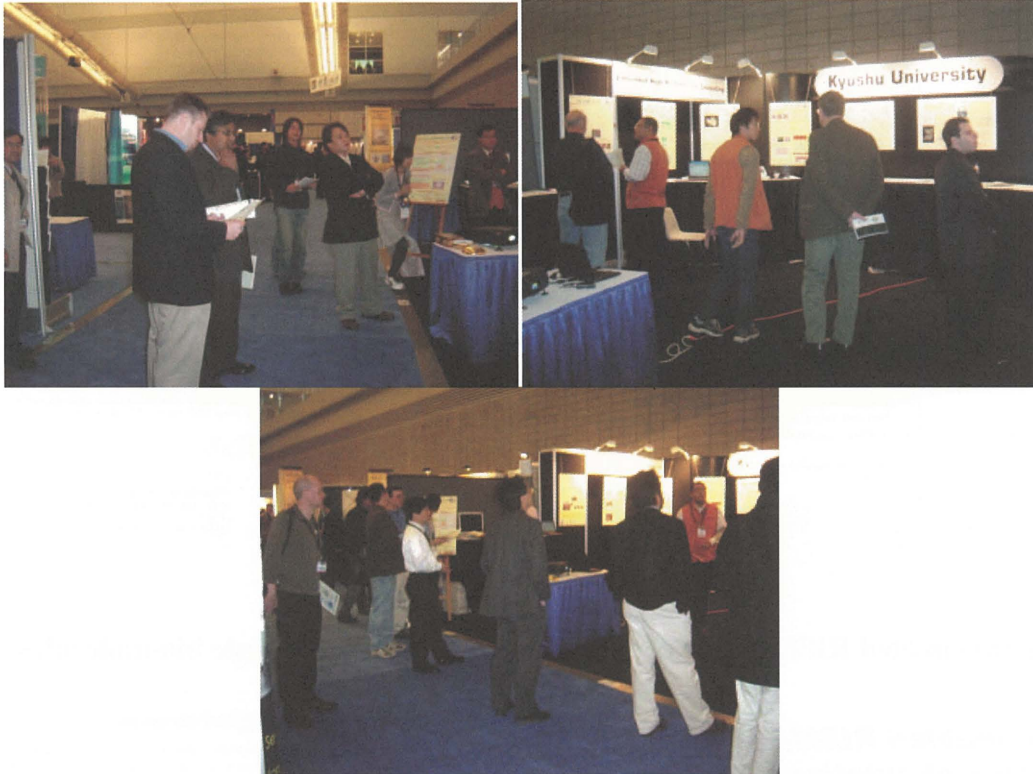
5.1.4 アジアへの展開・おわりに

拠点大学プロジェクトのテーマの一つである「次世代インターネットを用いた遠隔医療の開発と臨床応用」は、非常に活動的であり、九州大学のP&P資金の支援の元、日韓の研究活動をアジアに展開しています。情報基盤センターもこの「超高速インターネットを利用したアジアにおける医療ネットワーク拠点の形成」というテーマに加わっており、ネットワーク技術支援で協力しています。現在すでに中国北京の清華大学、上海の交通大学病院、タイマヒドン大学病院との遠隔医療実験を行っており、今後はシンガポール、台湾やオーストラリアにも展開してゆく予定です。

九州大学は大学病院の遠隔医療の活動を中心にして、韓国とはじめとするアジア各国との次世代インターネットを利用した実証実験を行なってきていますが、他国・組織との違いは、例えば、韓国のソウル大学大学病院が実験をする時、ソウル大学のキャンパスネットワークではなく、それとは別の研究・開発ネットワークが用いられています。それは、そちらの方が高品質なネットワークの利用が容易だからです。それに対して、九州大学では、情報基盤センターが運用・管理をしているギガビット級キャンパスネットワークを用いています。これは、情報基盤センターと病院キャンパスの間の距離の問題もありますが、情報基盤センターが導入し、大学に提供しているネットワークをあえて使用することにより実用化のための多くの問題を実証実験を通じて解決できるという意味があります。実際、日韓の実験が始まった頃、多くの問題がありましたが、最近ではそのような問題の解決方法も見つけ、専用の研究用ネットワークを用いているのと同様な品質で実験をすることが可能になっています。この実証実験で得た技術、知識、ノウハウは、新キャンパスでのネットワーク構築にも活かされています。このように情報基盤センターでは、日韓・アジアとの次世代インターネットに関する研究を進めながらそこで得たものを大学のネットワーク運用にもフィードバックしています。

