年報 第 8 号

2006 年度

Annual Report No.7

2006

東京大学情報基盤センター

Information Technology Center

the University of Tokyo

目　　次

巻頭言 …… 1

ＰＡＲＴ１　概況 …… 3

組　織 …… 5

組織図 …… 5

現員 …… 5

教職員 …… 6

平成１８年度中の人事異動 …… 11

東京大学情報基盤センター運営委員会委員名簿 …… 14

予　算 …… 15

収入・支出 …… 15

外部資金 …… 15

補助金等 …… 17

平成１８年度科学研究費採択状況 …… 17

平成１８年度受託研究費受入状況 …… 18

平成１８年度共同研究費受入状況 …… 19

平成１８年度奨学寄附金受入状況 …… 20

ＰＡＲＴ２　研究活動報告 …… 21

情報メディア教育研究部門 …… 23

山口　和紀 …… 25

田中　哲朗 …… 27

尾上　能之 …… 29

安東　孝二 …… 33

関谷　貴之 …… 35

図書館電子化研究部門 …… 43

中川　裕志 …… 45

二宮　　崇 …… 57

吉 田　　稔 …… 69

清田　陽司 …… 79

キャンパスネットワーキング研究部門……89
若原　恭……91
中山　雅哉……98
加藤　朗……102
中村　文隆……109
中村　誠……114
関谷　勇司……116
スーパーコンピューティング研究部門……127
金田　康正……129
石川　裕……138
佐藤　周行……148
佐藤　文俊……151
黒田　久泰……161
松葉　浩也……170
PKI プロジェクト……181
西村　健……183

ＰＡＲＴ３　教育・サービス活動……187
情報メディア教育部門……189
概要……191
教育用計算機システム運用報告……193
平成 18 年度情報教育研究集会……201
IC カード機能付学生証を用いた出欠管理システムの構築と運用……204
IC カードを使用したパスワード通知書発行システム……208
オープンソースを利用したウェブキオスク端末の構築……212
東京大学 ECCS プリンタ課金システムにおけるポストペイ決済の導入……216
平成 18 年度名古屋大学総合技術研究会……220
学内組織向けメールサーバ(MAILHOSTING)運用報告……221
WEB PARK サービス運用報告(2006 年 4 月－2007 年 3 月)……223
遠隔講義支援サービス運用報告(2006.4－2007.3)……234
CFIVE 運用報告(2006 年 4 月－2007 年 3 月)……238
図書館電子化部門……243
概要……245
デジタルコンテンツサービス……246
学術情報リテラシー支援……251

図書館関係システム運用・管理……258
キャンパスネットワーキング部門……263
概要……265
東京大学キャンパスネットワークシステム(UTnet3)運用管理……267
セキュリティ対応……275
東京大学情報システム緊急対応チーム(UT-CERT)との連携……277
学内ソフトウェアライセンス……279
学内共同利用３部門共通 講習会等……281
講習会・研究会開催報告……283
スーパーコンピューティング部門……287
スーパーコンピューティング業務部門概要……289
２００６年度のシステム整備状況……290
２００７年度のシステム整備計画……294
超並列型スーパーコンピューターシステムの調達……299
専門委員会……301
システムの利用状況……302
プログラム指導員と刊行物……305
職員による研究開発活動……307
スーパーコンピューターにおけるバッチジョブのスケジューリングについて……309
PKI プロジェクト……315
プロジェクト概要……317
運用報告……317
講習会・研究会開催報告……320

ＰＡＲＴ４　その他……321
平成１８年度委員会委員等……323

巻 頭 言

東京大学情報基盤センター長　米 澤 明 憲

本センターは、（1）情報メディア教育部門、（2）図書館電子化部門、（3）キャンパスネットワーキング部門、（4）スーパーコンピューティング部門、の４部門を擁しており、教員と職員の部門別の人員もほぼ同数です。しかしながら、学内外にあたえる全般的なイメージは、スパコンセンターのそれが先行しているようで大変残念なことです。他の３部門によるサービス業務は、学内の教育・研究・事務の情報インフラの大変大きく重要な部分の基盤となっております。またどの部門においても、教員と職員の密接な連携によって円滑で高度なサービスが可能になっています。本年報では、各部門での教員の研究活動と業務上の活動は、分離した形で、前者は「研究活動報告」として述べ、後者としては職員との連携により実施している活動も含めて「業務活動報告」で報告しています。

スーパーコンピューティング部門も含めて、４部門における教員の研究活動はこのところ年々活発さを加えてきています。それらの中には、ユーザへのサービス業務の中から興味深い課題を見つけ、その課題を研究し、得られた成果をサービス業務の高度化につなげるという、センターの研究としていわば理想的なサイクルに入っているのものが、幾つか見られます。また、多々ある研究成果の中には国際水準にあるものも少なくありません。

ここにお届け致します東大情報基盤センターの年報は、昨年以来その編集や体裁を少し変え、成果内容や業務内容を専門の方でなくてもわかる部分をより多くし、センターの活動が学内外で理解・評価され易いように図ったつもりです。しかしながら、まだまだ十分なものとは言いがたい部分も多く、今後さらに努力をつづけますので、ご理解を賜りたく宜しくお願い申し上げます。

PART 1

概　況

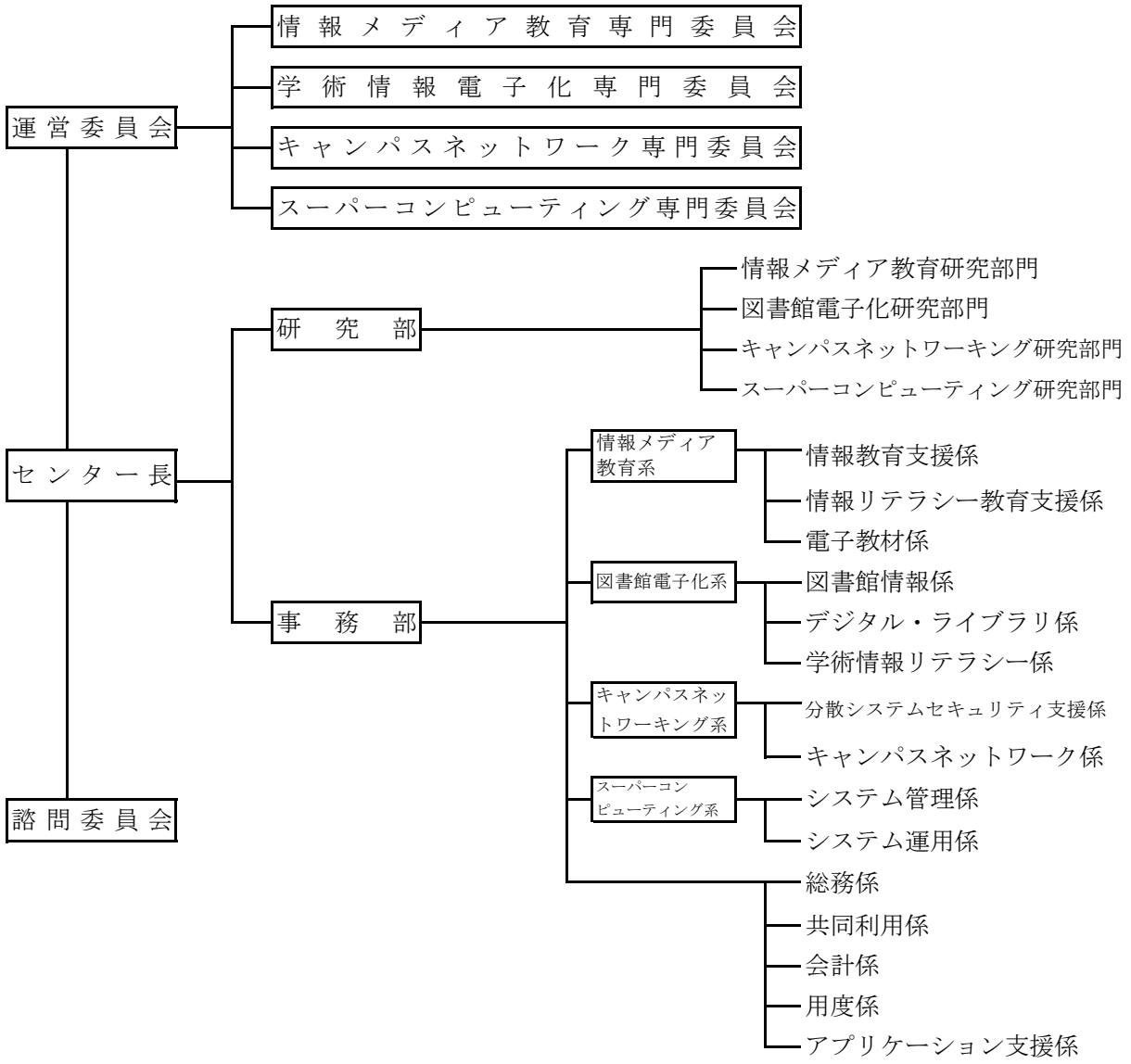
組　織

予　算

補助金等

組 織

組織図



現 員

(2006. 4. 1)

センター長			兼務
研究部	教員	教授	4
		助教授	5
		講師	2
		助手	8
	小計		19
事務部	職員	事務職員	25
		技術職員	23
	小計		48
合計			67

教職員

（１）2007 年 3 月 31 日現在

情報基盤センター長　　米　澤　明　憲
（情報理工学系研究科教授）
秘書（事務補佐員）渡　邉　かがり

研究部

情報メディア教育研究部門

教　授	山　口　和　紀
助教授	田　中　哲　朗
講　師	尾　上　能　之
助　手	安　東　孝　二
助　手	関　谷　貴　之
教務補佐員	寺　脇　由　紀

図書館電子化研究部門

教　授	中　川　裕　志
講　師	二　宮　　崇
助　手	吉　田　　稔
助　手	清　田　陽　司
特任助手	清　水　伸　幸
事務補佐員	吉　冨　美　樹
事務補佐員	赤　田　沙　織

キャンパスネットワーキング研究部門

教　授	若　原　　恭
助教授	中　山　雅　哉
助教授	加　藤　　朗
助　手	中　村　文　隆
助　手	中　村　　誠
助　手	関　谷　勇　司
事務補佐員	川　崎　しのぶ
事務補佐員	吉　澤　文　代

スーパーコンピューティング研究部門

教　授	金　田　康　正
教　授（兼務）	石　川　　裕
助教授	佐　藤　周　行
助教授	佐　藤　文　俊
助　手	黒　田　久　泰
助　手	松　葉　浩　也
事務補佐員	亀　田　文美代
事務補佐員	高　橋　弘　美

PKI プロジェクト

特任助手	西　村　　健
事務補佐員	伊　東　雅　美

事務部

事務長	菊　池　　力
副事務長	小　出　幸　夫
専門員 （企画調整担当）	大　塚　浩　一
主査 （情報業務担当）	佐　藤　安一郎
専門職員 （スーパーコンピューティング担当）	大日方　一　男

総務係

係　長	米　山　　浩
主　任	灰　塚　毅　弘
技術職員	佐々木　一　孝
事務補佐員	麦　谷　真　弓

共同利用係

係　長	西　澤　明　生
主　任	木　崎　信　一

会計係

係　長	野　呂　清　隆

主　任　　小　林　拓　志

用度係

係　長　　冨　田　正　明
主　任　　草　開　泰　之
主　任　　岡　戸　彰　二
一般職員　　三　上　　匠

［事務電算化支援系］　チーフ　佐　藤　安一郎

アプリケーション支援係

係　長　　（主査　兼務）
主　任　　入　江　健　司
一般職員　　大　島　大　輔
技術職員　　牛　丸　　守

［情報メディア教育支援系］　チーフ　平　野　光　敏

情報教育支援係

係　長　　平　野　光　敏
主　任　　山　本　和　男
技術専門職員　　伊　藤　真　之
技術補佐員　　小　薗　隆　弘
技術補佐員　　宮　北　美　保
派遣職員　　石　原　里　佳

情報リテラシー教育支援係

係　長　　櫻　田　芳　男
主　任　　前　田　光　教
技術職員　　友　西　　大
技術補佐員　　安　部　達　巳

電子教材係

係　長　　柏　　芳　美
主　任　　小　川　大　典
技術職員　　佐　藤　孝　明
技術補佐員　　加　藤　康　一
技術補佐員　　永　岡　陽　香

［図書館電子化支援系］　チーフ　本　多　　玄

図書館情報係

係　長	本　多　　玄
一般職員	中　村　恭　子
一般職員	坂　牧　一　博

デジタル・ライブラリ係

係　長	鴫　　邦　宏
一般職員	西　村　昭　子
一般職員	赤　津　愛　美

学術情報リテラシー係

係　長	吉　野　明　美
一般職員	濱　田　智恵子
一般職員	相　澤　奈津子
事務補佐員	古　木　景　子

［キャンパスネットワーキング支援系］　チーフ　早　野　裕　士

分散システムセキュリティ支援係

係　長	早　野　裕　士
主　任	佐　野　　徹
技術職員	坂　井　朱　美

キャンパスネットワーク係

係　長	中　村　昇　平
主　任	佐　島　浩　之
主　任	佐々木　　馨
技術職員	松　本　浩　一

［スーパーコンピューティング支援系］　チーフ　大日方　一　男

専門職員	丹　下　藤　夫

システム管理係

係　長	井　爪　健　雄
主　任	石　﨑　　勉

技術職員　　滝　本　友　彦

システム運用係
係　長　　有　賀　　浩
主　任　　田　川　善　教

平成１８年度中の人事異動

（転入・昇任・配置換等）

2006年 4.1	米澤 明憲	情報基盤センター長（兼務）
2006年 4.1	二宮 崇	図書館電子化研究部門講師／新規（任期　18.4.1～23.3.31）
2006年 4.1	菊池 力	事務長／農学系総務課長から0
2006年 4.1	野呂 清隆	会計係長／国立オリンピック記念青少年総合センター基金部管理課管理係長から
2006年 4.1	鴫 邦宏	デジタル・ライブラリ係長／国立情報学研究所開発・事業部コンテンツ課学術情報形成第一係長から
2006年 4.1	柏 芳美	電子教材係長／文科省大臣官房国際課専門職／（併）内閣府国立公文書館アジア歴史資料センター事業第２係長
2006年 4.1	坂牧 一博	図書館情報係／東京外国語大学情報図書館課総務係から
2006年 4.1	赤津 愛美	デジタル・ライブラリ係／新規
2006年 4.1	櫻田 芳男	情報リテラシー教育支援係長／システム管理係長から
2006年 4.1	本多 玄	学術情報リテラシー係長（免）／学術情報リテラシー係長（命）
2006年 4.1	吉野 明美	学術情報リテラシー係長／デジタル・ライブラリ係長から
2006年 4.1	小林 拓志	会計係主任／会計係から
2006年 4.1	前田 光教	情報教育支援係主任／情報教育支援係から
2006年 4.1	佐々木 馨	キャンパスネットワーク係主任／キャンパスネットワーク係から
2006年 4.1	田川 善教	システム運用係主任／システム運用係から
2006年 4.1	渡邉 かがり	センター長秘書事務補佐員／新規
2006年 5.18	赤田 沙織	図書館電子化研究部門技術補佐員／新規
2006年 6.1	石川 裕	スーパーコンピューティング研究部門助教授（兼務）
2006年 7.1	米山 浩	総務係長／人事部職員課勤務環境チーム係長から

2006年 7. 1	有賀　浩	システム運用係長／国立教育政策研究所教育研究情報センター情報支援課データベース・情報提供係長から
2006年 7. 1	井爪健雄	システム管理係長／国立特殊教育総合研究所研修情報課情報管理係長から
2006年 7. 1	佐野　徹	分散システムセキュリティ支援係主任／人事部分析グループ人事業務改善チーム主任から
2006年 7. 1	三上　匠	用度係／農学系総務課附属演習林北海道演習林会計係から
2006年 7. 1	西村昭子	デジタル・ライブラリ係／附属図書館情報管理課資料契約係から
2006年 7. 1	大塚浩一	専門員総務係長（兼務免）／専門員総務係長（兼務命）から
2006年 7. 1	丹下藤夫	専門職員／システム運用係長から
2006年 7. 1	伊藤真之	情報教育支援係（技術専門職員）／情報リテラシー教育支援係（技術専門職員）から
2006年 7. 1	前田光教	情報リテラシー教育支援係主任／情報教育支援係主任から
2006年 7. 1	友西　大	情報リテラシー教育支援係／情報教育支援係から
2006年 7. 1	佐々木一孝	総務係／システム運用係から
2006年 7.16	石川　裕	スーパーコンピューティング研究部門教授(兼務）／スーパーコンピューティング研究部門助　教授（兼務）から
2006年 8. 16	松葉浩也	スーパーコンピューティング研究部門助手／新規（任期　18.8.16～24.3.31）
2006年 9. 11	安部達巳	情報リテラシー教育支援係技術補佐員／新規
2006年 10. 1	王　玉馨	図書館電子化研究部門教務補佐員／新規
2006年 12. 1	高橋弘美	スーパーコンピューティング研究部門事務補佐員／新規
2007年 1. 1	清水伸幸	図書館電子化研究部門特任助手／新規（任期　19.1.1～22.3.31）
2007年 3. 1	相澤奈津子	学術情報リテラシー係／新規（任期　19.3.1～20.3.31）

（転出・退職等）

2006年	4. 1	越塚　登　キャンパスネットワーキング研究部門助教授／大学院情報学環助教授へ
2006年	4. 1	野内公一　会計係長／財務部決算グループ財会システムチーム係長へ
2006年	4. 1	堀井英夫　電子教材係長／文科省大臣官房国際課専門職／（併）内閣府国立公文書館アジア歴史資料センター事業第２係長へ
2006年	4. 1	下田哲郎　キャンパスネットワーク係主任／国立情報学研究所開発・事業部ネットワーク課運用係長へ
2006年	4. 1	飯塚亜子　学術情報リテラシー係／附属図書館総務課企画渉外係／文科省研修生へ
2006年	4. 1	加藤さつき　図書館情報係／東京外国語大学情報図書館課情報サービス係長へ
2006年	7. 1	中山仁史　分散システムセキュリティ支援係主任／国立教育政策研究所教育研究情報センター情報支援課データベース・情報提供係長へ
2006年	7. 1	宮嵜　洋　情報リテラシー教育支援係主任／国立特殊教育総合研究所研修情報課情報管理係長へ
2006年	7. 1	小山憲司　デジタル・ライブラリ係／国立情報学研究所開発・事業部コンテンツ課コンテンツ企画係長へ
2006年	7. 1	石田卓也　総務係／人事部職員課勤務条件チームへ
2006年	7. 1	奥原由美　用度係／財務部調達グループ移転・役務チームへ
2007年	3. 31	小出幸夫　副事務長／定年退職
2007年	3. 31	安東孝二　情報メディア教育研究部門助手／辞職

東京大学情報基盤センター運営委員会委員名簿

任期：平成１７年４月１日～平成１９年３月３１日

平成１８年４月１日現在

氏　　名	所　　属　・　職　　名	選　考　理　由
米　澤　明　憲	情報基盤センター長	規則第３条第１号
金　田　康　正	情報基盤センター・教授	規則第３条第２号
中　川　裕　志	情報基盤センター・教授	規則第３条第２号
若　原　　恭	情報基盤センター・教授	規則第３条第２号
山　口　和　紀	情報基盤センター・教授	規則第３条第２号
喜連川　　優	生産技術研究所・教授	規則第３条第３号
太　田　勝　造	大学院法学政治学研究科・教授	規則第３条第３号
佐　藤　隆　夫	大学院人文社会系研究科・教授	規則第３条第３号
川　合　　慧	大学院総合文化研究科・教授	規則第３条第３号
相　田　　仁	大学院新領域創成科学研究科・教授	規則第３条第３号
平　尾　公　彦	大学院工学系研究科・教授	規則第３条第３号
堀　　浩　一	大学院工学系研究科・教授	規則第３条第３号
青　木　秀　夫	大学院理学系研究科・教授	規則第３条第３号
山　口　　泰	大学院情報学環・教授	規則第３条第３号
矢　島　美　寛	大学院経済学研究科・教授	規則第３条第３号
西　郷　和　彦	附属図書館長	規則第３条第４号

オブザーバー　武川　正吾　総長補佐（大学院人文社会系研究科教授）

予　算

収入・支出

2006 年度決算額

収入

区　　分	決算額（千円）
寄 附 金 収 入	６，９３５
受託研究等収入	１２３，９４６
自　己　収　入	２５，８９６
計	１５６，７７７

支出

区　　分	決算額（千円）
人　　件　　費	４９１，５２３
物　　件　　費	３，２１３，８６６
計	３，７４５，８６４

外部資金

１）奨学寄附金受入（2006 年度）

内　　訳	受入件数	受入額（千円）
情報メディア教育研究部門	０	０
図書館電子化研究部門	０	０
キャンパスネットワーキング研究部門	２	５，４３５
スーパーコンピューティング研究部門	１	１，５００
計	３	６，９３５

２）受託研究（2006 年度）

内　　訳	件　　数	受入額（千円）
情報メディア教育研究部門	０	０
図書館電子化研究部門	２	１７，３３８
キャンパスネットワーキング研究部門	４	８４，８００
スーパーコンピューティング研究部門	１	１５，８０８
計	７	１１７，９４６

3）共同研究（2006 年度）

内　　訳	件　数	受入額（千円）
情報メディア教育研究部門	0	0
図書館電子化研究部門	0	0
キャンパスネットワーキング研究部門	2	1，５００
スーパーコンピューティング研究部門	3	4，５００
計	5	6，０００

4）科学研究費補助金（2006 年度）

内　　訳	件　数	受入額（千円）
情報メディア教育研究部門	0	0
図書館電子化研究部門	4	２１，９１０
キャンパスネットワーキング研究部門	0	0
スーパーコンピューティング研究部門	2	2，２００
計	6	２４，１１０

補助金等

平成１８年度科学研究費採択状況

研究代表者　　教授　中川　裕志
研究種目　　特定領域研究
研究期間　　平成 18 年度
研究課題　　多言語Ｗｅｂテキストからの知識マイニングに関する研究
研究費　　5,900,000 円（平成 18 年度）

研究代表者　　教授　中川　裕志
研究種目　　基盤研究（A）
研究期間　　平成 17 年度～19 年度
研究課題　　Ｗｅｂテキストからの知識抽出支援システムに関する研究
研究費　　13,910,000 円（平成 18 年度）

研究代表者　　助教授　佐藤　周行
研究種目　　基盤研究（C）
研究期間　　平成 17 年度～19 年度
研究課題　　最適化検証つきコンパイラの理論と実証のための展開
研究費　　1,300,000 円（平成 18 年度）

研究代表者　　助手　吉田　稔
研究種目　　若手研究（B）
研究期間　　平成 17 年度～18 年度
研究課題　　ＨＴＭＬ文書からの輪理構造自動推定に関する研究
研究費　　1,000,000 円（平成 18 年度）

研究代表者　　助手　清田　陽司
研究種目　　若手研究（B）
研究期間　　平成 18 年度～19 年度
研究課題　　Ｗｅｂ上の情報資源と図書館情報資源を統合的に利用する情報探索支援システム
研究費　　1,100,000 円（平成 18 年度）

研究代表者　　助手　松葉　浩也
研究種目　　特別研究員奨励費
研究期間　　平成 17 年度～18 年度

研究課題　　ＰＣクラスタを単一のシステムとするオペレーティングシステムの設計と実装
研究費　　900,000 円（平成 18 年度）

平成１８年度受託研究費受入状況

研究代表者　　助教授　加藤　朗
相手機関名　　（株）日本レジストリサービス
研究期間　　平成 18 年度
研究課題　　BGP　Anycast 環境下における DNS の運用と測定
研究費　　4,500,000 円（平成 18 年度）

研究代表者　　教授　若原　恭
相手機関名　　情報・システム研究機構
研究期間　　平成 18 年度
研究課題　　最先端学術情報基盤構築事業
研究費　　68,600,000 円（平成 18 年度）

研究代表者　　教授　中川　裕志
相手機関名　　富士通研究所
研究期間　　平成 18 年度
研究課題　　高度なテキスト分析支援技術の開発
研究費　　3,000,000 円（平成 18 年度）

研究代表者　　教授　石川　裕
相手機関名　　科学技術振興機構
研究期間　　平成 18 年度
研究課題　　並列・分散型組込みシステムのためのﾃﾞｨﾍﾟﾝﾀﾞﾌﾞﾙｼﾝｸﾞﾙｼｽﾃﾑｲﾒｰｼﾞ　OS
研究費　　15,808,000 円（平成 18 年度）

研究代表者　　助教授　中山　雅哉
相手機関名　　科学技術振興機構
研究期間　　平成 18 年度
研究課題　　都市基盤広域センシングに関する手法・データ伝送と処理・総合リスク評価 そしてフィールド適用に関する研究
研究費　　5,200,000 円（平成 18 年度）

研究代表者　助教授　中山　雅哉
相手機関名　科学技術振興機構
研究期間　平成 18 年度
研究課題　大規模データ収集･発信のための高次ネットワーク基盤の構築
研究費　6,500,000 円（平成 18 年度）

研究代表者　教授　中川　裕志
相手機関名　科学技術振興機構
研究期間　平成 18 年度
研究課題　日中・中日言語処理技術の開発研究
研究費　14,338,000 円（平成 18 年度）

平成１８年度共同研究費受入状況

研究代表者　教授　若原　恭
相手機関名　沖電気工業株式会社
研究期間　平成 18 年度
研究課題　P2P　トラヒック弁別に関する研究
研究費　1,000,000 円（平成 18 年度）

研究代表者　助教授　佐藤　周行
相手機関名　（株）KDDI
研究期間　平成 18 年度～19 年度
研究課題　大学電子認証基盤におけるプライバシー保護技術の研究
研究費　2,000,000 円（平成 18 年度）

研究代表者　助教授　中山　雅哉
相手機関名　（株）三菱総合研究所
研究期間　平成 18 年度
研究課題　Ipv6 ネットワークに関する研究
研究費　500,000 円（平成 18 年度）

研究代表者　教授　金田　康正
相手機関名　(株)本田技研研究所基礎技術研究センター　高度情報科学技術研究機構
研究期間　平成 18 年度
研究課題　ナノシミュレーションにおける大規模、高速アルゴリズムの研究
研究費　1,500,000 円（平成 18 年度）

研究代表者　　教授　金田　康正
相手機関名　　（独）国立環境研究所
研究期間　　平成 18 年度
研究課題　　スーパーコンピュータを用いた GOSAT 観測データ処理システムに関する研究
研究費　　1,000,000 円（平成 18 年度）

平成１８年度奨学寄附金受入状況

研究代表者　　助教授　加藤　朗
相手機関名　　Community Foundation Sillicon Valley
研究費　　4,535,385 円（平成 18 年度）

研究代表者　　教授　金田　康正
相手機関名　　本田技研研究所
研究費　　1,500,000 円（平成 18 年度）

研究代表者　　教授　若原　恭
相手機関名　　富士重工・スバル技術研究所
研究費　　900,000 円（平成 18 年度）

PART 2

研究活動報告

情報メディア教育研究部門

図書館電子化研究部門

キャンパスネットワーキング研究部門

スーパーコンピューティング研究部門

PKI プロジェクト

情報メディア教育研究部門

教育向けシステムならびにデータ分析手法に関する研究

山口和紀

1 概要

教育用計算機システムと学習管理システムの機能について応用的研究を行うとともに、データ分析手法について原理的研究を行った。

2 研究内容題目:教育用計算機システム

2.1 背景

ポストペイカードによる印刷では、カード残高が不足する場合は印刷を2回に分けてしなければならなかったり、ポストペイカードの販売機がなければ印刷できないという問題があった。.

2.2 内容

リコー、NTT コム、DC カードと共同研究を実施し、IC カード学生証／職員証のポストペイ機能で決済ができるシステムを構築した。

2.3 具体的成果

開発したシステムを導入し、2006 年 9 月からサービスを開始した。(学内広報 1346)

3 研究内容題目:学習管理システム

3.1 背景

教育での ICT 活用の進展にともない学習管理システムへのニーズが増大している。

3.2 内容

学習管理システムに必要とされる機能の研究を行った。([査読付 1]、[査読付 4])プレゼンテーションを簡単に収録／公開したいという要望に対しては、リコーが開発した MPMeister と情報基盤センターが開発した CFIVE をベースにして、リコーとの共同研究でシステムを開発した。([報道 1])

3.3 具体的成果

システムを情報基盤センターの遠隔講義室に仮設置し、セミナーの収録・公開を行った。開発したソフトウェアはオープンソースとして公開している。[公開 1]

4 研究内容題目:データ分析手法

4.1 背景

様々な分野でデータの量が増えてきており、多量のデータから本質的な要素を抽出することが重要になってきている。

4.2 内容

シラバス、ウェブ、自然画像に対してデータ分析手法を開発した。

4.3 具体的成果

主成分を抽出するアルゴリズムと情報量最小化及び不動点アルゴリズムの関係を発見し、自然画像に適用して確認した。([査読付 2] , [査読付 3])非有基的集合論を用いたウェブ構造の分析手法を2つ開発した。([査読付 5])用語の意味を用語集合で捉える方法を考案し、シラバスの分析に適用した。([発表 1])

5 成果要覧

査読付論文リスト

[査読付 1] Case Study of Information Systems Cooperation Centered in Open Source LMS, CFIVE, Takayuki Sekiya and Kazunori Yamaguchi, Workshop on Free and Open Source Learning Environments and Tools: FOSLET 2006, Como, Italy, 2006, pp.21-28.

[査読付 2] The InfoMin Principle for ICA and Topographic Mappings, Yoshitatsu Matsuda and Kazunori Yamaguchi, ICA2006, Charleston, USA, LNCS 3889, Springer-Verlag, pp.958-965.

[査読付 3] A Fixed-Point Algorithm of Topographic ICA, Yoshitatsu Matsuda and Kazunori Yamaguchi, ICANN2006, Athens, Greece, LNCS 4132, Springer-Verlag, 587-594.

[査読付 4] 学習管理システム CFIVE の継続的な開発と運用の評価, 関谷貴之, 佐藤孝明, 寺脇由紀, 尾上能之, 山口和紀, 情報教育シンポジウム SSS2006, 2006, pp.339-346.

[査読付 5] 高階ランクを用いたウェブ構造の分析, 堀江郁美, 山口和紀, 柏原賢二, 情報処理学会論文誌, 47(9), 2006, pp.2830-2839.

公開ソフトウェア

[公開 1] 情報基盤センター、CFIVE, 2003-2007.

その他の発表論文リスト

[発表 1] “Ontology and context based repository of learning materials,” T. Sekiya and K. Yamaguchi, The Computer Assisted Learning (CAL) biennial conference, CAL ’07, poster.

特記事項

[特記 1] APRU DLI(Distance Learning and the Internet) 2006 を東大で開催したが、その実施に協力した。

報道関連

[報道 1] 2007/2/28 の日経産業新聞(7 面) 動画を完全自動収録 リコー、東大とシステム 講義の資料と連動.

ゲームプログラミングに関する研究

田中哲朗

1　概要

ゲームプログラミング研究の一環として、2003 年から他研究室と共同でオープンソース将棋プログラムの開発を行っている。さまざまな改良を加えた結果、世界でもトップレベルに達することができた。

2　オープンソース将棋プログラムの作成

2.1　背景

1997 年に、チェスプログラムが初めて人間の世界チャンピオンを破るという事件が起きたが、将棋もあと数年で同様のレベルに達することが期待され、コンピュータ将棋に関する研究は一番面白い時期を迎えようとしている。

しかし、コンピュータ将棋の研究を始める際に超えなくてはいけない障壁は小さくない。まず、将棋は駒の種類がチェスより多く、独自の持ち駒ルールや、打ち歩詰め、二歩等のルールなどのため、ルール通り指させるだけでも大変である。

また、ある程度の強さを持つプログラム上で評価しないと、意味のある結果が得られないが、終盤の詰め将棋など十分な強さを持つものをバグなしに作るのは難しい。

そのため、コンピュータ将棋研究の全体的なレベルアップを狙って、最初から公開する前提で将棋プログラムの開発を 2003 年からおこなっている。将棋プログラム名は GPS 将棋で、開発メンバーは田中研究室の学生の他に、総合文化研究科の教員、学生、会社員などが加わっている。

2.2　内容

この将棋プログラム自体は 2003 年度から開発を開始し公開もしているが、2006 年度は

- 並列化
- 必至ルーチン
- 統計的学習による評価関数のパラメータ調整

などをテーマに改良をおこなった。

2.3　具体的成果

ライブラリは BSD ライセンスで公開し、将棋プログラム全体は GPL で公開する形にした。これにより、純粋に研究目的の人も、あるいは商品を売って儲けようという人もどちらも自由に使えるようにしながら、「コンピュータ将棋選手権何位」といった売り文句は使えないという形で公開することができた。

作成したプログラム GPS 将棋 (図 1) は、5 月の世界コンピュータ選手権で決勝には進めなかったものの全体で 10 位という好成績をおさめた。

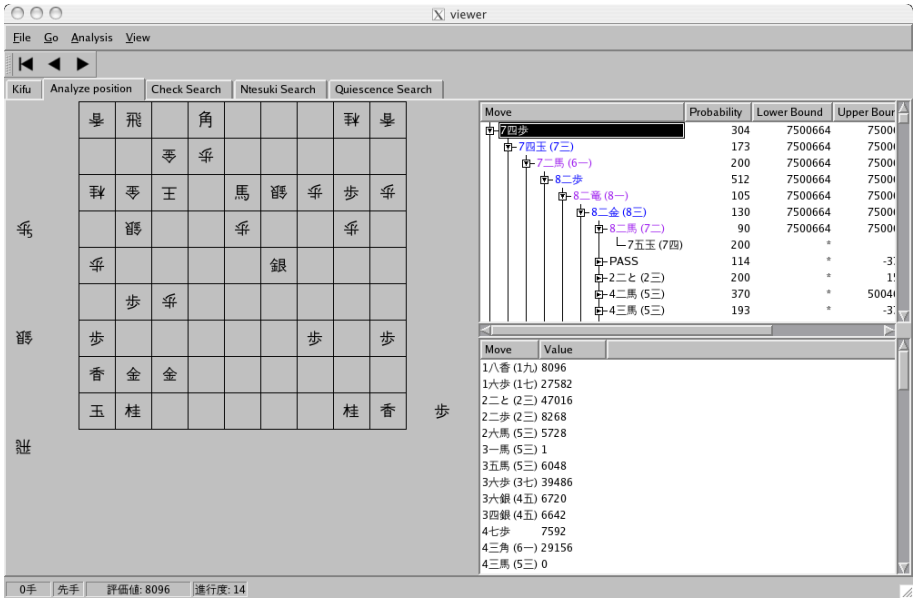


図 1: GPS 将棋の操作画面

3 成果要覧

その他の発表論文リスト

[発表 1] 副田俊介, 美添一樹, 田中哲朗: 証明数を用いたλ探索の効率的な実装, 第 11 回ゲームプログラミングワークショップ予稿集 (GPW-06), pp. 48 - 55(2006)

公開ソフトウエア

[公開 1] 金子知適、田中哲朗、副田俊介、林芳樹、竹内聖悟: osl-for-csa (バージョン 0.3), コンピュータ将棋協会のコンピュータ将棋選手権使用可能ライブラリ, 2007 年 1 月

特記事項

[特記 1] 田中哲朗: 平成 18 年度情報処理学会山下記念研究賞, 「ボードゲーム「シンペイ」の完全解析」(2006-GI-15)

[特記 2] GPS 将棋: 世界コンピュータ将棋選手権第 10 位, コンピュータ将棋協会,2006 年 5 月

情報システムを用いた教育研究支援

尾上 能之

1 概要

教育用計算機システムや学習管理システムを運用しサービスとして提供することで、利用者がより少ない負担で効率的に教育研究活動を行なえるための支援を行なっている。

2 学習管理システム CFIVE

2.1 背景

当部門では日本ユニシス株式会社と共同で開発した学習管理システム (LMS: Learning Management System) である CFIVE(Common Factory for Inspiration and Value in Educational) を、2004 年から運用し学内の利用者にサービスとして提供している。2006 年には各学期それぞれ 30 前後の講義で利用されるようになり、以前より利用者や利用機会が増えたことにつれて、システムに内在する新たな問題が表面化することもでてきた。このような際は CFIVE の運用に携わっている教員 2 名と職員 1 名で対処にあたるが、システムの内部の修正については主に教員が行なっているのが現状である。

2.2 内容

今季 CFIVE で起きた主な障害や不具合と、行なった対応は以下の通りである。

- 課題提出が重なった際にサーバ負荷が著しく高くなる件の調査、対応
  - 学生が課題提出時にその講義履修者に対するすべての同期をデータベースと WebDAV サーバの間でとろうとしエラーになる件を調査し、改修を発注していた業者に修正依頼した
  - 同提出時に履修者すべてのコメントファイルを WebDAV サーバから　取得し負荷が上がる件を調査し修正した
- テストを受けても「未提出」扱いになってしまう件の調査、対応
  具体的には、ソースコードの修正ならびにその適用を行ない、データベース上で未提出者の状態を「提出済み」に変更し、合計点を再計算の上修正した。
- LDAP 上で uid と cn が違う場合受験結果が反映されない件の調査、対応
- 課題情報取得エラーが起きる件の調査

また CFIVE では利用者からのバグ報告や改善要求をバグトラッキングシステムで管理し、随時システムの改良を行なっている。今季行なった機能拡張、バグ修正のうち主なものを以下に挙げる。

- 各種一覧画面を CSV, Excel, HTML の表としてダウンロードする機能

- ユーザ登録数が多い場合正しく表示できなかった「ユーザ ID 一覧」に対するページャー機能の実装
- 掲示板に登録する記事のプレビュー機能
- テスト問題を再利用する機能
- 主に開発者を対象にした、開発標準ならびに開発手順書の策定

これ以外にも、今後利用者数がさらに増加したときの影響も考慮し、サーバ設備の増強や稼働しているソフトウェアをより効率的に動かすための工夫等も随時行なっている。[発表 1]

CFIVE はオープンソースのソフトウェアとして広く利用してもらえるよう一般に公開しているが、大学で運用することを念頭に設計された本格的なシステムであることもあって、利用者の中には構築の際につまづく事も多いようである。そこで当部門で提供しているサポート窓口に来た質問に対し、適宜アドバイスを与えることによって、より多くの場所で CFIVE を活用してもらえるよう働きかけている。中には CFIVE の運用を予定している機関の例の一つとして放送大学があり、インストールなどの環境構築に協力をしている。

2.3 具体的成果

図 1 は CFIVE に申請のあった講義数 (各月毎と累積) であるが、この図からもわかるように利用講義数ならびに利用者数は増えてはいるものの、全学の規模からするとまだ僅かであると言えよう。今後も新たな機能を導入し利用者の利便性を改善しつつ、安定したサービスを提供できるよう努力していきたい。

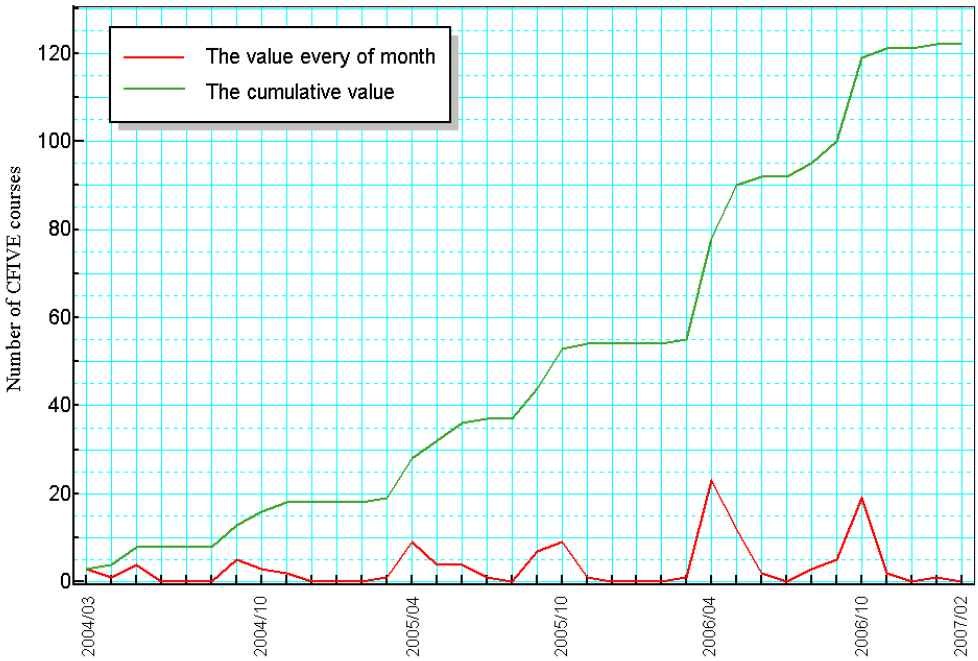


図 1: CFIVE 利用状況

3 教育用計算機システム

3.1 背景

当部門では現在、2004 年に導入した教育用計算機システム (Educational Campuswide Computing System, ECCS) の運営を行っており、教養学部における前期課程教育 (全科類で必修)、各学部 / 研究科における専門課程教育のための環境を提供しているだけでなく、電子メール等学内のすべての構成員の教育活動の支援をしている。これらのシステムを合わせると端末、サーバを合わせて 1,400 台を超える大規模分散システムとなっており、これを登録者数約 30,000 人、実利用者数約 20,000 人のユーザが利用している。また、近年のインターネット及びその上のサービスである 電子メール、World Wide Web (以下 WWW と記す) 等の普及から、以前のように授業時間帯だけの利用ではなく、24 時間 365 日のサービスを提供している。

このような大規模なシステムを 10 名前後の技術職員だけで運用していくことは難しく、教員の協力が不可欠な状態である。

3.2 内容

今季行なったユーザサポートや障害への対処のうちのいくつかを以下に紹介する。

- 部局から依頼のあったソフトウェアの導入・更新
  ECCS ではシステムの調達時以外に、講義で用いる要望のあったアプリケーション (部局負担で用意) をシステム内に入れることで、限られた予算で多くの利用者に満足してもらえるサービスを提供している。例として、経済学部からは MATLAB, TSP, STATA といった統計処理のソフトウェアライセンスを提供していただき、ECCS に入れることで講義や自習で広く利用してもらっている。

- DNS ホスティングサービスの開始
  これまでに当部門ではメールや Web といったネットワークの基幹サービスに対し、サーバ代行サービス (ホスティング) の提供を行なっている。しかし各部局にはサーバを管理できる人材が少なくなってきており、DNS においてもセンターに管理を任せたいという声が挙がっていた。そこで管理者の委譲機能をもった DNS サーバのアプライアンス (専用機器) を導入し、これまで部局で用意していたサーバ機能をセンター側で代行するサービスを実験的に開始した。

- ハードウェア、ソフトウェアの監視、障害対応
  レンタルに含まれるものは保守の対象に入っており障害があっても修理依頼を行なうだけで済むが、現システムは多様なサービスを提供していることもあり保守に入っていないサーバも多数稼働している。また保守に入っているものでも、大学での用途に合わせて独自な環境設定をしていると、業者でも修理対応できないものもあり結局はこちらで対処が必要なこともある。

  具体的には、ユーザ情報発信用 WWW サーバや講義用 WWW サーバでハードディスク障害などによって停止したサーバの復旧作業を行なった。これらのサーバは講義で使用されることもあり、利用者が増えても対応できるよう大容量のハードディスクが搭載されている。もちろん RAID などの障害対策を行なっているものの、それでも障害が起きることもあり、ディスク容量が大きい分対処に時間がかかってしまうのが現状である。このような対処は、勤務時間に制限のある職員が行なうことは難しく教員が中心となって行なっている。

また現在のシステムは 2007 年度でレンタル契約が終了し、2008 年度からは新たなシステムを運用する必要があるため、次期システム調達への準備を今年度から開始している。その中で担当した箇所を以下に挙げておく。

- 市場動向の調査
- 一般の利用者からの意見を聞く公聴会のための資料作成
- 導入作業部会の開催、ならびに導入説明書の作成

3.3 具体的成果

ECCS は利用者数や端末台数からみてもわかるように巨大なシステムで、このようなシステムを運用している大学は国内でも数少ない。こうした ECCS が大きなサービス停止もなく、利用者に様々なサービスを提供できていることが一番の成果と言えるであろう。

4 成果要覧

その他の発表論文リスト

[発表 1] 関谷貴之, 佐藤孝明, 寺脇由紀, 尾上能之, 山口和紀: 学習管理システム CFIVE の継続的な開発と運用の評価, 情報教育シンポジウム SSS2006, pp. 339–346, 2006.