5.2 国際会議 SC2004 における研究展示

SC は、毎年米国で行われている計算機とネットワーク技術に関する国際会議であり、特に近年は出席者数が 5000～7000 人と、この分野としては非常に規模が大きく、最新技術の情報交換を行う貴重な場となっている。扱われるトピックは多岐にわたるが、最も重点がおかれているのは、大規模科学技術計算のための高性能計算機に関する話題である。会期中の主な行事は研究発表と企業や研究機関による Exhibit(展示) であり、他に計算機の最新技術に関する講習 (チュートリアル) やパネルディスカッション、招待講演等も行われる。



2004 年の SC04 は 11 月に米国ピッツバーグ市で開催された。本センターからも、会場の展示ブースを借りて研究活動に関する展示を行い、世界の企業や研究機関からの参加者と情報を交換した。研究展示は以下の 5 つのテーマで行った。

- Shaping the Future with Computational Scientists in Japan
- Grid enabled RISM-FMO coupled simulations on large scale bio-molecules
- Showing bandwidth to CJK
- Preemptive Job Scheduler for Emergency Jobs on Grid
- Web Mining Activities in Kyushu University

なお今回の展示は、EHPC(Embedded High Performance Computing) プロジェクト (代表: 村上和彰 九州大学情報基盤センター長) との共同展示である。

5.2.1 Shaping the Future with Computational Scientists in Japan

Shaping the Future with Computational Scientists in Japan

Computing and Communications Center, Kyushu University

Our Mission

The Computing and Communications Center of Kyushu University offers its large computation capability to computational scientists in all academic institutions in Japan. To this end, we put our efforts together to design the most useful and reliable computer systems with the latest hardware, to provide users with the most friendly software environment, to maintain the systems to their best of the performance 24 hours a day, and to keep up with the ever-growing computer technology.

Our Machines



(Photo by Courtesy of IBM Japan, Ltd.)

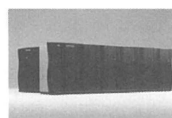
We offer two large computer systems: a *scalar-parallel computer system** and a *vector-parallel supercomputer system*.

Scalar-Parallel Computer: IBM eServer p5 595
With great expectations, we are now waiting for the shipment of seven sets of IBM's latest scalar-parallel server, eServer p5 595.

The most powerful node will consist of 64 POWER5 processors and 512GB memory. This will help those users who want to pursue the best performance of their large-scale OpenMP programs. The second class nodes will have 64 processors and 256GB memory for serving those programs which are smaller but dominant in number. They are also expected to offer a platform for hybrid parallel programs in OpenMP and MPI. The smallest node will be equipped with 32 processors and 128GB for testing use.

This new set of powerful scalar-parallel servers will add a new *cutting edge* for future computational science studies.

* ... The new scalar-parallel computer system is scheduled to start its services in March 2005.



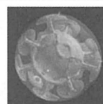
(Photo by Courtesy of Fujitsu, Ltd.)

Vector-Parallel Supercomputer: Fujitsu VPP5000/64

Fujitsu VPP5000/64 is a distributed-memory multiprocessor consisting of 64 processing elements (PEs). Each PE has 8-16 GB main memory and is capable of 9.6 billion floating-point operations per second (in other words, its peak performance is 9.6 GFLOPS). These PEs are connected via a high-speed crossbar interconnection network.

This kind of vector-parallel supercomputers are still the most powerful tool for researchers in certain fields such as fluid dynamics or seismic simulation.