ECCS プリンタにおけるポストペイ決済の導入

安東孝二

1　概要

東京大学情報基盤センターの運用する教育用計算機システム(以下 ECCS)では、印刷に関してプリペイドカードによる課金を行っている。一方、東京大学では職員証、学生証に IC カードの導入を進めており、希望すればクレジットカードの与信枠に基づいたポストペイ機能が利用できる。プリントに関する課金処理をポストペイ機能を利用して行うためのシステムを導入した。

2　プリント課金処理とポストペイ

2.1　背景

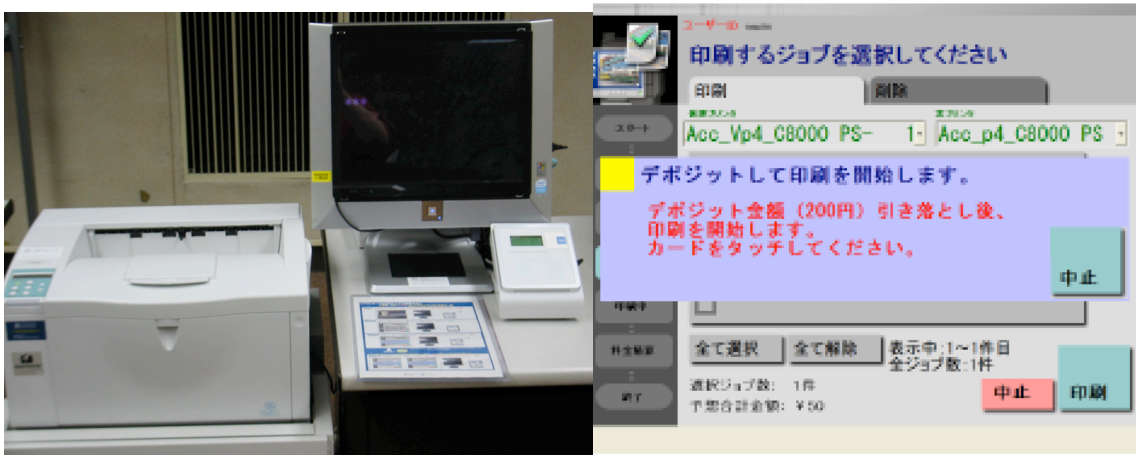
プリペイドカードによる課金はうまく機能しているがいくつかの解決すべき問題がある。セキュリティ上の理由からプリペイドカードの最高額面を 1,000 円に設定しているが、2 枚を使っての精算が不可能なため、どうしても残額が少ないカードの使い勝手が悪い。また、大量の印刷が一気に出来ないなど、課金処理のための不便がいくつか生じていた。

2.2　内容

新たに追加導入したプリンタシステムではポストペイ課金が可能な仕組みを構築した。従来からのオンデマンド印刷を踏襲しているが、IC カードに記録されている東京大学内での全学 ID と ECCS アカウントの連携を取ることにより、ユーザが ECCS アカウントとパスワードの入力をすることなく、IC カードを非接触リーダにかざすだけでオンデマンド印刷を可能にしている。また、非接触インターフェイスを利用するため、デポジットの概念を導入し課金処理に問題が起きないように工夫した。

2.3　具体的成果

本郷地区2台、駒場地区2台のポストペイ対応プリンタを導入し、サービスに供している。



3　成果要覧

受賞関連

[受賞 1] 安東孝二: 平成 18 年度情報処理学会山下記念研究賞,「大学における印刷管理の新しい試み」(2005-DSM-038)

その他の発表論文リスト

[発表 1] 安東 孝二,山口 和紀,関谷 貴之: 東京大学 ECCS プリンタ課金システムにおけるポストペイ決済の導入,平成 18 年度情報処理教育研究集会予稿集(2006)

教育支援システムの開発と運用 — 教育用計算機システム，学習管理システム CFIVE, 講義データベース —

関谷 貴之

1 教育用計算機システム

1.1 背景

高等教育機関において教育用の計算機環境を提供するに当たっては，多数の利用者に対して，各種のアプリケーションが動作する，一様なハードウェア・ソフトウェア環境を，安価な導入費用や低い運用コストで構築する必要がある．本センターでは，NetBoot によって Mac OS X が動作する端末を中心とした教育用の計算機環境 (ECCS2004 という) を 2004 年 3 月より運用している．この ECCS2004 の運用を通じて，今後の教育用計算機環境のあり方について研究を行っている．

1.2 内容

1.2.1 ECCS2004 の設計

ECCS2004 の設計にあたっては，主に以下に述べる要求仕様を設定している．

多数の端末を提供

コンピュータやネットワークが情報の発信や伝達の手段として身近になった現在において，教育用の計算機環境は，コンピュータの仕組みや活用方法を学ぶ授業の場で利用されるだけでなく，教育や研究を支える道具として時間や場所を問わずに利用される．

そこで，教養学部を中心とする講義での端末需要を満たすべく，数十から 200 台程度の端末を設置した教室が 10 室以上提供できるだけの端末を用意し，更に図書館のような公共の場にも端末を設置する事とした．結果として，ECCS2004 全体では 1,000 台以上の端末を提供する事が要求仕様の一つとなった．

多数の利用者に基本的なインフラを提供

WWW や電子メールが教育や研究における情報発信・情報共有・情報交換の基本的な手段として普及し，教育用計算機システムがこれらの手段を提供している．そこで，ECCS2004 では端末上で WWW ブラウザやメーラが動作させる事が ECCS2004 の要求仕様の一つとなった．

また本センターでは，教育用計算機システムとは別に，任意のメールドメインを利用できるメールサーバ環境を，学内向けのサービスとして提供している[1]．ECCS2004 はこの学内向けのメールサーバ環境も同一のシステムとして提供する事となったため，任意のメールドメインを作成可能で，アカウントの管理等をメールドメイン毎に個別の管理者に委譲できるシステムを導入する事が要求仕様の一つとなった．

各種のアプリケーションを実行可能

教養学部前期課程の情報リテラシー教育においては，OS には UNIX 系の OS を，プログラミング

[1]学内組織向けメールサーバ (MailHosting)
http://mh.itc.u-tokyo.ac.jp/

言語には Java を用いてきた．また，プログラミング作業は，Emacs などのエディタ上でファイルを作成して，ターミナル上でコンパイルするというものである．また，従来より教育用計算機システムでは，GNU ソフトウェアを利用してきた．従って，GNU ソフトウェアなどの各種オープンソースソフトウェアや Java VM が動作する事が要求仕様の一つとなった．

一方，情報リテラシー以外のいくつかの講義で利用するために，統計処理ソフトウェア SAS，表計算ソフトウェア Excel，そして CAD として AutoCAD が動作する事が求められた．

多様な用途に配慮

教育用計算機システムの利用目的は多岐に及ぶようになった．例えば，教育用計算機システムのアカウントを所持しない教職員や学生を対象として，特定のソフトウェアが動作するコンピュータを使った講習会や授業を開きたいという要望があった．

このような要望を実現するために，教育用計算機システムの通常の利用者以外であっても，一つの授業やある一日だけなど一時的にシステムを利用可能に設定した上で，教育用計算機システムの標準的な環境とは異なるソフトウェアが動作する環境を提供し，終了後は簡単に元の状態に戻せる事が要求仕様の一つとなった．

また，各学部や学科で必要な計算機環境を用意するが，当該組織に所属する学生や教職員の教育用計算機システムの認証情報を参照したいとの要望もあった．そこで，利用者の所属等に基づいて柔軟に利用権限を設定可能で，各種の OS やアプリケーションの認証に利用可能な仕組みを導入する事も要求仕様の一つとなった．

端末上のハードディスクを利用しない

1999 年から運用してきた教育用計算機システムでは，ディスクレスの端末を導入する事で，それ以前のシステムよりも端末の故障率を減らし，故障時の端末の交換作業も容易なものとなった．そこで ECCS2004 においても，ハードディスクを使わない端末を導入する事とした．

端末上で主に動作する OS は UNIX とする

本センターでは，UNIX 系の OS の管理の経験やノウハウが蓄積されており，経験の乏しい他の OS の場合に比べて，障害からの復旧も容易であると考えられる．そこで，ECCS2004 においても，UNIX 系の OS を端末の OS とする事を設計方針とした．

1.2.2 ECCS2004 の構築と運用

1.2.1 節で述べた要求仕様を踏まえて，ECCS2004 は NetBoot による端末を中心としたシステムを導入した図 1 は，ECCS2004 の大まかなシステム構成を示している．主な構成要素を以下に説明する．

端末

授業や教育・研究の情報基盤として中心的な役割を果たす iMac 端末 1,149 台と，Windows を必要とする講義用のディスクレス Windows 端末 227 台で構成される．iMac 端末は端末上のハードディスクのコネクタを取り外して動作しない状態にして，Xserve をブートサーバとする NetBoot によって起動する．

印刷

端末から印刷ジョブを送った後，プリンタのそばにある専用の端末を操作し，利用者自身がプリペイドカードで印刷料金を支払うシステムを導入した．更にこのシステムでは，プリンタの管理を外部業者に委託しており，プリペイドカードの料金収入に基づいて，トナーやプリンタ用紙の補給やメンテナンスを実施してもらっている．

認証

認証情報は LDAP サーバから各サーバや端末に対して提供する．マスターサーバ 1 台に対して，スレーブサーバが 3 台と Windows 用の Active Directory サーバが 3 台存在する．これによって，全ての端末や各種のサービスで認証情報を共有することを可能としている．

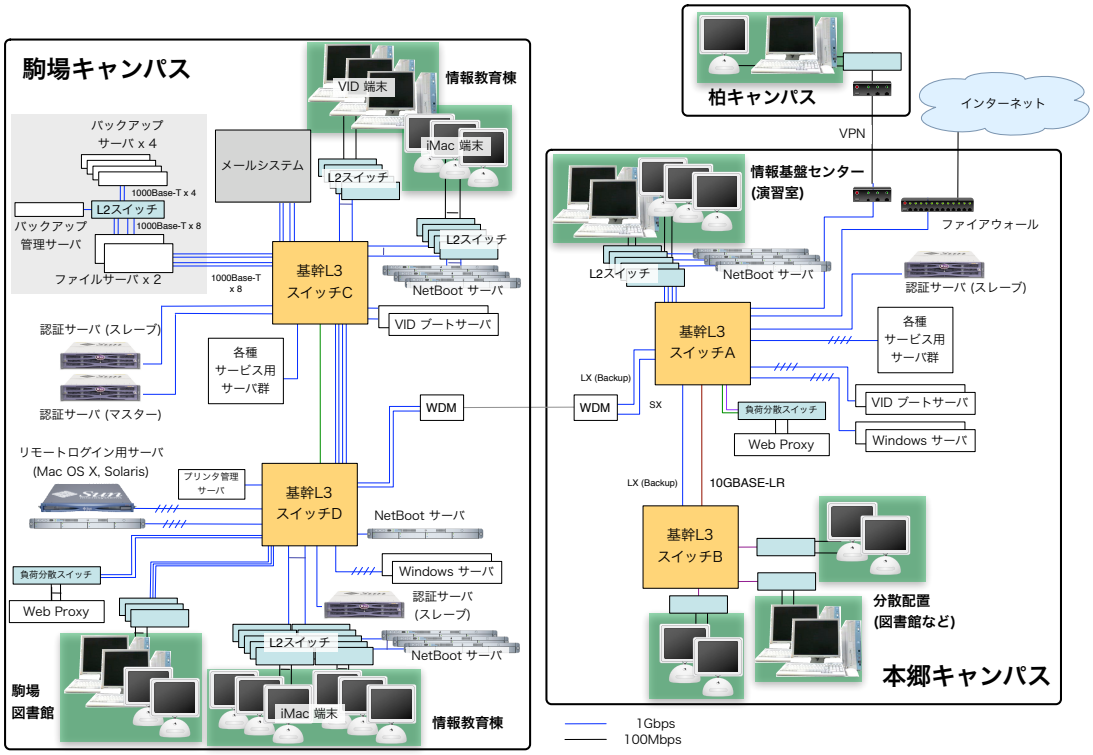


図 1: ECCS2004 のシステム構成

1.3 成果

教育用計算機システムは，世の中の計算機のハードウェアならびにソフトウェアの進歩に追いつくため，4，5 年を目安に定期的に更新している．次回の更新は 2008 年 3 月に迫っており，既にシステム更新のための準備作業を行っている．

ECCS2004 の最も大きな成果は，NetBoot を用いた Mac OS X の端末を教育用計算機システムとして導入した事である．当初 Windows に比べると一般的には馴染みの薄い OS を用いた事による話題性が強調されたが，システムを安定して運用できる事実が次第に正当に評価された．その後，ECCS2004 と同様のシステムが他大学でも導入されるようになった．

一方，実際の運用を通じて得られたノウハウ等は多数あるが，そのうちの一つとして端末のハードウェアの故障率については，次の知見が得られた．

ECCS2004 の iMac 端末は，端末上のハードディスクを利用せずに NetBoot で起動するのは当然のことながら，通常店頭で販売している iMac に搭載済みのハードディスクが動作しないように，一度筐体を開けてディスクに接続されている電源コネクタを外している．これによって，機械的に動作する部品を減らし，端末の故障率を下げる効果を期待した．

iMac 端末においては，全 1,149 台のうち，システム導入時の初期不良で交換した端末を除くと，2 年間に 53 台 (約 4.6%) の iMac を交換している．一方，ディスクレス Windows/UNIX 端末においては，そもそも筐体内にハードディスクが存在しない状態で出荷されている．こちらは全 227 台のうち，約 2 年間に 12 台 (約 5.3%) の端末本体と 15 台 (約 6.6%) の液晶モニタが故障のため交換されている．iMac 端末が液晶モニタが本体に接続された一体型であることを考慮すると，むしろ iMac 端末の故障率の方が小さい．

端末の交換作業自体は比較的容易で短時間で完了するが，障害のある端末を特定して，準備した代替機に交換するまでには時間を要し，その間当該端末は利用できない．従って， 障害のために利用できない端末を減らすべく，より故障率の小さい iMac 端末を中心とした構成にした事は有効であった

といえる．

但し，故障率の絶対的な値自体は，ハードディスクを使っていない割には高いのではないかと我々は考えている．一般的な PC のハードウェアの故障率が近年低下しており，デスクトップ PC では最初の 1 年目の年平均故障率は 5%と Gartner 社は報じている．調査対象の一般的な PC と教育用計算機システムの端末では動作環境も大きく異なるため，両者を単純に比較することはできないが，故障率を下げるという目的の実現に当たっては，ハードディスクを停止させた効果の有無は判断しにくい．

ハードディスクを使わない端末を導入したことで，故障時の端末の交換作業は容易で，ECCS2004 の端末の運用コストは小さい．しかし，端末上のハードディスクをキャッシュや一時的な保存領域として利用することで，交換作業の手間を増やさずにハードディスクを有効に活用できた可能性もある．従って，今後の教育用計算機システムの端末選定時には，ハードディスクを全く使わない事に拘らずに，総合的に判断するべきであると考えられる．

2 学習管理システム CFIVE

2.1 背景

情報基盤センターは，2004 年に日本ユニシス・ソリューションと共同でオープンソースの LMS(Learning Management System) である CFIVE(Common Factory for Inspiration and Value in Education) を開発し，同年 4 月より学内の教職員・学生向けのサービスとして，CFIVE の実行環境を提供してきた．また，サービスの利用者である教職員や学生からの要望や，本センターでの運用上の工夫などを取り込みながら，新しい機能の開発や不具合の修正を行い，現在も継続して CFIVE を改良している．その一方で，実際に授業で CFIVE を利用することで，設計開発時には想定していなかった問題も，次第に明らかになっている．

2.2 内容

2.2.1 システムの改良

2004 年 4 月に CFIVE をリリース後，約半年から一年毎に大きな変更を続けており，2005 年 5 月の時点でバージョンは 1.2 となっている．変更点は，不具合の修正と，実運用を通じて，利用者から要望のあった機能の実装である．主な改良点を以下に紹介する．

- テストのプレビュー
  テストを作成する教員や TA が，学生と同じ立場で，つまり WWW アプリケーションとして学生の利用者と同じ画面で，自身が作成したテストに解答することで，学生の使用感や設問の妥当性等を確認したいとの要望があった．そこで，これを実現するプレビュー機能を実装した．

- 数式や Wiki 書式への対応
  テストの問題文や課題の記述に際し，文中に数式やソースコードを入れたり，箇条書きや整形済みのテキストを入れたいとの要望が教員から寄せられた．そこで，mimeTeX[2]を用いた数式の画像を取り込んだり，Wiki 書式でテキストを整形できる機能を実装した．

- ダイレクト URL
  CFIVE 上の掲示板やテストなどを直接参照できるリンクを，通常の WWW ページに貼りたいとの要望が，講義に関する情報を提供する Web ページを公開している教員から寄せられた．そこで，ダイレクト URL と呼ぶ CFIVE 上のページを参照可能な URL を生成する機能を実装した．

[2]http://www.forkosh.com/mimetex.html

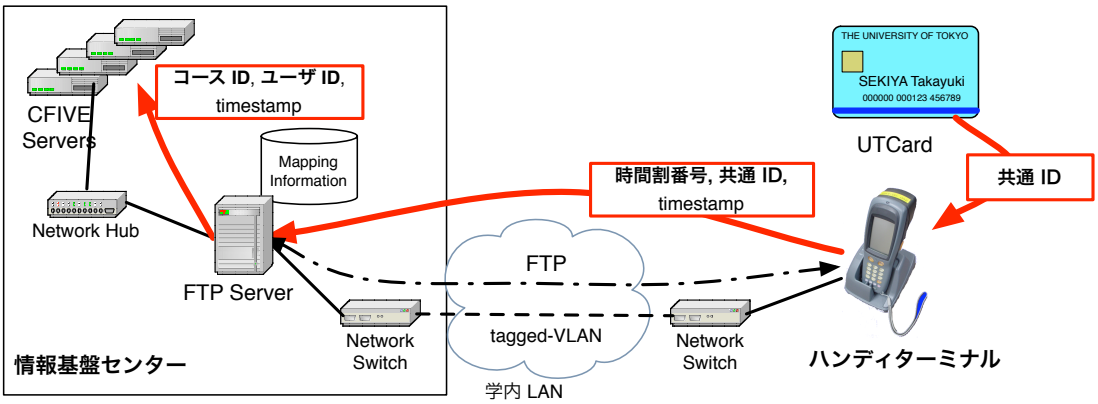


図 2: 出欠管理システム

- WebDAV 対応
  レポートの一括ダウンロードや，コメントを付けたレポートを提出した学生に返却するといった作業を，容易に行えるように，WebDAV サーバの機能を実装した．

2.2.2 システム連携

学務システムとの連携

オープンソースとしての CFIVE の変更以外に，本学内の教育環境の必要に応じて，他のシステムとの連携の仕組みを構築している．

本学の教養学部は，2004 年より UTask-Web と呼ぶ学務システムの開発を開始した．UTask-Web は，教養学部前期課程の学生の履修登録や成績管理を行うシステムだが，本センターが教育用の計算機システムや CFIVE のために用いているようなアカウント管理の枠組みがこれまで教養学部には無かった．そこで，UTask-Web と CFIVE との連携を行う事となった．

CFIVE と UTask-Web との間では，次の 3 つの情報を交換している．

1. センターが発行する学生の認証情報を UTask-Web に提供
2. UTask-Web を利用する教員の認証情報を CFIVE に提供
3. UTask-Web の履修者情報を CFIVE に提供

両システム間は，ハードウェアベースの VPN を用いて接続し，更にその上で SSH でファイルをコピーする事で，情報を安全にやり取りしている．

IC カードベースの出欠管理システム

学生が所持する IC カード機能付き学生証の共通 ID をハンディターミナルと呼ぶ携帯型の端末を用いて読み込む事で，教室で取った出欠情報を CFIVE で管理可能な IC カード機能付き学生証による出欠管理システムを，カードののベンダーである NTT コミュニケーションズと，ハンディターミナルのベンダーである DENSO と共に 2005 年に開発した．図 2 は出欠管理システムの概念図を示す．

出欠管理システムは次のように動作する．まず，受講者である学生が所持する学生証から共通 ID を，教室でハンディターミナルを用いて読み込む．次に授業終了後に，専用の通信ユニットにハンディターミナルを置き，共通 ID の情報を管理する専用のサーバに送信する．共通 ID 管理サーバでは，共通 ID に対応する学生の氏名や学生証番号と，時間割番号に対応する CFIVE 上のコースの情報を保持する．この情報に基づいて，共通 ID のリストから CFIVE のアカウントのリストを生成する．更に，上記のリストを CFIVE の出欠情報として登録する．CFIVE のインタフェースを通じて，教員は出欠情報を管理し，学生は自身の出欠状況を確認する事ができる．

この一連のシステムは，約一ヶ月で開発する事ができた．これは，CFIVE 上の出欠情報に関するデータ形式が公開されており，SQL コマンドを実行するだけで，CFIVE のデータベースへのデータ入力が可能であったからである．

2.3 成果

CFIVE は 2004 年 4 月に学内向けにサービスを開始してから，これを利用する講義や学生の数は着実に増加している．

学務システム UTask-Web と CFIVE との連携業務を行っているが，特に問題は起きていない．本センターにとっては，UTask-Web の履修者情報がオンラインで得られる事で，かつてフロッピーディスクを教養学部教務課に受け取りに行った状況に比べると，手間を省く事ができた．出欠管理システムも大きな問題は起きていない [査読付 1, 査読付 2].

3 講義データベース

3.1 背景

教材や論文のレポジトリ，オンラインシラバスという形で電子化されたコンテンツが大量に存在する現在，コンテンツの検索や類似度を算出する研究は広く行われている．これに対して，コンテンツが対象とする分野の範囲や，その分野の中でのコンテンツの位置づけを全体的に把握する研究は比較的少ないと考えられる．そこで，コンテンツの情報から，対象分野の用語や用語間の関係からなるドメイン・オントロジーを構築し，コンテンツの検索・俯瞰に活用するための研究を行っている．

3.2 内容

2006 年度は，予備的な実験として，情報処理教育関連の過去の論文や東京大学教養学部のシラバスからの用語の抽出を試みている．

専門的な用語の抽出にあたっては，論文等のテキストに含まれる専門用語に関して，以下が推測される．

- 同一の論文には互いに関連のある複数の専門用語が用いられる．
- 関連のある複数の論文に共通して用いられる専門用語は，互いに関連がある．

また，用語の多くは多義的で，用語の意味が文脈依存であると考えている．このような前提の下で，本研究では単一の用語を扱うのではなく，用語の集合を意味のある概念として扱い，ガロア束を用いて「専門用語の集合」を抽出する．更に論文のクラスタリング結果と対照させることで,「意味の有りそうな専門用語の集合」を行う．具体的には次の手順を実施する．

1. 論文のテキストから，mecab を用いた形態素解析によって用語を抽出する．
2. 論文の集合と論文に含まれる用語の集合に対して，ガロア束を適用する．これによって，用語の部分集合と論文の部分集合の対応を示すコンセプトが得られる．この用語の部分集合を専門用語の集合の候補とする．
3. 1 の処理結果から，各論文に含まれる用語の頻度に基づいて，isomap を用いた論文のクラスタリングを行う．
4. 2 のコンセプトについて，その論文が 3 で得られた論文のクラスタに偏在する場合，そのコンセプトに対応する用語の部分集合は「意味の有りそうな専門用語の集合」と見做す．

実験として，平成 14 年度情報処理教育研究集会論文集のうち，207 本の論文から約 13,000 語の単語を抜き出した上で，専門用語の集合を求める事を試みている．図 3 は前述の手順の第 2 段階を実施した結果である．図 3 中のクラスタ A に出現する傾向のある専門用語の集合を，表 1 に示す．

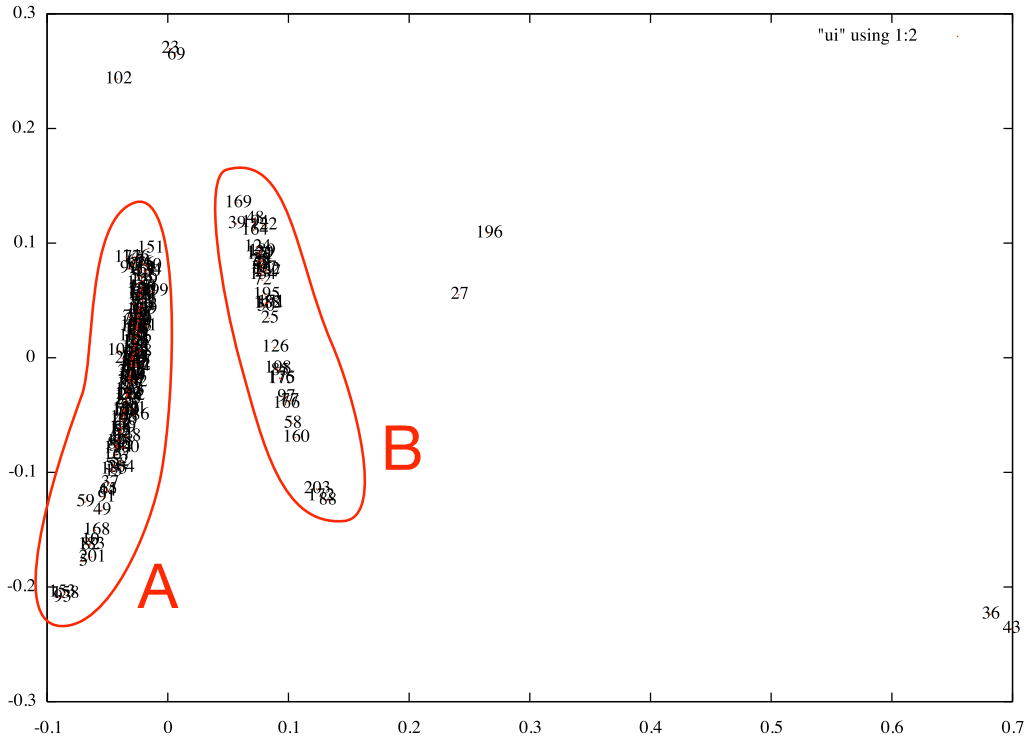


図 3: isomap による論文のクラスタリング

表 1: クラスタ A に出現する傾向のある専門用語の集合

{563:情報処理教育, 1072:カリキュラム }
{563:情報処理教育, 625:情報処理, 1072:カリキュラム }
{477:取り組み, 502:情報教育 }
{1197:パソコン, 3213:タイピング }
{1614:速度, 3213:タイピング }
{563:情報処理教育, 1072:カリキュラム, 1805:ビジネス }
{255:デジタル, 1456:画像 }
{1352:映像, 1397:データ }
{167:文書, 1197:パソコン }
{502:情報教育, 1197:パソコン }

3.3 成果

現時点では，対象としたコンテンツの集合に対して，ある程度意味のありそうな専門用語の集合を抽出できたと考えている．今後シラバスに対して同様の手法を適用した上で，用語集合同士の関係に基づいて構造化したシラバスを可視化する仕組みを構築する予定である．なお，基本的な考え方について，CAL'07 においてポスター発表を行った [発表 1].

4　成果要覧

査読付論文リスト

[査読付 1] T.Sekiya and K.Yamaguchi. Case Study of Information Systems Cooperation Centered in Open Source LMS, CFIVE, FOSLET 2006, Proceedings of Workshop on Free and Open Source Learning Environments and Tools, pp. 21-28, Como, Italy, 2006.

[査読付 2] 関谷 貴之, 佐藤 孝明, 寺脇 由紀, 尾上 能之, 山口 和紀: 学習管理システム CFIVE の継続的な開発と運用の評価, 情報教育シンポジウム SSS2006, pp. 339-346, 2006.

その他の発表論文リスト

[発表 1] T.Sekiya and K.Yamaguchi. Ontology and context based repository of learning materials, CAL'07, Poster Presentation, Dublin, Ireland, 2007.

図書館電子化研究部門

Web テキストからの知識マイニングに関する研究

中川裕志

1 概要

本年度は以下のテーマについて研究を行った。

(1)統計的機械学習を用いた自然言語処理:このテーマに関しては、検索エンジンに人名を質問して得られたページを実体として異なる人物のページごとにクラスタリングするWeb名寄せシステムについて研究した。[査読付 4] [査読付 5] [公開 2]

(2)多言語テキストからの情報と知識のマイニング:このテーマに関しては、多言語のWebページから重要語を抽出するシステム「言選Web」について引き続き研究した。特に本年度は、中国語を対象にしたシステムの開発、評価を行った。一方、(1)にも関連するテーマとして、より深い知識抽出として、シーケンシャルマイニングを工夫し、テキストコーパスから主語、述語、目的語などの構造を効率的に抽出するアルゴリズムの研究を行った。[招待 1] [招待 2] [受賞 1] [査読付 3] [公開 1] [発表 2] [発表 4] [発表 5] [発表 6]

(3)Blog および Web2.0 関連の応用システム:多言語(日中韓英)の blog の収集および検索システム、ならにび話題となる単語に共起する単語群の提示を行うシステムを開発した。さらに、blog 記事に現れている感情を抽出する方法についても検討した。[受賞 2] [査読付 1] [査読付 6] [発表 3] [発表 7] [発表 8]

(4)自然言語処理を利用した語学教育システム:Webからダウンロードした英文テキストに対して、自然言語処理技術によって穴埋め 4 択問題を自動生成するシステムの開発と評価を行った。[査読付 2] [発表 1]。

以下では、紙数の関係により(1)(2)に関連した成果としてWebテキストの効率的利用を目指した知識マイニング、とりわけパタンマイニングアルゴリズム、および(3)に関連した成果として blog 関連の Web2.0 に関連する automated folksonomy について述べる。その他の成果については成果要覧に記した。

コーパスからのWebテキストの効率的利用を目指した知識マイニングを推進するにあたっては、大量ないし多様なテキストを処理する必要がある。このような目的に沿って、以下の基本的テキスト処理技術の探求、Web テキスト処理ツールの開発を行った。すなわち、大量の Web テキストから比較的軽い処理で知識の候補となるイベントを網羅的に抽出する半構造マイニングの手法を考案して、実装した。これについて2. で述べる。

最近、“Folksonomy” とよばれる、ユーザーのタグ(自由に選んだキーワード)付けによる分類法が注目を集めている。この手法では、ユーザーはタグでアイテムの中身を把握できるため、効率的な閲覧法を提供できる。最近、膨大な量が発信されている blog に対して Folksonomy を自動化するシステムを試作した成果について3. で述べる。

2 係り受け関係を考慮したテキストマイニング用の半構造マイニング手法

2.1 背景

テキストマイニングにおいて、精密な知識をマイニングするために、1 文内の単語の共起関係だけではなく、係り受け構造のような単語間の関係性を考慮することによって、文が持つ意味構造を含めたマイニングを行う重要性が高まっている。一般に、係り受け構造は、文節単位で係ることから、文節を節点に持つ木構造で表現されるので、FREQT[1][3]のような従来の木構造マイニング手法が適用可能である。しかし、係り受け構造を木構造で表現する場合、次のような問題が生じる。例えば、 図 2.1 にある(a)(b)2 つの木構造に共通する部分構造を抽出する場合を考えると、(c)のような部分構造が抽

出される。(c)が抽出されたとしても、節点数が少ないことから、得られる情報量が少ないという問題が生じる。これは、文節を節点とした木構造の場合、助詞や表記上のぶれ(例では、「に」と「を」の違い、「首相」と「総理」の違い)などによって、ラベルの異なり数が増加することに起因する。

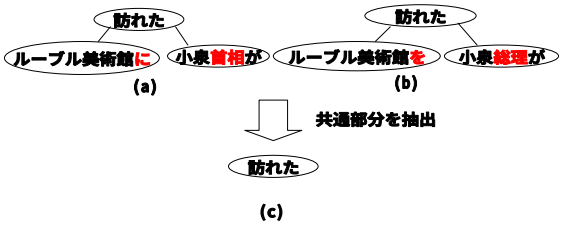


図 2.1 木構造表現による共通部分木の抽出例

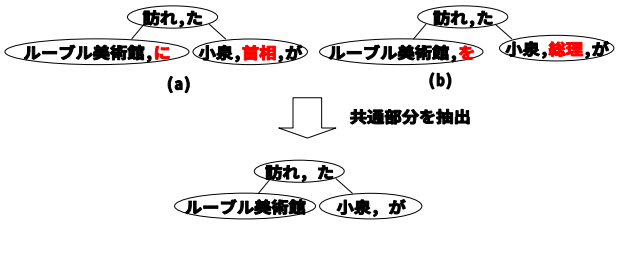


図 2.2 提案データ構造による共通部分構造の抽出例

よって、節点数の少ないパターンが抽出され、単語間の関係(例では、「ルーブル美術館」「小泉」「訪れる」などの関係)が抽出できない場合がある。本研究では、係り受け構造を表現する新しいデータ構造を提案し、そのデータ構造に対するマイニング手法を提案することで、上記の問題を解決する。提案するデータ構造は、単語(形態素)を1つのアイテムとし、木構造の各節点に複数のアイテムを保持する半構造である。図 2.2 に具体例を示す。

各節点は、複数のアイテムを保持している。節点間の比較は、文節単位ではなく、形態素単位で、共通する部分構造を抽出する。これにより、図 2.2 の(a)(b)からは、(c)のような部分構造が抽出される。ここで重要なのは、含む節点数が多いパターンが抽出できることである。図 2.1 の抽出された部分木の節点数が1であるのに対し、図 2.2 では、節点数が 3 のパターンが抽出される。これにより図 2.1 と比べると、図 2.2 では、情報量の多い頻出パターンが抽出されていると考えられる。提案するデータ構造は、厳密には木構造ではないので、従来の木構造マイニングアルゴリズムは適用できないが、シーケンシャルパターンマイニングアルゴリズム PrefixSpan を改良することで、提案するデータ構造に対する半構造マイニングアルゴリズムを実現する。

2.2 内容

2.2.1 シーケンシャルパターンマイニング

$I=\{i_1,i_2, \cdots, i_n\}$ をアイテムの集合とする。このアイテムの集合を**エレメント**と呼ぶ。エレメントは、アイテムを'()'で囲むことで表現する。a,b をアイテムとすると、(a,b)などがエレメントである。エレメント中のアイテムの順番は考慮しないため、予めエレメント中のアイテムを辞書順にソートしておく必要がある。しかし、アイテムの元来持っている順序を考慮する必要がある場合、エレメント中のアイテムを辞書順にソートしないでおけば、エレメント中のアイテムの元々の順序を考慮した処理が可能である。

シーケンスとは、順序を保持するエレメントの列である。例えば、(a,b)(a,d) ≠(a,d)(a,b)である。

シーケンス s を s=<$\mathbf{e_1}$,$\mathbf{e_2}$,…,$\mathbf{e_l}$>と表記する。e_k(k=1,2,…,l)は任意のエレメントである。例えば、<(a b) (a c) (b d) (a b)>がシーケンスである。

シーケンス中のアイテムの個数をシーケンスの長さとする。

あるシーケンス s_1 中のすべてのアイテムが、別のシーケンス s_2 中に存在し、その順序も保持している場合、s_1 を s_2 のサブシーケンスと呼び、このときの s_1 と s_2 の関係を $s_1 \subseteq s_2$ と表記する。

シーケンスデータベース S とは、シーケンス ID(sid)とシーケンス s のタプル(sid,s)の集合である。

$S=\{(sid_1,s_1),(sid_2,s_2), \cdots,(sid_n,s_n)\}$

シーケンス α のシーケンスデータベース S における**サポート値**とは、S 中のすべてのシーケンスのうち、シーケンス α を含むタプルの数である。$support_S(\alpha) = |\{(sid,s)| (sid,s) \in S \wedge \alpha \subseteq s\}|$

シーケンシャルパターンマイニングとは、シーケンスデータベース S から、**最小サポート値**と呼ばれる任意の正の整数 ξ に対し、$support_S(\alpha) \geqq \xi$ となるような頻出シーケンス α を全て抽出する問題である。

PrefixSpan[4]は、シーケンシャルパターンマイニングの代表的アルゴリズムであり、**射影(Prefix-Projection)** という操作を、深さ優先で再帰的に行うことで、頻出シーケンスを効率的に抽出する。射影とは、射影するアイテムの **Postfix** を新たにシーケンスデータベースとする操作である。例えば、シーケンス s=<(c,d)(b)(a,d)(c,d)(b,c,a)> を a で射影すると、<(_d)(c,d)(b,c,a)>(= s における a の Postfix)となる。つまり、シーケンス中に最初に現れる a 以降のアイテム列を Postfix とし、この Postfix を新たにシーケンスデータベース(**射影データベース**と呼ぶ)とすることである。エレメントの内外にあるアイテムを区別するために'_'という Prefix をつける。

2.2.2 提案手法:pFREQT

係り受け構造を、文節をラベルとする節点を保持する木構造で表したのでは、助詞や表記上のぶれにより、節点数の少ない部分木が抽出されてしまう。そこで係り受け構造を表現する新しいデータ構造を提案し、そのデータ構造に対するマイニング手法を提案することで、上記の問題を解決する。

提案するデータ構造は、厳密には木構造ではないので、従来の木構造マイニングアルゴリズムが適用できない。

しかし、図 2.3 に示すように、深さ優先探索で反時計周りに前順走査で各節点を走査し、節点内はアイテムの順序に従って各アイテムを列挙し、index を割り当てれば、シーケンシャルパターンマイニングの適用が可能なデータ構造となる。エレメントは、節点に割り当てられたアイテム集合となる。ただし、各アイテムは、順序を保持していると考え、集合ではなく順序付きの列とする。各エレメントに、木構造における情報(親，長子，次弟)のエレメント番号(すなわち index)を保持させる。存在しない場合は、index を-1 とする。上記の操作によって得られるシーケンスを**準木構造シーケンス**と呼ぶことにする。なお、ある節点の「長子」とは、一般には複数存在するその子節点のうちの最左節点である。また、ある節点の「弟」とは兄弟節点のうち、自分の右側にある節点であり、「次弟」とは、弟のうちで最左、すなわち自分の隣接する弟である。図 2.3 の四角で囲まれた部分が、シーケンシャルパターンマイニングのデータ構造となっている。よってシーケンシャルパターンマイニングアルゴリズム PrefixSpan が適用可能となる。ただし、単純に適用したのでは、図 2.3 の(a,b,c)(c,b)のような非連結なパターン(木構造となっていないパターン)も抽出されてしまうため、制約が必要となる。

pFREQTと名付けた本提案手法は、上記のような変換されたデータ構造に対して、連結されたパターン(部分木)のみ抽出するための制約付きの射影(**Tree-Projection**)を行うアルゴリズムである。以下にアルゴリズムで使う用語を定義する。

射影したアイテムを積んでいくスタックを**「射影スタック」**と呼ぶ。あるアイテム i で射影したとすると、アイテム i は射影スタックに積まれる。アイテム i での射影が終了すると射影スタックから削除される。射影スタック中のアイテム集合を **Proj-Items** と表現する。この Proj-Items が抽出される頻出パターンであり当初目的としたものである。 エレメント I と A の間に経路が存在し、エレメント I の方が A よりも深い位置にいるとする。I と A の経路上のエレメント数(節点数)が k-1 のエレメントであるとき、 エレメント A をエレメント I の「k 代前の祖先」と呼ぶ。エレメント I の親は、1 代前の祖先となる。エレメント I

の 0 代前の祖先をエレメント I 自身とする。例えば、図 2.3 の index=2 のエレメント(c,b)の 1 代前の祖先は(d,a)、2 代前の祖先は、(a,b,c)である。

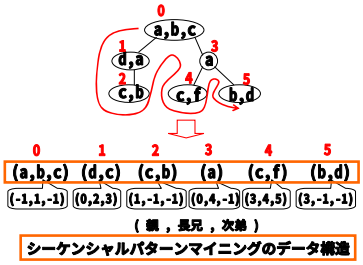


図 2.3 データ構造の変換例

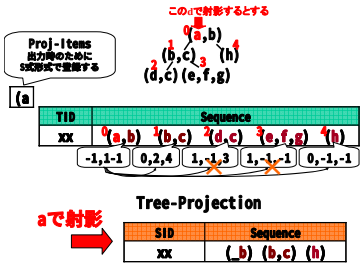


図 2.4 index=0 の a による Tree-Projection の動作例

準木構造シーケンス S のアイテム i による **Tree-Projection** とは、以下の制約を満たす射影である。

制約:射影データベース $S_{<i>}$に含まれるアイテムは、アイテム i の Postfix の中で、射影するアイテム i または Proj-Items 中のアイテム n との間に枝を持つアイテム、つまり、連結するアイテムのみ射影データベースとすることを意味する。図 2.4 に具体例を示す。図 2.4 は、index=0 のエレメント(節点)中のアイテム a による Tree-Projection の動作例を示している。

Tree-Projection は、以下の 3 つの射影からなる。

element-projection

child-projection

Level k sibling-projection

ただし、各射影において、次に射影するアイテムを選択し、Tree-Projection が再帰的に呼び出される。射影したアイテムは、射影スタックに積まれていく。射影が終了すると、スタックから削除される。

以下、それぞれの射影を説明する。

element-projection

準木構造シーケンス S のアイテム i による element-projection とは、アイテム i と同一のエレメント(節点)内にあるアイテムのみ射影データベースに含める射影である。PrefixSpan では、'_'という prefix をつけることで、エレメント内外にあるアイテムを区別して抽出するが、このような射影を別に設けることでも実現できる。

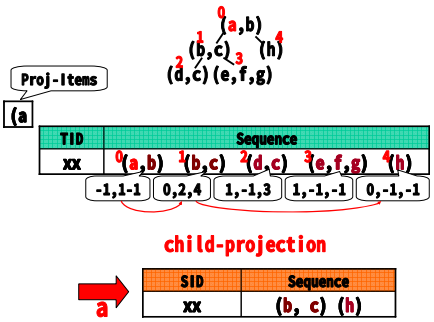


図 2.5 index=0 の a による child-projection の動作例

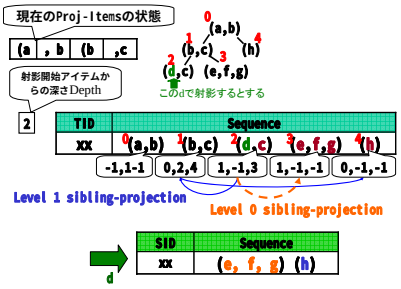


図 2.6 index=2 の d による sibling-projection の動作例

射影したアイテムは、射影スタックに積まれていく。射影が終了すると、スタックから削除される。

以下、それぞれの射影を説明する。

element-projection

準木構造シーケンス S のアイテム i による element-projection とは、アイテム i と同一のエレメント(節点)内にあるアイテムのみ射影データベースに含める射影である。PrefixSpan では、'_'という prefix をつけることで、エレメント内外にあるアイテムを区別して抽出するが、このような射影を別に設けることでも実現できる。

child-projection

準木構造シーケンス S のアイテム i による child-projection とは、アイテム i の子となるアイテムのみ射影データベースに含める射影である。

具体的には、以下の操作を行う。

アイテム i が属する節点の長子である子節点の index を取得する。

長子の次弟の index を取得する。

順次、次弟の index を取得していく。

以上により、アイテム i と子の関係にあるアイテム集合を取得できる。これらのアイテムのみ射影データベースに含める。動作例を、図 2.5 に示す。

Level k sibling-projection

準木構造シーケンス S のアイテム i による Level k sibling-projection とは、アイテム i の k 代前の祖先の弟となるエレメントのアイテムのみ射影データベースに含める射影である。具体的には、以下の操作を行う。

k 代前のエレメントの index を取得する。

k 代前の祖先の次弟の index を取得する。

順次、次弟の index を取得していく。

以上により、アイテム i の k 代前の祖先の弟となるアイテム集合を取得できる。これらのアイテムのみ射影データベースに含める。Proj-Items に 1 番最初に積まれたアイテムからの深さを **Depth** とすれば、上記の操作を k=0 から Depth－1 まで繰り返す。動作例を、図 2.6 に示す。

2.3 具体的成果

データセットは、(株)日本航空インターナショナルで収集されている航空安全レポート(個人情報は削除して、個人が特定できないようにした。)を用いた。1 文を単位とし CaboCha を用いて係り受け解析を行った。単一アイテムを節点とする木構造は、従来の文節を節点とした係り受け木とし、複数のアイテムを節点とする木構造は、文節をエレメントとし、単語(形態素)をアイテムとした。

提案したデータ構造に対するマイニング手法が従来存在しないため、提案手法の抽出速度は、厳密には評価できない。しかし、提案したデータ構造は、木構造をサブセットとして含む。よって、係り受け構造を、文節を節点とする木構造として表現したデータに対する速度評価を行う。

便宜上、本提案手法を pFREQT(projection-based FREQT)と以下呼ぶことにする。

図 2.7 に、文節を節点とする木構造データに対し、木構造マイニングアルゴリズム FREQT (http://chasen.org/ taku/software/freqt/) [1][3]と提案したマイニングアルゴリズム pFREQT を適用させ

た場合の速度の比較を示す。最小サポート値を 2 とした。図 2.7 からもわかるように、pFREQT は、FREQT よりも抽出時間が早い。これは、射影を節点間の関係によって分けているためパターンの探索時に、pFREQT のメモリ使用量が FREQT よりも少なくて済むからであると考えられる。

図 2.8 は、抽出されたパターンの節点数の統計量である。最小サポート値を 3 とし、節点が 2 以上のパターンを抽出した。2000 個の係り受け木を持つデータセットを 5 つ作成して評価を行った。

文節を節点として FREQT を適用させた場合と提案手法を適用させた場合で、抽出されたパターンの節点数の平均値、中央値、最大値の比較を行った。横軸は、各データセットであり、縦軸は、各々の節点数の統計量である。

図 2.8 により、本提案手法は、従来手法に比べて、抽出される部分木は、平均値、中央値、最大値すべてにおいて節点数の多い木である。より節点数の大きい部分木が抽出できれば、それだけ係り受け関係を持つ単語の関係を多く抽出できると言える。よって、本提案手法は、従来手法では抽出できないパターンを抽出可能であると言える。ただし、こうして取り出された節点は多数にのぼるため、知識として意味のある節点を選択する問題に取り組む必要がある。

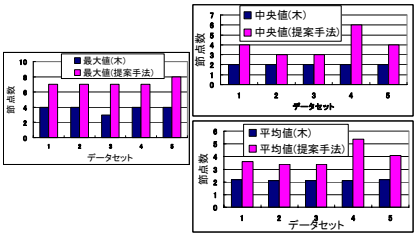


図 2.5 抽出した節点の統計量

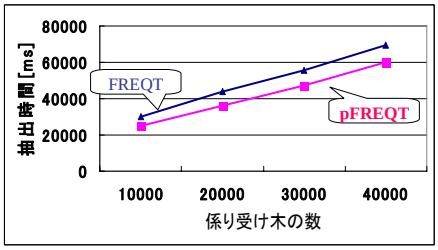


図 2.6 抽出時間の比較

参考文献

[1] Kenji Abe, Shinji Kawasoe, Tatsuya Asai, Hiroki Arimura, Setsuo Arikawa : Optimized Substructure Discovery for Semi-structured Data, Proc. of PKDD 2002(2002)

[2] R.Agrawal and R.Srikant : Mining Sequential Patterns, In Proc. of ICDE1995, IEEE Press , pp.3-14(1995)

[3] Tatsuya Asai, Kenji Abe, Shinji Kawasoe, Hiroki Arimura, Hiroshi Sakamoto, Setsuo Arikawa : Efficient Substructure Discovery from Large Semi-structured Data : In Proc. of SDM 2002(2002)

[4] J.Pei, J.Han, B.Mortazavi-Asl, H.Pnto,Q.Chen, U.Dayal, and M.Hsu : PrefixSpan:Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth, In Proc. of ICDE2001, IEEE Press , pp.215-224(2001)

3 Blog を対象にした Automated Folksonomy

3.1 背景

最近、Folksonomy[1]と呼ばれる新しい整理法が注目を集めている。Folksonomy[1] とは、個々のユーザーが自身の視点でタグ（自由に選んだキーワード）をアイテムに付与することで、システム全体と

[1] Folksonomy とは folk と taxonomy の合成された新語である。

してみれば多様なタグが個々のアイテムに付与されるというもので、ユーザーの語彙や視点が分類に反映されるという特徴をもつ。Flicker や del.icio.us といったサービスで採用され、これらのサービスの人気の主な要因といわれている。Flicker の画面を図 3.1 に示す。

画面下半分にユーザが与えた多くのタグが表示されており、そのうちのひとつをクリックすると、そのタグに分類された写真の一覧が表示される。(例えば、Family をクリックすると家族の写真のサムネール画像一覧が表示される。)

これらに示される、Folksonomy を用いた典型的なサービスについての概観を図 4 に示す。その構造と機能をまとめると以下のようになる。ユーザーは何らかのアイテム(写真や URL など)をサービスに登録し、他のユーザーと共有する。そして、ユーザーは自分のアイテムだけでなく他人のアイテムにもタグを付け、自分の好きな形で整理することができる。各々が自分の整理のために付けたタグは共有されるので、全体としてみると個々のアイテムにはたくさん視点を反映したタグが付く。

Folksonomy の特筆すべき特徴は、ユーザーによる分類という点である。現在、専門家による分類や著作者による分類が一般的だが、Folksonomy では一般ユーザーの多様な視点を反映した分類を行える。一人一人の個人に適切な分類能力があるとは考えにくいが、個人が集まり、たくさんの視点を組み合わせることで、分類の精度は高まる。タグはユーザーの視点でユーザーの語彙から選ばれたものであるため、その分類基準や表現は一般ユーザーの理解しやすいものとなる。また、語彙やその指し示す概念は時間が経つにつれて変化するが、従来の固定されたカテゴリ体系による分類では語彙やその概念の変化への対応は遅れがちである。Folksonomy では、図 3.2 に示すように、ユーザーが常に最新の語彙や概念に基づいた新しいタグをつけていくので、これらの変化に俊敏に対応できる

図3.1　Flickerの画面例

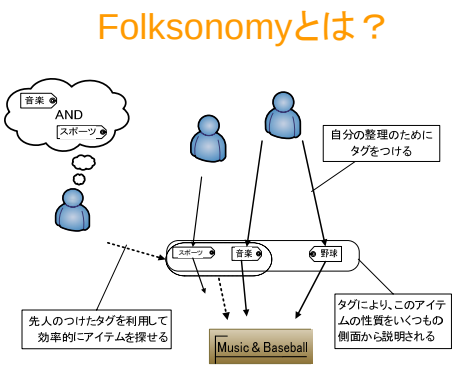


図 3.2:　Folksonomy の概念

このように、Folksonomy は
1. タグはユーザーの語彙から選ばれる
2. 分類には多数の個人の視点が反映される
3. 語彙やその指し示す概念の時間的変化に適応する
という特徴を持つので、Folksonomy では閲覧による受動的な情報検索を実現できる。
さて、情報検索には主に下記の 2 種類の方法がある。
・クエリを用い、その語を含む文書を探す（能動的）
・閲覧しながら、興味のある物を探す（受動的）
クエリによる方法は現在のインターネットでは一般的な方法である。インターネットの創成期には初期の Yahoo のような閲覧による方法も一般的であったが、現在では衰退している。
しかし、この 2 種は利用できる場面が異なり、ユーザーは両者揃った時に情報を最も有効活用できる。Folksonomy では、タグは個々のユーザーが整理するために付与するものであるので、通常、分類対象を記述する説明的な単語が選ばれる。このため、タグを利用した閲覧システムは、興味のあるものを探すのに有用である。
以上、述べたような Folksonomy の特徴は、最近、インターネットで盛んになってきた Weblog(blog、あるいは blog)の検索において重要な技術となる。blog では、個人的な日記や、事件や事実、知識に関する個人的な視点からの記述がなされることが多いため、専門家の分類にはなじまないところがある。よって、一般ユーザの使う用語でタグ付けし分類されることによって、blog は、高度な専門用語を知らない利用者からも容易にアクセスできるようになると期待される。
内容
Folksonomy のスケーラビリティは専門家によるそれよりもはるかに高いが、人手に依る以上、膨大な情報にはやはり対応できない。膨大な情報に対応するためには、Folksonomy を機械化する必要がある。この際、上記の Folksonomy3 つの特徴を失わないようにしなくてはならない。
上記の特徴 1 を満たすため、タグは blog のカテゴリ名を元にした。多くの blog サービスや blog ツールでは、ユーザー自身がカテゴリ体系を決め、その体系で自分の記事を分類できる。このため、blog で用いられているカテゴリ名はユーザーの語彙に基づいており、またユーザーが分類のためにつけた名前であるので、タグの候補となり得るものが多いと考えられる。また、特徴 2 を満たすため、マルチタグを付与することとした。さらに、個々のタグは複数の著者の blog から得た記事群を用いて学習させた分類器を用いて付与することとした。さらに、特徴 3 を満たすため、タグ付与器の学習
は繰り返し（毎日 1 回など）行うこととした。対象とするカテゴリ数が多いため、用いる機械学習アルゴリズムは適用だけでなく学習も高速であることが必要となる。
また、今回はタグを付与する対象としても blog 記事を利用した。これは、日本国内に限っても、去年 1 年間に新しく投稿された blog 記事は 1 億件に達するとみられ、人手による Folksonomy では網

羅できない規模であり、Folksonomy の自動化を試みる対象として適切だと考えたためである。そこで図 3.3 のようなシステムを提案することにした。これは blog 記事を対象としたマルチタグ付与システムであり、1 つ 1 つの記事に複数のタグを付与する。付与するタグや、そのタグの概念は収集した多数の blog 記事から抽出する。システムの各要素の詳細は文献[査読付 6]を参照されたい。

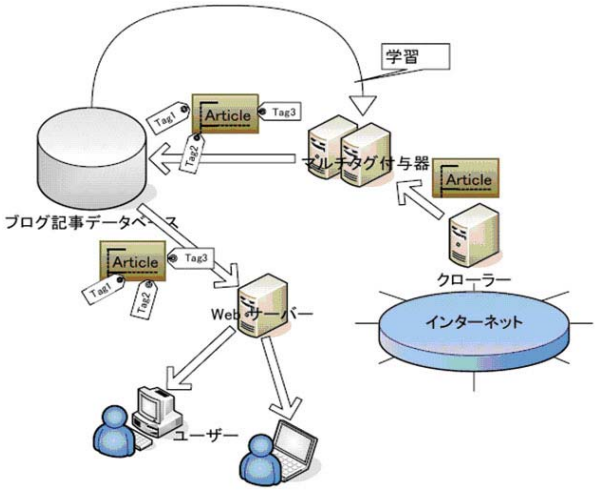


図 3.3: 機械化された Folksonomy システムの概観

3.2 具体的成果

上記のアイデアに沿って開発したシステムで、カテゴリ名あたりの blog 記事とタグ付け精度の関係として、サンプリングデータで図 3.4 のような結果が得られた。

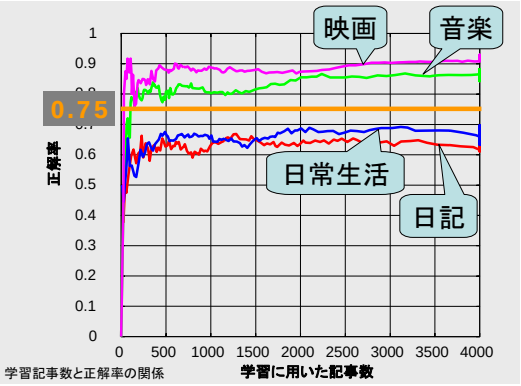


図 3.4 カテゴリごとの分類精度の違い

これによれば、ここでは 2.3 節で述べたバイアスのかかっていない SVM 出力で記事を分類した（SVM の実装として TinySVM4 を利用した）。ここで、精度は学習記事数の 1/3 の記事（学習とは異なる記事）で測定した。この図より、説明力のある語(日記, 日常生活) の精度は高い水準で、説明力のない語(音楽, 映画) は低い水準に収束することが分かる。そこで、実用的な性能を得られるカテゴリとして、図 3.4 の精度が 75%以上のカテゴリを選択して実験したところ、カテゴリ名あたりの blog 記事数が400以上あれば、精度 80% 、適合率 95%以上という実験的評価を得、実用に耐えうる性能をもつことが分かった。最後の試作したシステムの操作画面例を示す。

図3.5: 初期検索画面

ここで、「音楽」タグをクリックすると、音楽に関連するブログの一覧を示す次図のような結果画面が表示される。

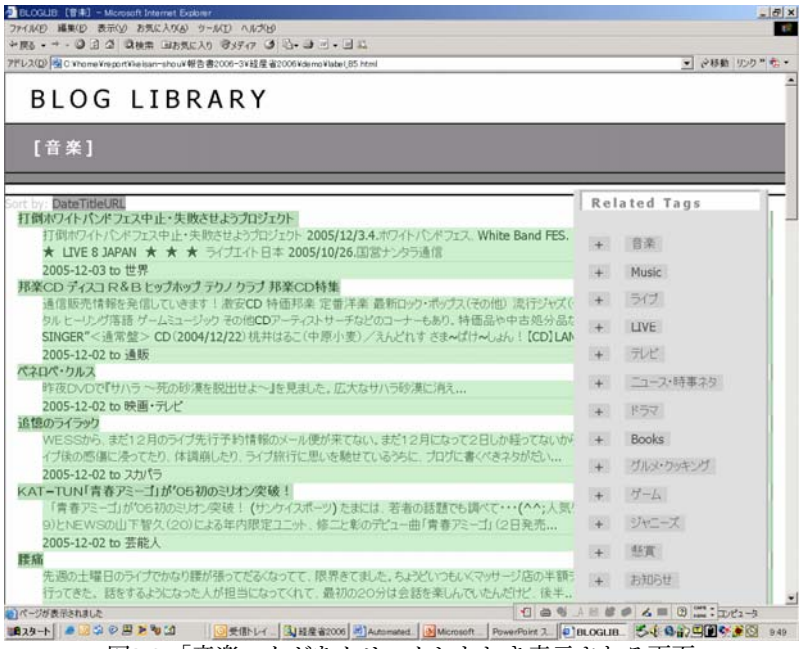


図3.6:「音楽」タグをクリックしたとき表示される画面

この画面でさらに Related Tags のうち「テレビ」をクリックすると、「音楽」と「テレビ」という2つのタグを持つ結果の一覧が表示される。

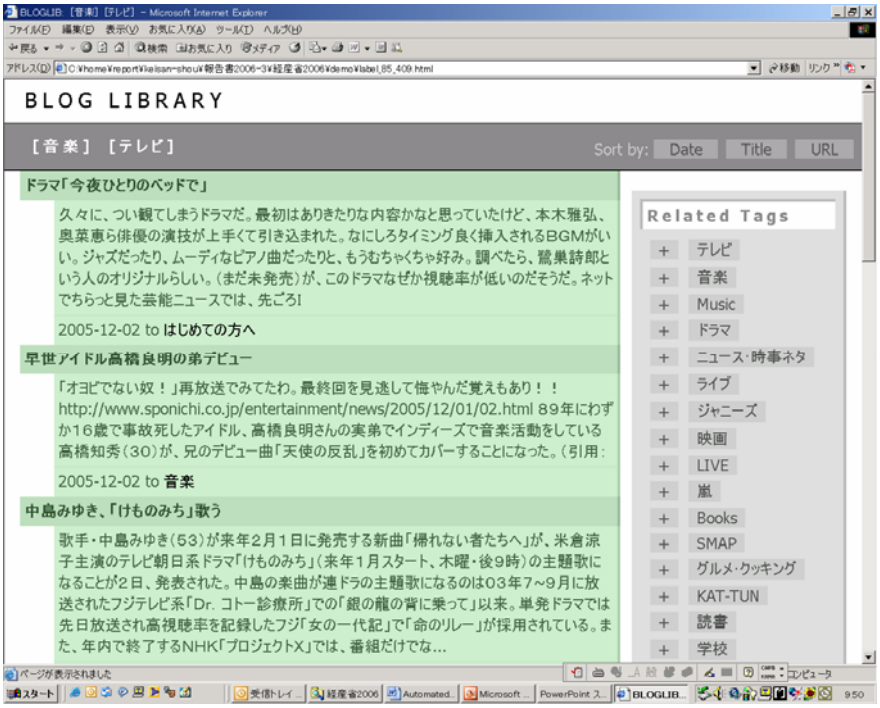


図 12: 「音楽」タグと「テレビ」タグを持つブログ一覧表示の画面例

4 成果要覧

招待講演／招待論文

[招待 1]中川裕志:Web ページからの用語抽出システム：言選 Web, DBWeb2006, 2006 年11月

[招待 2] 中川裕志：中国語テキストにおける専門用語，「機械翻訳のイノベーション」シンポジウム，独立行政法人情報通信機構、独立行政法人科学技術振興機構主催, 2007 年 3 月

受賞関連

[受賞 1] 佐藤一誠, 中川裕志:研究発表奨励賞 DBWS2006, 電子情報通信学会データ工学研究会，情報処理学会データベースシステム研究会,日本データベース学会, 2006 年 7 月

[受賞 2]福原知宏、村山敏泰、中川裕志、西田豊明 ： 2006 年度全国大会優秀賞，人工知能学会,2006 年 5 月

査読付論文リスト

[査読付 1] Tomohiro Fukuhara, Hiroshi Nakagawa, Toyoaki Nishida : Understanding Sentiment of People from News Articles: Temporal Sentiment Analysis of Social Events", ICWSM-2007 -- Int. Conf. on Weblogs and Social Media, pp.271-272, Boulder, Colorado, USA, Mar 26-28, 2007

[査読付 2] Ayako Hoshino, Hiroshi Nakagawa. "Assisting cloze test making with a web application", SITE 2007--Society for Information Technology & Teacher Education International Conference, pp.2807-2814, San Antonio, Texas, USA; March 26-30, 2007

[査読付 3]佐藤一誠、中川裕志: "係り受け関係を考慮したテキストマイニングのための半構造マイニング手法の提案 2, 日本データベース学会 Letters, Vol.5, No.2, pp.53-56, 2006 年9月

[査読付 4] Youji Kiyota, Hiroshi Nakagawa, "A Domain Ontology Production Tool Kit Based on Automatically Constructed Case Frames", LREC2006, pp.1482-1487, May 25-27, 2006, Genova, Italy

[査読付 5] Shingo Ono, Minoru Yoshida, Hiroshi Nakagawa. "NAYOSE: A System for Reference Disambiguation of Proper Nouns Appearing on Web Pages", in Information Retrieval Technology (Third Asia Information Retrieval Symposium, AIRS2006), pp. 338-349, Springer LNCS 4182, 2006, Singapore

[査読付 6] Tsutomu Ohkura, Youji Kiyota, Hiroshi Nakagawa. "Browsing System for Weblog Articles based on Automated Folksonomy", WWW2006 Workshop on the Weblogging Ecosystem, May, 2006, Edinburgh, Scotland

公開ソフトウェア

[公開 1]前田朗、小島浩之、中川裕志:言選 Web 中文版、2007 年 4 月 2 日

[公開　2] 小野真吾、中川裕志:Nayose, Web ページの名寄せシステム(同姓同名の人の Web ページを実際の同一人物ごとにまとめるシステム)、2007 年 7 月

その他の発表論文リスト

[発表 1] Ayako Hoshino, Hiroshi Nakagawa. "Sakumon: An assisting system for English cloze test", SITE 2007--Society for Information Technology & Teacher Education International Conference (Demo), pp.2816-2816, Texas, USA; March 26-30, 2007

[発表 2] 田村 悟之 , 清田 陽司, 増田英孝,中川 裕志 : 図書館における自動レファレンスサービスシステムの実現に向けて ～Web 上の二次情報と図書館の一次情報の統合～, 情報処理学会 第 86 回情報学基礎研究会, pp.1-8, 2007 年 3 月

[発表 3] 福原知宏, 宇津呂武仁, 中川裕志 : 複数言語間の語彙出現傾向比較による言語横断型ウェブログ関心解析システムの開発, 言語処理学会第 13 回年次大会ワークショップ「大規模 Web 研究基盤上での自然言語処理・情報検索」,pp.27-30,2007 年 3 月

[発表 4] 清田 陽司, 田村 悟之 , 中川 裕志 , 増田 英孝: Reference Navigator: 異種オントロジーの統合ブラウジングツール ～図書館の分類体系と Wikipedia カテゴリの対応付け～ , 言語処理学会第 13 回年次大会ワークショップ「言語的オントロジーの構築・連携・利用」, pp35.-38,2007 年 3 月

[発表 5] 王玉馨,小島浩之,中川裕志,前田朗: 中文版言選 Web 的評価与分析, 第三届文学与信息技術国際検討会論文集, pp.39-44,2007 年 3 月

[発表 6] 佐藤一誠, 中川裕志: 係り受け関係を考慮したテキストマイニングのための半構造マイニング手法の提案, DBWS2006 (57), 主催;電子情報通信学会データ工学研究会,情報処理学会データベースシステム研究会,日本データベース学会, 2006 年 7 月

[発表 7] 福原知宏、中川裕志、西田豊明 :感情表現と用語のクラスタリングを用いた時系列テキスト集合からの話題検出, 第 20 回人工知能学会大会 2E1-02, 2006 年 5 月

[発表 8] 福原知宏、村山敏泰、中川裕志、西田豊明 :Weblog から社会の関心を探る, 第 20 回人工知能学会大会 3D2-01, 2006 年 5 月

自然言語構文解析の研究とテキスト情報検索への応用 —HPSG 構文解析の高速化と高精度化および 大規模テキストに対する意味構造検索への応用—

二宮 崇

1 概要

本年度は、英語、日本語などの自然言語の構文解析の高速化および高精度化の研究を主に行った。構文解析は、文を入力とし、入力文の構文構造や単語間の依存構造を出力することを目的としたタスクであり、例えば、英語の主語や目的語を同定したり、名詞句や動詞句、関係節といった言語学的に規定される句構造を同定することを目的としている。構文解析により、世の中に大量に存する Web テキストや新聞、論文、図書から種々の知識が構文構造または述語-項関係の形式で獲得できるため、テキストからのより高度な情報抽出やより高精度な情報検索を行うための基礎技術として期待されている。本研究では、主辞駆動句構造文法 (HPSG) と呼ばれる言語学的に厳密に定義された文法を用いた構文解析の高速化および高精度化の研究を行った。具体的には、1) ビームサーチと呼ばれる探索を HPSG 構文解析に導入し、2) 単純ながら非常に有用な構文解析の確率モデル、および、従来の確率モデルと新しい確率モデルを混合させたモデルの提案をした。1) の研究により単純なビタビサーチに対し 1000 倍の高速化を達成し、2) の研究により 1) に比較しさらに 3〜4 倍の高速化、および、F スコアで約 2.0%ポイントの精度の上昇を得た。また、構文解析の応用として開発した情報検索システムについても間単に報告する。具体的には下記の研究内容について報告する。

ビームサーチによる HPSG 構文解析の高速化 [査読付 2] 本研究は HPSG 構文解析器の高速化のために、ビームサーチによる高速化手法と、深い構文解析のための様々な高速化手法の適用を行った [査読付 2]。ビームサーチの再現率向上のためにイテレィティブ構文解析を適用し、クイックチェック、 大句構造の抑制、および CFG チャンクパーザーとのハイブリッド構文解析による高速化を実現した。構文解析の性能評価によく用いられる Penn Treebank (Wall Street Journal) に対し実験を行い、それぞれの手法を精度と解析速度の点から評価した。

極語彙化モデルによる HPSG 構文解析の高速化と高精度化 [査読付 1] HPSG 構文解析のための新しい確率モデルの研究を行った。提案するモデルは、句構造の特徴を一切用いず、単語品詞 n グラムによる辞書項目の選択確率の積で単純に定義される。辞書項目の選択確率しか用いていないが、実験により、従来の句構造ベースの確率モデルと同等の精度を達成できることが確認された。複雑な句構造ベースの確率モデルを用いずとも、高い精度の構文解析が単純な確率モデルにより実現可能になったといえる。また、従来の確率モデルとの混合モデルにより、3〜4 倍程度の高速化を達成し、また、およそ 2% ポイントの F スコアの向上を確認した。

構文解析された医学生物学論文からの意味検索システム [査読付 3, 査読付 4] およそ 1500 万件の医学生物学論文のリファレンスと 700 万件の概要からなる MEDLINE と呼ばれる大規模なテキストベースからの意味検索システムを構築し、その性能を評価した。このシステムの構築のために、まず MEDLINE 全テキストに対し HPSG 構文解析、専門用語タグ付けを行い、統語構造と述語項構造および医学生物学オントロジーの意味タグが付与されたテキストベースを構築した。このテ

キストベースに対し領域代数と呼ばれる検索系を用いることにより、医学生物学で重要と考えられている遺伝子-疾患関係などの深い関係概念を高速かつ高精度に検索できることを実験により示した。

上記以外に、自然言語文要約のための文圧縮の研究も行った [査読付 5]。文圧縮とは、文中に含まれる余分な単語や句を自動的に切り捨てる技術のことであり、より短くより読みやすい文を生成することを目的としている。また、過去に提案したコーパス指向文法開発の手法についてチュートリアル講演を言語処理学会第 13 回年次大会で行った [特記 1]。

2 ビームサーチによる HPSG 構文解析の高速化

2.1 背景

構文解析は文を入力とし、入力文の構文構造や単語間の依存構造を出力することを目的としたタスクであり、例えば、単一化文法は言語学的形式化と計算効率の観点から広く研究をなされてきた。しかし、それらは非常に精緻な言語学的構造を文に対して与えるものの、その精緻な記述のために計算効率は非常に悪いとみなされてきた。高速化は実用アプリケーションのための主たる目的の一つであるため単一化文法の構文解析効率のための研究が非常に多くなされてきた。それらの研究により解析効率が大幅に改善されたものの、構文解析速度はまだ実用的なアプリケーションのためには十分ではなかった。

近年、広範な解析能力を持つ単一化文法のための確率モデルが提案され、確率値を参照することにより、より効率的な構文解析の探索を行う手法の可能性がでてきた。つまり、一般に最も確率値の高い構文木だけが解析結果として要求されるので、最終的な解析結果に貢献する可能性の高い部分解析だけ計算すればよい。この方法はビタビ構文解析やビームサーチ構文解析に代表される確率付き CFG 構文解析で広く研究されてきた。しかしながら、今まで、大規模コーパスに対し単一化文法の確率モデルを用いてビームサーチ構文解析を行った例はなく、本研究は大規模コーパスに対する確率的 HPSG 構文解析にビームサーチを適用する手法の考案、およびその性能について報告をする。また、深い構文解析のための種々の高速化手法の性能についても実験で確認する。

2.2 内容

2.2.1 深い構文解析のための高速化手法の適用

クイックチェック 単一化文法である HPSG は全ての文法規則の適用に単一化が行われるため、単一化の高速化は構文解析全体の高速化に大きく貢献する。クイックチェックは、単一化に失敗する素性構造の組み合わせを除去するフィルタリング手法の一つであり、単一化に失敗する可能性の高いパス (=素性ラベル列) を事前に求めておき、単一化実行時には、それらのパスの値の単一化可能性を最初に調べることにより不要な単一化の実行を避けることができる。

大句構造の抑制 大句構造の抑制は、文頭と文末に隣接しない長さ 20 単語以上の句構造を生成しない、というヒューリスティックスである。つまり、文中の大きな句構造は最終的な構文解析結果に貢献することが稀であるという経験則を応用した手法である。

ハイブリッド構文解析 ハイブリッド構文解析により深い構文解析を高速化する手法が近年報告されている。前処理として、再現率は低いが適合率の高い高速な浅い構文解析を行い、深い構文解析の探索空間を大きく絞り込む。解析速度と解析精度の間にはトレードオフがあるため Penn Treebank のような実世界の大規模テキストに対してハイブリッド構文解析の効用を評価する必要がある。本研究は、非常に高速で適合率の高い CFG チャンクパーザーを HPSG 構文解析の前処理として用いた。CFG チャンクパーザーにより出力された句構造を句境界として用い、その句境界をまたがる句構造の生成を禁止することにより探索空間を大幅に絞り込む。

表 1: Section 23 (≤40 単語) に対する精度と解析速度

	適合率	再現率	F スコア	解析時間
イテレィティブ	87.7%	86.8%	87.2%	355 ミリ秒/文

表 2: Section 24 (≤15 単語) に対する精度と解析速度

	適合率	再現率	F スコア	解析時間
ビタビ	87.8%	87.5%	87.7%	94,374 ミリ秒/文
ビームサーチ	88.6%	82.5%	85.4%	90 ミリ秒/文
イテレィティブ	87.4%	87.0%	87.2%	101 ミリ秒/文

2.2.2 ビームサーチによる高速化

ビタビアルゴリズム 解析途中の句構造は、一般に、⟨ 句構造の左端の位置, 句構造の右端の位置, 句構造を表す素性構造 ⟩ で表現される。HPSG のための CYK 構文解析アルゴリズムにおいては、等価な句構造の集合を一つの句構造に縮退することにより、計算量の爆発を防いでいる。この等価な句構造の縮退の際に、句構造集合がもつ確率値のうち最も高い確率値を縮退される句構造に付与することにより確率付き HPSG 構文解析のためのビタビアルゴリズムが実現する。ビタビアルゴリズムでは最適解が保証されている。

ビームサーチ ビームサーチは最適解が保証されないかわりに大幅な高速化を実現する手法である。同じ左端の位置、右端の位置の句構造集合の中で、1) もっとも確率値の高い句構造の確率値からビーム幅 w を引いた値以上の確率値を持つ句構造、および 2) 確率値が高い n 個の句構造のみを解析の対象とする。近似計算により求めた外側確率を用いて句構造全体の中でもっとも高い確率値からビーム幅 w' を引いた値以上の確率値を持つ句構造のみ解析の対象とする global thresholding も実装し評価した。

イテレィティブ構文解析 ビームサーチアルゴリズムでは過剰に枝刈りが行われ解が生成されないことがあるため再現率が低くなる。本研究は、解析結果が出力されるまでビーム幅を広げながら繰り返し構文解析を行う手法を提案、実装した。CFG 構文解析の場合は、計算結果を記憶しておくコストと再計算コストがほぼ等価であるため、繰り返し計算を行う場合は以前に計算した結果を全て消去するほうがよいが、HPSG 構文解析の場合は、単一化による再計算はコストが高いため、以前計算した結果を再利用する。

2.2.3 評価

実験は Penn Treebank の Section 02-22 を用いて生成された HPSG 文法を用い、Penn Treebank の Section 23,24 を評価対象とした。2.4 GHz の AMD Opteron CPU を搭載するサーバーを使用した。

まず、上述の全ての手法を組み合わせた総合性能を評価した。表 1 は Section23 の 40 単語以下の文長に対する依存構造の適合率、再現率、F スコア (= 適合率と再現率の調和平均) と一文あたりの平均解析時間を表している。

次に、ビタビ、ビームサーチ、イテレィティブ構文解析の性能を比較した。表 2 は、Section24 の 15 単語以下の文長に対する依存構造の適合率、再現率、F スコアと一文あたりの解析時間をビタビ、ビームサーチ、イテレィティブ構文解析の比較を表しており、図 1 は文長に対する解析時間の片対数グラフである。この表から、本研究の提案する HPSG 構文解析でのビームサーチの適用手法は、最適解を出力するビタビアルゴリズムと比較して、F スコアで 0.7% 低くなるだけで、およそ 1000 倍の高速化を実現していることがわかる。また、同様に本研究の提案するイテレィティブ構文解析により再現率がおよそ 5%ほど上昇し、F スコアでおよそ 2%ほど上昇することが確認できた。

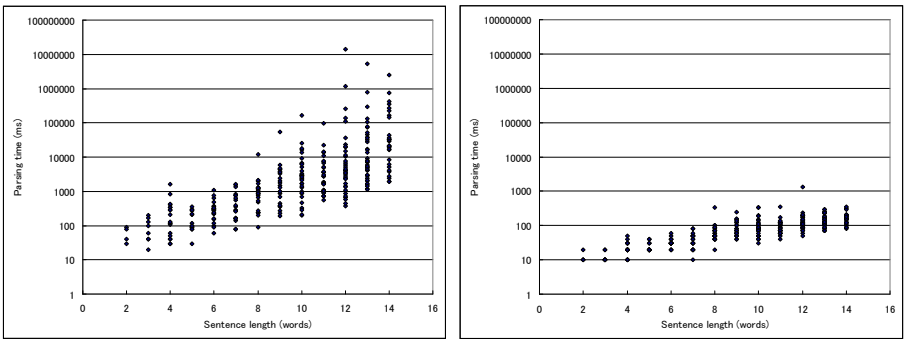


図 1: Section 24 (≤15 単語) に対するビタビの解析時間 (左) とイテレィティブの解析時間 (右)

表 3: 各手法の効果

	適合率	再現率	F スコア	解析時間
全ての高速化手法	85.4%	84.1%	84.7%	419 秒/文
ビームサーチのみ	85.5%	84.8%	85.1%	1100 ミリ秒/文

最後に上述のそれぞれの手法の効果を Section24 の 40 単語以下の文章に対して評価した。表 3 と図 2 はその結果を表している。表は深い構文解析のための高速化手法を用いた場合 (全ての高速化手法) と用いなかった場合 (ビームサーチのみ) の結果を表しており、一文あたりおよそ 2.5 倍の高速化を達成していることがわかる。図の左側は、ビームサーチによる効果を表していて、ビーム幅を確率値の幅で調整したもの (width)、ビーム幅を句構造の数で調整したもの (num)、それらの組み合わせ (num+width)、global thresholding (global)、イテレィティブ構文解析 (iterative)、global thresholding とイテレィティブ構文解析の組み合わせ (iterative+global) を用いたそれぞれの場合の解析時間と F スコアの関係を表している。各手法を組み合わせれば組み合わせるほど良い効果、すなわち F スコアを落とすことなく解析時間が短くなっていることがわかる。図の右側はチャンクパーザー (chp) と組み合わせたハイブリッド構文解析の効果を表しており、全ての手法に対して、ハイブリッド化による性能向上を確認できた。

2.3 具体的成果

本研究では、HPSG 構文解析においてビームサーチを含む様々な手法の効果を実験により比較した。PCFG で用いられれているビームサーチを単純に適用した場合はビタビアルゴリズムに比較して 1000 倍程度の高速化が得られたが再現率が減少した。イテレィティブアルゴリズムは再現率を大きく落とすことなく同程度の高速化を達成できた。クイックチェック、大句構造の抑制、ハイブリッド構文解析を用いることによって、F スコアが 0.4%減少するかわりに、構文解析速度がおよそ 2.6 倍になった。大句構造の抑制の効果は小さく、クイックチェックとハイブリッド構文解析の効果が大きかった。40 単語以下の Penn Treebank Section 23 に対し実験したところ、一文平均 355 ミリ秒の解析時間および F スコアで 87.2%の精度で構文解析を行うことができた。

3 極語彙化モデルによる HPSG 構文解析の高速化と高精度化

3.1 背景

十数年以上前ではおもちゃ文法 (toy grammar) と非現実的な人工的文章に対してのみ構文解析の研究が行われていたが、近年、現実に存在する文章に対する高精度かつ高被覆[1]な構文解析の研究が多くな

[1]構文解析可能な文の割合が高いこと。特定のテキストに対して構文解析結果を出力する文の割合や依存構造の割合のことを被覆率という。

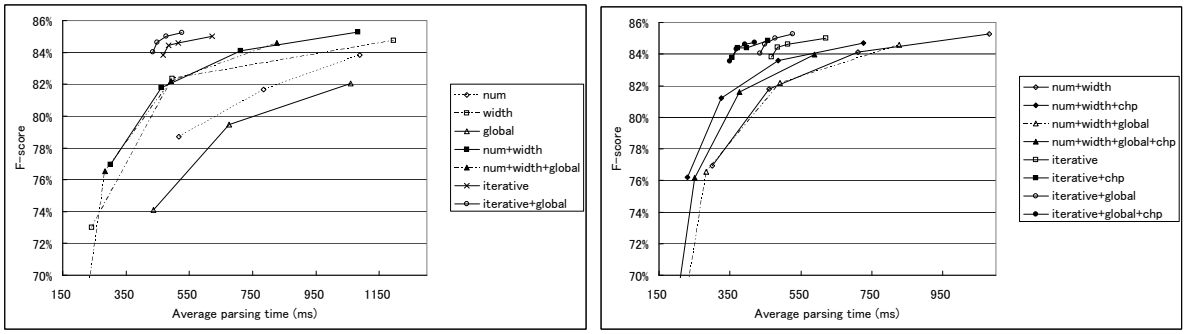


図 2: Section 24 (≤40 単語) に対する解析時間と F スコア

されるようになってきた。このような現実の文章に対して構文解析を行う際の大きな問題は、構文構造の曖昧性とその解消であり、例えば、最先端の構文解析技術では、一文あたり数千から数百億の構文構造候補が出力されるが、正しい構文構造はその中に一つしかない。曖昧性解消の際には、構文構造候補の句構造の形が曖昧性解消を行う大きな手がかりになると考えられるため、多くの構文解析の研究では句構造に対して、確率事象を定義することが多かった。例えば、確率的 CFG (PCFG) では文法規則に対して確率が定義され、主辞駆動句構造文法 (HPSG) や組み合わせ範疇文法 (CCG) では解析結果の句構造に対して確率が定義されてきた。今までに様々な構文解析の研究が行われてきたが、基本的なアイデアは句構造や文法規則の特徴を捉えることであり、一方、句構造の一部である語彙情報を利用することは構文解析の精度を大きくあげることが理解されてはいるものの補助的に用いられるだけであった。