Our Users



(Image by Courtesy of Dr. Atsushi Suzuki)

Our primary mission is to support leading computational scientists through our large-scale computing services. Now, let's peek at a preview of the future -- *the future being shaped by our users!*

The upper image is a visualization of a finite element computation of Earth's mantle convection problem. This computation has been executed on the current scalar-parallel computer Fujitsu GP7000F Model 900.

The lower image is a visualization of a large-scale numerical analysis of unsteady three-dimensional separated and vortical flow structure in an axial flow compressor rotor. This analysis has been carried out on our vector-parallel supercomputer Fujitsu VPP5000/64.



(Image by Courtesy of Prof. Masato Furukawa)

Our Efforts

To bridge the gap between computational science and computer science, we have the Research Division consisting of *computational scientists* and *computer scientists* working together. The unique collaboration of these two types of experts is the key to our cutting edge to assist our users to carve out the future.

You can find several examples of the latest research and development results of the Research Division during SC2004 in our exhibition booth #2727.



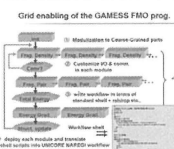
Computing and Communications Center, Kyushu University
6-10-1 Hakozaki, Higashi, Fukuoka 812-8581 JAPAN
<http://www.cc.kyushu-u.ac.jp/index-e.html>

5.2.2 Grid enabled RISM-FMO coupled simulations on large scale bio-molecules

Grid enabled RISM-FMO coupled simulations on large scale bio-molecules

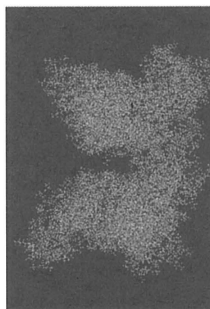
Computing and Communications Center, Kyushu University

Grid Enabling of the GAMESS FMO prog.



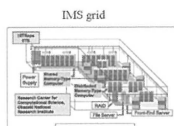
In this study, FMO method is implemented and modified to carry out large scale bio-molecular simulations on distributed grid computational environments. Original code (as a part of GAMESS program) is divided into large granularity tasks such as computational modules of fragment calculations, fragment pair calculations, and so on. Then we describe data-exchange of each inter-module as to the UNICORE file transfer. Using the NAREGI workflow tool, FMO workflow can be submitted to the NAREGI workflow engine.

Demonstrate the performance of Grid FMO

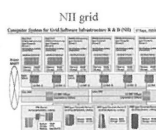


Serum Albumin is the carrier protein of the fatty acid.

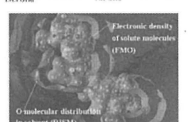
To demonstrate the performance of Grid FMO, we selected a Serum albumin, consisted of 18242 atoms, and 578 residues. Serum albumin play an important role as the carrier protein of the fatty acid. Each protein molecule can carry seven fatty acid molecules. Fatty acid without bind the albumin is water-insoluble and shows even toxicity if they are in your blood. However, it is interesting to note that they bind in deep crevices in serum albumin, burying their carbon-rich tails safely away from the surrounding water.



Computer resources used in our Grid FMO simulations.



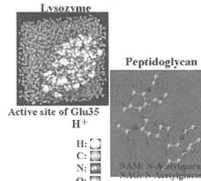
Mediator-based RISM-FMO coupled Simulations



It is increasingly important to analyze multi-scale and interdisciplinary problems in Nano- and Bio-technology, in which complex physical phenomena in different spatial and temporal scales are inter-related. In effort to provide solution to these problems, we would like to explain a new grid middleware in the form of mediators that allows various kinds of scientific simulations to be coupled efficiently over Grids.

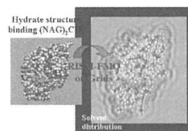
An example scenario is a RISM-FMO coupled simulation using mediator for structure and functionality analysis of Nano-scale molecules immersed in solvent. RISM simulation computes solvent distribution, which is coupled with FMO simulation that calculates electronic structure of the solute molecules.