一方、構文解析の曖昧性解消問題に対する非常に興味深い手法として、スーパータグ付けというアイデアが語彙化木接合文法 (LTAG) の文法枠組の中で Bangalore らによって 1999 年に提案された。スーパータグ付けは、品詞タグが付与された文に対し、スーパータグを付与するタスクであり、ここで、スーパータグというのは、LTAG では基本木のことであり、HPSG や CCG では辞書項目に相当する。スーパータグ付けは、軽いプロセスであるため、重い構文解析前に辞書項目候補を枝狩りするためによく使われる。Bangalore らは、「もし正しくスーパータグを付与することができれば、構文解析はほぼ自明である」と主張している。これの意味するところは、スーパータグには下位範疇化フレームなど構文構造の情報も含まれているため、もし、スーパータグを正しく付与することができるのなら、それにより構文構造がほとんど決定する、ということである。2004 年 Nasr らは、正しいスーパータグが付与された文に対しては非常に高い精度で構文解析が行えることを実証し、同 2004 年 Clark らは CCG 構文解析で構文解析速度を非常に大きく向上させることに成功した。スーパータグ付けのアイデアは非常にシンプルであり、構文解析の前処理モデルとして非常に面白く、その効果も実証されている。しかしながらこれら今まで行われてきたスーパータグ付けの手法は、スーパータグ付け器を外部タグ付け器とし用い、辞書項目の候補を枝狩りしメインの構文解析を補助する目的でのみ用いられてきた。

本研究では、HPSG 構文解析の極端なケースとして、スーパータグ付けの確率のみ、つまり、辞書項目を選択する確率のみで HPSG 構文解析を行うモデルを提案する。また、このモデルと従来の HPSG 構文解析のための確率モデルとの混合モデルについても提案する。

3.2 内容

本研究では、辞書項目の選択確率のみで HPSG 構文解析を行う確率モデルを提案する。ここではこのモデルを極語彙化モデルと呼ぶ。また、極語彙化モデルと従来の確率モデルとの混合モデルについても報告する。

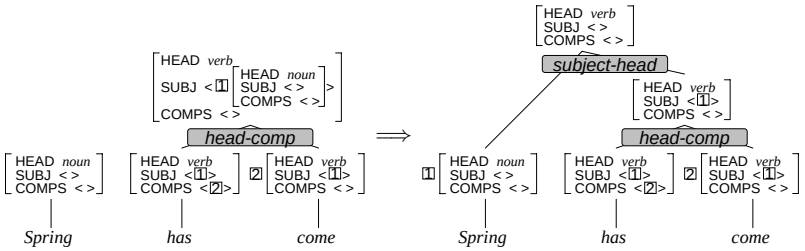


図 3: HPSG 構文解析の例

極語彙化モデルは、単語品詞 n グラムによる単純な辞書項目の選択確率の積で定義される。つまり、この確率モデルは句構造の特徴を一切用いない。この辞書項目の選択確率の計算は、品詞タグ付けと同様の技術で実現可能であり、実験では、単語 3 グラム、品詞 5 グラムを用い、機械学習における選別手法として有名なエントロピー最大化法により計算された。このモデルでは、辞書項目の選択確率しか用いていないが、単純に確率の高い辞書項目だけが選択されるのではなく、文法により適格な辞書項目のうち最も確率の高い辞書項目列が選択される。辞書項目の選択確率しか用いていないが、実験により、従来の複雑な句構造ベースの確率モデルと同程度の高い精度を達成できることが確認された。つまり、複雑な句構造ベースの確率モデルを用いずとも、高い精度の構文解析が単純な確率モデルにより実現可能になったといえる。

また、極語彙化モデルと従来の句構造ベースの確率モデルとの混合モデルについても実験を行った。この混合モデルでは、構文木全体の確率は単純に極語彙化モデルの確率と句構造ベースの確率モデルの確率の積で定義される。実験により、3〜4 倍の高速化と 2%の精度の向上を達成することを確認した。これは、辞書項目の選択確率を句構造ベースの確率モデルに導入することにより、より性能の良い確率モデルが得られることを意味している。

まず、HPSG 構文解析と従来の確率モデルについて説明を行い、その後に、極語彙化モデルと混合モデルについて説明する。

3.2.1 HPSG 構文解析と従来の確率モデル

HPSG は制約に基づく語彙化文法と呼ばれる統語理論であり、少数の非常に一般的な生成規則と単語特有の性質を記述した非常に多くの辞書項目からなる。生成規則と辞書項目は素性構造と呼ばれるグラフ構造により記述され、文の構造は、それら辞書項目と生成規則の組み合わせにより説明される。文の構造の間の制約は素性構造の単一化により実現される。図 3 は “*Spring has come*” に対する HPSG 構文解析の例を示している。まず、“*has*” と “*come*” に対する語彙項目が “*Head-Complement Schema*” と呼ばれる生成規則の子ノードに相当する素性構造と単一化され、親ノードが生成される。より大きな句構造は同様に生成規則を繰り返し適用することにより得られ、最終的に文全体に対応する句構造が構文解析結果として出力される。

素性構造で定義される文法枠組に対して、確率モデルを定義することは単純ではないが、Abney、Johnson ら、Riezler ら、Malouf と van Noord、Kaplan ら、Miyao らによって、HPSG を含む単一化ベースの文法に対する確率モデルが、対数線形モデルまたはエントロピー最大化モデルと呼ばれる確率モデルとして提案された。文 $\mathbf{w} = \langle w_1, \ldots, w_n \rangle$ が与えられたとき、構文解析結果 T に対する確率は次のように与えられる。

$$p_{hpsg}(T|\mathbf{w}) = \frac{1}{Z_{\mathbf{w}}} \exp\left(\sum_u \lambda_u f_u(T)\right), Z_{\mathbf{w}} = \sum_{T'} \exp\left(\sum_u \lambda_u f_u(T')\right)$$

ただし、λ_u はモデルパラメータであり、f_u は構文構造 T の特徴を表す素性関数であり、$Z_{\mathbf{w}}$ は与えられた文に対する全ての可能な構文構造に対する確率の和である。つまり、構文構造 T にあらわれる特

表 4: 素性

$$
\begin{aligned}
f_{binary} &= \left\langle \begin{array}{l} r, d, c, \\ sp_l, sy_l, hw_l, hp_l, hl_l, \\ sp_r, sy_r, hw_r, hp_r, hl_r \end{array} \right\rangle \\
f_{unary} &= \langle r, sy, hw, hp, hl \rangle \\
f_{root} &= \langle sy, hw, hp, hl \rangle \\
f_{lex} &= \langle w_i, p_i, l_i \rangle
\end{aligned}
$$

r	name of the applied schema
d	distance between the head words of the daughters
c	whether a comma exists between daughters and/or inside daughter phrases
sp	number of words dominated by the phrase
sy	symbol of the phrasal category
hw	surface form of the head word
hp	part-of-speech of the head word
hl	lexical entry assigned to the head word
w_i	i-th word
p_i	part-of-speech for w_i
l_i	lexical entry for w_i

combinations of feature templates for f_{binary}

$\langle r, d, c, hw, hp, hl \rangle, \langle r, d, c, hw, hp \rangle, \langle r, d, c, hw, hl \rangle,$
$\langle r, d, c, sy, hw \rangle, \langle r, c, sp, hw, hp, hl \rangle, \langle r, c, sp, hw, hp \rangle,$
$\langle r, c, sp, hw, hl \rangle, \langle r, c, sp, sy, hw \rangle, \langle r, d, c, hp, hl \rangle,$
$\langle r, d, c, hp \rangle, \langle r, d, c, hl \rangle, \langle r, d, c, sy \rangle, \langle r, c, sp, hp, hl \rangle,$
$\langle r, c, sp, hp \rangle, \langle r, c, sp, hl \rangle, \langle r, c, sp, sy \rangle$

combinations of feature templates for f_{unary}

$\langle r, hw, hp, hl \rangle, \langle r, hw, hp \rangle, \langle r, hw, hl \rangle, \langle r, sy, hw \rangle,$
$\langle r, hp, hl \rangle, \langle r, hp \rangle, \langle r, hl \rangle, \langle r, sy \rangle$

combinations of feature templates for f_{root}

$\langle hw, hp, hl \rangle, \langle hw, hp \rangle, \langle hw, hl \rangle,$
$\langle sy, hw \rangle, \langle hp, hl \rangle, \langle hp \rangle, \langle hl \rangle, \langle sy \rangle$

combinations of feature templates for f_{lex}

$\langle w_i, p_i, l_i \rangle, \langle p_i, l_i \rangle$

徴 f_u に対応する正規化された重み $\exp \lambda_u$ の積として定義される。モデルパラメータ λ_u は訓練データに対し、対数尤度を最大化するよう推定される。

また、宮尾らはこれに “予備的確率モデル” と呼ぶ確率 $p_0(T|\mathbf{w})$ を参照分布として導入している。予備的確率モデルは構文解析を必要とせずに全体の確率を近似する確率モデルを想定しており、パラメータ推定を容易にするために導入されている。この確率が低い構文構造の計算はパラメータ推定の段階で省略されるため、パラメータ推定が高速化される。この予備的確率モデルを導入した構文構造全体に対する確率モデルは次のように定義される。

(従来の確率的 HPSG)

$$p_{hpsg'}(T|\mathbf{w}) = p_0(T|\mathbf{w}) \frac{1}{Z_{\mathbf{w}}} \exp\left(\sum_u \lambda_u f_u(T)\right)$$

$$Z_{\mathbf{w}} = \sum_{T'} p_0(T'|\mathbf{w}) \exp\left(\sum_u \lambda_u f_u(T')\right), p_0(T|\mathbf{w}) = \prod_{i=1}^{n} p(l_i|w_i)$$

ただし、l_i は T 中に出現する単語 w_i に対して割り当てられる辞書項目であり、$p(l_i|w_i)$ は w_i に対して l_i を選択する確率である。

評価ではこのモデルと比較をする。このモデルで使われた素性 (=機械学習のための特徴ベクトルの要素) は表 4 のとおりである。素性テンプレート f_{binary} と f_{unary} は二分分岐と単分岐に対応し、f_{root} は構文構造のルートに対応し、f_{lex} は予備的確率モデルの語彙項目に対応する。

3.2.2 極語彙化モデル

実験では、従来の確率的 HPSG と本研究で提案する 2 種類の確率モデルを比較する。最初のモデルはもっとも単純な確率モデルで、辞書項目の選択確率のみから定義されており、単純に辞書項目の選択確率の積で定義される。つまり、従来の確率的 HPSG で用いられていた句構造に対する素性を使っていない。

辞書項目集合 L と、文 $\mathbf{w} = \langle w_1, \ldots, w_n \rangle$、辞書項目の選択確率 $p(l_i \in L|\mathbf{w}, i)$ が与えられたとき、最初のモデルは次のように定義される。

(極語彙化モデル)

$$p_{model1}(T|\mathbf{w}) = \prod_{i=1}^{n} p(l_i|\mathbf{w}, i),$$

ただし、l_i は構文構造 T 中に出現する単語 w_i に割り当てられる辞書項目であり、$p(l_i|\mathbf{w}, i)$ は w_i に対して l_i を選択する確率である。

表 5: 辞書項目の選択確率のための素性

$$f_{exlex} \quad = \quad \left\langle \begin{array}{l} w_{i-1}, w_i, w_{i+1}, \\ p_{i-2}, p_{i-1}, p_i, p_{i+1}, p_{i+2} \end{array} \right\rangle$$

combinations of feature templates
$\langle w_{i-1} \rangle, \langle w_i \rangle, \langle w_{i+1} \rangle,$
$\langle p_{i-2} \rangle, \langle p_{i-1} \rangle, \langle p_i \rangle, \langle p_{i+1} \rangle, \langle p_{i+2} \rangle, \langle p_{i+3} \rangle,$
$\langle w_{i-1}, w_i \rangle, \langle w_i, w_{i+1} \rangle,$
$\langle p_{i-1}, w_i \rangle, \langle p_i, w_i \rangle, \langle p_{i+1}, w_i \rangle,$
$\langle p_i, p_{i+1}, p_{i+2}, p_{i+3} \rangle, \langle p_{i-2}, p_{i-1}, p_i \rangle,$
$\langle p_{i-1}, p_i, p_{i+1} \rangle, \langle p_i, p_{i+1}, p_{i+2} \rangle$
$\langle p_{i-2}, p_{i-1} \rangle, \langle p_{i-1}, p_i \rangle, \langle p_i, p_{i+1} \rangle, \langle p_{i+1}, p_{i+2} \rangle$

表 6: Section 23 に対する実験結果

	Section 23 ($\leq$ 40 単語)			
	適合率	再現率	F スコア	解析時間
従来の確率的 HPSG	87.7%	87.0%	87.3%	468 ミリ秒/文
極語彙化モデル	87.5 %	86.9%	87.2%	111 ミリ秒/文
混合モデル	89.8 %	89.0%	89.4%	132 ミリ秒/文

二つ目のモデルは、極語彙化モデルと従来の確率的 HPSG の混合モデルであり、従来の確率的 HPSG の予備的確率モデルが極語彙化モデルに置き換えられている。

(混合モデル)

$$p_{model3}(T|\mathbf{w}) = \frac{1}{Z_{model3}} p_{model1}(T|\mathbf{w}) \exp\left(\sum_u \lambda_u f_u(T)\right)$$

$$Z_{model3} = \sum_{T'} p_{model1}(T'|\mathbf{w}) \exp\left(\sum_u \lambda_u f_u(T')\right).$$

実験では、句構造に対するモデルパラメータは、従来の確率的 HPSG で用いられたモデルパラメータをそのまま用いた。

辞書項目の選択確率 $p(l_i|\mathbf{w}, i)$ は次のように定義される。

(辞書項目の選択確率)

$$p(l_i|\mathbf{w}, i) = \frac{1}{Z_w} \exp\left(\sum_u \lambda_u f_u(l_i, \mathbf{w}, i)\right), Z_w = \sum_{l'} \exp\left(\sum_u \lambda_u f_u(l', \mathbf{w}, i)\right).$$

ただし、Z_w は w_i に対する可能な全ての辞書項目候補に対する確率の和である。実験で用いられた素性テンプレートは表 5 に記載されており、単語 3 グラムと品詞 5 グラムからなる。

3.2.3 評価

従来の確率的 HPSG、極語彙化モデルおよび混合モデルによる構文解析の解析速度と精度を評価した。文法は、Enju 2.1 と呼んでいる Penn Treebank の Section 02〜21 (39,832 文) を用いて辞書項目を獲得した HPSG 文法を用いた。この文法の辞書は、10,536 語に対する 3,797 辞書項目からからなる。確率モデルも同様に Penn Treebank の Section 02〜21 を用いて学習し、Section 22 を開発用テストセットとして用い、Section 23 を用いて性能評価を行った。精度の評価は、HPSG 構文解析の結果出力される述語項構造の依存構造に対する適合率、再現率およびそれらの調和平均である F スコアで評価した。計算機は、AMD Opteron 2.4-GHz のサーバーを用いた。

Section 23 に対する結果は、表 6 に示されている。本研究で提案する極語彙化モデルと混合モデルは速度だけでなく、高い精度も達成している。混合モデルはおよそ 3〜4 倍程度の高速化と F スコアで 2%ポイントの向上を達成している。実験結果が示す非常に興味深い結果として、極語彙化モデルは非常に単純な確率モデルであり、語彙構造に対する情報しか用いていないにもかかわらず、従来の確率的 HPSG とほぼ同等の精度を達成しており、非常に大きい高速化を達成している。

3.3 具体的成果

本研究では、HPSG 構文解析の高速化、および高精度化のための極語彙化モデルと従来の確率的 HPSG との混合モデルを提案した。極語彙化モデルは辞書項目の選択確率だけからなるという非常にシンプルなモデルでありながら、従来の確率モデルと同等の精度および 3〜4 倍の高速化を達成した。従来の構文解析のモデルでは句構造の特徴を基本として確率モデルが構築されてきたが、この実験により、従来のような複雑な句構造をベースとしたモデルではなくても十分な精度が達成されうるということが示された。また、極語彙化モデルと従来のモデルと混合することにより、3〜4 倍の高速化だけではなく F スコアで 2%ポイントの高精度化も達成できることが実験により示された。

従来の構文解析では、動的計画法による構文構造の生成とその確率計算が行われてきたが、動的計画法に依存するために、文法や確率モデルは、一般に局所性と呼ばれる親子間のノードや隣接するノードといった局所的構造に依存してしまうという問題があった。つまり、動的計画法のために実装が複雑になってきたこと、および局所性のために、文法の記述や素性の設計、および意味構造の設計に大きな制限がかかってきたことを意味している。これは、単一化ベースの確率モデルが、全ての構文構造候補の中から、解を一つ選択するというモデルになっているため、数千から数十億におよぶ構文構造を動的計画法で計算してきたということに起因している。一方、極語彙化モデルにより、句構造の情報を用いずとも句構造ベースの確率モデルと同等の精度を達成できることが示せた。今後の研究の発展として、従来のように動的計画法を用いるのではなく、例えば、極語彙化モデルと決定的構文解析手法を組み合わせることにより、動的計画法をまったく用いない構文解析の実現が期待される。

4 構文解析された医学生物学論文からの意味検索システム

4.1 背景

Web や電子化された論文に代表されるように、急速なテキスト量の増大により莫大な量のテキストからの情報検索が要求されている。特に論文から科学的事実や知識を検索する場合では Web 上での検索のように関連するテキストをユーザーに提示するだけでは不十分で、より高精度かつ関係概念のような深い関係を検索する必要がある。医学生物学の分野では、10%の年率で増大する MEDLNE と呼ばれる 1500 万件のリファレンスが電子的に存在し、これらから関連する事象、例えば、蛋白質-蛋白質関係や、遺伝子-疾患関係を検索することが要求されている。しかし、このような深い関係を抽出するためには、述語で記述される実体間の関係を同定する必要があり、キーワードの出現頻度を基本とする従来の検索システムでは、これらの関係概念を正確に検索することは困難である。

本研究では、巨大なテキストから深い関係概念をリアルタイムで検索するための情報検索システムを提案する。このシステムでは、事前に全テキストを HPSG 構文解析し、述語項構造と呼ばれる意味構造をテキストに付与する。ユーザーのリクエストはオンフライでこの意味構造に変換され、意味構造上での検索を行う。意味構造に対して検索が行われるため上記のような深い関係を検索することが可能であり、事前に全テキスト構文解析するため現実的な使用に供せられる速度で検索することが可能になった。

4.2 内容

本研究で作成したシステムを MEDIE と呼ぶ。図 4 は MEDIE の中での処理の全体像を表している。まず、テキストが与えられ、そのテキストに対し節 2, 3 で説明した HPSG 構文解析器により構文解析が行われ、また、同時に専門用語認識器により、専門用語がテキストに付与される。HPSG 構文解析器と専門用語認識器による意味構造の付与は、検索時ではなく、オフラインで行われる。実際にクエリーが入力されたときには、この意味構造が付与されたテキストに対し、領域代数検索エンジンと呼ぶシステムによって検索が行われる。

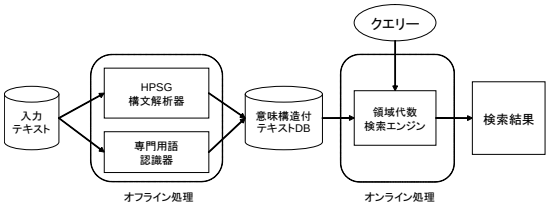


図 4: MEDIE の情報検索プロセスの全体像

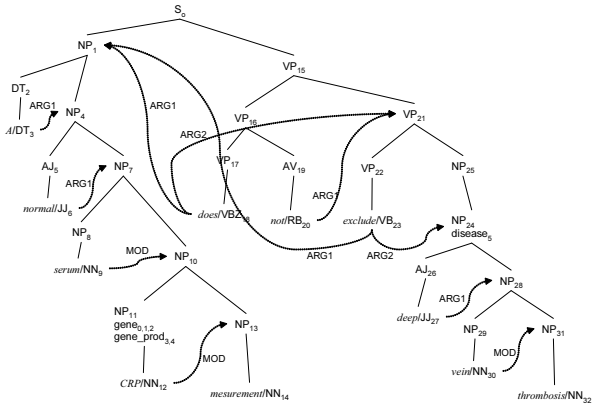


図 5: 構文解析結果の例

このシステムは MEDLINE と呼ばれる、およそ 4500 の生物医学論文ジャーナルに掲載された論文リファレンスを検索対象としており、そのおよそ半数は概要をもっている。本研究では 2004 年末の MEDLINE を対象とする。14,785,094 件のリファレンスからなり、その半数のおよそ 700 万件が概要をもつ。

4.2.1 オフライン処理

オフライン処理では節 2,3 で説明した HPSG 構文解析器により述語項構造とよばれる意味構造と統語構造が解析される。構文解析器は十分に速くなったとはいえ、大量の文章を解析するにはまだ不足しており、本研究では東京大学の本郷キャンパス、柏キャンパスに分散する 340 CPU と、それらを制御するために田浦らによって開発された分散処理システム GXP を用いて、およそ 9 日間かけて MEDLINE を解析した。専門用語認識は、医学生物学用語に対し、遺伝子、疾患といった医学生物オントロジーの意味ラベルを付与するタスクであり、本研究では、鶴岡らによって開発された辞書ベースの専門用語認識器を用いた。辞書は、koike と takagi によって開発された GENA と UMLS と呼ばれる巨大なメタシソーラスを用いた。

図 5 は “A normal serum CRP mesurement does not exclude deep vein thrombosis” という一文に対する意味構造付テキストの例である。専門用語認識器により、CRP が遺伝子であること、また、deep vein thrombosis が疾患であることがわかる。構文解析器により、文の統語構造関係と破線で表示される述語項構造関係がわかる。述語 exclude に注目すると、その主語 (ARG1) が “A normal serum CRP mesurement” であることがわかり、その目的語 (ARG2) が “deep vein thrombosis” であることがわかる。また、これらの関係は、not で修飾されているため、否定関係にあることがわかる。

4.2.2 オンライン処理

検索は、増田らによって開発された領域代数検索エンジンを用いた。実際に検索したい単語や意味ラベルは、例えば、主語となる名詞句のどこかに存在して、構造的な位置を一意に指定できるわけではなく、近接関係にある。一方、述語と主語、目的語の関係のように構造的に強い制約を課したいという目的もあるため、近接関係と構造関係を同時に指定できる領域代数がこの検索の目的に適している。領域代数は、AND、OR といったブーリアン検索のクエリーと、CONTAINING, CONTAINED といった、領域間の構造検索のクエリーを同時に指定できる。例えば、

```
[sentence] >> ([word arg1="$subject" arg2="$object" base="cause"] &
               ([phrase id="$subject"] > gene) & ([phrase id="$object"] > disease))
```

というクエリーを与えたとき、これは、動詞の基本形が cause であり、主語となる名詞句に gene の意味ラベルが付与された名詞が存在し、かつ、目的語となる名詞句に disease の意味ラベルが付与名詞

表 7: クエリー

クエリー No.	クエリー
1	*something* inhibit ERK2
2	*something* trigger diabetes
3	adiponection increase *something*
4	TNF activate IL6
5	dystrophin cause *disease*
6	macrophage induce something
7	*something* suppress MAP phosphorylation
8	*something* enhance p53 (negative)

表 8: 実験結果

クエリー No.	キーワード検索				提案システム				
	検索結果数	検索時間 (最初の解)	検索時間 (全て)	適合率	検索結果数	検索時間 (最初の解)	検索時間 (全て)	適合率	相対再現率
1-1	252	0.00	1.5	74%	143	0.01	2.5	96%	69%
1-2	348	0.00	1.9	61%	174	0.01	3.1	89%	69%
1-3	884	0.00	3.2	50%	292	0.01	5.3	91%	42%
2-1	125	0.00	1.8	45%	27	0.02	2.9	85%	38%
2-2	113	0.00	2.9	40%	26	0.06	4.0	85%	48%
2-3	6529	0.00	12.1	42%	662	0.01	1527.4	76%	19%
3-1	287	0.00	1.5	20%	30	0.05	2.4	80%	30%
3-2	309	0.01	2.1	21%	32	0.10	3.5	81%	29%
3-3	338	0.01	2.2	24%	39	0.05	3.6	82%	33%
4-1	4	0.26	1.5	0%	0	2.44	2.4	-	-
4-2	195	0.01	2.5	9%	6	0.09	4.1	83%	22%
4-3	2063	0.00	7.5	5%	94	0.02	10.5	95%	40%
5-2	287	0.08	6.3	73%	116	0.05	14.7	97%	51%
5-3	602	0.01	15.9	50%	122	0.05	14.2	96%	46%
6-1	10698	0.00	42.8	14%	1559	0.01	3014.5	65%	71%
6-3	42106	0.00	3379.5	11%	2776	0.01	5100.1	61%	45%
7	87	0.04	2.7	39%	15	0.05	4.2	67%	29%
8	1812	0.01	7.6	19%	84	0.20	29.2	87%	37%

が存在する文が検索される。HPSG 構文解析を用いているため、例えば、これが受身であったり、関係節であったりしても同様に検索ができる。つまり、受身では主語、目的語が by の前置詞により逆転してしまうが、by の前置詞句にある名詞が本来の主語 (上記の arg1)、主語が本来の目的語 (上記の arg2) として解析されるため、より正しい述語項構造関係のみが検索される。

4.2.3 評価

評価は、表 7 に記載されているクエリーを用いて行った。クエリー 1, 3, 4 は遺伝子-蛋白質関係の検索クエリーを表しており、クエリー 2,5 は疾患に関する検索クエリーを表している。クエリー 6 は遺伝子や疾患に関しないクエリーを表しており、クエリー 7 は複雑なクエリーの例になっている。実験結果は表 8 に示されている。実験は、意味構造を使わない場合の検索としてキーワード検索による評価と、提案システムによる評価を比較した。表 8 の中のクエリー No. の添字は X-1 が専門用語タグを使わない場合で、X-2 が専門用語タグを用いた場合で、X-3 が専門用語タグと動詞に対する小規模なオントロジーを用いた場合の結果を表している。キーワード検索の精度は表中の適合率の列に記載されており、出力された検索結果のうち 100 件を人手で判定してその正解率を測った。提案システムの精度は適合率と相対再現率により評価した。相対再現率はキーワード検索の結果の中に含まれる正解を真の正解とした場合の再現率である。真の再現率を測るのは非常に困難であるため、この尺度を用いて評価した。

実験結果より、提案システムによる検索では適合率が大きく向上していることがわかる。ほとんどのクエリーで 80%の適合率を超えており、クエリー 4 と 5 ではほぼ 100%に近い。相対再現率はおおよそ 30%から 50%であった。検索時間はキーワード検索に比べ遅いものの、クエリー 6 をのぞきほぼキーワード検索に近い速度で検索できている。

4.3 具体的成果

本研究では、MEDLINE に対する大規模な意味構造検索システムの構築を行った。オフラインで HPSG 構文解析を行い、統語構造、述語項構造および専門用語タグを付与することにより、意味構造の検索

が可能となった。再現率は低いが、意味構造を付与することにより、非常に高い精度を実現できることが実験により示された。

5 成果要覧

査読付論文リスト

[査読付 1] Takashi Ninomiya, Takuya Matsuzaki, Yoshimasa Tsuruoka, Yusuke Miyao and Jun'ichi Tsujii: Extremely Lexicalized Models for Accurate and Fast HPSG Parsing, In Proc. of EMNLP 2006, pp. 155–163, 2006.

[査読付 2] Takashi Ninomiya, Yoshimasa Tsuruoka, Yusuke Miyao, Kenjiro Taura and Jun'ichi Tsujii: Fast and scalable HPSG parsing, Journal of Traitement Automatique des Langues (TAL), 46(2), pp. 91–114, Association pour le Traitement Automatique des LAngues, 2006.

[査読付 3] Yusuke Miyao, Tomoko Ohta, Katsuya Masuda, Yoshimasa Tsuruoka, Kazuhiro Yoshida, Takashi Ninomiya and Jun'ichi Tsujii: Semantic Retrieval for the Accurate Identification of Relational Concepts in Massive Textbases, In Proc. of COLING/ACL 2006, pp. 1017–1024, 2006.

[査読付 4] Tomoko Ohta, Yusuke Miyao, Takashi Ninomiya, Yoshimasa Tsuruoka, Akane Yakushiji, Katsuya Masuda, Jumpei Takeuchi, Kazuhiro Yoshida, Tadayoshi Hara, Jin-Dong Kim, Yuka Tateisi and Jun'ichi Tsujii: An Intelligent Search Engine and GUI-based Efficient MEDLINE Search Tool Based on Deep Syntactic Parsing, In Proc. of the COLING/ACL 2006 Interactive Presentation Sessions, pp. 17—20, 2006.

[査読付 5] Yuya Unno, Takashi Ninomiya, Yusuke Miyao and Jun'ichi Tsujii: Trimming CFG Parse Trees for Sentence Compression Using Machine Learning Approaches, In Proc. of the COLING/ACL 2006 Main Conference Poster Sessions, pp. 850–857, 2006.

特記事項

[特記 1] 二宮 崇, 宮尾 祐介: 自然言語処理における文法開発の軌跡と展望, 言語処理学会第 13 回年次大会 (NLP2007) チュートリアル講演, 2007 年 3 月 19 日.

テキスト検索高度化に向けて
—自動同義語抽出および自動異義語判定に関する研究—

吉田 稔

1 概要

WWW ページや電子ジャーナル、組織内に蓄積されたレポート等、今日、様々な場面で多種多様な文書が電子的に蓄積されており、それらを活用して有用な知識を発見するための技術は、益々重要性を増しつつある。大量の文書を扱うための基礎技術としてテキスト検索技術があるが、本年度は、より高度なテキスト検索を実現するため、キーワード検索を補助するための手法、およびキーワード検索と異なる新たな検索手法、について研究を行った。具体的なテーマとしては、以下のものが挙げられる。本稿ではこのうち、最初の 2 テーマについて解説を行う。

- テキストマイニングの為の高速な同義表現抽出手法の開発 [発表 5, 発表 1, 発表 3]

  文書中に出現する異なる文字列が同じ意味を持つ場合、その対応付けを行うことは、網羅性の高い検索の実現には不可欠の技術である。本研究では、このような同義表現を自動的にかつ高速に抽出するための手法の開発を行う。

- Web 名寄せ (同名曖昧性解消) システムの開発 [査読付 1]

  文書中において、同一の文字列が異なる実体を指す現象 (例：同姓同名の人物) は、特に人名・地名等の固有名詞を扱う場合に問題となる。本研究では、検索エンジンに固有名詞をクエリとして与えた結果の Web ページ集合を、同一実体毎に分類して表示するためのシステム構築を目指し、その手法を開発する。

- 表形式の検索システムの開発 [発表 4]

  表形式は、属性・属性値の形で表現できる事象を表現する手段として、特に Web 文書で頻繁に用いられる形式であり、これを検索対象とすることにより、例えば「行先がヨーロッパのツアーを探したい」等、単純なキーワード検索ではない、より構造化された検索が、タグ付け等の人手による構造化を経ずに可能となる。本研究では、検索結果の Web ページ集合から、「意味のある」表形式を抽出し、ユーザーに提示するシステムの開発を目指す。

- 数値による検索システムの開発 [発表 2]

  テキスト中の数値情報は、意味が特定しやすく、また、複数の数値間で順序付けが行えるという点で、特に複数テキスト間を関係付けるための手掛かりとして重要な情報源となる。本研究では、そのような「数値によるテキスト間対応付け」の一例として、新聞記事データを「売上高」等の特定の数値データでランク付けして表示するためのシステムの開発を行う。

- Web 文書構造抽出 [発表 6]

  Web 文書のレイアウトを解析することにより、文章の形では表現されない、「文字の位置情報により表現される単語間の意味関係」を解析することを目指す。これにより、Web 文書の検索結果

(Snippet) を、単なる単語の羅列でない、より構造化された表現とし、ユーザーに分かり易く提示することが可能となる。

2 テキストマイニングの為の高速な同義表現抽出手法の開発

2.1 背景

本研究の目的は、業務レポート等、小規模かつある程度話題の制限された文書集合を対象としたテキストマイニングである。

自然言語処理の分野において、同義語抽出は、主に単語意味関係の推定や言い換え抽出のタスクの一環として、様々な研究が行われてきている。しかしながら、これらの研究では、大規模コーパスに対してある程度の時間を掛けて処理を行うことを前提としており、事前に同義表現の候補 (単語単位、名詞句単位、等) を規定しておく必要がある。これに対し、本研究の提案は、文字列が入力された後に動的に文書集合内から同義語を検索するという手法であり、これにより、例えば単語の一部分や、主語・動詞の連接等、様々な種類の文字列に対する、柔軟な同義表現抽出が可能となる。

本研究では、リアルタイムに同義表現を検索するシステムを提案する。同義表現とは、表層上は異なるが同一の意味を持つ単語や句等のことであり、たとえば「HDD」と「ハードディスク」、「住所」と「所在地」等が例として挙げられる。これらの同義表現は、あらゆる種類の文書に現れる。何故なら、同一の概念を表現する方法は複数存在することが普通であり、異なる著者による文書は、異なる表現を使うことが常であるからである。従って、同義表現を抽出するシステムは、検索やテキストマイニング等、文書を対象とした情報処理において、必要不可欠のものであると言うことができる。

本研究では、ユーザがクエリを入力し、それに対してリアルタイムに同義表現を返してくれるシステムの開発を目指す。システムは、同じ意味の言葉は類似した周辺文脈のもとに現れるという仮定に基づき、「クエリ q の周辺文脈を検索」→「検索された周辺文脈をさらに検索」→「周辺文脈に囲まれた表現を同義表現候補として取得」という順序で動作する。われわれは、このような動的な同義表現検索を可能にするデータ構造として、Suffix Array を採用した。Suffix Array は、文書に対し任意の部分文字列で検索を可能にするインデックス構造であり、これにより、ユーザーは、従来の同義表現抽出のような、特定の文法的カテゴリに限定されない、あらゆる表現をクエリとして用いることができるようになる。

同義表現を高速に検索するためには、検索に用いる「周辺文脈」の数をなるべく減らすことが望ましい。我々は、周辺文脈として、「tail string」を用いることを提案する。Tail string は、あるクエリ q の「右側に連接する可変長文字 n-gram」と定義される。

2.1.1 関連研究

同義表現抽出に関しては、従来より様々な観点から研究が行われている。例えば、「スパゲティー」と「スパゲッティ」のような「表記のゆれ」を対象とした研究、「東大」と「東京大学」のような「文字の省略」を対象とした研究、あるいは、フレーズ内の文法構造を、意味を変えずに変換する研究、さらに、類似単語のクラスタリングや、文章等、句より大きな単位も対象にした「言い換え」の研究等もある。

このうち、表記のゆれや省略に対しては、たとえば表層文字列の類似度を計測する等、効率的な手法が存在する。文法構造変換の研究に関しても、使用される単語が共通しているという点では、表層上の類似度が用いられていると考えられる。これに対し、我々が目指すものは、表層上全く異なる文字列に対しても同義表現を求めることができるシステムである。他にも、事前に与えられた言い換え集合からの学習手法、パラレルコーパスや辞書を用いた研究等があるが、これはそのような特殊なリソースが存在する一部の文書にのみしか適用できない。

「類似した意味の表現は、類似した周辺文脈のもとに現れる」という仮定を同義表現抽出に用いる手法も盛んに研究されている。動詞と目的語の関係等、文法的情報を用いる手法に関しては、構文解

析の時間的計算量あるいは構文情報の索引付けの空間的計算量のコストを考慮し、動的同義語抽出を目指す本研究では採用しない。左右に連接する文字列を文脈として使う手法に関しては、長い文字列を用いると再現率が下がり、短い文字列を用いると適合率が下がるという問題が報告されているが、本研究ではこれを、文字列長を可変にすることで解決することを目指す。また、連接しない周辺単語を特徴ベクトルとして用いる手法もあるが、連接しない文字列同士の関係を索引付けすることが困難であるため、やはり本システムでは採用していない。

2.2 内容

2.2.1 手法の概要

一般に、同義語抽出においては、「似た意味の単語は、似た周辺文脈のもとに現れる」という仮定を用い、周辺文脈の類似度に基づいて単語の類似度が測定される。このとき、文脈情報としては、主に「連接文字列」「周辺文字列」「動詞係り受け」の 3 種類の手掛かりが使われる。

連接文字列は、主に文法的カテゴリを特定することに役立つ。例えば、直後に助詞が連接する場合、例えば「が」や「を」といった文字が直後に来る場合、文字列は名詞である可能性が高い。また、一般に、周辺文字列は、意味カテゴリを特定することに貢献する。例えば、「食べた」「レストラン」という単語が周辺に出現するような文字列は、食べ物である可能性が高い。動詞係り受けの情報は、「助詞」と「動詞」の２種類の文脈情報を同時に持っており、文法カテゴリと意味カテゴリをともに規定する効果を持つが、文法的なまとまりを持つ単語やフレーズでなければ扱えないという制約がある。

本研究では、このうち「連接文字列」を、意味類似度の測定に用いる。さらに、連接文字列の長さを可変とすることにより、「周辺文字列」の情報も一部「連接文字列」の中に含むことを目指す。例えば、「に行った。赤門では」という文字列が直後に連接するような文字列は、「東京大学」あるいはその同義語である可能性が高い。これは、「に（行った）」の部分文字列が文法的カテゴリを、「赤門」という部分文字列が意味的カテゴリを規定する効果を持つためである。本研究では、この「連接文字列」の情報をもとに、動的に類似表現の抽出を行う。

しかしながら、一般に、長い連接文字列は、コーパス内の出現数が少ないという性質があるため、それらを同義語抽出の手掛かりとして用いた場合、発見できる同義語の数も少なくなるという欠点がある。この欠点を克服するため、複数の連接文字列を用い、各連接文字列の出現頻度分布を計測する。連接文字列を複数用いることにより、同義語を見逃すリスクを低減する。さらに、単純な出現頻度ではなく、より詳細な情報である連接文字列の出現頻度分布を類似度測定の根拠として用いることで、意味類似度測定の精度を向上させる。

2.2.2 アルゴリズム

「任意の文字列に対する類似文字列の高速検索」という目的を達成するため、インデックス構造として Suffix Array を採用した。Suffix Array は、文書集合から「任意の文字列を」「高速に」検索するためのデータ構造である。

アルゴリズムの概要を以下に示す。システムは、対象となる文書集合の Suffix Array を事前に構築する。ユーザーはこのシステムに対して文字列をクエリとして入力し、システムはそれに対して同義語の候補を、クエリへの類似度に基づいてランク付けして表示する。

連接文字列取得

- クエリ周辺に出現する文字列を、文書集合から取り出す。
- 取り出された文字列をもとに、連接文字列の出現頻度を計測する。

連接文字列選択

- 頻度からスコアを計算し、スコアの高いN個の連接文字列を選ぶ。

TF \ TN	10	50	100	300	500
10	0.230	0.271	0.272	0.272	0.272
50	0.387	0.421	0.416	0.417	0.417
100	0.390	0.433	0.431	0.423	0.423
300	0.434	0.471	0.473	0.474	**0.474**
500	0.430	0.457	0.460	0.459	0.459

表 1: 平均精度による評価

同義語候補取得

- 選ばれた連接文字列を検索し、連接文字列の周辺文字列を取り出す。
- 取り出された周辺文字列から、同義表現候補の頻度を計測する。

同義語候補ランク付け

- 計測された頻度分布情報に基づき、KL-Divergence によるスコアを計算する。
- スコアの高い順に同義表現候補を出力する。

2.3 具体的成果

2.3.1 実験

以上のシステムを実装し、(株) 日本航空インターナショナルの所蔵する機長レポートを対象に、同義語抽出実験を行った。また、同様に、(株) 日本航空インターナショナルにおいて制作された同義語辞書を使用し、性能を測定した。クエリ及び正解同義語は、頻度 100 のものに限定した。クエリ数は 226、各クエリに対する正解同義語数は、平均で 1.18 個となった。

性能の測定には、「平均精度」という尺度を用いた。これは、辞書内の情報をどの程度再現できたか (再現率)、および、ランク上位に提示された単語がどの程度辞書内の情報と一致したか (適合率)、を、統一的に測ることのできる尺度であり、

$$\frac{1}{|S|}\sum_{1\leq k\leq n} r_k \cdot precision(k)$$

として定義される。ここで、$precision(k)$ は、上位 k 個の解における正解率、S は、正解同義語の集合、r_k は、ランク k の出力が正解なら 1、不正解なら 0 となる値である。

実験の際、システムの設定パラメータとして、TN および TF を設定した。TN は、上位何個の連接文字列を使用するかを設定するパラメータであり、TF は、連接文字列の頻度上限を示すパラメータである。

表 1 に、各設定における平均精度を示す。

パラメータの設定は、(TF,TN)=(300,500) において、最高精度を達成した。また、クエリ毎の平均実行時間は、(TF,TN)=(10,10) で 0.038 秒, (TF,TN)=(500,500) で 0.332 秒であった。

比較対象として、単語頻度ベクトル及びコサイン類似度に基づく、単純なベクトル空間モデルによるアルゴリズムをベースラインとして実装し、同様に精度を測定した結果、精度は 0.404 であった。本手法とベースラインとの精度差は、主に、ベースラインで関連語 (例えば、クエリ「医師」に対する同義語候補「治療」等) のスコアが高くなることに起因するものであった。