Analyze protein structure of Lysozyme in solvent molecules



The coupled simulation is realized over NAREGI grids. RISM distributed on the IMS grid and FMO distributed on the NII grid are executed simultaneously and data communication between simulation components is supported by mediators. The mediator provides semantic transformations between physical quantities associated with different discretization model such as mesh and particle methods, and different physical model such as partial charge on molecules.

Hydrate structure of Lysozyme



Hydrate structure of Lysozyme calculated from molecular dynamics simulation. One of the active structures for proton donor is chosen and there a peptidoglycan exists as the substrate binding to a cleavage in Lysozyme. Using RISM-FMO coupled simulation, we calculate whole electronic structure of Lysozyme in solvent distribution in a self-consistent manner. We show large variation of electronic structure in peptidoglycan according to the position of proton transfer from glutamate acid 35 to glycosidic oxygen of peptidoglycan.

Thank you for your attention.



Computing and Communications Center, Kyushu University
6-10-1 Hakozaki, Higashi, Fukuoka 812-8581 JAPAN
E-mail: kyushu@cc.kyushu-u.ac.jp
This research work is supported by NAREGI project, and R&D are proceeded in collaboration with Hitachi group & Fujitsu group co., Ltd.

5.2.3 Showing bandwidth to CJK

Showing bandwidth to CJK
(China, Japan, Korea)

Computing and Communications Center, Kyushu University

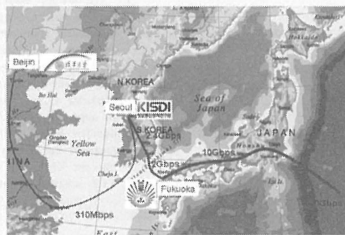
Abstract

We have established high bandwidth Internet among China, Japan Korea and are developing the advanced applications using this advanced Internet. In this showing, we will introduce our activities using this Internet among these 3 Asian countries using practice applications such as advanced multimedia streaming tools i.e. HD over MPEG2. And we will show our Internet's bandwidth using measurement tools and visualized measurement data. Finally we will try to get high band width among China, Japan and Korea. Getting more than 100Mbps even through 3 countries is surely interesting from many views, Internet operations, Traffic Engineering and Political issues. We will display the recent history of our Internet, how we get high bandwidth Internet among our countries, CJK.

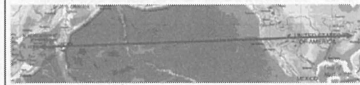
The Advanced Internet of CJK

Japan – Korea
2Gbps
since April 2003

Korea – China
310Mbps
since August 2004



The Bandwidth Challenge and Demonstration



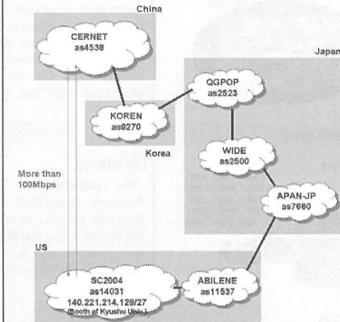
This may be the first trial of Bandwidth Challenge between "Asia (Japan, Korea, China...)" and US using advanced Internet. And the media transmission from Asia using DV and HD will be shown as typical applications using advanced Internet.

Network Topology

CERNET (The China Education and Research Network) is the first nationwide education and research computer network in China. The CERNET project is funded by the Chinese government and directly managed by the Chinese Ministry of Education. It is constructed and operated by Tsinghua and other leading universities.

KOREN (Korea Advanced Research Network) was founded for the purpose of expanding the technological basis of Korea and for providing a research environment for the development of high-speed telecommunications equipment and application services. Established in 1995, KOREN is a not-for-profit research network that seeks to provide universities, laboratories and industrial institutes with a research and development environment for 6T related technology and application services based on a subsidy from the Ministry of Information and Communications.

OGPOP (Kyushu GigaPOP Project) aims to build a dedicated R&D Internet over Kyushu region in parallel with commodity Internet, focusing on Internet's end-to-end principle and new features like IPv6, multicasting, and Mobile IP. We are a consortium of universities, industries, and non-profit research institutions to share our R&D infrastructure among the research community acting as a part of the global research Internet over the world.



Many Universities, Institutes and Organization of China, Korea, Japan and US cooperate and support network operations for this Bandwidth Challenge and Demonstration.



Computing and Communications Center, Kyushu University
6-10-1 Hakozaki, Higashi, Fukuoka 812-8581
JAPAN
E-mail: oka@cc.kyushu-u.ac.jp

5.2.4 Preemptive Job Scheduler for Emergency Jobs on Grid

Preemptive Job Scheduler
for Emergency Jobs on Grid

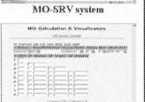
Yusaku UMEZU, Hirofumi AMANO, Takeshi NANRI,
Mutsumi AOYAGI (Kyushu University)

Preemptive Job Scheduling

We are developing a preemptive job scheduler which can serve jobs having different priorities efficiently on Grid.

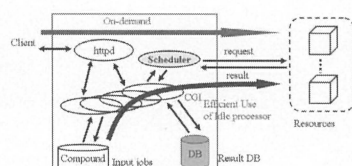
- Scheduling objective
 - To maximize the utilization of resources
 - To serve high-priority (emergency) jobs immediately
- Scheduling architecture
 - Select one low-priority job according to a policy
 - Abort the job and re-execute it later

Sample System

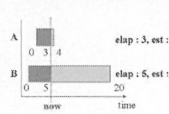


Sample System MO-SRV

MO-SRV is a system to calculate the molecular properties and visualize it. Currently, a client can request jobs to this system and receive the results on-demand. MO-SRV is designed to create computation jobs automatically for filling up the molecular database and to assign individual jobs to idle resources on Grid. Therefore when an on-demand job arrives, the scheduler needs a policy to select a job to abort the execution.



Proposed Policies



We have the following three policies to select a job to be aborted.

- Shortest-elap policy
- Smallest Progress-ratio policy (elapsed time / estimated time)
- Shortest-estimated policy

The left diagram describes two example jobs. In Policy 1, A is selected because A started running more recently. In Policy 2, B is selected because B has the smaller ratio of progress. Under Policy 3, A is selected because A has the shorter estimated time. Via simulation, we are now evaluating the policies.

Simulation Result and Discussion

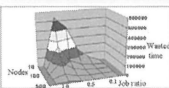


Fig.1 wasted time vs. job ratio (X-axis) and number of nodes (Y-axis)

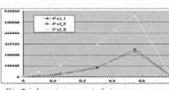


Fig.2 job ratio vs. wasted time (low-priority jobs : 100,000 nodes : 50)

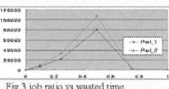


Fig.3 job ratio vs. wasted time (low-priority jobs : 100,000 nodes : 50, uncertainty of estimated time : 0.5-3.0)

Outline of Simulation

This study focuses on the victim selection scheme so as to minimize the wasted execution time.

This simulation takes the followings parameters.

- number of low-priority jobs
- number of emergency jobs
- ratio of different priority jobs
- number of processing nodes
- uncertainty of estimated time

Discussion

The left figures show the simulation result of each policy. They show that the Policy3 wastes more time than the others obviously. This is due to the fact that it does not take into account the already executed time. The interesting point is that the performance curves of Policy1 and Policy2 cross around 0.5 of job ratio. It indicates that Policy2 generates less wasted time as the job ratio increases because Policy2 tends not to abort the long jobs.

Fig.3 shows the comparison of Policy1 and Policy2 with the uncertain estimated time. We found that in this case, Policy1 performs better than Policy2, which uses estimated time.

Customizing OpenPBS

Through the more thorough simulations, the best policy will be implemented as a customized scheduler for Open PBS.

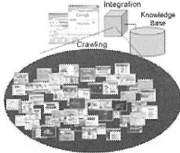
- We have newly added the following functions to OpenPBS.
 - Deleting a running low-priority job automatically at the time an emergency job arrive at the server.
 - Obtain the estimated time and elapsed time of the running jobs automatically.