以上、実験の結果、単純なベクトル空間モデルに基づくベースライン手法と同等以上の性能を達成した。また、7MByte の文書集合に対して、平均して 0.5 秒以内の時間で同義語を出力することができた。

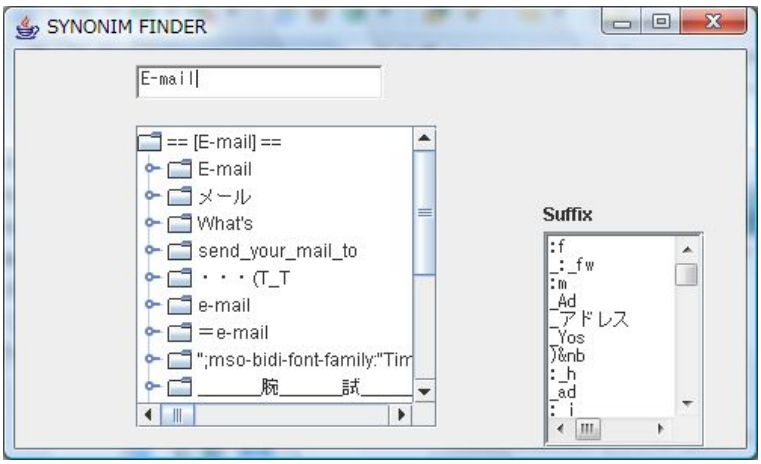


図 1: 同義語検索システムのスクリーンショット

2.3.2 試作システム

本研究のアルゴリズムは Java で実装され、ユーザーは Swing のインターフェースを通じて同義表現の検索を行える形式となっている。図 1 にスクリーンショットを掲載する。左側の 2 つのウインドウが、クエリ入力部及び同義表現候補出力部となっている。同義表現は、部分文字列の関係に応じて Tree 状で表現されている。(具体的には、s が t の部分文字列で、かつ s のランクが t よりも高い場合、だった場合、t は s の子ノードとして格納される。) また、右側に「Suffix」と表示された部分は、使用された連接文字列のリストを表示している。

3 Web 名寄せシステムの開発

3.1 背景

組織内のデータベースに複数個所に出現する同一人物を統合することを**名寄せ**という。この作業において、組織内に同姓同名の人物が複数存在した場合、人手による分離は可能であるが、その同姓同名の人物を自動的に分離するのは難しい問題であり、効果的な解決策はまだ見つかっていない。

本研究の目標は、Web においてこの名寄せの作業を自動的に行うシステムを構築することである。Web 上では Google や Yahoo!などの検索エンジンを用いることで、検索したい文字列 (これを**クエリ**と呼ぶ) を含む文書を集めることができる。検索エンジンの研究は大いに進んでおり、検索エンジンが返す結果は非常に高い精度を持っている。ところが、検索エンジンの返す結果を用いて実際に目的の内容が書かれているページを探すためには、人手による判定が必要であり、先に挙げたような同姓同名である人名の区別や、同じ名前を持つ地名の区別といったことを自動的に行うことができない。例えば、検索エンジン Google で “富士見町” を検索すると約 51 万ものページが見つかるが、各ページに出現する “富士見町” は必ずしも同一の場所を指してはいない。実際に各ページを見てはじめて判定できる。

名寄せを行うことを、別の言葉で言い換えると、次のようになる。

名前がなにか実世界の実体を指し示すことを、**参照**という。また、本論文では、同一の文字列を持つ固有名詞のことを同名、あるいは同じ名前と呼ぶことにする。名寄せとは、すなわち、同名の参照の曖昧性の解消を行っていることになる。これは、**共参照問題**(co-reference task) と呼ばれる問題の一種であり、自然言語処理の分野で扱われている問題である。

共参照問題は、以下の 2 つの子問題を持つ。

- 異なる名前が同じ実体への参照を持つ場合

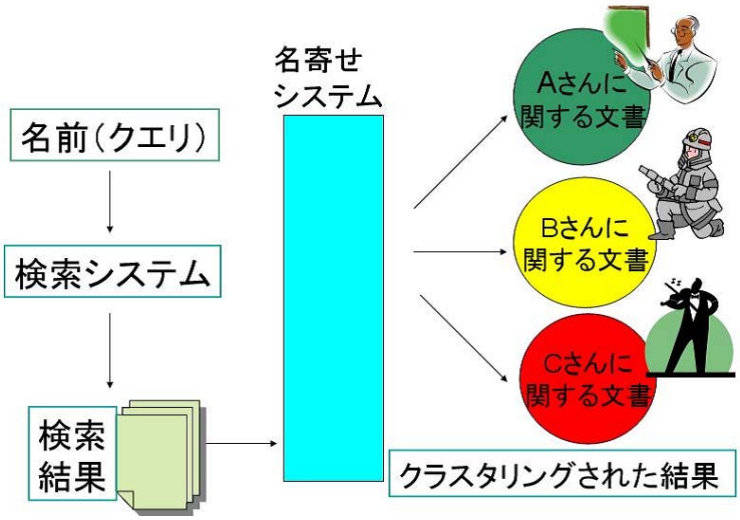


図 2: 名寄せシステムのイメージ図

- 同じ名前が異なる実体への参照を持つ場合

前者の具体例としては、言い換えや同義語、偽名や通称などの解決といった問題が挙げられる。これらの問題については、古くから多くの研究がされている。一方、後者は最近になって活発に研究が行われるようになったもので、まだそれほど研究が進んでいない。本研究では、この問題の解決を目的としている。

この研究が最近になって活発となった背景に計算機の発展がある。名寄せの問題はそもそも、大規模な文章に対して複雑な処理を行うことが前提になる。大量のコーパスを機械的に収集し、それらを高速に解析するためには、高性能な計算機があり、対象となるテキストが揃っているという 2 つの点が必要不可欠である。

名寄せの問題が解決できることにより、様々な応用が考えられる。

まず、機械的に同姓同名を同定するシステムは、まだ実用に至っていない。たとえば、大手の通販サイトである Amazon では、本の著者の検索を行うことができるが、同姓同名の著者がいる場合は分離せずに表示させている。これを同一著者ごとに分離して表示することができるようになる。

同様の問題として、名前がキーとなるようなシステムでの問題がある。たとえば、人間関係ネットワークを発見するシステムのようなソーシャルネットワークに関する研究において、同姓同名が出現し、それらを同定できない場合、システムの性能を大きく下げてしまうことが予想される。

他の応用例としては、個人プロファイリングへの利用が挙げられる。これは、人名を検索対象として、コーパス中から自動的に情報を抽出、選択して、目的の人物の業績リストや略歴などを作成するものである。この作業を自動的に行うツールは現在存在せず、そのような要求に対しては人手で対処している状況である。

以上に示したように、名寄せを自動的に行うシステムは大変有用なものであるが、未だ、実用に至るシステムは存在していない。本研究では、以上のような応用例を踏まえた上で、様々な言語や分野に適用できる一般的な手法を用いたシステムの構築を目的とする。これは、検索エンジンとともにこのシステムを用いることで、クエリを入力すると、通常の検索結果に加えて結果を実体ごとに分類して表示するようなシステムを想定している。人名を対象にした場合の名寄せシステムのイメージ図を図 2 に示す。

先に述べたとおり、実用に至るシステムは存在しないが、このように参照する実体ごとに文書を分類することを目的とした研究がいくつか発表されている。本研究において開発したシステムは、実験

に用いたデータが異なるため直接の比較はできないが、これらの先行研究と同程度の性能が得られた。ただし、特筆すべき点として、従来研究は人工的なデータを対象としたものが多かったが、本研究では検索エンジンの出力結果という実在の Web ページ集合を対象としている。

3.2 内容

文書クラスタリングにおいては、各文書を単語の頻度により表現し ("bag of words" モデルと呼ばれる)、その頻度ベクトルの類似度を計算することで文書間の類似度を測定するという手法（ベクトル空間法と呼ばれる）が主に用いられている。本研究では、この bag of words によるモデル化の際に失われる情報として、単語の位置に関する情報、および、単語の意味的重要度に関する情報に着目し、これらを利用した以下の 2 つのアルゴリズムを提案する。

3.2.1 近隣文字列マッチング

先述のとおり、一般的なベクトル空間法を用いる際には、単語の位置情報が失われてしまう。そこで、その欠点を補うべく、単語の位置情報を強く利用する方法を提案する。

近隣文字列マッチングのアルゴリズム

提案するアルゴリズムは、次のようになる。

procedure 近隣文字列マッチング

step 1. クエリ q を含むすべての文書 d_j $(1 \leq j \leq k)$ について、以下を行う。

step 1.1. 文書 d_j $(1 \leq j \leq k)$ におけるクエリ q の出現位置 p_i $(1 \leq i \leq n_j)$ をすべて検索する。ただし、n_j は文書 d_j におけるクエリ q の出現回数であり、また単語の位置は文書の先頭からの単語数で計算する。

step 1.2. $1 \leq i \leq n_j$ であるすべての i について、$p_i - \alpha$ から $p_i + \alpha$ にある単語 w を集める[1]。なお、α はあらかじめ設定したマッチングに用いる単語数である。集めた単語の集合を S_j とする。

step 1.3. S_j に属す単語 a に以下の重み w_{aj} をつける。

$$w_{aj} = 1/n_j$$

step 2. すべての文書のペア d_x, d_y $(1 \leq x < y \leq k)$ について、以下の方法により、d_x, d_y においてクエリ q が示す実体が等しいかどうかの判定を行う。

step 2.1. S_x、S_y に共通して含まれる単語と、その重みを用いて、近隣文字列の類似度 $\mathrm{sim}_{\mathrm{NSM}}(d_x, d_y)$ を次のように計算する[2]。

$$\mathrm{sim}_{\mathrm{NSM}}(d_x, d_y) = \sum_{a \in S_x \cap S_y} w_{ax} w_{ay}$$

step 2.2. $\mathrm{sim}_{\mathrm{NSM}}(d_x, d_y) \geq \theta_{\mathrm{NSM}}$ であるならば、2 つの文書 d_x, d_y に出現するクエリ q は同一の実体を参照する。なお、θ_{NSM} はあらかじめ定めた閾値である。

この方法においては、規則をもとにしたマッチング (この場合は近隣文字列の一致) により、文書がクラスタリングされると見なすことができる。そこで、この方法を**近隣文字列マッチング**(NSM; Neighbor String Matching) と呼ぶ。

[1] このとき $p_i - \alpha$ が 0 以下になったら 1、$p_i + \alpha$ は文書の出現単語数を超えたら文書の出現単語数とする。

[2] この計算法は、ベクトル空間モデルを用いて近隣文字列の集合 S_x, S_y を表現し、その内積で類似度を計算することに相当する。

人名を例に考えると、文脈中に人名が出現した場合、肩書き、年齢などといったその人に関する情報は、人名の出現位置の近くに出現すると考えるのが自然である。すなわち、近隣文字列マッチングでは人名の出現位置のごく近傍に出現する語に着目することで、特殊な情報抽出の方法を用いることなしに、名寄せを行う上で重要な素性を抽出し、比較が行われることが期待できるのである。

近隣文字列マッチングによる分類は、それ単独で用いることも、また先述の階層併合的クラスタリングと組み合わせて用いることも可能である。階層併合的クラスタリングと組み合わせる場合においては、先に近隣文字列マッチングをし、その結果について階層併合的クラスタリングをする場合と、先に階層併合的クラスタリングをし、その結果について近隣文字列マッチングをする場合を考えることができる。

3.2.2 Named Entity を用いたマッチング

本節では、文書中に出現する特徴的な語に注目して類似度を計算する方法を提案する。

Named Entity(固有表現、以下 NE) とは人名や組織名など事物に特有の表現である。NE を検出する際は、NE をタグ付けするツールを使用することができる。例えば、構文解析器である CaboCha についても形態素解析の機能の他に NE をタグ付けする機能を持っており、簡単に NE の情報を利用することが可能である。

NE は一般的な語と比較して、より文書を特徴付けていると考えられ、NE を用いることができれば、文書の意味を幾分か利用可能である。そこで、NE を用いて同一人物の同定をすることを考える。

3.2.3 Named Entity を用いたマッチングのアルゴリズム

NE のマッチングのアルゴリズムを以下に示す。

procedure NE マッチング

step 1. クエリ q を含むすべての文書 d_j $(1 \leq j \leq k)$ について、NE tagger によってタグ付けされた人名、および地名を抽出する。なお、人名に関しては姓名が揃って出現するものに限る。

step 2. NE の類似度 $\mathrm{sim}_{\mathrm{NE}}(d_x, d_y)$ を次のように計算する。

$$\begin{aligned}\mathrm{sim}_{\mathrm{NE}}(d_x, d_y) = \quad & \beta * (d_x, d_y\text{に共通して出現した人名の数}) \\ & +\gamma * (d_x, d_y\text{に共通して出現した地名の数})\end{aligned}$$

なお、β, γ はそれぞれ人名の共起、地名の共起に対する重みである。

step 3. $\mathrm{sim}_{\mathrm{NE}}(d_x, d_y) \geq \theta_{\mathrm{NE}}$ であるならば、2 つの文書 d_x, d_y に出現するクエリ q は同一の実体を参照する。なお、θ_{NE} はあらかじめ定めた閾値である。

なお、NE に関しても、ベクトル空間法同様文書が長くなった場合に、離れた位置に出現する NE が悪影響を及ぼすことが考えられるので、適当な window size を設定する。また、近隣文字列のマッチング同様、単独で用いることも、階層併合的クラスタリングと組み合わせて用いることも可能であり、近隣文字列マッチングと組み合わせることも可能である。

3.3 具体的成果

3.3.1 実験

実験データとして、37 個の人名、地名を検索エンジンへのクエリとして使用し、各クエリについて最大 200 個の URL を収集した。このうち実際にクエリを含む Web ページを集め、最終的に 3,859 個の Web ページを得た。同一クエリにより収集したページを、人手により分類し、正解集合とした。

評価の指標としては、F-measure，Precision，Recall を用いた。これらは、それぞれ以下のように定義され、いずれも 1 に近いほど性能が良いという評価になる。

- Precision は分類した結果が正解をどれだけの精度で含んでいるか
- Recall は分類した結果が正解をどれだけ再現しているか
- F-measure は分類の総合評価 (Precision と Recall の調和平均)

3.3.2 性能比較の方法

収集した各クエリごとに、ベクトル空間法 (VSM), 近隣文字列マッチング (NSM), NSM と VSM を併用 (NSM,VSM), NE を用いたマッチング (NE), NE と VSM を併用 (NE,VSM), NE と NSM を併用 (NE,NSM), NE と NSM，VSM を併用 (NE,NSM,VSM) の 7 種の手法により分類を試み、指標の値を計算する。なお、NE や NSM と VSM を併用する際は、NE や NSM によるマッチングの後に VSM を用いたクラスリングを行うことにする。

3.3.3 実験結果

37 クエリについて実験を行った結果を表 2 に示す。NE のマッチングと近隣文字列マッチングを併用する方法 (NE,NSM) が最も高い性能を示し、次いで NE のマッチングのみを用いる方法 (NE)、NE のマッチングと近隣文字列マッチングを行い、さらにベクトル空間法でクラスタリングする方法 (NE,NSM,VSM) となっている。

NE,NSM では基準となる VSM に対してすべての指標が上回り、F-measure において約 0.22 ポイントの向上が見られることから、名寄せの問題の解決策として効果があると言える。

NE に関しても同様のことが言える。さらに、NE については F-measure が 2 番目に高いことに加え、Precision が高いことも特筆すべき点である。Precision は先述のとおりクラスタ内に正解とは異なる要素が含まれることにより低下する。NE においては、Recall も VSM 以上の値を達成していることから、正解の要素のみを高い精度で集める性能があると評価できる。

NE,NSM,VSM と、NSM,VSM や NE,VSM といった VSM を併用する方法では、併用しない場合と比較して F-measure の低下が見られる。これらはどれも共通して、VSM を用いない場合よりも Precision が低下し、Recall が向上している。

表 2: 実験結果 (全 37 クエリの平均)

適用した手法	平均		
	F	P	R
VSM(Baseline)	**0.4596**	0.6821	0.5154
NSM	**0.5871**	0.8302	0.5852
NSM,VSM	**0.5510**	0.6889	0.6597
NE	**0.6311**	0.9484	0.5585
NE,VSM	**0.5874**	0.7126	0.6579
NE,NSM	**0.6834**	0.7991	0.7357
NE,NSM,VSM	**0.6225**	0.6639	0.7811

3.3.4 試作システム

本システムは Web 上で動作するシステムとして実装されている。図 3 にスクリーンショットを示す。

4 成果要覧

査読付論文リスト

[査読付 1] Shingo Ono, Minoru Yoshida, Hiroshi Nakagawa: NAYOSE: A System for Reference Disambiguation of Proper Nouns Appearing on Web Pages, Proceedings of AIRS2006(Asia Information Retrieval Symposium), LNCS 4182, pp.338-349, 2006.

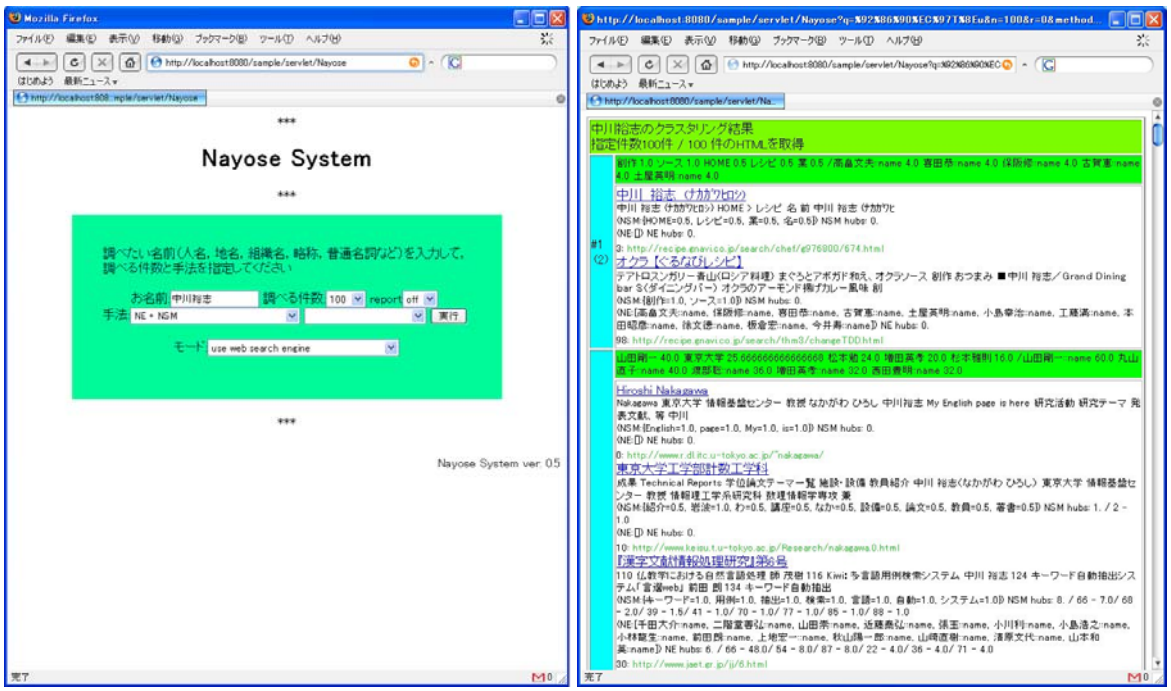


図 3: 名寄せシステム

その他の発表論文リスト

[発表 1] 寺田 昭, 吉田 稔, 中川 裕志: 同義語の類似度に関する考察, 言語処理学会 第 13 回年次大会, B5-2, 滋賀, 2007 年 3 月

[発表 2] 杉浦 隆博, 吉田 稔, 山田 剛一, 増田 英孝, 中川 裕志: 新聞記事の数値による情報検索システムの提案と実装, MuST(動向情報の要約と可視化に関するワークショップ) 2006 年度成果進捗報告会, 京都, 2007 年 3 月

[発表 3] 寺田 昭, 吉田 稔, 中川 裕志: 文脈情報による同義語辞書作成支援ツール, 情報処理学会研究報告「自然言語処理研究会」 No.176, pp. 87–94, 2006

[発表 4] 横川 智浩, 吉田 稔, 山田 剛一, 絹川 博之, 中川 裕志: 機械学習を用いた Web 上の表情報の例示検索方式, 第 5 回情報科学技術フォーラム (FIT2006), D-049, 福岡, 2006 年 9 月

[発表 5] 吉田 稔, 中川 裕志, 寺田 昭, 齋藤 隆: Suffix Array を用いた同義表現抽出手法の研究, NLP 若手の会 第 1 回シンポジウム, 愛知, 2006 年 9 月

[発表 6] Minoru Yoshida: Web Document Structuring by Header Extraction, Japanese-German Workshop on Natural Language Processing, Tokyo, June 2006

多様な情報資源を統合的に活用した インタラクティブ情報探索支援

清田陽司

1 概要

人間とのインタラクションを行うことを前提としたコンピュータシステムの研究においては、システムの性能をどう評価するかという問題が常につきまとう。人間とのインタラクションを抜きにしてシステムを評価する場合には、入力に対してどのような出力が望ましいのかを定義した評価用データコレクションを準備すれば客観的に性能を測定することができる。しかし、人間とのインタラクションが発生する場合、人間はシステムの出力に影響を受けて次の入力を行うため、単純な「入力→(システム)→出力」というモデルでは十分な評価ができず、「システム-人間」という系全体をみて評価を行うことが必要となる。

Web サーチエンジンや図書館蔵書検索システムを代表とする情報探索システムの研究分野では、「系全体を評価する」という観点がこれまであまり重視されておらず、「入力→(システム)→出力」モデルによる性能評価が中心であった。具体的には、入力としてのユーザ質問(キーワードや自然文)に、出力として適切な情報集合を対応づけたテストコレクションを利用した性能評価(精度/再現率)が採用されてきた。しかし、情報探索という問題を扱う際にはこの方法論では不十分である。なぜなら、情報探索の過程で「人間がもっている漠然とした情報要求が具体化される」という側面を無視しているからである。例えば、Web サーチエンジンで「今日のディナーに行くお店」を探す場合を考えてみよう。最初から「○○料理が食べたい」という要求が明確になっているわけではなく、適切な検索キーワードをなかなか思いつけないが、それでも何とかキーワードをひねり出して「本郷　グルメ」などと入力することも多いのではないだろうか。そこで表示されるページのリストを眺めていて、イタリア料理店のページを見つけてはじめて「イタリア料理にしようかな？それじゃあ『本郷　イタリアン』でもう一度探してみようか」と思い至って検索キーワードを修正することもあるだろう。現実に日々発生する情報探索では、いかに効率的にシステムが人間に情報探索のヒントを提示し、容易に人間が情報要求クエリを修正できるかが、情報探索全体の効率を大きく左右する。

ここで示したような「情報要求のインタラクションによる具体化」という過程は、テストコレクションでは評価することが困難であり、未だに適切な評価手法は確立されていない。私は、この過程を評価するには「実世界において多数の利用者にシステムを利用してもらい、人間とシステムの振る舞いを分析する」ことが不可欠だと考えている。ゆえに「分析結果に基づきシステム設計に変更を加え、その変更が振る舞いに与える影響を調べる」という試行錯誤を繰り返すことで、適切な評価手法を徐々に確立することを目指している。

私は上記の目的意識に基づき、以下の研究テーマに取り組んでいる。今年度は、実世界において各種システムを運用するための基礎となる情報資源の整備と、システムのプロトタイプ作成を中心に行った。来年度は実際にシステムを図書館のレファレンスサービスや企業のテキストマイニングの現場で運用し、その結果を詳細に分析することを目標とする。

- 図書館における情報探索支援システム [発表 1]

図書館において行われているレファレンスサービスを補完し、利用者自身が多様な情報資源(図書館や Web)を使いこなすことを支援するシステムの実現を目指している。今年度は、東京大学附属図書館におけるレファレンスサービスの実態を調査し、その結果に基づいてシステムの要件定義と概要設計を行った。さらに来年度のシステム実運用に向けてプロトタイプシステムを作成した。

- **Web の分類体系と図書館の分類体系の対応付け** [査読付 1] [発表 2] [発表 3] [発表 4]

  Web の情報資源と図書館の情報資源を統合的に活用した情報探索を実現するには、両者のそれぞれの情報資源を分類している体系を適切に対応づけることが必要とされる。本研究テーマでは、Web の分類体系として Web ディレクトリと Wikipedia、図書館の分類体系として件名標目表・書架十進分類法を対象とし、両者を精度良く対応づけるための方法論を考察し、オントロジー対応付けの実験と精度評価を行った。また、両者の分類体系をブラウジングするためのインタフェース「Reference Navigator」を試作した。

- **事例レポートを対象としたドメインオントロジー作成支援** [査読付 2] [発表 5]

  航空会社などにおいてテキスト形態で蓄積されている膨大な事例レポートを解析し、事故の未然発生防止に生かすには、human factor に関するドメインオントロジーが必要とされる。本年度は、昨年度開発したドメインオントロジー作成支援ツールキットに関する研究発表を行った。また、いくつかの human factor について、作成されたドメインオントロジーを用いた時系列解析を行った。

2 図書館における情報探索支援システム

2.1 背景

現在の図書館はさまざまなサービスを利用者に提供しているが、その中の一つにレファレンスサービス(参考調査業務)がある。レファレンスサービスとは、利用者の資料・情報探索行動を図書館員が援助する人的支援サービスである。「図書館のコンシェルジェ・サービス」とも例えられており、非常に一般的で広範な質問に対して応答し、資料提供などを行っている。例えば、次のような質問が寄せられる。

- **利用案内**:図書館内の施設やシステムの利用方法を問う質問
  - 例:OPAC の使い方を教えてください。
- **所蔵調査**:資料の所蔵有無を問う質問
  - 例:「歌舞伎登場人物辞典」は東大図書館にありますか。
- **文献調査**:資料の探し方を問う質問
  - 例:昭和 20 年代当時の東大の雰囲気を知ることができる資料が欲しい。
- **事項調査**:特定事項の事実関係及び調査方法を問う質問
  - 例:大学いもの「大学」は東京大学のことを指しているのですか。

しかし、現状では必ずしもレファレンスサービスが利用者に有効活用されているとはいえない。その原因を整理すると以下のようになる。

- レファレンスサービスの存在自体が利用者に認知されていない。富山県でのアンケート調査によれば、図書館来訪者の 8 割以上がレファレンスサービスを一度も利用したことがないと答えている。

- Web で調べ物を済ませてしまう利用者が少なくない。これはサービスの利便性の問題が大きい。Web サーチエンジンは 24 時間どこからでも利用可能であるのに対し、図書館の現行サービスには時間的・空間的な制約が存在する。図書館においても各種のサービスを Web 上で提供していこうという試みはなされているが、個々のサービスの連携不足のため、ハイパーリンクという Web の特性が十分に生かせていない。また、Web 上に存在する有用な情報探索リソースとの連携も不十分である。
- 情報探索のノウハウが十分に共有化されておらず、それぞれの図書館員のスキルに負う部分が大きい。共有化の試みとして、図書館員がもっているさまざまな探索のノウハウをパスファインダーとよばれる形式で提供するサービスもなされているが、それぞれの分野についてのパスファインダーを人手で作成しているため、十分なカバレッジを得られていない。

2.2 内容

上に示した問題を解決するためには、図書館のもつさまざまな情報資源と Web 上の情報資源を統合的に探索できるシステムを利用者向けに提供することが必要であると考えている。言い換えると、「図書館の Yahoo!」と呼べるようなポータルシステムが必要である。

これを実現するのに必要な知見を得るため、東京大学附属図書館の協力を得て、現状のレファレンスサービスの業務分析を行った。具体的には、以下の 3 種類のレファレンスサービスに関する調査を行った。

- カウンターにおける質問応答

  本郷キャンパス総合図書館の協力を得て、1 階総合カウンターに寄せられる質問の傾向を分析した。統計的データの調査だけではなく、実際にカウンターに入って具体的な質問内容を調べた。

- 文書を介した質問応答

  東京大学総合図書館に FAX、郵便、E-mail で寄せられるレファレンス質問の分析を行った。また、図書館員が質問に応答するために利用している情報資源についても調べた。

- ASK サービスにおける質問応答

  ASK サービスとは、東京大学附属図書館が 2005 年 3 月より運用を開始したデジタルレファレンスサービスである。利用者はレファレンス質問を Web から申し込み、メールで回答を得ることができる。運用開始時より 2006 年 8 月までのすべてのレファレンス質問について、質問内容などの調査を実施した。

上記の調査の結果に基づき、本研究の目標を満たすシステムの要件定義、概要設計を行った。さらに、一部のコンポーネント（インタフェース、情報探索メタ知識、統合オントロジー）について詳細設計を行い、システムのプロトタイプを作成した。

2.3 具体的成果

業務分析の結果、以下の事実が明らかになった。

- カウンターにおける質問の約 6 割は、利用案内に関するものであった。また、利用案内に関する質問のうちの大部分は頻繁になされるものであり、それらへの回答は附属図書館ホームページやパンフレットに掲載されているものであった。しかし、一般利用者にはホームページやパンフレットによって提供されている情報が十分に行き渡っていない傾向が明らかとなった。
- 文書を介した質問の約 8 割は、所蔵調査に関するものであった。そこで、所蔵調査に関する質問において図書館員が探索に利用している情報資源について調べたところ、調査までの手

順はほぼパターン化されていることが明らかとなった。具体的には、OPAC などで質問に該当する図書十進分類を求めた後、実際にその分類に対応する参考図書コーナーに足を運んで有用な手がかりを探している図書館員が多かった。

- ASK サービスにおける質問応答のうち約 6 割は電子ジャーナルのトラブルに関する質問が占めており、ASK サービスの提供者側が想定していた文献調査や事項調査に関する質問は極めて少なかった。その原因としては、ASK サービスにレファレンスインタビュー(利用者との対話)の機能が備わっていないことが考えられる。利用者自身が質問を文章の形で明確化しなければならないため、レファレンスサービスの重要な特徴である「図書館員のインタビューによる利用者質問の明確化」というプロセスが欠落している。

上記の調査結果に基づき、本研究の目的を達するためのシステムの要件定義を下記の通り行った。

- 図書館の利用に関する基本的な知識を自然言語の形式で持っている。
- 図書館員がもっている探索に関するノウハウを体系化して保持している。具体的には、探索事項別に有用な情報リソースを整理して持っている。
- 利用者の質問文に含まれるキーワードをより一般的な主題キーワード(件名標目や図書十進分類)に置き換えるためのオントロジーを備えている。
- 対話的インタフェースを備えており、利用者に対するレファレンスインタビューと同様の問い返しを実現することができる。

さらに、要件定義に基づいて実際にシステムの概要設計を行った。システムの概要図を図 1 に示す。

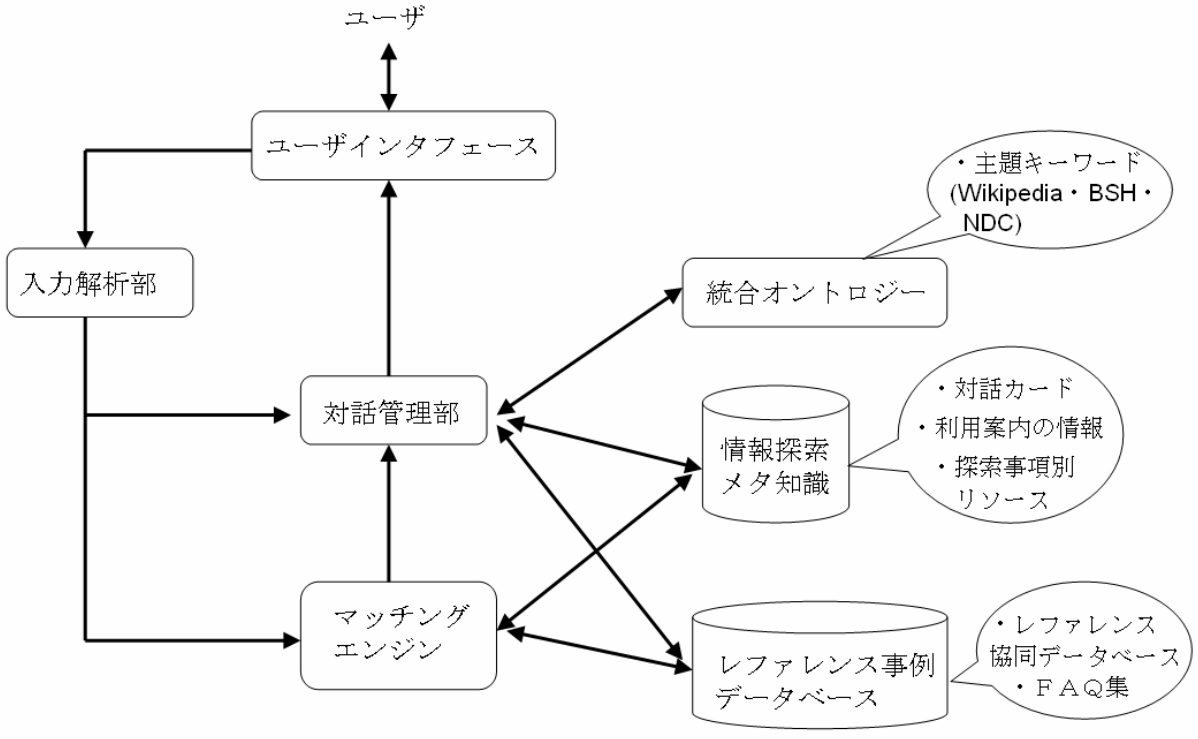


図 1　対話型情報探索支援システムの概要図

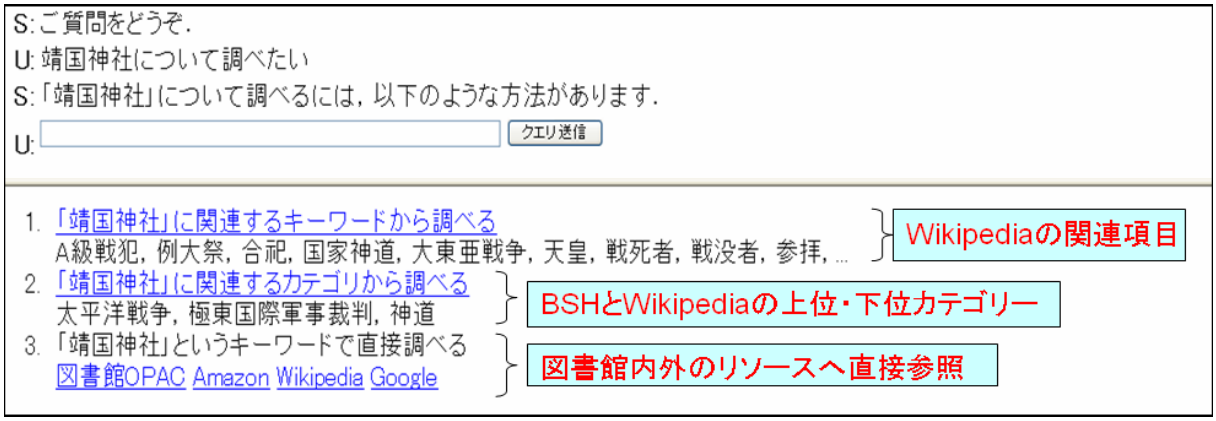


図 2　対話型情報探索支援システムのユーザインタフェース

続いて、システムを構成する各コンポーネントの詳細設計を行い、プロトタイプシステムを作成した。図 1 のコンポーネントのうち、入力解析部、対話管理部、マッチングエンジンについてはマイクロソフト「ダイアログナビ」の研究成果[1]を参考にした。また、対話型インタフェース(図 2)の実装も行った。統合オントロジーについては次の節で詳しく述べる。

ここに述べた研究成果は、2006 年 11 月に開催された日本図書館情報学会研究大会にて発表した[発表 1]。来年度はプロトタイプシステムを実際に東京大学附属図書館にて公開運用し、その運用ログを分析してシステム全体の評価を行うことを予定している。

3　Web の分類体系と図書館の分類体系の対応付け

3.1　背景

インターネットが一般に普及する以前、情報探索を行うにあたって中心的な役割を果たしていたのは図書館であった。図書館には、膨大な資料を整理・組織化し、利用者の情報探索に役立てるためのさまざまなツールが存在する。代表的なツールとしては、日本十進分類法(NDC)や基本件名標目表(BSH)などの分類体系があげられる。

現在では図書館にかわって Web が情報探索の中心的な役割を果たしつつある。まず Google などの Web サーチエンジンを利用することで、たいていの検索質問に対しては何らかの Web ページが見つかるという状況になっている。しかし、Web 上の膨大な情報は十分に組織化されていないため、情報要求に適したページを見つけることは必ずしも簡単ではない。

近年、Web 技術を活用することで膨大な知識を多人数の共同作業によって組織化しようという取り組みが盛んに行われるようになってきている。その代表例としては、Wikipedia[2]や、各種の Folksonomy サービス(Flickr[3]など) があげられる。その結果として、Web 上に一種の情報探索用オントロジーとよべるものが存在するという状況が生まれつつある。これらのオントロジーには、多数の人々の多様な観点が反映されているという大きな利点があるが、限界もある。Web 上に存在するオントロジーは浅い組織化にとどまりがちであり、探索を深く掘り下げていくという用途には必ずしも適していない。

一方で、Web の世界での変革に対応して、図書館の世界においても新たな潮流がみられる。例えば、特定のトピックごとに図書館が提供できる関連資料をリスト化したパスファインダーとよばれる情報資源を Web 上で公開する取り組みも盛んに行われるようになっているが、図書館が人手をかけて

[1] 清田 陽司, 黒橋 禎夫, 木戸 冬子: 大規模テキスト知識ベースに基づく自動質問応答 ーダイアログナビー, 自然言語処理, Vol. 10, No. 4, pp. 145-175, July, 2003.
[2] http://ja.wikipedia.org/
[3] http://www.flickr.com/

構築する必要があるため、Web 上に存在するオントロジーに比較すればごく一部のトピックをカバーしているにすぎない。図書館はまだまだ Web の変革のスピードに追いついていないというのが現状である。

3.2 内容

上で述べた状況に対する解決策として、図書館の分類体系と Web で構築された各種オントロジーを対応付け、情報探索のインフラとして活用することを提案する。具体的には、図書館における分類体系(NDC、BSH)と Wikipedia を対応づけ、情報探索に活用する方法を研究している。

情報探索ツールとしての図書館分類体系と Wikipedia の特徴を整理すると、下記のようになる。

- 図書館分類体系

  専門家による慎重な作業を経てトップダウン的に構築されていることから、安定的に利用することができる、深い組織化がなされているという長所がある。また、特に BSH、LCSH(米国議会図書館件名標目表)、NDLSH(国立国会図書館件名標目表) などの各種件名標目表は、情報探索に拡がりをもたせるためにきわめて有用なツールである。

  反面で、改定の周期が長いため新しくできた概念には対応できない、一般的に上位概念は 1 つしかもてないため概念の多様な側面を反映しづらいという欠点をもっている。

- Wikipedia

  Web 上で Wiki システムによって共同編集されているオンライン百科事典であり、誰でもどこからでも自由に編集することができる。現在、200 以上の言語で作成されており、日本語版では 30 万項目以上の記事が存在する。多数の人が編集に参加していることから、世界中のほとんどの概念を網羅していると思われる。また、ハイパーリンクを容易に張ることができるという Wiki の特徴を生かし、項目の組織化の取り組みも行われている。膨大な項目から効率的に一覧を生成するために「Wikipedia カテゴリ」というボトムアップ的な組織化の仕組みが導入されており、項目を多様な観点から分類することが可能となっている。例えば、「価格」という項目には「マーケティング」「経済学」「市場」といった複数のカテゴリを付与することができる。また、カテゴリ自体にも上位概念(親) のカテゴリを付与することが可能である。複数の親をもたせることができるのは、従来の分類体系にはない大きな特徴である。

  反面で、頻繁な編集が行われるため体系が安定していない。信頼性に欠ける情報も少なくないといった欠点がある。

このように、図書館分類体系と Wikipedia は対照的な特徴を有している。本研究では、両者のもつ共通部分を対応付けることによって、情報探索の出発点として Wikipedia を用い、そこから概念を一般化することによって図書館の分類体系に導いていくという統合的活用法を提案する。

図 3 に情報探索への活用例を示す。まず、情報探索の出発点としての Wikipedia の活用について説明する。ここで、Wikipedia の記事「阪神・淡路大震災」には、カテゴリとして「日本の経済史」「地震の歴史」が付与されている。さらに、カテゴリ「日本の経済史」には上位カテゴリとして「経済史」が、カテゴリ「地震の歴史」には上位カテゴリとして「災害と防災の歴史」「地震」が付与されている。このように、Wikipedia の記事を一つとりあげてみると、関連するカテゴリ群をツリー構造として取り出せることがわかる。このツリー構造をたどることで、「阪神・淡路大震災」について調べる際の切り口が明確になっていく。例えば、「日本の経済史」→「経済史」とたどっていくと、「阪神・淡路大震災が日本経済に与えた影響」について調べることが可能だとわかるし、「地震の歴史」→「地震」とたどっていくと、「阪神・淡路大震災を引き起こした地震(すなわち 1995 年兵庫県南部地震)」について調べることが可能だとわかる。結果として、「阪神・淡路大震災」という探索主題を、「経済史」「災害」「地震」に一般化して調べることが可能となる。