5.2.5 Web Mining Activities in Kyushu University

Web Mining Activities
in Kyushu University

Computing and Communications Center,
Graduate School of Information Science and Electrical Engineering,
Kyushu University

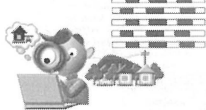
Our Goal



Huge amounts of information in various formats are available on the Web. We are constructing an information integration system utilizing Web pages on specific topics. In order to realize such an information integration system, we need to realize the following subtasks: (1) **discovery** of specific topic documents from the web, (2) **collection** of specific topic documents, (3) **extraction** useful record parts, and (4) **information integration** into a knowledge base.

Our Approach

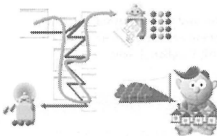
● Discovery



Discovery

We researched two problems. One is page discovery from Web space, and the other is document structure discovery from collected pages. We developed an **intelligent topic crawler** for the first problem, and **common pattern discovery algorithms** for the second.

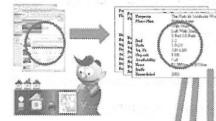
● Collection



Collection

We developed a topic keywords self-learning crawling algorithm and implement it in our topic crawler. The crawler can collect topic specific Web pages effectively.

● Contents Extraction



Extraction

We research attribute name/value string extraction technique using text pattern frequency. Target pages are a series of Web documents, which have the same appearance when a user views them with a browser.

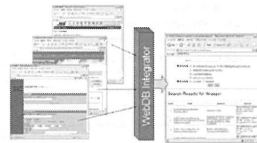
● Information Integration



Information Integration

We study information integration system. The system integrate some series of Web documents, written in heterogeneous format, but same topic. Integrated data are described in XML format.

Our Applications

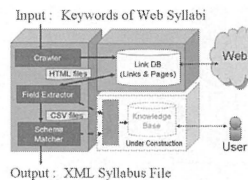


WebDB Integration

We develop semi-automatic **WebDB integration system**, such as meta-search engine. Key technologies are parameter name and type specification from `Locm` part in HTML file, parameter matching degree estimation in heterogeneous WebDB sites. We also develop WebDB site categorization into a category tree.

Web Syllabi Integration

We develop **Web syllabi integration system** for an e-Learning application using our techniques. The crawler collects Web syllabi, and stores them into **LinkDB**. LinkDB can work as a web proxy. Then, the field extractor extracts records from series of web pages. The extractor applies template discovery algorithm to separate *record* and *template*, and then decides *label* or *data*. Finally, the **schema matcher** integrate a series of Web syllabi files into an XML file. We chose *NIAD syllabus XML schema* for integrated XML file schema.



九州大学情報基盤センター 渡川研究室

Computing and Communications Center, Kyushu University
6-10-1 Hakozaki, Higashi, Fukuoka 812-8581, JAPAN
<http://vega.cc.kyushu-u.ac.jp/>

Computing and Communications Center, Kyushu University
6-10-1 Hakozaki, Higashi, Fukuoka 812-8581, JAPAN
<http://vega.cc.kyushu-u.ac.jp/>

5.3 海外出張記録

氏名	用務	用務先	日程
青柳 睦	米国IBM POWERアーキテクチャ研究フォーラムに参加のため	オーストラリア	7/26~7/29
青柳 睦	SC2004に出席し情報収集	アメリカ	11/6~11/15
天野 浩文	グリッドコンピューティングに関する研究打合せ	韓国	8/16~8/18
天野 浩文	SC2005に出席し情報収集	アメリカ	11/5~11/14
伊東 栄典	SC2004に出席し情報収集	アメリカ	11/5~11/14
伊東 栄典	拠点大学交流事業セミナー参加のため	韓国	11/24~11/28
伊東 栄典	ICADL2004に参加し情報収集	中国	12/14~12/17
伊東 栄典	フォーマル・コンセプト・アナリシス国際会議参加	フランス	2/12~2/16
伊東 栄典	コンピューテーション・インテリジェンス国際会議参加	トルコ	2/24~3/4
岡村 耕二	拠点大学ネットワーク基盤共同研究打合わせのため	韓国	4/15~4/17
岡村 耕二	共同研究打合せ	韓国	8/20~8/22
岡村 耕二	拠点大学ネットワーク共同研究	韓国	9/17~9/18
岡村 耕二	拠点大学ネットワーク共同研究打合せ	韓国	10/14~10/17
岡村 耕二	SC2005に出席し情報収集	アメリカ	11/6~11/14
岡村 耕二	拠点大学交流事業セミナー参加のため	韓国	11/24~11/28
岡村 耕二	拠点大学ネットワーク共同研究	韓国	12/1~12/4
岡村 耕二	拠点大学交流事業共同研究のため	韓国	1/15~1/17
岡村 耕二	拠点大学ネットワーク共同研究打合せのため	韓国	1/15~1/17
岡村 耕二	第19回APAN会議に出席	タイ	1/23~1/29
岡村 耕二	拠点大学交流事業共同研究のため	韓国	3/11~3/13
岡村 耕二	AINA2005 に出席、情報収集	台湾	3/27~3/30
岡村 耕二	拠点大学交流事業、共同研究、IPv6ネットワーク情報収集	タイ	3/15~3/20
笠原 義晃	The2004 ICCS & A に出席し研究発表	イタリア	5/13~5/18
笠原 義晃	第15回玄海会合及びAPAN会合に参加	オーストラリア	7/1~7/8
笠原 義晃	共同研究打合せ	韓国	8/20~8/22
笠原 義晃	拠点大学ネットワーク共同研究打合せ	韓国	10/14~10/17
笠原 義晃	SC2004に出席し情報収集	アメリカ	11/6~11/14
笠原 義晃	拠点大学交流事業セミナー参加のため	韓国	11/24~11/28
笠原 義晃	第19回APAN会議に参加し大規模データストリームからの超高速データマイニングの研究に関する資料収集を行う	タイ	1/23~1/30
笠原 義晃	拠点大学交流事業共同研究のため	韓国	3/11~3/13
中藤 哲也	グリッドコンピューティングに関する研究打合せ	韓国	8/16~8/19
中藤 哲也	AIRS2004に参加	中国	10/17~10/21
中藤 哲也	DIALOGUES国際会議出席	イタリア	10/1~10/10
中藤 哲也	AIRS2004に出席し情報収集及び研究発表	中国	10/17~10/21
中藤 哲也	拠点大学セミナー参加	韓国	11/24~11/27
中藤 哲也	アクティブ・マイニング・ワークショップに参加し、情報収集	ベトナム	12/3~12/9
南里 豪志	グリッドコンピューティングに関する研究打合せ	韓国	8/16~8/19
南里 豪志	国際会議CIC2004への参加及び研究打合せ	韓国	10/25~10/27
南里 豪志	SC2004にて情報収集	アメリカ	11/5~11/15
南里 豪志	国際会議ITRCにて研究発表	韓国	11/24~11/27
南里 豪志	拠点大学交流事業セミナー参加のため	韓国	11/24~11/27
南里 豪志	国際会議International Workshop on Innovation Architecture において研究発表	アメリカ	1/16~1/20
南里 豪志	Global Grid Forum に参加し情報収集、研究打合せ	韓国	3/13~3/16
廣川 佐千男	DIALOGUES国際会議出席	イタリア	10/1~10/10
廣川 佐千男	AIRS2004に出席し情報収集	中国	10/17~10/20
廣川 佐千男	インターネット技術研究会	韓国	11/24~11/28
廣川 佐千男	アクティブ・マイニング・ワークショップ参加のため	ベトナム	12/3~12/9
廣川 佐千男	拠点大学交流事業セミナー参加のため	韓国	11/24~11/28
廣川 佐千男	フォーマル・コンセプト・アナリシス国際会議参加	フランス	2/12~2/16
廣川 佐千男	コンピューテーション・インテリジェンス国際会議参加	トルコ	2/24~3/1
廣川 佐千男	研究調査	オーストラリア	3/13~3/17
渡部 善隆	ハワイ大学にて計算機援用証明法に関する共同研究	アメリカ	1/2~1/7