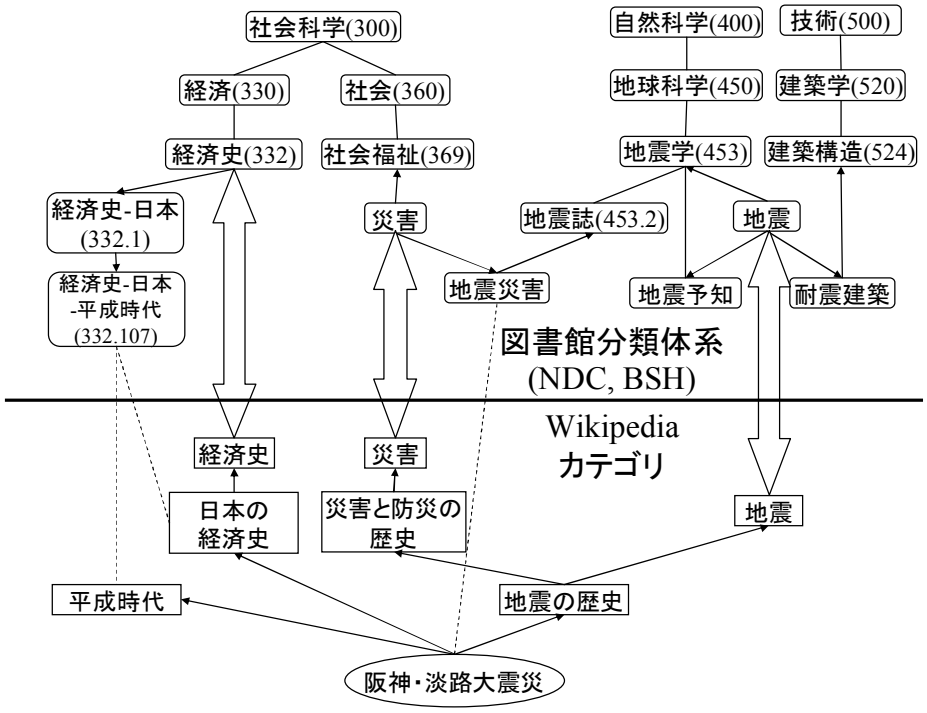


図 3　図書館の分類体系と Wikipedia カテゴリの対応付け

次に、Wikipedia カテゴリと図書館の分類体系(BSH、NDC) の対応付けと、図書館の分類体系の活用について説明する。Wikipedia カテゴリと図書館の分類体系の間には、後述するようにカテゴリ名が一致するものが存在する。図 3 では、「経済史」「災害」「地震」が一致している。この一致を利用することにより、Wikipedia の活用によって一般化された探索主題を、図書館の分類体系によってさらに掘り下げていくことが可能となる。例えば、「経済史」→「経済史-日本」→「経済史-日本-平成時代」とたどっていくと、NDC 分類記号 332.107 によって「日本経済事典」(日本経済新聞社、1996 年) などを探すことができるし、「地震」→「耐震建築」とたどっていくと、件名「耐震建築」によって「地震とマンション」(西澤英和ほか、筑摩書房、2000 年) などを探すことができる。

3.3　具体的成果

上で提案した統合的活用法の有用性を検証するため、異種オントロジーの統合ブラウジングツール「Reference Navigator」を試作した。本ツールのインタフェースを図 4 に示す。本ツールは、Wikipedia、NDC9、BSH4 のオントロジー構造に加えて、Open Directory Project で構築された Web ディレクトリ構造をリレーショナルデータベースに格納しており、統合的に探索することが可能である。上のテキストボックスに探索主題(Wikipedia の項目名) を入力すると、その上位カテゴリ名などがハイパーリンクとして表示される。ここでいずれかのカテゴリ名をクリックすると、さらにそのカテゴリの上位カテゴリ名、下位カテゴリ名などが表示される。同一のカテゴリ名が Wikipedia 以外にも含まれる場合には、それらの上位カテゴリ名、下位カテゴリ名も共に表示される。クリックを繰り返していくことで、情報探索を深めていき、図書館の情報資源を探すための手掛かりを得ることができる。

本年度は、Reference Navigator の概要についての発表[発表 2][発表 3]を中心に行い、Reference Navigator を東京大学総合図書館における参考調査業務にて試用する取り組みも並行して行った。また、本研究テーマに関連して、Wikipedia と Open Directory を精度良く対応づける研究[発表 4]や、一般の人々に馴染みのある分類体系を Blog カテゴリをもとに自動付与する研究[査読付 1]にも取り組んだ。来年度は、Reference Navigator のデータベースを前節で述べた図書館における情報探索支援システムに利用し、その有用性を検証することを計画している。

図 4　Reference Navigator

4　事例レポートを対象としたドメインオントロジー作成支援

4.1　背景

現在、航空、鉄道、医療などの各分野において、事故や失敗例を事例レポートとして収集するシステムが整備されつつある。例えば、米国連邦航空局(FAA)は NASA と共同で Aviation Safety Reporting System (ASRS)[4]というレポート収集システムを運営し、航空関係者から事故やインシデント、人的エラーなどに関する事例を収集している。また、日本においても科学技術振興機構(JST)が科学技術分野における事故や失敗の事例と、得られる教訓をデータベース化し、「失敗知識データベース」[5]として公開している。

航空分野では、事故やインシデントの 80%以上は人的要因によって引き起こされるといわれている。人的要因としては、判断ミス、不注意、コミュニケーション不足、組織上の問題など、多種多様なものが存在する。事例レポートを専門家が解析することで、どのような人的要因が関与していたのかを知ることができる。しかし、事例レポート中に出現する表現のうちどのようなものが人的要因の存在を示すのかは自明ではないため、計算機による事例レポートの解析はそれほど易しくない。

4.2　内容

本研究テーマでは、大量のテキストから自動構築された格フレームを利用して、人的要因に関するレポート分析を実現する枠組みを提案した。具体的には、格フレームを媒体とした人的要因に関するドメインオントロジー(辞書)の作成を支援するツールキットを開発した。システムによる関連格フレームの検索と、ユーザによるフィードバックを繰り返すことにより、高精度なドメインオントロジーを作成することが可能である。詳しい内容に関しては、2005 年度の年報に掲載しているので参照してほしい。

また、ツールキットの評価実験として、3 つの人的要因(「運航スケジュール遵守に伴う心理的圧力」「乗客の迷惑行為」「適切な支援による問題解決」)についてのドメインオントロジーを作成する実験を行った。さらに、作成されたドメインオントロジーを実際の事例レポートに適用し、分析を行った。

[4] http://asrs.arc.nasa.gov/
[5] http://shippai.jst.go.jp/

図 5　JAL 航空安全レポートにおける human factor 辞書の構築

4.3　具体的成果

JAL 航空安全レポート集合(1992 年-2003 年)の約 6、600 個の事例レポートに提案手法を適用し、3 つの人的要因に関してのドメインオントロジー作成実験を行った結果を図 5 に示す。上半分には実際に構築されたドメインオントロジーの一部を示した。ここから多様な表現(格フレーム)が収集されていることがわかる。

また図 5 の下半分には、ドメインオントロジーの表現収集のプロセスを示した。図の縦軸は格フレーム収集プロセスの繰り返し回数(被験者が格フレームを追加した回数)、横軸は提案手法によるランキングの順位(すでにその時点において追加されている格フレームを除いたもの)を意味している。また、“o”、“N”、“*”は関連格フレームを示す: “o”は次のプロセスにおいて被験者が入力ボックスに追加した格フレーム、“N”はそのプロセスにおいて初めて 50 位以内にランキングされた格フレーム、“*”はそれ以外の関連格フレームである。“o”はほとんどの場合 10 位以内にランキングされていることから、本研究において提案したランキング手法は有用であるといえる。また、“N”は被験者からのフィードバック回数が 10 回目以上でも見つかっていることから、フィードバックの繰り返しは有用だといえる。

本研究の成果については、言語資料の利用に関する国際会議 LREC2006 (Genova, Italy, 2006 年 5 月)において発表[査読付 2]した。また、航空関係者が集まるシンポジウムにおいても発表[発表 5]を行った。

5　成果要覧

査読付論文リスト

[査読付 1] Tsutomu Ohkura, Yoji Kiyota and Hiroshi Nakagawa: Browsing System for Weblog Articles Based on Automated Folksonomy. In Proceedings of the WWW2006 Workshop on the Weblogging Ecosystem, Edinburgh, Scotland, May, 2006.

[査読付 2] Yoji Kiyota and Hiroshi Nakagawa: A Domain Ontology Production Tool Kit Based on Automatically Constructed Case Frames. In Proceedings of The Fifth International Conference on Language Resource and Evaluation (LREC 2006), pp. 1482-1487 (P7-T), Genoa, Italy, May, 2006.

その他の発表論文リスト

[発表 1] 國安 結, 清田 陽司, 綾部 輝幸: 東京大学附属図書館におけるレファレンスサービス業務分析 －自動レファレンスサービスシステムの実現を目指してー. 日本図書館情報学会 第 54 回研究大会発表要綱, pp. 101-104, 九州女子大学, 北九州市, October, 2006.

[発表 2] 清田 陽司: Wikipedia と図書館資料の分類体系との対応づけ ～Web 上の二次情報と図書館の一次情報の統合的活用に向けた取り組み～. NLP 若手の会 第 1 回シンポジウム, 名古屋大学, 名古屋市, September, 2006. (発表資料は http://sslab.nuee.nagoya-u.ac.jp/nlp-youth/ に掲載)

[発表 3] 清田 陽司, 田村 悟之, 中川 裕志, 増田 英孝: Reference Navigator：異種オントロジーの統合ブラウジングツール ～図書館の分類体系と Wikipedia カテゴリの対応付け～. 言語処理学会 第 13 回年次大会 ワークショップ「言語的オントロジーの構築・連携・利用」論文集, pp. 35-38, 龍谷大学, 大津市, March, 2007.

[発表 4] 田村 悟之, 清田 陽司, 増田 英孝, 中川 裕志: 図書館における自動レファレンスサービスシステムの実現に向けて ～Web 上の二次情報と図書館の一次情報の統合～. 情報処理学会 研究報告 情報学基礎(SIG-FI) 86-1, pp. 1-8, 化学会館, 東京都, March, 2007.

[発表 5] 清田 陽司, 中川 裕志: 航空安全情報分析ツール －human factor に着目したレポート分析手法の提案－. 第 44 回飛行機シンポジウム CD-ROM 講演集, 2D9, pp. 501-507, 大宮ソニックシティ, さいたま市, October, 2006.

キャンパスネットワーキング研究部門

ネットワークの設計・運用・保守・管理の合理化を図る技術

― アドホックネットワーク、アドレス偽造の検知、P2P トラヒックの弁別、及び次世代衛星ネットワーク―

若原　恭

1　概要

情報通信ネットワークは、そのアプリケーションの急激な発展とともに、転送情報量の規模的な拡大に加え、機能面の質的な変革も必要となる。このような認識のもと、情報通信に係わる各種リソースの最大限の利活用によって、情報通信ネットワークの設計・運用・保守・管理を合理的なものにすることを目標にしている。具体的には、以下に示す諸技術に焦点を当て、理論と実験の両面から研究を行っている。

- アドホックネットワークの運用と制御

移動可能な無線端末のみから構成されるアドホックネットワークは、今後の本格的なユビキタス社会において必須になると考えられている。そのようなアドホックネットワークでは、運用と制御に係わる多くの基本技術が課題として山積している。そのうち、新規参加ノードへのアドレス割当方式、情報転送ルートにおける情報転送能力導出法、ネットワーク転送容量の最大化技術等を検討しており、アドホックネットワークの運用と制御に係わる実用的な基本技術の確立を目指す。

- アドレス偽造の検知

現在のコンピュータネットワークの弱点の一つは、送信側のアドレスを容易に偽造できる点にある。このため不正アクセスや不正侵入等の攻撃においては、送信元アドレスを偽造することが多く、その結果攻撃ホストの特定が困難で匿名性を許すこととなっており、大きな社会問題の一つとなっている。そこで、ネットワークで転送されるパケットの送信側アドレスの偽造を検知する技術を開拓し、攻撃ホストを迅速に特定可能とすることによって安全な情報社会の実現を目標とする。

- P2P トラヒックの弁別

近年、P2P(Peer to Peer)技術に基づくファイル交換ソフトウェア等のアプリケーションが注目を浴び利用が急増している。その結果、トラヒック量の急激な増大だけでなく、著作権のあるファイルの不法な交換や機密情報の漏えい等、社会的な問題にも結びついている。そこで、このようなトラヒックをネットワーク上で弁別するため、パケットのヘッダ情報から P2P アプリケーションのトラヒックであるか否かを判別する方法の検討を進めており、P2P トラヒック量の定量的把握と P2P アプリケーションの利用ホストを迅速に特定可能とすることを目指す。

- 次世代衛星ネットワークの設計と運用

低軌道周回衛星システムは、地震やテロ等の大災害に強く、グローバル性や同報性等の特長を持つことから、次世代の衛星ネットワークとして有力視されている。そのようなシステムの運用では、衛星が常に動いているため、地上の端末がアクセスする衛星が変更となるハンドオーバが頻繁に

発生する。ハンドオーバには、多くの制御情報の授受が必要となったり、パケット転送品質が劣化する等の問題がある。そこで、授受する制御情報を抑え、転送品質の劣化を低減する方法の検討を進めている。また、衛星が故障した場合の対策方法についても検討を進めている。これらの方法によって、安定した運用が可能で衛星障害に強い低軌道周回衛星システムの実現を目指す。

2 アドホックネットワークの運用と制御

2.1 背景

アドホックネットワーク MANET(Mobile Ad-hoc Network)は、ルータやアクセスポイント等のインフラ設備を用いることなく、移動可能な無線端末ノードのみから構成されるネットワークで、例えば、イベント会場や旅行先での情報交換、走行中の車同士の間での情報交換等、近い将来に実現が予想される本格的なユビキタス社会において必須になると考えられている。このため近年研究が活発になってきており、例えば、ルーティングについては、IETF(Internet Engineering Task Force)において、AODV(Ad Hoc On Demand Distance Vector Algorithm)、DSR(Dynamic Source Routing Protocol)、OLSR(Optimized Link State Routing Protocol)等の標準化が進む等実用化のための技術開発も進んでいる。

しかし、例えば移動端末を識別するためのアドレスを割り当てる方法、所要の品質を維持したパケット転送法、無線リソースの有効利活用によるパケット転送情報量の拡大等、未解決の技術課題が多く、これらの解決が急務となっている。

2.2 内容

本研究では、まず MANET にノードが新たに参加する場合に必要な処理としてアドレス割当法を検討した。その結果、予備ノードを含め任意の 2 個のノードで集中的にアドレスを管理し割り当てる方式を考案した。この方式では、一つの MANET が複数個に分割したり、逆に複数個の MANET が一つに統合したりした場合に特別なアドレス管理処理とアドレス再割当処理が必要となるが、総合的にはアドレス割当時間の短縮とネットワーク負荷の削減を達成できるという特長が得られることを実験によって確認した。特にネットワーク負荷の増大は、無線リンク等のリソースが限られている MANET では深刻な問題になることがあるため、本考案方式は、そのような問題に対する有力な解決策と期待できる。

MANET を利用して情報を転送する際には、ノードの移動状況に応じて転送情報の品質(QoS: Quality of Service)が動的に変化するため、転送開始時に所望の品質が達成できるか否かを判定できることが望ましい。その実現を目指し、与えられた転送ルートにおける情報転送能力を高速に計算可能な方法を考案して具体化するとともに、TDMA(Time Division Multiple Access)ベースの MANET におけるタイムスロット割当アルゴリズムを詳細に設計してプログラム化し、実験評価を行った。その結果、ネットワーク規模の大小に依らず高い精度で転送能力が計算できること、及び計算時間が極めて小さく実用上問題にならないことを確認した。また、各無線リンクのタイムスロットの情報転送方向を意識してタイムスロットを割当ることによって、情報転送能力を拡大する方法を考案し、この方法を上述のタイムスロット割当アルゴリズムに含めるよう拡張する検討を開始した。

次に、MANET 全体として同時に転送可能な情報量を最大とする制御法の検討を行った。具体的には、FDM(Frequency Division Multiplex)ベースの MANET を対象として、チャネルの空間的な再利用効率を向上させるチャネル割当法を検討し、利用可能なチャネルのうち最も近傍で使用中のものを選択し割り当てることによって可能な限りチャネルを敷き詰めて使うという原理に基づく方法を考案した。更に、その第 1 段階の評価として、ノードが移動せず固定という条件下で実験を行い、例えば 20%以上の転送可能情報量の向上が達成できることを確認した。

2.3 具体的成果

前節で述べた諸技術は、今後の期待が大きいアドホックネットワーク MANET が現在抱えている基本問題の解決を目指すもので、それぞれ効果が明確であることを確認できており、MANET の実用と発展に寄与するものと期待される。

アドレス割当法については、代表的な従来技術と比較して、割当時間の短縮とネットワーク負荷の削減が可能であることを実証でき、その成果を学会の研究会や大会で発表した[発表 1,10,14,15]。

与えられたルートにおけるパケット転送能力の導出方法については、精度が高くかつ導出に要する計算量が少ないことを検証でき、導出法の原理とアルゴリズム及びその定量的評価結果を国際学会や研究会・大会等で論文発表した[査読付 3,4 発表 8,11,18]。

MANET 全体としてのパケット転送能力の拡大法については、MANET を構成するノードが移動しない場合について評価を行って、従来手法に比較して十分効果があることを確認し、その方法の原理と定量的な効果を学会の大会で発表した[発表 16]。

3 アドレス偽造の検知

3.1 背景

現在広く普及しているインターネットでは、転送するパケット情報に送信元を識別するためのアドレスが含まれている。しかし、この送信元アドレスは任意の値に設定することが可能であり、結果的に匿名性を許す設計・運用となっている。昨今情報ネットワーク社会の安全性を脅かす大きな脅威として、分散型サービス攻撃 DDoS (Distributed Denial of Service)や不正侵入等ネットワーク越しでの様々な攻撃があるが、これらの攻撃では、この匿名性を悪用して攻撃者が送信元アドレス情報を偽装することが多く、攻撃者の特定や追跡を困難にしている。これが現在の情報ネットワーク社会における最も大きな技術的課題の一つになっている。

この課題の解決を図るためには、受信したパケットに含まれる送信元アドレスの偽装を検知する方法、受信した攻撃パケットが転送されたルートを受信ホストから逆にたどるトレースバックを実施することによって送信者即ち攻撃者を特定する方法等がある。一般に、前者の方法では、アドレスに関して何らかの制約条件を設け、その条件下でアドレス偽造を検知できるというレベルに留まっている。また、後者の方法では、ネットワーク内のルータが連携してトレースバックを可能とするための特別な機能を追加する必要があるという制約がある。このように、いずれの方法でも制約や適用可能な範囲が限られており、決定的に有効な対策が未だ得られていない。

3.2 内容

本研究では、送信元アドレスの偽装を検知する方法として、ネットワーク内のルータ間で授受しているネットワークトポロジー情報を利用して、各ルータで受信するパケットが持つ送信元アドレスとして許されるものを受信ポート毎に作成して管理することとし、実際に受信したパケットの送信元アドレスが許容されたものか否かを検査することによってアドレス偽造を判定する方法を考案した。本考案法の有効性、特にアドレス偽造検知率を定量的に明らかにするため、またこの偽造検知機能をどのようなルータに具備するべきかを明らかにするためシミュレーション実験を実施した。その結果、一般に送受信回線が多いルータに本考案機能を実装することによって効率よくアドレス偽造を検知でき、高いアドレス偽造検知率を達成できることを明らかにするとともに、その検知能力は従来手法に比較して遥かに優れていることを実証した。

また、本アドレス偽造検知法を現在のコンピュータネットワークで実現する具体的な実装方法を検討した。対象としたルーティングプロトコルは、中規模以上のネットワークで広く採用されている OSPF(Open Shortest Path First)で、OSPF が持つエリア構造を前提として概要設計を行った。その結果、OSPF で定義されている制御メッセージの若干の拡張等によって考案方法が実現できることを確認した。

3.3 具体的成果

考案したアドレス偽造検知方法には、ルータ間授受情報を利用するため信頼性が高いこと、ルータに追加すべき新機能が少ないこと、送信と受信とでルートが異なる非対称ルートにも適用できること等の多くの特長があることを明らかにした。それら成果を、考案方法の原理や方式説明を含め、国際学会や研究会・大会等で論文発表した [査読付 5 発表 2,3,7,13,19]。

4 P2P トラヒックの弁別

4.1 背景

インターネットの発展に伴って、ネットワークアプリケーションの増大と利活用が著しいが、特に、最近は機能を集中したサーバを持たず、ホストが基本的に対等な関係となる P2P (Peer to Peer)型アプリケーションの利用が急増している。例えば、ファイル交換ソフトウェアとして広く普及している Napstar, Gnutella, WinMX, Winny 等がある。このようなアプリケーションは、現在のインターネットで転送されている全トラヒックの過半数を占めているという報告もあり、そのネットワーク設備設計のためにも実態を定量的に把握することが急務であるうえ、著作権のあるファイルの不法な交換や機密情報の漏えい問題を引き起こすこともあるため、そのような P2P トラヒックを弁別できることが強く期待される。そこで、ネットワーク上転送されるトラヒックをモニタし、P2P トラヒックを弁別する技術の検討を進めた。このモニタでは、利用者のデータ自体を対象とせずヘッダ情報のみを対象とすることによって、検閲やプライバシ侵害に係わる問題が生じない方針を採用した。

4.2 内容

これまでの研究の結果、任意のホストが情報を交換する相手先ホストの多様性を表す新規接続率等を特徴量として P2P トラヒックを識別する方法を考案しており、今年度はまずこの方法を分析評価した。その結果、高い新規接続率という判断で P2P トラヒックを高い精度で弁別することは、昨今の FTTH(Fiber To The Home)による高速伝送路の普及に伴って困難になる場合があることを明らかにした。

そこで、新規接続率とは異なるあらたな特徴量について多面的に調査検討した。その結果、P2P アプリケーションのホストは互いに対等であることから、ホスト間で相互に接続が頻繁に行われるという性質に着目し、短い一定時間内に相互接続が一定回数以上行われるという特徴量によって弁別する新しい方法を考案した。そして、実際のネットワークにおける実験を実施し、考案したこの新 P2P トラヒック弁別法を評価した結果、かなり高い精度で弁別できる見通しを得ることができた。一方、この提案方法には、弁別に要する処理時間が、例えば数 10 分程度であり、比較的短かくて済むという特長があることも明らかにした。

4.3 具体的成果

考案した弁別法については、特許出願を行うとともに、原理と具体的アルゴリズムおよび定量的な効用等を、研究会や大会等で論文発表した。[特許出願 1　発表 4,5,6,12,17]

これまではファイル交換ソフトウェアとして、主に Winny を対象として検討してきた。以上に示した成果によって、今後、Share や Donkey 等の他の P2P ファイル交換ソフトウェアを対象とするべく拡大を図れるとの見通しも得た。

5 次世代衛星ネットワークの設計と運用

5.1 背景

地震やテロ等の大災害に対する強靭性・グローバル性等の理由から低軌道衛星(LEO: Low Earth Orbit)周回通信システムが次世代のネットワークとして有望視されている。パケット転送方式に基づく

LEO(Low Earth Orbit)衛星通信システムを実用化するためには、地上端末がアクセスするサービス衛星を切り替えるというハンドオーバが頻繁に生じることから、地上端末とサービス衛星との対応関係を維持し管理するための移動管理、ハンドオーバに伴うパケット転送品質劣化の最小化等を果たす技術の確立が必須である。地上端末の移動管理については衛星ネットワークのパケット転送容量が限られているため地上のネットワークで採用されている方式を適用すると、管理のために授受する制御情報の量が大半を占めてしまうので、管理のために授受する制御情報の最小限化が極めて重要となる。また、ハンドオーバ時には、一般に、転送パケットの損失・順序逆転を伴ったり、パケット転送が中断したりしてパケット転送に係わる品質が劣化する傾向にあるので、ハンドオーバ時でも高いパケット転送品質を維持できる技術が特に必要となる。端末の移動管理やハンドオーバ時の品質維持技術については、地上のコンピュータネットについての研究は極めて多いが、端末だけでなくネットワークノードとなる衛星までも移動するため移動管理負荷やパケット転送品質劣化が極めて顕著な LEO 衛星通信システムに関する研究はあまり進んでいない。一方、LEO 衛星通信システムの運用という観点からは、衛星が故障した場合の対処技術を開拓することも必須である。この対策については、衛星コストが小さくないことから、予備衛星を持てない場合や予備衛星の故障時の対策も必要となるが、このような対策についてはあまり系統的な研究がなされていないのが現状である。

5.2 内容

LEO 衛星通信システムにおける地上端末の移動管理方法を検討した結果、各地上端末の位置情報をデータベースに集中して蓄積管理することを原則とし、各地上端末がアクセスするサービス衛星は、位置情報と衛星の軌道情報を基にした計算して導出するものとし、導出した衛星と実際のサービス衛星とが異なる場合のみ、サービス衛星を蓄積管理する方式を考案した。一般に、地上端末の移動速度は、衛星に比べるとはるかに小さいので、衛星の移動に伴ってハンドオーバが頻繁に発生してもデータベースが蓄積している地上端末の位置情報を更新する必要はほとんどない。つまり、移動管理に伴う制御情報量は極めて少なくて済み、ネットワーク全体の情報転送容量の過半を占めることはない。シミュレーションを実施し、本考案方法の定量的効果も確認でき、その結果移動管理に関する実用上の問題は基本的に解決できた。

ハンドオーバに伴うパケット転送品質維持方法については、最小限のバッファを各衛星に用意してパケットの損失と順序逆転を完全に回避できる方式を設計するとともに、ハンドオーバの制御処理シーケンスを詳細に分析して、衛星間リンクの伝播遅延時間等を考慮することによってパケット転送中断時間を最小限にできる方式を考案した。考案した方式を、地上のネットワークについて研究した結果得られている方式と比較したところ、パケット転送中断時間を大幅に削減できる等確実に優れていることを実証できた。

LEO 衛星通信システムにおける衛星故障時には、故障衛星の近傍で正常に動作している衛星を幾つか移動し、故障した衛星がカバーしていたエリアをカバーできるようにするとともに、移動した衛星が持つべき衛星間伝送リンクを再設定する衛星再配置法の検討を進めた。このような衛星再配置法の検討は過去極めて少なく、研究アプローチ自体の開拓も研究内容の一部に位置づけて検討を進めた。具体的には、衛星の再配置の最適化をまず行い、その結果に基づいて衛星間リンクの最適化を図るという 2 ステップ設計法を考案した。第 1 ステップでは、特にサービスエリアの最大化を図る方針としたが、移動する衛星個数、サービス停止時間や停止回数等も考慮して総合判断する方針とした。第 2 ステップの設計では、衛星間リンクの距離、ルーティングの容易さ、情報転送容量等を考慮して総合判断する方針とした。

また、以上の設計アプローチを、我が国で研究開発を進めてきた LEO 衛星通信システムである NeLS(Next generation Low earth orbit Satellite System)を対象として、一つの衛星が故障した場合に適用し、評価を進めた。NeLS は 120 個の衛星と 240 本の衛星間リンクで構成される。評価の結果、故障衛星と同一軌道上の前後にある 2 個の衛星と 8 本の関連衛星間リンクのみを再配置することによってサービスカバー率をほぼ 100%に維持した上で衛星間リンク距離やルーティング上のホップ数の増大を抑えることが可能であることを明らかにし、本設計アプローチの有効性を確認した。

5.3 具体的成果

地上端末の移動管理とハンドオーバ時のパケット転送品質維持技術に関する研究成果ついては、国際学会で論文発表を行った[査読付 1,2 発表 9]。また、衛星故障時の対策技術に関する研究成果については、2007 年 4 月に開催される国際学会での論文発表を予定している。

特許出願

[1] 若原恭、中村文隆、松田崇、田中良明、千田浩一、加藤圭、藤原雄彦：P2Pトラフィック監視制御装置及び方法、特願 2006-221578(平成 18 年 8 月 15 日)

査読付論文リスト

[査読付 1] Yasushi WAKAHARA, Kazumasa SATO and Jun-ichi MIZUSAWA, “Proposal of Seamless Handover for Low Earth Orbit Satellite Systems”, 24th AIAA International Communications Satellite Systems Conference (ICSSC), AIAA2006-5357, San Diego California (June 2006)

[査読付 2] Yasushi WAKAHARA, Hiroko TSUCHIYA and Jun-ichi MIZUSAWA, “Efficient Mobility Management of Earth Terminals for Low Earth Orbit (LEO) Satellite Systems”, 24th AIAA International Communications Satellite Systems Conference (ICSSC), AIAA2006-5335, San Diego California (June 2006)

[査読付 3] Jianping LI and Yasushi WAKAHARA; End-to-end Bandwidth Guarantee in a Mobile Ad Hoc Network、情報処理学会、マルチメディア・分散・協調とモバイルシンポジウム(DICOMO 2006)、pp.389-392（2006 年 7 月）

[査読付 4] Jianping Li and Yasushi Wakahara：Time Slot Assignment for Maximum Bandwidth in a Mobile Ad Hoc Network, Proc. Of International Symposium on Wireless Pervasive Computing, ISWPC#1568999587, Puerto Rico（February 2007）

[査読付 5] Toshinori Ohtsuka, Fumitaka Nakamura, Yuji Sekiya, Yasushi Wakahara：Proposal and Efficient Implementation of Detecting and Filtering Method for IP Spoofed Packets, International Conference on Information Communication Technology (ICICT), (March 2007)

その他の発表論文リスト

[発表 1] 粕田 大樹，水澤 純一，若原 恭：モバイルアドホックネットワークにおけるアドレス構造化を用いたアドレス管理手法、電子情報通信学会、次世代ネットワークソフトウェア研究会、pp.13-21（2006 年 6 月）

[発表 2] 大塚俊範、中村文隆、関谷勇司、若原　恭；ネットワークトポロジ情報を用いた効果的なIP偽装パケット検知手法、電子情報通信学会技術報告IN2006-30、pp.79-84（2006 年 6 月）

[発表 3] 大塚俊範、中村文隆、関谷勇司、若原　恭：IP偽装フィルタリング機能の効果的な実装、電子情報通信学会技術報告NS2006-75、pp.27-32（2006 年 9 月）

[発表 4] 中村文隆、松田　崇、若原　恭、田中良明：トラヒック特徴量解析とアプリケーション弁別[招待講演〕、電子情報通信学会技術報告NS2006-80、pp.57-62、IN2006-60、pp.67-72、CS2006-26、pp.25-30（2006 年 9 月）

[発表 5] 松田　崇、中村文隆、若原　恭、田中良明：待ち受けポートの挙動に着目した P2P トラヒック弁別手法、電子情報通信学会、通信ソサイエティ大会、B-6-1、（2006 年 9 月）

[発表 6] 松田　崇、中村文隆、若原　恭、田中良明：通信の相互性に着目した PureP2P トラヒック弁別手法、電子情報通信学会技術報告 NS2006-134、pp.61-66（2006 年 11 月）

[発表 7] 大塚俊範、中村文隆、関谷勇司、若原　恭：ISPネットワークにおけるＩＰ偽装パケット検知機能FSNの実装、電子情報通信学会技術報告CS2006-90、pp.115-120（2007 年 1 月）

[発表 8] Jianping Li and Yasushi Wakahara : Time Slot Assignment in a CDMA/TDMA Ad Hoc Network、電子情報通信学会技術報告IN2006-158、pp.121-126（2007 年 1 月）

[発表 9] Yasushi Wakahara: Network Control for LEO (Low Earth Orbit) Satellite Systems, Proceedings of COE Symposium on Advanced Electronics for Future Generations, pp.263-269 (January 2007)

[発表 10] 粕田 大樹、水澤 純一、若原 恭：モバイルアドホックネットワークにおけるアドレス構造化を用いた集中管理型アドレス自動割当方式、電子情報通信学会第 6 回次世代ネットワークソフトウェア研究会、pp.1-2（2007 年 2 月）

[発表 11] Jianping Li, Yasushi Wakahara : Optimal Time Slot Assignment for Wireless Ad Hoc Networks with Directed Links、電子情報通信学会技術報告IN2006-235、pp.327-332（2007 年 3 月）

[発表 12] 松田　崇、中村文隆、若原　恭、田中良明：相互接続における順逆接続間隔を利用した P2P トラヒック弁別手法、電子情報通信学会技術報告 NS2006-237、pp.415-420（2007 年 3 月）

[発表 13] 大塚俊範、中村文隆、関谷勇司、若原　恭：OSPFを利用したIP偽装パケット検知機能FSNの実現方法、電子情報通信学会技術報告NS2006-236、pp.409-414（2007 年 3 月）

[発表 14] 粕田 大樹、水澤 純一、若原 恭：モバイルアドホックネットワークにおけるアドレス構造化を用いた集中型アドレス割当方法、情報処理学会第 69 回全国大会 3U-7、pp.3-129-3-130（2007 年 3 月）

[発表 15] 田川 雅士、水澤 純一、若原 恭：アドホックネットワークにおけるアドレス分散化理法の提案、情報処理学会第 69 回全国大会 3U-8、pp.3-131-3-132（2007 年 3 月）

[発表 16] 田中 隼人、水澤 純一、若原 恭：複数チャネルを用いるアドホックネットワークにおける最適チャネル割り当て法の提案、情報処理学会第 69 回全国大会 3U-9、pp.3-133-3-134（2007 年 3 月）

[発表 17] 松田　崇、中村文隆、若原　恭、田中良明：PureP2P トラヒックにおける相互接続の近接性、電子情報通信学会総合大会 B-6-21、p.21（2007 年 3 月）

[発表 18] Jianping Li, Yasushi Wakahara : Formulation of Optimal Time Slot Assignment for Ad Hoc Networks with Directed Links：電子情報通信学会総合大会B-21-58, p.647（2007 年 3 月）

[発表 19] Toshinori Ohtsuka, Fumitaka Nakamura, Yuji Sekiya, Yasushi Wakahara : Detecting and filtering method for IP spoofed packets by making use of topology information、電子情報通信学会総合大会 BS-10-8、pp.S-107-S-108（2007 年 3 月）

[発表 20] 山登 庸次、木村 正二、若原 恭：ネットワークソフトウェア技術の研究動向、電子情報通信学会総合大会 B-19-1、p.563（2007 年 3 月）

広域分散環境の高度基盤技術に関する研究

中山雅哉

1 概要

Internet に代表される広域分散環境は、今日の社会生活に不可欠な存在となってきた。最近では、計算機だけでなく気象センサーや振動センサーなどの様々なデバイスがネットワークに接続される様になり、これらの小型ノードから得られる情報が社会生活に対する安心情報として提供される様にもなってきている。しかし、ネットワークに接続されるノードの規模が爆発的に拡大するにつれて、各ノードから得られた情報をデータ統合ノードに集約する際の中継ネットワークの輻輳や遅延に関する問題や、データ統合ノードで得られた集約情報を広域に分散した多数の利用者に提供するための技術が必要不可欠な状況になっている。

また、無線ネットワークの利用が一般的になってきたため、どこでも誰でも簡便に近くのノード間でのグループコミュニケーションを行うことができるアドホックネットワーク技術の研究が盛んになってきている一方で、匿名性を有した通信手段の登場により、DoS やフレーミングなどの通信妨害も頻繁に発生する様になり、不審者の検知を早期に行う方法の開発も重要な課題となっている。

本研究室では、これらのネットワーク基盤技術の研究に取り組んでいる。

2 自律分散ノード間での効率的な情報流通技術に関する研究

2.1 背景

今日の社会生活に不可欠な存在である広域分散環境 (Internet) には、最近は計算機だけでなく気象センサーや振動センサーなどの様々なデバイスが接続される様になってきた。

これらの小型センサーは、自律分散ノードとして観測されたデータを集約ノードなどに常時送信する機能を持っており、集約ノードに集められた情報に基づいて、社会生活に対する安心情報としてネットワーク利用者に提供される様になってきている。

例えば、気象庁が行っている緊急地震速報は、振動センサーから得られる地震に関する情報を、ネットワークを介して利用者に提供する仕組みの一例として挙げることができる。

2.2 内容

ネットワークに接続される自律センサノードの規模が爆発的に拡大すると、各ノードから得られた情報をデータ統合ノードに集約する際の中継ネットワークの輻輳や遅延に関する問題や、データ統合ノードで得られた集約情報を広域に分散した多数の利用者に提供するための技術が必要不可欠なものとなる。

各ノードの持つ有線ネットワークインタフェースの速度は、年々高速化しており 100Mbps 程度の速度は当り前になってきている。デスクトップ PC の多くや ノート型 PC でも、1Gbps の NIC が搭載されている機種が多くなっており、LAN 環境では、100Mbps や 1Gbps の速度で構成されることは一

般的になってきた。また、広域ネットワークでも、WDM 技術を用いた機器が一般化し始めたことで、ISP から加入組織に 100Mbps や 1Gbps での回線提供も行われる様になってきている。

しかし、いかに各ノード装置を広帯域なネットワーク環境で接続する環境が構築されていても、遠くに離れたノード間で TCP 方式に基づくデータ転送を行うと、帯域・遅延積の関係から、ネットワークの帯域を十分に使ったデータ転送を行うことができない問題が生じる。これは「long-fat-pipe 問題」と呼ばれている。

これに対して本研究室では、本来はエンドノード装置間で行われる一つのデータ転送を、中継ノードを用いて 2 つ以上のデータ転送に分割することで、ノード間の遅延を仮想的に短縮させる方法の提案を行っている [査読付 2, 発表 8]。

この方式は、広域に分散したノード同士でのデータ転送ほど効果が高く、LAN 環境での効果は低いものとなる。このため、JGN2 ネットワーク上で全国にノードを点在させた実験環境を構築して有効性を検証するための実験を行ったり、PlanetLab 環境を用いて、世界中にノードを配置した場合の実証実験を行ってきた。

また、各ノード間のデータ転送効率を検証するためには、各自律分散ノードでの時刻も揃えておく必要がある。このため、ネットワークを介した時刻同期方式に関する精度についても検証を行っており、高精度な SNTP サーバの構築とその効果に関しても研究を進めている [発表 2]。

一方、集約ノードに集められた情報を分散した場所の利用者に提供する技術として、災害時の安否情報提供システムの構築とその効果に関する研究 [発表 3] や、皆既日食や金星の太陽面通過のインターネット中継 [査読付 1] などを通じて、情報提供ノードの負荷分散技術に関する研究を行っている。

2.3 具体的成果

本研究は、[査読付 2, 発表 8] で研究成果の発表を行っており、IC2006 において研究奨励賞 [受賞 1] を受賞ことができた。

また、この研究内容は、気象センサーや振動センサーなど多様な自律分散ノードでの応用が可能となる技術である。気象センサーによる情報共有の研究を中心としている LIVE E! プロジェクトは、全国に広がる研究者で構成されており、全国各地の観測データを共有することで、幅広く教育・研究での成果も得られている [招待 1, 査読付 3, 発表 4]。

また、JGN2 ネットワークの上に構成した実験ネットワークにおける実証実験に対して、[受賞 2, 受賞 3] などの利用促進に関わる賞を受賞することができた。

3 安全で安定した社会基盤を構築するネットワーク技術の研究

3.1 背景

最近は、無線ネットワークの利用が一般的になっており、ノート PC をはじめとして、携帯型の端末を持ち歩きながら、どこからでもネットワークに接続して利用できる環境が都市部では当たり前になってきた。しかし、無線 LAN の基地局がカバーする範囲は狭いため、移動しながら通信ができるネットワーク環境がどこでも構築される状況には至っていない。

そこで、近くの無線ノード同士がお互いに他のノードの通信を中継し合うことで、簡便に近くのノード間でグループコミュニケーションを行うことができるアドホックネットワーク技術の研究が盛んに行われるようになってきた。しかし、グループ内の通信といえども他者の通信の中継を通常のノードが担うことから、不審者による通信妨害の検知や回避は不可欠であり、その解決は重要な課題となっている。

また、無線アドホックネットワークに限らず、匿名性を有した通信方法の普及にともなって、DoS やフレーミングなどの通信妨害も多く発生する様になってきている。社会基盤として安全で安定したネットワークを構成するための技術開発が広く望まれている。

3.2 内容

本研究室では、安定した社会基盤を構築するネットワーク技術として、以下の内容の研究を行っている。

アドホックネットワーク技術のうち、通信の中継を行うノード間で協調制御を行うことができるメッシュネットワークにおいて、同期式 MAC 制御を用いて隣接ノードが扱っているデータフローの情報を集めて、各々のデータフローの効率を自律的に制御する方式を提案している [発表 5]。この方式では、他のフローとの干渉関係を考慮してタイムスロットを割り当てるため、従来の制御方式と比較して高いデータ転送効率を達成することができる。

続いて、安心した社会基盤を構築するネットワーク技術として次の 3 つの内容の研究を行っている。

1 つめは、近くの無線ノード同士がお互いに通信制御を行って、他のノードの通信を中継するアドホックネットワークにおいては、他者の通信の中継を通常のノードが担うことから、不審者による通信妨害が発生する問題への対策に関する研究である。この問題を解決するために、各ノードが自身が中継する通信をモニタする IDS の機能を持たせ、階層的な通信監視機構を動的に構成することで不審ノードを早期に検知する方法の提案を行っている [発表 7]。

2 つめは、匿名性に起因する DoS 攻撃のうち、正常に動作する一般の DNS サーバなどに被害者のアドレスを送信元と偽装してパケットを送信するリフレクタ型 DDoS 攻撃では、従来の traceback 方式では、最初の攻撃元を被害者側で直接調べることができない問題への対策に関する研究である。この問題を解決するために、ネットワーク上を通過するパケットから確率的に ICMP メッセージを生成することで、被害者側でリフレクタの協力なしに DDoS の攻撃元を検出することが可能となる方法の提案を行っている [発表 6] 。

3 つめは、匿名性の高い BBS や SNS などにおいては、実社会における対面コミュニケーションに比べてフレーミングなどの通信妨害が多く発生する問題への対策に関する研究である。既存の BBS や SNS では、コミュニケーション相手の発話の意図や背景環境 (文脈) が不明なために生じる問題であり、これを提供する支援システムの構築 [発表 1] とアンケートによる評価を行った。

3.3 具体的成果

本研究室で行っている、安全で安定した社会基盤を構築するネットワーク技術の研究に関しては、それぞれ研究会や全国大会などでの対外発表を行っている段階である。

4 成果要覧

招待講演／招待論文

[招待 1] Masaya Nakayama, Satoshi Matsuura, Hiroshi Esaki, Hideki Sunahara: ”Live E! Project: sensing the Earth”, Proc. of AINTEC 2006, LNCS 4311, pp.61-74, 2006.

受賞関連

[受賞 1] 小林弘和、中山雅哉: 研究奨励賞, インターネットコンファレンス 2006 実行委員会, 2006/10/24.

[受賞 2] JB プロジェクト 東京大学: 利用促進賞 (地域貢献優秀賞), 次世代高度ネットワーク推進会議幹事会, 2007/1/17.

[受賞 3] 地域間廣帯域コンテンツ流通基礎実験 東京大学: 利用促進賞 (地域貢献優秀賞), 次世代高度ネットワーク推進会議幹事会, 2007/1/17.

査読付論文リスト

[査読付 1] 荻原文恵、尾久土正己、富田晃彦、半田利弘、中山雅哉:「金星太陽面通過のインターネット中継とその教材化」, 地学教育, Vol.59, No.6, pp.201-212, 2006/12.

[査読付 2] 小林弘和、中山雅哉:「仮想リンクを利用した送受信ノード間の RTT 分割による転送スループットの向上」, インターネットコンファレンス 2006 論文集, pp.105-112, 2006/10/23-24.

[査読付 3] Satoshi Matsuura, Hiroki Ishizuka, Hideya Ochiai, Shinichi Doi, Shinichi Ishida, Masaya Nakayama, Hiroshi Esaki, Hideki Sunahara: "Live E! Project: establishment of infrastructure sharing environmental information", Proc. of SAINT2007, 2007/1/16.

その他の発表論文リスト

[発表 1] 湯浅晃:「文脈による電子コミュニケーション支援に関する研究」, 信学技報 Vol.106, No.52, SITE2006-4, pp.19-24, 2006/5/19.

[発表 2] 金子晋丈、中山雅哉:「オンデマンド時刻取得に基づく SNTP サーバの開発」, 第 13 回天網の会ワークショップ集録, pp.63-71, 2006/6/12.

[発表 3] 中山雅哉:「災害時安否確認システムの現状の課題と今後の展開-相互接続など」, FIT2006, 2006/9/6.

[発表 4] 江崎浩、砂原秀樹、相原玲二、中山雅哉、松本敏文:「Live E! プロジェクト – ディジタル百葉箱による自律的な気象データの共有 –」, 電子情報通信学会ソサイエティ大会論文集, BS-7-5, 2006/9/19.

[発表 5] 齋藤健太郎、中山雅哉:「無線メッシュネットワーク向け TDMA ベース帯域予約方式における、フロー間の干渉関係を考慮したスロット割り当て方式」, 信学技報 Vol.106, No.358, IN2006-104, pp.91-96, 2006/11/16-17.

[発表 6] 高田友則、中山雅哉:「リフレクタを用いた DDoS 攻撃への対策法」, 電子情報通信学会全国大会, B-7-122, 2007/03/20.

[発表 7] Muhammad Arifin Ritonga、中山雅哉:「モバイルアドホックネットワークにおける侵入検知システム」, 電子情報通信学会全国大会, B-7-207, 2007/03/20.

[発表 8] Hirokazu Kobayashi, Masaya Nakayama: "Improvement of Transfer Throughput with Split Connection and Discovering A Relay Node", 電子情報通信学会全国大会, BS-10-4, 2007/03/21.

ネットワーク運用に関する研究

加藤 朗

1 概要

コンピュータネットワークは，単に構築することではその価値は発揮されず，適切に運用されて初めてそれを利用した活動が可能にある．特に研究・教育ネットワークは商用ネットワークとは異なる側面がある．いまや 10Gbps 級のネットワークリンクは商用ネットワークでも日常的に使われているが，それを一人が占有するような使い方はされていない．そのため，このような研究・教育ネットワークの，通常とは多少異なる利用方法を考慮した運用やそれに関連するユーザサポートなどが必要である．ここでは，それらに関する以下の研究の概要について紹介する．

- **広帯域ネットワーク運用に関する研究**

  光技術を用いて保証された帯域を提供する lambda networking を題材に，関連ネットワークとの調整やパケットの振る舞いの微視的解析を通じて，その有効性を確認するとともにユーザのサポート方式について考察する．

- **研究・教育ネットワークにおける経路制御の運用に関する研究**

  近年比較的密に接続されてきている国際的な研究・教育ネットワークにおいて指摘されている最適ではない経路に関して，現状を分析し，障害発生時においても妥当な経路が得られる経路制御運用方式について考察する．

- DNS サーバ運用に関する研究

  DNS サーバの運用やそのトラフィック情報の収集・解析を通じて，現在用いられている Anycast の評価を行うとともに，より安定なサービス提供のための運用方式について考察する．

2 広帯域ネットワーク運用に関する研究

2.1 背景

広帯域ネットワークを用いて大容量のデータ転送を効率的に行うためには，帯域を単一アプリケーションによって占有する，いわゆる Lambda Networking が有効であると考えられている．特に大規模科学では，大量のデータの生成源が特定の場所に限定されており，その消費先が地理的に離れている場合が多い．通常のインターネットで用いられているトランスポートプロトコルのように公平性を重視するため，積極的な再送要求などによる転送性能向上はできないが，専用の帯域が確保されている場合には，その限りではない．また理想的には競合するトラフィックがないため，それらによる帯域の削減や中間ノードによるバッファ容量の問題などは発生しない．

しかし，転送ノード間に専用線を敷設できればいいが，費用などの問題は深刻である．特に，専用線が必要で，かつそれを十分に活用できることを示すことが必要であるため，いくつかの国際的な広

帯域ネットワークの運用を行っているグループは GLIF[1] – Global Lambda Integrated Facility – という活動に参加している．これらのネットワークの帯域は，主にその国の政府からの予算によって調達されているため，それらの活動が優先されて然るべきであるが，そのような活動がない場合，それを相互に接続することによって，短期間のデモンストレーションや実験を可能にすることに同意し，活動を行っている．

2.2 内容

GLIF では，広帯域ネットワークの相互接続地点での運用が非常に重要であると考えている．これらの相互接続地点は GOLE – GLIF Open Lambda Exchange – と呼ばれており，光 Path の物理的・電気的配線や機材の各種設定の変更を，デモンストレーションなどの要求に応じて変更する必要があることから，GOLE の運用者の協調は非常に重要になっている．そのため，各担当者は毎月（2006 年 10 月からは隔月）で電話会議を行い，可能な経路やタイミング等に関する調整を実施しており，筆者は東京に於ける実験的 Lambda Exchange である T-LEX[2] の運用に携わっており，可能な限り，座長役を含めこの電話会議に参加している．

また，GLIF では年に一回（Tech および制御プレーンに関してはその間に中間ミーティングを共同で開催しているため年二回），ワークショップを行い，現在の Lambda Networking の運用の状況の情報交換やそれを用いた研究成果の発表，さらに今後の活動の調整などを行っている．2006 年は NICT，WIDE Project，および NII の共催によって 9 月 11 日から 2 日間秋葉原のダイビルでこのワークショップが開催された．またこれに先立ち，NASA/ESNET の主催による，光ネットワークの現状および将来方針に関するまとめを行い，アメリカ政府等へ報告書を提出する目的のワークショップ ONT3[3] が NICT と NASA の共催により開催された．また，9 月 13 日には，日本の研究者・運用者に対して GLIF の活動を宣伝普及する目的で Open Lambda Symposium が開催された．筆者はこれらの一連の会合に関して，ローカルアレンジメントおよびプログラム委員として貢献し，我が国の光ネットワークの現状について講演を行った．

ところで，Lambda Network を用いた場合，データの発信元と宛先の間が完全に一本の光 Path で接続された場合には，低い確率で発生するビットエラーがネットワーク的には問題になるに過ぎない．実際には，送受を行うホスト間の往復伝搬遅延時間が米国宛で 100∼150ms，ヨーロッパでは 200∼250ms に達するため，送受ホスト間のフロー制御が効かないことによる問題が深刻であるが，これはアプリケーションユーザ側の問題であると考えることもできる．

しかし，GLIF のような枠組みを用いても，送受ホスト間に純粋な layer-1 を設定することは困難である．その理由は，特に我が国やアメリカ国内のネットワークでは，SONET/SDH をそのまま提供することは稀で，仮に帯域が占有できたとしても SONET/SDH ではなく 10GE（10Gigabit Ethernet）LANPHY を提供していることが多い．海底ケーブルでは SONET/SDH であるため，POS を用いるか，10GE WANPHY を使用する必要があるが，機器のコストなどから最近では後者が用いられることが多い．いずれにせよ，リンク速度が異なるものが途中に介在することになるため，Layer-2 スイッチを介在させて，これらのリンク速度の調整をしなければならない．ところがこの Layer-2 スイッチが，バッファ長の問題や，回線は独立だが同一スイッチを経由する，いわゆるクロストラフィックによって影響を受け，パケット損失やジッタの発生源になっていることが分かってきた．

10Gbit 級ネットワークによって効率的な長距離伝送を実施するためには，これらの機材のマクロな性能だけでは十分ではなく，微視的なパケットの処理の振る舞いが問題になってくる．装置に手をいれることは一般に困難であるため，光タップによって信号の一部を取り出し，観測することが必要である．10Gigabit でパケットの観測ができ，比較的正確な時間情報を付加することは，いわゆるネッ

[1] http://www.glif.is/
[2] http://www.t-lex.net/
[3] http://www.nren.nasa.gov/workshops/workshop9.html

トワークアナライザで実現可能であるが，価格面や遠隔操作などの点で容易な方法とは言えない．また定常的な現象を観測するには，バッファ容量の制約の範囲内で可能であるが，例えば一時間に一回程度しか現れないパケット損失を観測するのには適していない．観測可能な Lambda Exchange として Cisco Systems の University Research Program の資金提供が得られため，本学情報理工学研究科平木教授の研究室で開発している 10GE 用実験機材である TGNLE-1 を用い，本センター若原研究室の修士課程の学生である室田氏と共同でパケットの振る舞いの微視的観測を可能にするための開発を行った．

2.3 具体的成果

GLIF2006 等の会議のホストの一員として貢献したことは前述の通りである．また，GLIF/GOLE の枠組みを活用し，平木教授のグループによる Internet2 Land Speed Record[4] の記録更新に参加した．

LSR は広帯域ネットワーク上での長距離データ伝送の記録を認定するプログラムであり，2004 年までは U.S. の研究者がその記録を保持していたが，2005 年の暮れから 2006 年にかけて，平木教授のグループによりその記録が更新され，IPv4 の単一 TCP ストリームで 8.80Gbps，IPv6 では 6.96Gbps というのがその記録であった．なお，LSR は単一 TCP ストリーム部門の他に複数 TCP ストリーム部門もあるが，単一 TCP ストリームの記録は複数 TCP ストリーム部門の記録とも認定されるため，全ての LSR は平木教授のグループによって保持されていることになる．

2006 年秋に IPv6 の記録打破の可能性を平木教授と協議し，同 12 月の GOLE 電話会議で幾つかの可能性に関して，各国のネットワーク担当者と調整を行った．その結果，大西洋区間は SURFnet の，太平洋区間は JGN2 および IEEAF の OC-192c 回線を用いることにし，アメリカ大陸縦断は 10GE LANPHY で運用されている TransLight を用い，StarLight の Layer-2 スイッチのみは往復の経路で同時に使用されるが，それ以外は同一 Layer-2 スイッチは経由しない経路が使用可能になった．この経路のセットアップをそれぞれの GOLE の担当者に依頼し，関係する機器の設定を行った結果，12 月下旬には実験環境が整った．これによって，IPv6 の単一 TCP を用いた場合，7.68Gbps の記録が得られ，また翌日には送受ホストで mmap() を用いることによってオーバヘッドを軽減したプログラムを使用して 9.08Gbps の記録が得られた．これらは直ちに平木教授によって Internet2 LSR 委員会に申請され，2007 年 3 月に LSR として認定された．2007 年 4 月に Internet2 から，また 5 月に本学からプレスリリース予定である．

また，この LSR には間に合わなかったものの，TGNLE-1 の FPGA プログラムを書き換えることで 10GE のパケットの微視的観測が可能になった．これによって LSR で用いられたのと同一の機材で，T-LEX 折り返しの負荷試験を実施したところ，同一スイッチでもスイッチングを Layer-2 と Layer-3 で行う場合にそのパケットの送出パターンに大きな差があることが確認された．これに関してはさらなる追試が必要であるが，FPGA によって直接のパケットを観測し，到着間隔の分布や TCP/RTP のシーケンス番号を追跡することでパケット損失を検出し，その結果および問題の瞬間の前後のパケット到着の状況を遠隔で入手する機構が有用であることが確認された．

3 研究・教育ネットワークにおける経路制御の運用に関する研究

3.1 背景

前述の 10Gigabit 級でなくても，近年の研究・教育ネットワークの整備の進捗によって，従来は米国をハブとする単純な構成であったものが，近年特にアジア域内での相互接続が追加され，また TEIN-2[5] の運用開始により，帯域は 622Mbps と限られているものの，米国を経由せずともヨーロッパの研究・

[4] http://www.internet2.edu/lsr/
[5] http://www.tein2.net/

教育ネットワークに到達できるようになり，一般的に通信ホストが双方とも研究・教育ネットワークに接続されていれば，商用インターネットを経由するよりも帯域等の面で有利なことが多い．

しかしながら，研究・教育ネットワークは，主に利用できるのはその国の学生や大学院生を含む研究者に限られ，他の２国間のトラフィックの搬送は，ポリシとして制限しているケースが多い．また多くの研究・教育ネットワークでは帯域は限られているものの商用インターネットへのアクセスを持っている場合も多い．そのため，研究・教育ネットワークでの経路制御のポリシ設定は，商用ネットワークの peer 経路と customer 経路による区別以上に複雑なものになる．このため，正しい設定をしたと思っていても，例えば我が国から韓国の研究・教育ネットワークである KREONET に到達するために Singapore 経由になるなど，思わぬところで最適ではない経路になってしまう例が多く見受けられた．

上記問題に関してはトラフィック通過ポリシを厳格に設定することで一応は回避できるが，バックアップ経路を考えた場合には十分ではない．通過トラフィックは許容できない場合でも，リンクやルータの障害などの場合には backup としてなら通過を許容しているケースも多い．そのため，ルータに於ける経路広告および受理の設定は非常に複雑になる．

このように経路制御ポリシの策定とのその実装に関しては，最新の注意が必要であり，さらに研究・教育ネットワーク相互間の国際的な調整が不可欠になってきている．

3.2 内容

経路制御に関する調整作業は，それぞれの研究・教育ネットワーク独自の運用上の制約や通過ポリシに大きく依存するが，それらについて相互に調整する場がなかった．商用ネットワークに関しては，経路制御ポリシは主に経済的関係で決まり，また NANOG[6] や JANOG[7] のようなオペレータ間の情報交換のフォーラムが設けられているため，これに習い，Indiana 大学の Jin William 氏，Chris Robb 氏に依頼して RENOG メーリングリストを設定，また Internet2 メンバーミーティングの際に会合を持つことにした．また，アジア太平洋地域の研究・教育ネットワークは APAN として年に２回が開催されていたため，これを利用して 2006 年 7 月および 2007 年 1 月の APAN Meeting において，経路制御調整に関するワークショップを主催した．

それぞれの研究・教育ネットワークでの通過ポリシを策定し，それを交換する作業が必要なのは勿論であるが，最適ではない経路をユーザから指摘される前に検出し，オペレータに通知して対策を依頼することも必要である．そのため，APAN の主な研究・教育ネットワークの主要ルータから経路提供を受け，それらを記録解析することが提案された．また，アジア各国からヨーロッパとデータ転送を行う際，帯域を優先すれば米国経由になるが，遅延時間が短いことを優先する場合には帯域は限られているものの TEIN-2 経由の方が好ましい．これを実現するためには，商用インターネットでは好まれないプリフィックス長の長い経路を異なった経路で広告し，特定の範囲のアドレスのみ対して遅延時間の短い経路を採用するような方式，あるいは BGP コミュニティによって遠隔ルータにおける経路選択に優先度を設定する方式などが議論された．

2006 年 12 月 26 日の台湾沖地震では，殆んどの主要な海底ケーブルが損傷を受け，インターネットの接続性に関しても大きな影響を受けている．研究・教育ネットワークは，それ自身の接続性と商用インターネット経由の接続性の両方を利用できるため，商用インターネットに比べて接続性の確保では有利であり，またバックアップに関しては通過ポリシが比較的緩やかであることから，早期の回復が期待できたが，現実には必ずしもそうならなかった．これは，主に最適ではない経路になるのを防ぐため，比較的厳しい経路フィルタが設定されていたことが原因ではないかと考えられている．これに関しては，RouteView[8] 等で記録された経路情報に基づいて，地震発生の前後に於ける経路の変

[6] http://www.nanog.org/
[7] http://www.janog.gr.jp/
[8] http://www.routeviews.org/

化の状況の解析を行っている．

3.3 具体的成果

経路制御調整に関しては，現在 APAN JP のオペレータと経路記録サーバの運用方針に関して調整中であるが，経路制御プログラム zebra の後継である quagga の multiple view 機能を用いることで，各研究・教育ネットワークの主要ルータからの経路情報を IBGP で供給をうけることに関しては概ね問題がないことが確認された．しかし，複数の view を横断的に観察する機能がないため，quagga の修正を行っており，現在作業中である．

また，台湾地震前後の，利用可能な経路および実際に利用可能だった経路に関する解析に関しては，MRT 形式で記録された経路データから必要な部分の経路情報を抜き出し，その関係を図示するところまでは完成しているが，この時間的変動について図示すること，およびその当時に利用可能だった発信元と宛先を指定した場合の経路に関する解析に関しては，現在作業中である．

4 DNS サーバ運用に関する研究

4.1 背景

DNS は現在のインターネットにおいて名前解決を行い，資源を IP アドレスではなく名前で指定することができる枠組みを提供している非常に重要な基幹的サービスの一つである．Root DNS サーバや TLD DNS サーバなど，階層的に定義された名前空間の上位ノードに対応する名前解決を提供するサーバは，インターネットの実用に大きく関わっており，その安定運用は非常に重要である．

特に DNS では，通常用いられている UDP に於ける DNS のメッセージ長の制限によって，あるノードに対応したサーバ数には上限がある．これを回避する方法として，局所的には負荷分散器を用いて複数のサーバでサービスを提供することにより，サーバのサービス容量を改善することができ，また少数のサーバの障害や保守による影響を抑制することができる．さらに Anycast，つまり同一サービスアドレスによって同一サービスを提供するサーバを複数インターネットのトポロジ上の異なった地点に配置し，経路制御が最寄りと考えるサーバに問合わせパケットを誘導する方法が活用されている．

4.2 内容

Anycast を用いた場合，個々のサーバのサービス品質の確認は，外部からは経路制御によって決定されたサーバにパケットが誘導されるため，容易ではない．そのため，サーバクラスタ内部に監視用の PC を設置し，適当な間隔でサービス品質を監視，異常時にはメール等で通報する方式の運用を 2005 年より行っており，ルータの意図しない障害時に経路が他のサーバに向けて安定するまでの期間に発生する短いサービス中断を除けば，非常に高いサービス品質が得られている．

Anycast を用いた場合の一つの問題として，TCP によるサービスがある．通常 DNS のトランザクションは単一の UDP による問合わせメッセージと，それに対する単一の UDP による応答メッセージから構成される．そのため，サーバがインターネット上に複数存在し，経路が変動したとしても，サーバとクライアントの間に双方向の接続性が確保されている限り問題は発生しない．しかし，もし TCP を用いたとすると，単一のメッセージのやり取りでは収まらず，一トランザクション中に，クライアント側からサーバに対して，複数のパケットが送られる．このトランザクション中に経路変動があった場合，あるいは ISP が回線に対する主に負荷分散を目的としてパケット単位での負荷配分を行った場合，全てのパケットが同一サーバに送られるとは限らない．つまり，Anycast を用いることによって TCP によるサービス提供ができなくなる可能性がある．

Web などのセッションが比較的長時間続く可能性がある場合には，最初のトランザクションによって Anycast によらない実アドレスへの redirect を発行し，実際のデータ転送は経路変動に影響しない実アドレスを利用するという方法が多用されている．この場合も，redirect を受け取るまでの間の

経路変動によって問題が発生するが，DNS のような基幹サービスではないため，その影響は限定的である．

DNS OARC[9] – Operation, Analysis, and Research Center – では，Days in the life プロジェクトとして，幾つかの DNS サーバについて，共通な一定の時間トラフィック情報を収集し解析を行っている．今年度は 2007 年 1 月の初旬の共通な 48 時間について， 13 ある Root DNS サーバのうち，筆者らが運用する M-Root DNS サーバを含む 5 つが参加した．M-Root DNS サーバでは，東京（3 個所），Seoul，Paris，及び San Francisco で稼働している全てのクラスタについて，データ収集に参加した．収集プログラムに若干のミスがあり，また一部スイッチにおけるメモリ問題で完全なデータを得ることはできなかったが，概ね十分なデータ収集を行った．その結果，約 1TB（圧縮した状態で 240GB）のデータを得ることができた．

4.3 具体的成果

収集データは非常に膨大であり，いろいろな角度からの解析が必要である．現在まだ解析作業中であるが，収集したデータのうち，TCP のパケットは約 7GB 分あり，これに対して解析を行った結果を概説する．

注目するべきことは，TCP のセッションが確立した直後，通常ならば問合わせメッセージが到着するが，問合わせを受信することなく直ちにセッションの解放を行うトランザクションが非常に多く，セッション数の 96.5% このタイプで占められていることが分かった．DNS サーバに対する計測・監視活動やポートスキャンなどが含まれていると想定されるが，このようなトランザクションを発生したホスト数は TCP による問合わせを行ったホスト数の 90% に達しており，DNS の実装上の問題であることが懸念される．

また，この 48 時間の先頭 32 時間分に関して解析したところ，TCP を送出したホストが合計 33.6 万あったのに対して，M-Root サーバの 6 個所あるクラスタのうち 2 個所で観測されたものは 3,700 に過ぎず，また 3 個所で観測されたものはわずか 72 ホストに留まっていた．これに関しては，RouteView などで記録されている経路制御の変動と付合わせることによってより詳細な分析ができるものと考えられている．

また，Anycast による運用は DDoS が発生した場合，その影響を一部のクラスタに限定することができ，インターネット全体に対するサービスに対する影響を一定以下に抑制できることが期待されていた．セキュリティ上の問題なため詳細を記すことはできないが，2007 年 2 月に発生した Root や一部の TLD DNS サーバを対象とした DDoS 攻撃では，回線の輻輳などによる応答性能の低下は 6 つのクラスタのうち 2 つに限定された．また商用ネットワークのオペレータとの連携で，比較的早い時期にフィルタの適用が可能だったため，インターネット全体に対するサービスの低下は限定的なものに止めることができた．

査読付論文リスト

[査読付 1] 小原泰弘・今泉英明・加藤朗・中村修・村井純：広範なトラフィック要求に対応する負荷分散経路計算アルゴリズム，情報処理学会論文誌，Vol.48，No.4，pp.1627-1640，2007 年 4 月

その他の発表論文リスト

[発表 1] Akira Kato：International Collaboration on Network Experiments，21st Century COE Workshop，Department of Frontier Informatics，Graduate School of Frontier Sciences，Kashiwa，April 2006

[9] https://oarc.isc.org/

[発表 2] **加藤 朗：**Global Lambda Networks，Interop2006 Conference，**幕張**，2006 **年** 6 **月**

[発表 3] Akira Kato：GLIF and Planet Lab，KRNET2007，Seoul，Korea，July 2006

[発表 4] Akira Kato：T-LEX update，22nd APAN Meeting，Singapore，July 2006

[発表 5] Akira Kato：T-LEX，3rd Optical Network Testbeds Workshops，Tokyo，September 2006

[発表 6] **加藤 朗：国内での** Lambda Networking，Open GLIF Symposium，**東京**，2006 **年** 9 **月**

[発表 7] Akira Kato：T-LEX Update，23rd APAN Meeting，Manila，January 2007

[発表 8] Akira Kato：T-LEX Update，GLIF Interim Tech/Control Plane Meeting，Minneapolis，February 2007

IP ネットワークにおける通信解析制御技術

中村 文隆

1 概要

- トラヒック特徴量解析

  広帯域環境における高効率なパケット弁別技術としてトラヒックの低位層情報を用いる特徴量解析手法を提案し、実トラヒックによる評価を行った。

- IP 偽装パケット検知技術

  サービス拒否 (DoS) 攻撃に常用される IP 送信元偽装パケットについてリンクステート型ルーティングにおけるトポロジ情報を利用した高精度かつ広適用性のある偽装検知技術を提案し、シミュレーションによる評価および実装手法の提案検討を行った。

2 トラヒック特徴量解析

2.1 背景

IP ネットワークにおけるトラヒックは、エンドノードの高性能化とアプリケーションの多様化、またそれらに並行するアクセスラインのブロードバンド化が相まって広帯域化の一途を辿っている。アプリケーションの多様化は、LAN から通信キャリアに至る様々なレベルの管理ポリシとの干渉を増大させる要因でもあり、望ましくないアプリケーショントラヒックによる帯域圧迫の軽減や著作権侵害などの反社会的行為の抑制がますます重要な課題となってきている。

しかし、エンドノード自体の高性能化とバックボーンネットワークの広帯域化は、望ましくないトラヒックの判別作業にかかるパケット処理量を増大させる傾向にある。それに加え、アプリケーションレイヤにおける広義のオーバレイネットワークの普及によってこれまで主流となっている (送信アドレス, 受信アドレス, 送信ポート, 受信ポート, プロトコル) のいわゆる 5-tuples による単純なマッチングによる判別は精度が低下する傾向にあり、パケットの高位層の内容 (payload) をデコードする DPI(Deep Packet Inspection) による判別が多用されるようになってきている。

DPI はパケットの高位層までのデコードを伴うため、5-tuples による低位層の処理に比較して処理負荷を更に増大させる性質をもっている。また、パケットの高位層を機械的にとはいえ参照することは通信の秘匿性に抵触するとの議論もあり、倫理的にも問題をはらんでいる。

以上の背景から、、本研究では従来用いられてきた 5-tuples と同じ第 4 層までの情報を用いた効果的なトラヒック判別の技術の開発と実トラヒックによる検証を進めることとした。

2.2 内容

あるホストの IP アドレス h が送信、または受信アドレスに現れるパケットの列を観測時刻 $t_i^{(h)}$ との対として $\{(t_i^{(h)}, \pi_i^{(h)})\}$ とする。以下では特定のホストアドレス h についてのみ考えるものとして (h) は省略して記述する。なお π_i は実トラヒックを観測して得るものとし、$t_1 < \cdots < t_n < \cdots$ とする。

従来のトラヒック解析は到着時刻情報 t_i を参照し、パケットの到着過程として種々の点過程をアプリケーションごとに仮定して行うものが主流であるが、トラヒック特徴量解析は t_i については積分してしまい、$\{\pi_i\}$ に含まれる情報のみを用いてトラヒックの特徴を抽出することを基本とする。パケットの到着時刻情報は一般に、途中の中継ノード、経由回線、エンドノード負荷などの多くの影響を受ける情報であるが、一定時間内に収集したパケットの統計情報を時間について積分することで上記諸要因からの影響を軽減することが核となる考え方であり、これは中村が考案し、松田・中村が Winny、BitTorrent、eDonkey の代表的 P2P アプリケーションについて具体化を行った。

2.3 具体的成果

昨年度までの研究においては $\{\pi_i\}$ のうち、h を受信アドレスとする SYN/ACK パケットの部分列 $\{\pi_i(\cdot, h, syn = 1, ack = 1)\}$ を取り出し、送信アドレス部分に現れる新規アドレス数の部分列総数に対する割合を新規接続率として、新規接続率の高いものを P2P に由来するとして弁別する手法の提案評価を行ってきた。これは、ホスト h から行った TCP 接続のうち、一定時間内に新規に行う TCP 接続の率が高くなるものが PureP2P であるという認識であり、中心サーバをもたない PureP2P の場合、オーバレイネットワークを形成維持するために新規ノードの発見とそれへの接続試行が絶え間なく行われることが必須であるとの理由による。この知見は中村により整理され、電子情報通信学会 NS/IN/CS 研究会合同開催の研究会において招待講演として発表された [招待 1]。

今年度は新規接続率に加えて、任意の SYN/ACK パケットの部分列 $\{\pi_i(\cdot, \cdot, syn = 1, ack = 1)\}$ を用い、送信または受信アドレスに h が現れるものを対象とした相互接続率について検討した。相互接続は、任意にとった一定時間間隔 Δt 内の $\{\pi_i(\cdot, \cdot, syn = 1, ack = 1)\}$ のなかで、$\pi_i(h, a, syn = 1, ack = 1)$ と $\pi_i(a, h, syn = 1, ack = 1)$ とがいずれも現れるような接続先 a がホスト h と相互接続を行った、として定義される。(b_j, h) もしくは (h, b) のみしか現れないホスト b についてはホスト h とは Δt 内に相互接続を行わなかったとみなす。

相互接続率は、ホスト h について、相互接続をしたホスト集合 $A = \{a_i\}$ と相互接続をしなかったホスト集合 $B = \{b_i\}$ を求め、$m = |A|/(|A| + |B|)$ として定義される。上述の新規接続率と同様の理由により、PureP2P ではオーバレイネットワーク維持のために双方向の接続が頻繁に行われるため、他のアプリケーションと比較して PureP2P の相互接続率は高い値を示す傾向にある。ただし、これはノードの役割が対称的であるような PureP2P ネットワークに限定され、例えば高速回線ノードの多い Winny ネットワークを弁別するために有利な指標であることがわかっている [発表 3][発表 6][発表 7]。

これらの検証結果については東京大学情報基盤センターにおける実トラヒックを匿名化したデータを最大限に活用しており、トラヒック収集、匿名化処理も中村が担当した。

また、今年度は、昨年度行った新規接続率の研究を沖電気工業との共同出願として特許出願を行った [特許 1]。

3 IP 偽装パケット検知技術

3.1 背景

サービス拒否 (DoS) 攻撃はサイバーテロの代表的な手法の一つであり、近年ではウィルスや不正侵入によって攻撃元ノードを多数確保することで個々の攻撃元ノードの特定を困難にする分散サービス拒否 (DDoS) 攻撃が常態化しつつある。DDoS 攻撃は経路的に攻撃先に近づくほどトラヒック帯域の異常値として検出できる可能性が増してくる傾向にあるが、攻撃の手法によっては正常トラヒックとの見分けを困難にするための仕組みを備えるものもあり、また攻撃先近傍での検知・遮断はネットワーク全体としてみれば不要トラヒックの増大・放置につながるため、攻撃元近傍での効率的な検知・遮断手法が求められている。

こうした攻撃元近傍での検知・遮断手法の代表的なものには ingress filtering があり、具体的な実現

法としてルータにおいて正規パケットの満たすべきルール (ACL=Access Control List) をオペレータが手動で実装する方法や、ルーティング情報を逆に参照することで自動的に偽装パケット排除ルールを設定する RPF(Reverse Path Forwarding) が知られている。

しかし手動実装は初期導入ならびにネットワークトポロジの変更時に要求されるオペレータの人的負荷が高く、かつヒューマンエラーによる漏れパケットの危険性を常にはらんでいる。また RPF は経路が対称的であって、あるインタフェースから転送されるべきサブネットからの逆方向パケットが必ず当該インタフェースを通ることが保障されていなければ、正規のパケットについても誤って遮断してしまう危険性を内包している。

RPF は ingress filtering を自動化する点で評価すべきものがあるが、上述の対称性に関する問題点をクリアする必要がある。そこで、AS(Autonomous System) 内で標準的に使用されているリンクステート型のルーチングプロトコルのもつトポロジ情報を利用して非対称経路においても誤遮断の可能性を排除して適用可能な ingress filtering の手法について提案検討した。

3.2 内容

リンクステート型ルーチングプロトコルはネットワーク内のルータがネットワークトポロジに関するデータベースとして重みつきの有向グラフ $G(V, E)$ を各ルータが保持し、それをもとに宛先ノードへの最短経路を計算する方式である。RPF では各ルータにおけるインタフェースでの部分グラフ $G_r(V, E)$ しか持たないため、完全な経路情報を決定することが原理的にできない。そのため、非対称経路やマルチパスをもつトポロジにおいて誤遮断の可能性がある。

一方、OSPF や ISIS などのリンクステート型プロトコルにおいては完全なトポロジ情報 $G(V, E)$ をもつが、$G(V, E)$ から経路表を作る際にはグラフの各ノードの情報のうち δ_+ すなわちノードからの出側の情報のみを利用した経路表を作成し、保持している。

Ingress filtering に用いる正規パケット表を作成するためには OSPF などが利用していない入側の情報 δ_- を基にした逆経路表に相当するものを計算する必要がある。この計算は通常の経路計算と同様に Dijkstra のアルゴリズムで効率的に計算することができ、その計算量のオーダは通常の経路計算の計算量と同程度で実現できる。経路計算アルゴリズムのシミュレータへの実装にあたっては関谷と共に中村が指導・開発を行った。

3.3 具体的成果

本方式を現実のネットワークに適用する場合、本方式を実装したルータである FSN ルータをどの程度の割合でどのようにネットワーク内に配置するかが、偽装パケット検知効率を決定する重要なファクタとなる。比較対称として同じ自動実装方式である RPF をとり、FSN,RPF の対応ルータの全ルータ数に対する割合に関して偽装パケット検知効率の推移をシミュレーションにより算出し、効率の検討を行った [発表 1]。

その結果、RPF では配備率 ($\in [0, 1]$) についてほぼ線形に検知効率が上昇するのに対し、FSN では指数的に上昇し、少ない配備率で高い偽装パケット検知効率を上げることができることが明らかとなった [発表 2]。

これは、RPF では誤遮断を避けるためにはネットワーク内のエッジのみに配備することが必須であることによる。A 個の偽装パケット生成ノードが一様ランダムにネットワーク内に存在しているとして、N_R 箇所の RPF ルータ配備可能な位置への配備率を上げていった際、各 RPF ルータがカバーする偽装パケット生成ノードの全体に対する割合は平均して一定の率 $\delta = A/N_R$ になるため、n_R 台の RPF ルータを配備した際の配備率 $0 \leq \rho = n_R/N_R \leq 1$ に対して検知率 $D_R(\rho)$ は

$$D_R(\rho) = \frac{\delta \times n_R}{A} = \rho \tag{1}$$

となることによる。

一方、FSN は完全な経路情報から計算される正規パケット表をもつためネットワーク内の任意の位置に配備することができる。この場合、N_F 台の FSN ルータを最終的に配備するとして、1 台の FSN ルータが平均してカバーする偽装パケット生成ノード数を $c = A/N_F$、k 台の配備済み FSN ルータがカバーする偽装パケット生成ノード数を Q_k、$k+1$ 台目の FSN ルータのカバーする偽装パケット生成ノード c のうち $\beta \cdot Q_k$ が k 台目までと重複するものとすれば、

$$Q_{k+1} = c + (1-\beta) \cdot Q_k \tag{2}$$

なので、$Q_0 = 0$ から一般項は

$$Q_k = c \cdot \sum_{i=0}^{k} (1-\beta)^i = c \cdot \frac{1-(1-\beta)^k}{\beta} \tag{3}$$

で、$k = N_F$ のとき $Q_{N_F} = A$ なので

$$\beta \cdot \frac{A}{c} = 1 - (1-\beta)^{N_F} \tag{4}$$

から β を求めることができる。N_R が十分大きく、$(1-\beta)^R \approx 0$ の時には $\beta \approx c/A$ とできるので、検知率 $D_F = Q_k/A$ は

$$D_F(k) = \frac{Q_k}{A} = 1 - (1-\beta)^k \tag{5}$$

となり、FSN ルータの配備数 k に対して指数的に 1 に近づくことがわかる。以上の解析は中村が担当し、部分的に ICICT2007 国際会議で発表された [査読付 1]。

また、Unix 上での OSPF 実装である zebra に FSN ルータ機能を実装し、実際の OSPF ネットワーク内で理論通りの性能を得られることを検証した。この実装および評価は関谷と共に中村が指導し、電子情報通信学会 NS/IN 研究会において発表された [発表 5]。

4 成果要覧

招待講演／招待論文

[招待 1] 中村文隆, 松田崇, 若原恭, 田中良明: ［招待講演］トラヒック特徴量解析とアプリケーション弁別, 電子情報通信学会技術研究報告,Vol.106,No.236,pp.67-72,Sep. 2006

査読付論文リスト

[査読付 1] T. Ohtsuka, F. Nakmaura, Y. Sekiya and Y. Wakahara: PROPOSAL AND EFFICIENT IMPLEMENTATION OF DETECTING AND FILTERING METHOD FOR IP SPOOFED PACKETS, ICICT 2007 Proceedings, paper ID picict179, Jan. 2007

特許申請／取得

[特許 1] 松田 崇, 田中 良明, 大崎 淳, 飯塚 正, 加藤 圭, 千田 浩一, 中村 文隆, 若原 恭: P2P トラフィック監視制御装置及び方法, 特願 2006-221578,15 Aug. 2006

その他の発表論文リスト

[発表 1] 大塚俊範, 中村文隆, 関谷勇司, 若原恭: ネットワークトポロジ情報を用いた効果的な IP 偽装パケット検知手法, 電子情報通信学会技術研究報告,Vol.106,No.118,pp.79-84,Jun. 2006

[発表 2] 大塚俊範, 中村文隆, 関谷勇司, 若原恭: IP 偽装フィルタリング機能の効果的な実装法, 電子情報通信学会技術研究報告,Vol.106,No.236,pp.27-32,Sep. 2007

[発表 3] 松田崇, 中村文隆, 若原恭, 田中良明: 通信の相互性に着目した PureP2P トラヒック弁別手法, 電子情報通信学会技術研究報告,Vol.106,No.355,pp.61-66,Nov. 2007

[発表 4] 大塚俊範, 中村文隆, 関谷勇司, 若原恭: ISP ネットワークにおける IP 偽装パケット検知機能 FSN の実装, 電子情報通信学会技術研究報告,Vol.106,No.450,pp.115-120,Jan. 2007

[発表 5] 大塚俊範, 中村文隆, 関谷勇司, 若原恭: OSPF を利用した IP 偽装パケット検知機能 FSN の実現方法, 電子情報通信学会技術研究報告,Vol.106,No.577,pp.409-414,Mar. 2007

[発表 6] 松田崇, 中村文隆, 若原恭, 田中良明: 相互接続における順逆接続間隔を利用した P2P トラヒック弁別手法, 電子情報通信学会技術研究報告,Vol.106,No.577,pp.415-420,Mar. 2007

[発表 7] 松田崇, 中村文隆, 若原恭, 田中良明: Pure P2P トラヒックにおける相互接続の近接性, 電子情報通信学会総合大会,B-6-21, 名古屋,Mar. 2007

成果報告

中村 誠

1 概要

UTnet の支線部分にあたる研究科・学部ネットワークを、所属専攻と、専攻内の研究室あるいは研究グループという階層的な組織構成に対応した階層化された管理・運用体制のもと、研究や教育が円滑に行われるための情報インフラとして、効率的にネットワークの運用・管理業務を行う事が可能なシステムの設計および実装を行っている。また、公開講演会のような広報活動を、プレゼンテーション資料を十分な品質とリアルタイム性を損なわずにインターネット中継するシステムの構築を行っている。

2 UTnet 支線ネットワークの運用管理システム

2.1 背景

UTnet の支線にあたる研究科・学部ネットワークは、研究室のような数名から数十名で構成されるグループ単位で自主的、独立に管理されるネットワークが専攻単位で集約し、さらに研究科全体で集約するといった階層的な管理体制で運用されている。そういった体制のもと、研究教育活動が円滑に行える情報インフラとして、利便性を損なう事なくネットワークインフラを効率よく構成し、運用を行う事が可能なシステムの設計と実装を行っている。

2.2 内容

物理的に分散して配置された研究室などでも、研究データ、各種サーバなどの情報資源を容易に共有出来るように仮想 LAN を構成できるようネットワーク機器を配置し、高速な通信の需要に耐えうるよう、全情報コンセントのギガビット化と研究科ネットワークのコアと主要なエッジスイッチ間の 10 ギガビット化を行った。無線ネットワークも有線ネットワーク同様、セキュリティ証明書を利用した認証機能を利用し、仮想 LAN を実現している。ほぼ全域にわたってネットワークが仮想 LAN を用いて論理的に構成されているため、研究室のネットワーク管理者などにスイッチの管理情報などを提供するシステムがトラブル対処やセキュリティ事故への対応において不可欠である。そのため、スイッチのトラフィック統計情報や、スイッチが学習している MAC アドレスや ARP エントリや経路情報、無線ネットワークのアクセスポイントに接続しているクライアントなどを検索するためのネットワーク情報閲覧検索システムを開発、運用している。

2.3 具体的成果

大学院理学系研究科・理学部ネットワークにおいて、エッジスイッチを一部更新し、ネットワーク全体のギガビット化を達成し、主要エッジスイッチの 10 ギガビット化をさらに進めた。無線ネットワークについては、学会等で大量のユーザを収容するためのアクセスポイントを導入し、その際にも仮想 LAN により、研究科のユーザは研究室のネットワークを利用し、ビジターには専用のネットワークを提供できるよう実装した。ネットワーク情報閲覧検索システムにより、専攻や研究室のネットワーク管理者が

自主的にスイッチの管理情報にアクセスし、クライアントのネットワーク接続トラブルへの対応や、セキュリティ事故の際のクライアント追跡などに活用されている。

3　インターネット中継システム

3.1　背景

公開講演会など広報活動の一環として、会場となる大学から遠い地域のユーザがインターネット上でイベントに参加できるよう、リアルタイムにイベントの様子をインターネット中継する事が求められている。その際、プレゼンテーション資料を、可読性を保ったままリアルタイム性を損なわずに提供する事が、ユーザの理解度を高め、満足してもらうために重要となる。

3.2　内容

Web ページ上でスクリプトを駆使し講演者の進行と同期したプレゼンテーション資料の提示を行うシステムを開発し、インターネット中継システムによる動画のリアルタイム配信と組み合わせる事で、インターネット上でイベントに参加するユーザに、会場に足を運んだ場合と同等の体験を提供する。

3.3　具体的成果

理学系研究科の公開講演会において、本システムを用いたインターネット中継を実施した。動画のみによる中継に対して、より少ない帯域使用率で、ユーザにより多くの情報を提供する事が出来た。

DNS の安定運用に向けた計測手法確立のための研究

関谷 勇司

1 概要

インターネットは社会生活の情報インフラとして確立し、日常生活ならびに研究活動に不可欠なものとなっている。インターネットの重要性が増すにつれて、その信頼性や耐障害性といったものがますます必要とされている。

DNS はインターネットの基盤技術であり、DNS が正常に動作しない環境においては、Web やメール、インスタントメッセンジャーといった、日々日常的に使われているサービスを利用することができなくなる。つまり、DNS を安定運用することは、インターネットにおけるサービスを安定して提供するための必要条件である。

そこで本研究では、インターネットの多くのサービスを支える基盤技術である、Domain Name System(DNS) の安定運用を目指した計測手法の確立をめざす。

DNS は分散環境であるため、ある単一の DNS サーバを監視するだけでは、DNS の障害を特定することは難しい。例えばある名前が解決できない場合に、その名前が解決できないのは、特定の DNS サーバに問い合わせを行った場合だけなのか、それともその名前空間を保持する DNS サーバの問題なのか、切り分けを行う必要がある。この際も単一の DNS サーバのみを見ていたのでは、障害を切り分けることはできない。

そこで、常時から DNS の運用状態を、単一の DNS サーバのみではなく、DNS というひとつのシステムとして監視するための手法が必要となる。常時の運用状態が把握できていれば、障害時の切り分けもスムーズに行うことができる。

本研究では、DNS サーバの集まりをひとつのデータベースシステムとして、その到達性を分析するための手法を確立することを目的とする。

本年度は具体的に以下の 2 つの項目について活動を行った。

- DNS サーバの安定運用を目指した分散計測手法に関する研究
- 計測アプライアンスを用いたボットネット式遠隔計測システムの構築に関する研究

2 DNS サーバの安定運用を目指した分散計測手法に関する研究

本研究は、インターネットの重要な基盤サービスである DNS の安定運用を目指すための分散型計測手法を確立し、運用の一助とすることを目的とする。

2.1 背景

インターネットの基盤技術のひとつとして DNS があげられる。メールやウェブといったインターネットにて一般的に利用されているサービスは、DNS が正常に動作しなければ通信ノードを特定することができず、通信を行うことができない。すなわち、DNS はインターネット上における通信をを支える

インフラとなっている。 インターネット上における多くのサービスを正常に運用するためには、DNS を正常に運用することが必要不可欠である。 つまり、DNS はインターネットという情報インフラを成立させている、インフラのためのインフラであると言える。

本研究では、この DNS の運用状況を複数の遠隔地点から計測し、運用の安定化を目指すための計測手法を確立することを目的とする。

2.2 内容

本研究は数年にわたって続けている研究であり、本年度は、前年度に確立した手法である、「ダイヤルアップゲリラ式分散計測手法」を用いたさらなる計測とその分析を行った。

2.2.1 ダイヤルアップによる計測手法

2006 年前半に集中的に行った dnsprobe による Root DNS サーバ群の計測結果を示す。dnsprobe[1] とは、計測ホストから DNS サーバ群までの RTT と DNS サーバ情報を計測するツールである。各種 UNIX 系 OS ならびに Windows にて動作する。

dnsprobe は、ホスト上にて動作させるソフトウェアであるため、基本的には、計測を行いたい拠点にホストが存在し、そのホストを利用できる権限あることが前提となる。

しかし、ある DNS サーバセットに対する世界中からの到達性と到達時間を計測するためには、世界中の多くの地点から dnsprobe を行う必要がある。すなわち、計測を行いたい地点に利用可能なホストが存在しない限り、計測は不可能となる。

この問題の解決策としては、dnsprobe をフリーソフトウェアとして公開することによって協力者を募り、数多くの地点からの結果を集めるという方法が考えられる。確かにこの方法にてある程度の協力者を得ることは可能である。しかし、必ずしも計測したい地点に協力者が現れるとは限らず、計測地点が偏ってしまう場合もある。

そこで本システムでは、測定するにあたって次の 2 種類の測定手法を採用した。

(1) native probe

(2) dialup probe

native probe とは、測定を行う地点にホストが存在し、そのホストにて dnsprobe を実行する手法である。 一方、dialup probe とは、測定を行う地点に存在する Internet Service Provider(ISP) のアクセスポイントにダイヤルアップ PPP 接続を行い、ダイヤルアップ接続したホストにて dnsprobe を実行する手法である。それぞれの実行形態を図 1 にまとめる。

dialup probe を利用することによって、利用できるホストが存在しない計測地点からも、手軽に計測を行うことが可能となる。しかし、native probe と dialup probe は明らかに計測手法が異なるため、dialup probe にて計測した結果と native probe にて計測した結果をそのまま比較することはできない。

多地点からの計測結果を基に、DNS サーバ群への到達生を分析するためには、native probe と dialup probe の計測結果を同様に扱えるように補正することが必要となる。次に、その補正方法について述べる。

2.2.2 測定値の補正

dialup probe にて測定したデータと native probe にて測定したデータを同列に扱うためには、dialup probe にて収集したデータを補正する必要がある。

このため、同地点から両方の方法にて計測したデータをもとに、測定値を補正する際の補正方法について述べる。

[1] http://mawi.wide.ad.jp/mawi/dnsprobe/

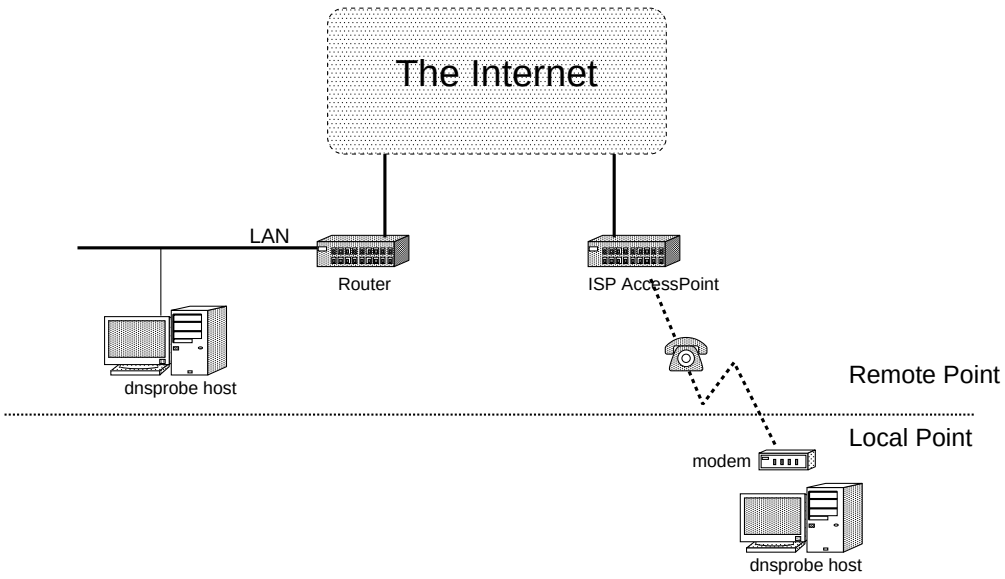


図 1: native probe と dialup probe

まず、native probe と dialup probe 結果を比較するために、米国 California 州 Los Angeles 市のデータセンタ内に、LAN によるインターネットへの接続性を有しており、かつモデムによるダイヤルアップ接続も受け付けているホストを用意した。 ルート DNS サーバ群に対して、このホストから native probe による測定と dialup probe による測定を行った。その結果を図 2 に示す。またこの結果を累積分布関数 (CDF) を用いて示した結果を、図 3 に示す。縦軸は累積分布を示し、横軸は応答が得られるまでの RTT を示す。なお、dialup probe は、神奈川県から V.34 方式をサポートしたアナログ回線モデムを用いて行った。

ルート DNS サーバ群は、a.root-servers.net から m.root-servers.net までの 13 個のサーバによって運用されている。このグラフでは、観測地点である Los Angeles 市のデータセンターから、それぞれ 13 種類の ルート DNS サーバまでの応答時間を示している。なお、これ以降本報告書にて示されるグラフにおいて、A から M までのアルファベットはそれぞれ a.root-servers.net から m.root-servers.net までのルート DNS サーバに対応するものとする。

この測定結果にて、ルート DNS サーバ n に対する dialup probe の RTT を α_n, native probe の RTT を β_n とすると、native probe と dialup probe の差 γ_n は、 $\gamma_n = \alpha_n - \beta_n$ となる。この結果にて導かれる γ_n は図 2 に示すように、ほぼ一定の値を示す。

このように、ダイヤルアップを用いて局所的にかつ集中的に計測し、その計測値を補正して利用する計測手法をダイヤルアップゲリラ式計測手法と定義する。

2.2.3 補正値

dialup probe にて行った計測結果を補正するにあたって、補正値 γ をどのように決定するかという問題がある。本研究では、補正方法として、以下の 2 つの方法を用いた。

(1) dnsprobe 実行ホストから、ダイヤルアップサーバまでの RTT を補正値 γ とする

(2) 基準 DNS サーバ群を定義し、dnsprobe を実行したホストから最も近い基準 DNS サーバへの RTT を補正値 γ とする

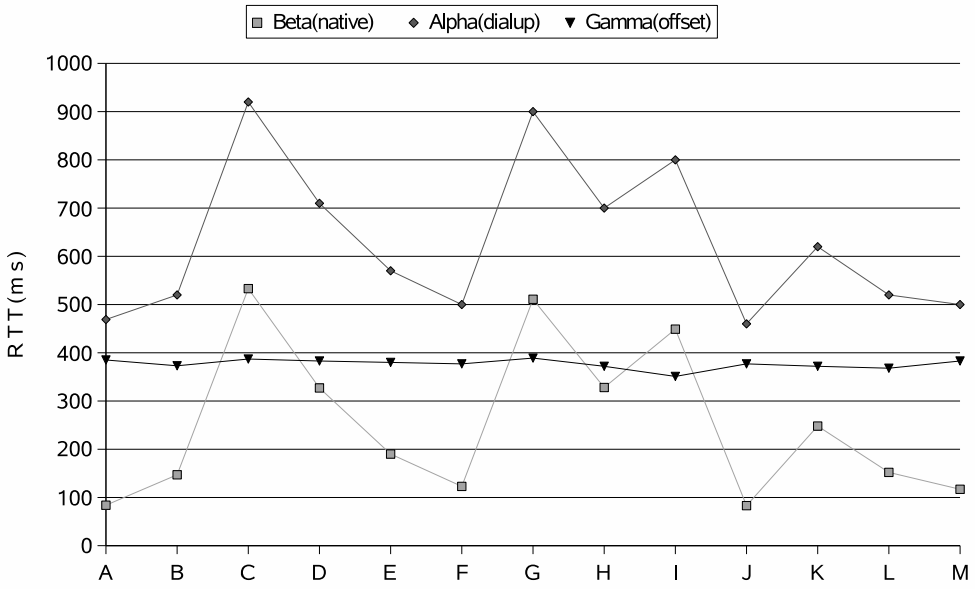


図 2: native probe と dialup probe の比較 (Los Angeles 市データセンタ内)

(1) の方法は、ISP のダイヤルアップサーバが設置された地点から、測定対象となる DNS サーバセットまでの RTT 値を算出することができる。すなわち、dialup probe において、ダイヤルアップ区間によって生じる RTT の差のみを除くための補正方法である。dialup probe に対してこの補正方法を用いると、ISP のダイヤルアップサーバがある地点において native probe を行った場合に近似した値が得られる。

しかし、ISP のダイヤルアップサーバは、ISP の中核となるバックボーンから数ホップ離れた位置に存在している場合がある。つまり、ISP 内部のネットワークトポロジは ISP 毎に異なるため、ISP のバックボーンからダイヤルアップサーバまでのネットワーク的な距離は同一条件であるとは言えない。

そこで、計測地点によるネットワーク構成の差異を吸収するため、基準 DNS サーバを用いて補正を行う方式を考案した。これが (2) の方法である。(2) の方法では、計測を行った地点から最もネットワーク的に近くにある基準 DNS サーバまでの RTT を補正値として用いる。

すなわち、図 4 に示すように、dnsprobe を行うホストからダイヤルアップサーバまでの RTT を RTT(a)、dnsprobe を行うホストから最も近い基準 DNS サーバまでの RTT を RTT(b) とすると、(1) の方法では補正値は $\gamma = median(RTT(a))$ で、(2) の方法では、$\gamma = median(RTT(b))$ である。

なお、この場合の RTT(a) の母集団は、dnsprobe にて測定を行う度にダイヤルアップサーバまで ping を行い、その RTT を記録したものである。RTT(b) の母集団は、最も近い基準 DNS サーバまでの dnsprobe の応答結果である。 また、補正値として中間値を用いているのは、突発的なネットワークの不安定による極端な値の振れの影響を除くために、平均値ではなく中間値を用いることとした。

つまり、DNS のパフォーマンスを測定するにあたって、同一測定地点からの複数の DNS サーバ群に対する測定結果を比較する場合や、測定地点からある DNS サーバ群までの RTT、すなわち応答が得られるまでの時間を調査したい場合には、補正法 (1) を利用すればよい。

一方、補正法 (2) は、複数の測定地点から、単一の DNS サーバ群に対しての測定結果を比較したい場合に用いる補正である。すなわち、ダイヤルアップ先の ISP 内部の構成の差異を補正するための方法である。また、ダイヤルアップに限らず、異なる測定地点からの測定結果を比較する際に、測定地点のネットワークの状況による測定結果の揺らぎを減らしたい場合に有効な手法である。

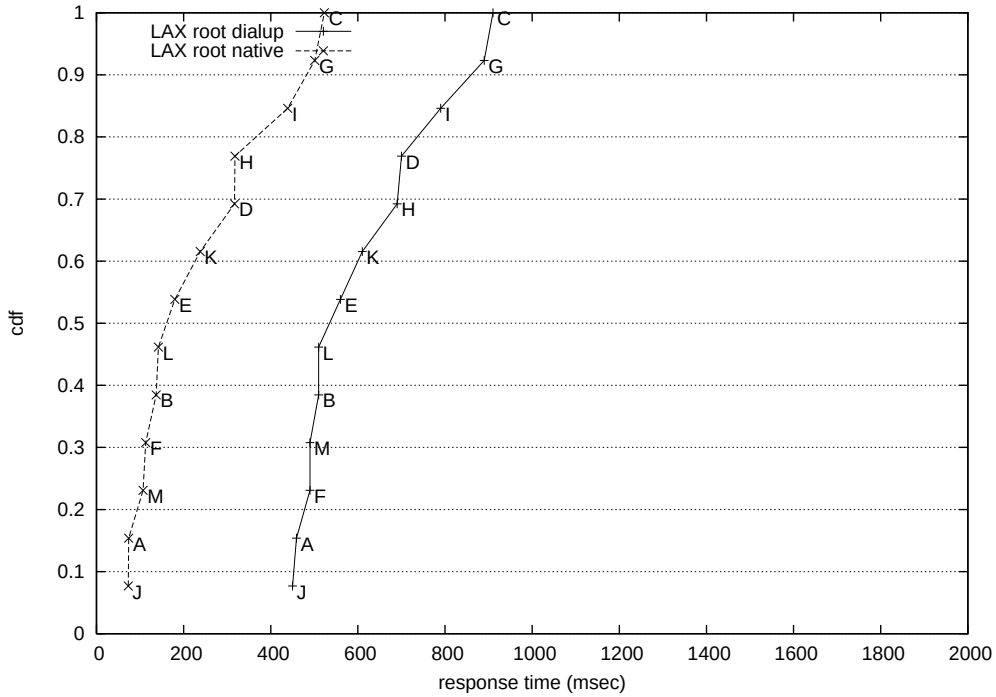


図 3: native probe と dialup probe の比較 (Los Angeles 市データセンタ内) - CDF

2.2.4 計測方法

今回の計測は、主にアジアの国々を中心に、2006 年前半に集中して行なった。前述したダイヤルアップゲリラ式計測手法を用いて、アジア各国の ISP のアクセスポイントに国際電話をかけ、計測を行った。なお、ISP は GoRemote ローミングに参加している、各国の現地の ISP を利用した。

計測は、ダイヤルアップゲリラ式に基づき、以下の条件にて行った。

- 国際電話を使ったアナログモデムによるダイヤルアップ接続
- 1 回のダイヤルアップ時間は約 30 分とする
- 1 つの国に対して複数回の試行を行う
- 計測ターゲットは Root DNS サーバ群とする
- 計測値の補正は補正法 (1) を用いる

2.2.5 計測地点

今回の計測では、主にアジア各国を計測地点に選んだ。計測を行った国ならびに都市と回数を表 1 に示す。

これ以外にも、次に示す国に対してダイアルアップを試みた。しかし、国際電話によるダイヤルアップがうまくつながらず、有効な計測結果が得られなかった。

- モンゴル
- カンボジア
- バングラディシュ
- パキスタン

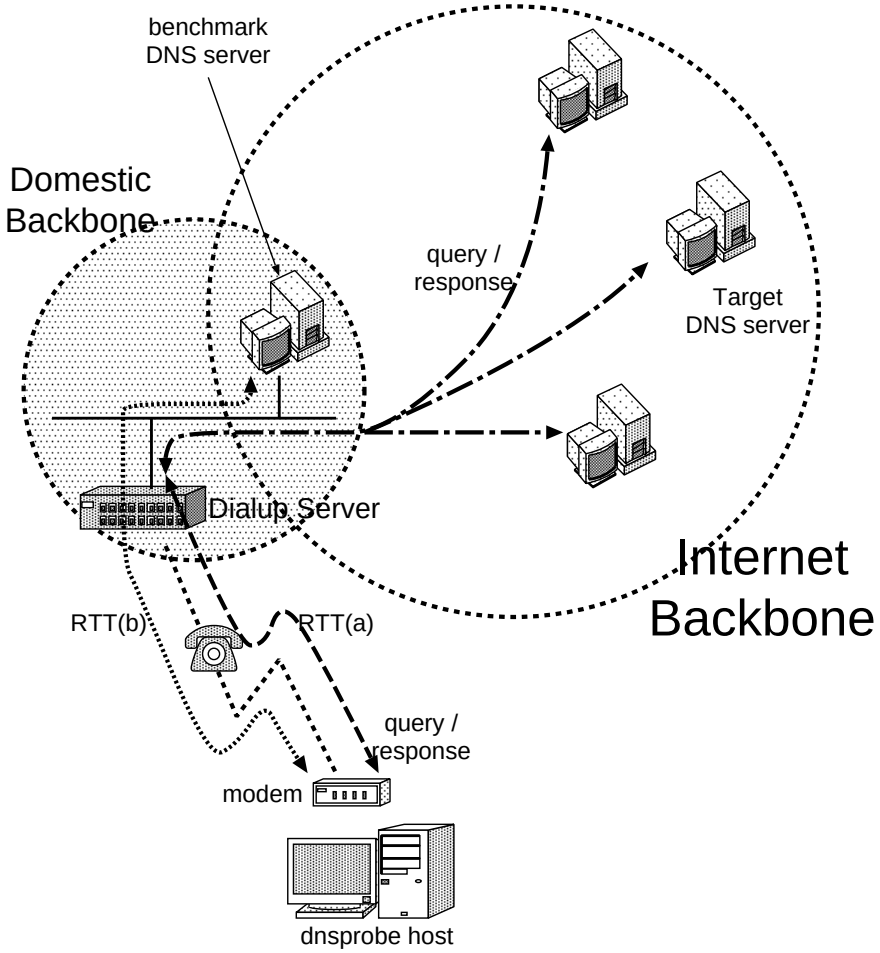


図 4: Dialup probe の補正方法

- ネパール

2.3 具体的成果

表 1 にあげた各国からの Root DNS サーバに対する計測結果を図 5 から図 13 に示す。各グラフは、横軸に Root DNS サーバまでの RTT を示し、1 つの CDF グラフで 1 回のダイヤルアップ結果を示す。また、CDF グラフ上の A から M までのアルファベットで Root DNS サーバの種類を示す。なお、1 回のダイヤルアップにて行った複数回の dnsprobe における中間値をグラフの値として用いた。

今回の計測は、アジア各国を中心に行った。その結果、有効な結果が得られた国のうち、スリランカとインドを除けば、最も RTT の小さい Root DNS サーバは 100ms 以下で応答することがわかった。これは予想よりも良好な結果であると言える。特に、シンガポールや香港、中国、韓国の結果から見てとれるように、エニーキャストによる Root DNS サーバの運用が有効に機能していることがわかる。

表 1: 計測地点

国名	ダイヤルアップした都市の数	ダイヤルアップ回数
中国	5	6
香港	1	2
インド	2	3
インドネシア	4	4
韓国	2	4
マレーシア	3	4
シンガポール	1	6
スリランカ	2	5
台湾	1	1
フィリピン	5	5

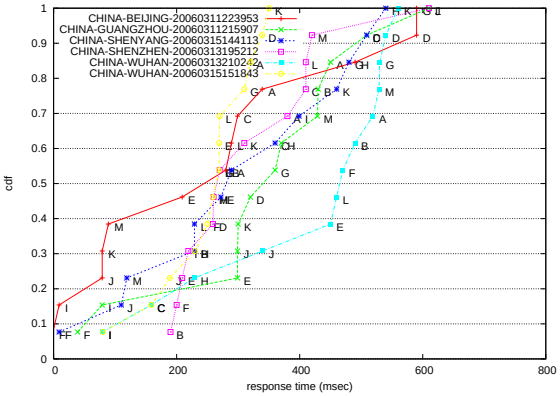


図 5: 中国からの計測結果

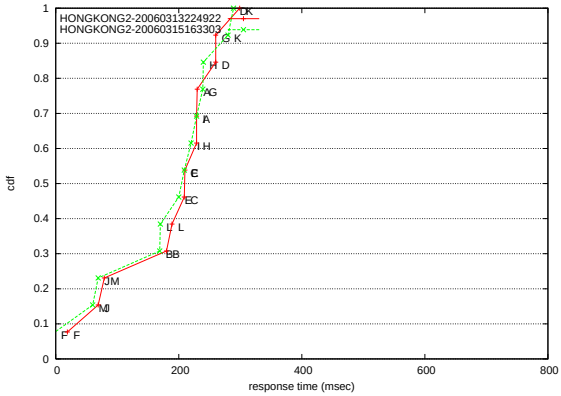


図 6: 香港からの計測結果

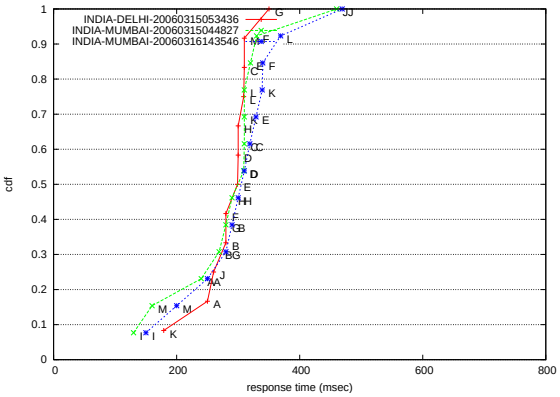


図 7: インドからの計測結果

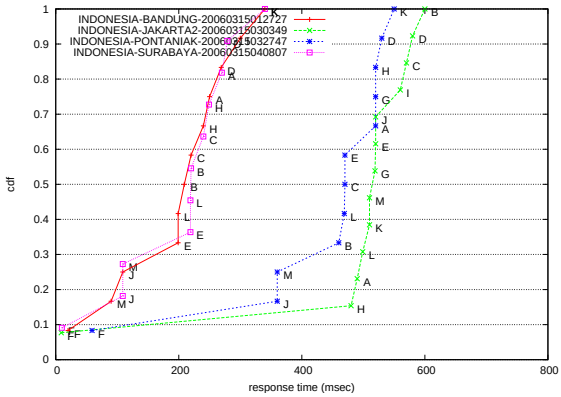


図 8: インドネシアからの計測結果

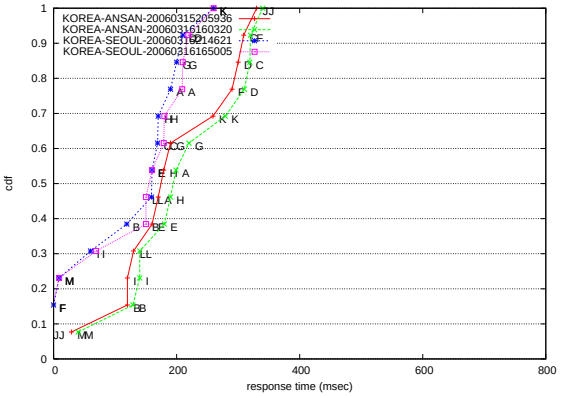


図 9: 韓国からの計測結果

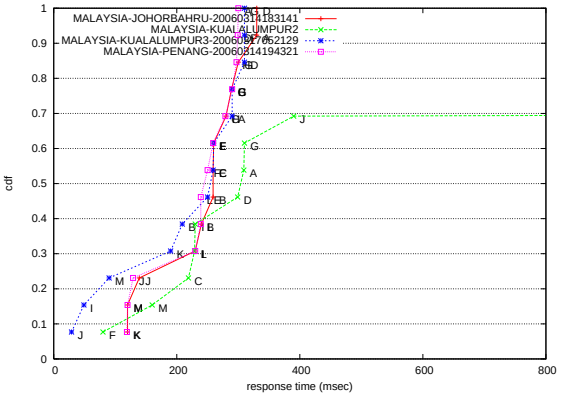


図 10: マレーシアからの計測結果

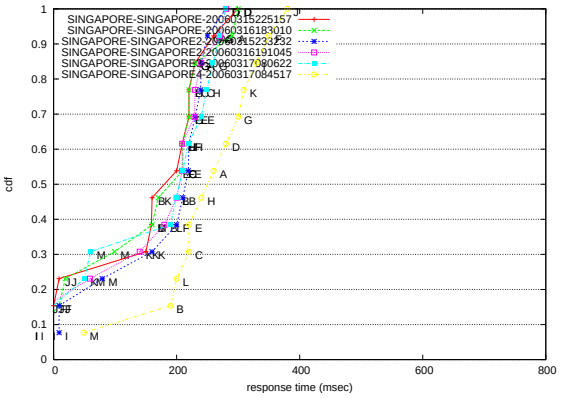


図 11: シンガポールからの計測結果

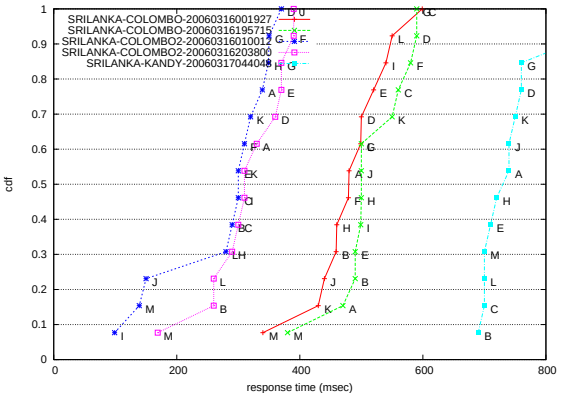


図 12: スリランカからの計測結果

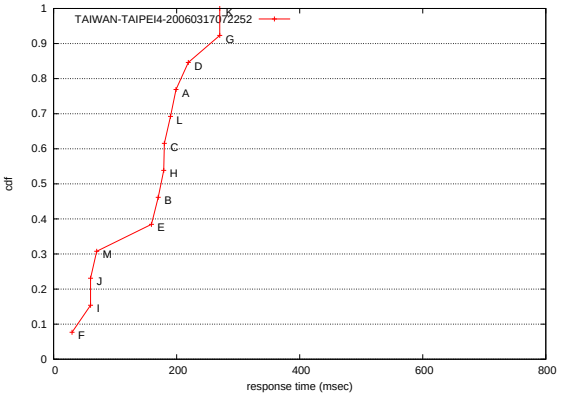


図 13: 台湾からの計測結果

3 計測アプライアンスを用いたボットネット式遠隔計測システムの構築に関する研究

3.1 背景

ダイヤルアップゲリラ式計測手法の弱点である、計測の継続性を補うために、アプライアンスを用いたボットネット式遠隔計測手法に関する研究とそのシステム構築を行った。

3.2 内容

ダイヤルアップを用いた計測には、以下の利点があげられる。

- 世界の複数地点からの計測を行うことができる
- 特別な機材を用いることなく、ISP へのダイヤルアップのみで行える
- 短時間で数多くの地点からの計測を行うことができる

逆に、以下の点が欠点としてあげられる

- 単一地点からの継続的な計測が困難である
- 同時に複数の地点から計測することが困難である
- 接続性が電話回線の状況に左右される

これらの欠点のうち、継続的な計測や同一時間に複数地点からの計測が行えないという点は、DNS の運用状態を分析するにあたって大きな制限となる。そこで、本研究で提唱したダイヤルアップゲリラ式測定手法に加え、定点から継続的に観測する手法を用いることで、さらなるデータの収集と、継続的変化の分析を行うことが可能となる。

しかし、定点からの計測を行うためには、機材を設置することが必要となる。世界中に機材を設置するにあたって、PC のような大型の機材を運搬もしくは送付し設置することは、金銭ならびに管理コストの面から見て現実的ではない。

その解決策として、Planet Lab を用いることを考えた。Planet Lab を用いることで、世界中に分散して存在している Planet Lab ノードから計測を行うことが可能となる。しかし、Planet Lab のノードの多くは、北米、EU 方面に集中しており、東南アジア、南米ならびにアフリカ方面への分布は少ない。DNS サーバの到達性と運用状態を分析するためには、インターネットコネクティビティが潤沢であるとは言えないこれらの地域からのデータが非常に重要となる。

そこで、本研究では Planet Lab の利用と同時に、計測専用のアプライアンスを作成し、これを世界に配布することで定点観測を行うこととした。

計測専用アプライアンスの製作にあたって、以下の点を目標とした。

1. 故障率の低い機材を用いる
2. 管理コストを削減する
3. 手軽に設置してもらえるような設計とする

条件 1 を満たすため、ハードディスクを搭載せず、フラッシュメモリのみで動作する機材を選び、そのファームウェアを改造することによって計測専用機器とした。

また、条件 2 を満たすために、すべての管理操作を遠隔から行えるような機構を組み込むこととした。具体的には XML-RPC と証明書による認証を組み合わせ、管理サーバからコマンドを発行する

ことによって計測の開始、停止ならびに機器の再起動からファームウェアアップグレードまで自動で行えるようソフトウェアを作成した。

さらに、条件 3 を満たすため、事前の設定を一切不要とし、DHCP によって IP アドレスを取得した後、管理サーバに自分の IP アドレスを通知して指示を待つ設計とした。計測の主導権はあくまで管理サーバ側にあり、計測専用アプライアンスはサーバの指示によって計測を行う設計とした。

3.3 具体的成果

これらの条件を満たした計測専用アプライアンスを作成し、世界各地への配布を行うべく、現在調整を行っている。2007 年 3 月現在、12 カ所への配布を行った。さらなる配布を目指し、2007 年度も活動を行う。

4 成果概要

受賞関連

[受賞 1] USAGI Project: 平成 18 年度情報通信月間総務大臣表彰受賞, http://www.soumu.go.jp/s-news/2006/060601_3.html, 2006 年 6 月

査読付論文リスト

[査読付 1] oshinori Ohtsuka, Fumitaka Nakamura, Yuji Sekiya and Yasushi Wakahara, “Proposal and Efficient Implementation of Detecting and Filtering Method for IP Spoofed Packets”, ICICT2007, Institute of Information and Communication Technology, BUET, Bangladesh, Mar. 2007

その他の発表論文リスト

[発表 1] 大塚俊範, 中村文隆, 関谷勇司, 若原恭: 「 OSPF を利用した IP 偽装パケット検知機能 FSN の実現方法」, 電子情報通信学会 NS 研究会, Vol.106, No.577, pp.115-120, 2007 年 3 月

[発表 2] Yuji Sekiya and Kenjiro Cho: “Remote Measurement System with Intelligent Management Framework”, RSSAC(DNS Root Server Advisory Committee) 27th, Prague, Czech Republic, http://www.icann.org/committees/dns-root/rssac-meetings/rssac27.pdf, Mar. 2007

[発表 3] 関谷勇司: 「Root DNS Servers における Anycast の現状」, インターネットウィーク 2006 DNS Day, パシフィコ横浜会議センター, http://www.nic.ad.jp/ja/materials/iw/2006/main/dnsday/index.html, 2006 年 12 月

[発表 4] Yuji Sekiya: “IPv6 Implementation in Linux”, NCHC IPv6 Workshop, National Center for High-Performance Computing, Hsinchu, Taiwan, http://www.nchc.org.tw/event/2006/ipv6/agenda_en.php, Nov. 2006

[発表 5] Yuji Sekiya: “Small Measurement Box Project”, 7th CAIDA-WIDE Workshop, San Diego Supercomputer Center, CA, U.S.A., http://www.caida.org/workshops/wide/0611/, Nov. 2006

[発表 6] Yuji Sekiya: “Large-scale Remote Measurement Box Management”, 2nd CNRS-WIDE Workship, CNRS, Paris, France, http://cnrs.wide.ad.jp/events-1/cnrs-inria-wide-meeting-20060918, Sep. 2006

公開ソフトウエア

[公開 1] USAGI Project: Linux における IPv6 プロトコルスタックの開発と公開, linux-kernel-2.6.11, linux-kernel-2.6.12, linux-kernel-2.6.13, linux-kernel-2.6.14, linux-kernel-2.6.15, linux-kernel-2.6.16, linux-kernel-2.6.17, linux-kernel-2.6.19, linux-kernel-2.6.20 における IPv6 プロトコルスタックのコード提供, http://www.kernel.org/, 2006 年 4 月〜2007 年 3 月

特記事項

[特記 1] NICT 受託研究: 「遠隔管理機構を利用した多地点からのリアルタイム DNS 運用計測システムの研究開発」, 慶應義塾大学との共同研究として受託, 2007 年 3 月

スーパーコンピューティング研究部門

スーパーコンピューターに基づいた 大規模数値計算に関する研究

金田　康正

1　概要

本研究報告は 2006 年度中におこなった、以下に示す 5 種類の研究テーマ（前 3 種類の研究テーマは主に研究室の大学院学生への研究指導としての研究テーマ）に関する研究概要報告である。

- $\pi(x)$ 計算の高速化
- 複雑な制御構造を持つプログラムの SIMD 命令セットによる最適化
- 4 倍精度積和演算の高速化
- pi.super-computing.org 上での π 並びに関連する情報公開
- スーパーコンピューターを用いた大規模高速計算

2　背景

本センタースーパーコンピューティング研究部門教員の研究における主たる役割として、利用者サービスに提供している現有のスーパーコンピューターの安定運用並びに、スーパーコンピューターの可用性向上の為の各種問題点の摘出、また利用者が将来要求する高性能計算需要をスーパーコンピューターの将来動向を見据えた上で安定的に導入・提供する事柄に関する情報収集や研究活動がある。

また教育における役割として、新領域創成科学研究科基盤情報学専攻の情報基盤センターに割り当てられている学生定員に対する兼担教員としての、同専攻修士・博士学生の教育指導も、前者に劣らず主要な役割となっている。ちなみに 2006 年度は修士 2 年 2 名、修士 1 年 2 名の教育指導をおこなった。なお、情報基盤センターに対する学生定員は、情報基盤センターになって初めて実現したということをここに記しておく。それまでは出身専攻の兼担として修士・博士課程の学生指導を引き受けることはあった。しかし毎年一定数の学生教育を引き受けるという状況ではなかった。学生定員が文部科学省定員として措置されてからは一定人数の学生が進学してくる様になり、中・長期的な研究テーマに取り組むことが可能となった。

研究室の指導方針として、研究テーマは各人の興味に基づき選択しているものであり、2005 年度以降の年報はそれまで発行していた年報の記述と変更になったが、それまでの年報の記載内容もそれぞれの教員の研究・教育活動報告であったことを考え、報告内容の継続性が重要であるということを考慮しつつ、これまでとは記載する研究テーマの順序を変更し、まずは研究指導の一環で研究室学生が得た成果を記載し、その次に学生が関係しない研究テーマの研究成果を記載するという流れて、2006 年度の研究・教育活動報告を以下に記す。

3 研究内容

3.1 $\pi(x)$ 計算の高速化

本研究は本研究室の修士 2 年の吉田仁君が修士課程の 2 年の間に行い、2007 年度に博士課程学生として私の研究室に進学した際におこなった研究である。彼の修士論文 [発表 1] の日本語アブストラクトには、

「本論文では, まず初めに素数個数関数 $\pi(x)$ についての解説と、それに関連するいくつかの関数、理論について述べている。次に、$\pi(x)$ の具体的な計算方法やアルゴリズムについても述べ、プログラムを実装するために十分な情報を記述している。 また、その中の 1 つのアルゴリズムである Gourdon の手法に基づいたプログラムを作成し、その計算時間に関連するパラメーターの挙動と並列性について実験をおこなった。その結果、c = 8.096 $\times$ 10^5、d = 1.864 とするパラメーターに従うと、$O(x^{0.805}/log^{6.64}x)$ で計算できることが分かった。 ここで、 パラメーター c と d は計算の範囲や時間に関わる別のパラメーターを調整する係数である。さらに、 並列計算を行うことで単独のプロセッサーでの計算の 11.4 倍まで計算速度が上昇することが判明した。以上の結果から、これまでの実験と同様に 8 プロセッサを使うと、 40 日前後の計算時間で未到達の値である $\pi(10^{23})$ を求めることができると予測される。」

と述べられている。具体的手法等についての報告は、修士論文そのものが英語で記載されていることもあり、2006 年度中に投稿論文の採録結果が届かなかったこともあるので、ここでは述べないこととするが、$\pi(x)$ は、一般に知られている円周率の値のことでは無く、x までの素数の数を与える $\pi(x)$ 関数であり、この関数の計算の高速化の観点から行っている、スーパーコンピューターの利用が必須となるこの研究である。これは引き続き博士課程で実施予定の大きな数の高速素因数分解の研究に向けての事前研究となっている。ちなみに大きな数の高速素因数分解は、$\pi(x)$ の計算とは比較にならない計算量の計算を必要とする計算であるが、現在までに行われている研究は、計算科学の専門家が行ってはおらず、数学研究を主とする研究者が行っており、計算科学の専門家が、計算機を使いこなす観点から、新しい計算アルゴリズムを発見し、実際にそのアルゴリズムに基づく高速計算プログラムを書くことが出来れば、大きな数の素因数分解の分野の研究が進展すると期待している。この修士課程での研究は、その為の事前トレーニングを兼ねた研究であった。

さて 2006 年度までに実施したこの研究でやり残した点は、今だ正確な値が求められていない $\pi(10^{23})$ を求める計算を実際に行うことであり、これは 2007 年度前半で完了させたいと考えている。もしこの計算が成功したならば、世界で初めて得た結果となるであろうことに間違いは無いが、問題は実際にこの計算を行う計算環境の準備である。この点に関しては、本学情報基盤センターの前進である大型計算機センターの時代において、教官やその教官の指導を受ける研究室の学生が研究で必要とする計算機リソースは潤沢に利用可能となっていた。しかし、法人後はその様な状況には無く、意欲を持つ学生にとっては実際に利用可能な計算機環境が身近にあるにも係わらず、研究経費がなければ計算が行えないという、実になさけない状況になってしまっている。この問題の改善無くして、スーパーコンピューターを利用する研究の進展、あるいは若い研究者の育成が順調に進むとは思えないということを記しておく。

3.2 複雑な制御構造を持つプログラムの SIMD 命令セットによる最適化

本研究は本研究室の修士 2 年で 2006 年度に修士課程を修了した廣松悠介君が修士課程 2 年の間におこなった研究である。彼が投稿して採録されている論文 [査読付 1] の日本語アブストラクトは、

「近年の汎用プロセッサーの多くは大量のデーターを扱うプログラムの高速化を目的として、SIMD（Single Instruction Multiple Data）命令セットという複数のデーターを一命令で演算可能な命令セットを搭載している。その SIMD 命令セットを利用してコンパイラーに最適化を行わせるための解析手法などが研究されているが、現状では限定的なコードを SIMD 並列化することが可能な程度である。SIMD 命令は全ての要素に対して同じ演算しか行うことしかできない。そのために、条件分岐やルー

プが含まれるコードを SIMD 命令で置き換えて正しい処理ができるようにするためには工夫が必要となる。本論文では、複雑な制御構造を持ったプログラムの SIMD 並列化を自動化するために、以下のような手法を提案する。

1. 分岐やループの制御構造や依存関係を単純化して扱うために、プログラムを解析して制御フローグラフを作成する。
2. 各ノードは依存関係順に SIMD 命令への置換を行う。
3. そして、グラフ構造上分岐やループに関する制御を扱っているノードには、正しい演算結果を取り出すための補正処理を追加する。

分岐やループを SIMD 並列化するためには余分な演算を追加しなければならないが、広範囲に渡るコードの SIMD 並列化によりプログラムの実行速度の向上を達成できる。この手法を COINS コンパイラーインフラストラクチャーで実現し、テストプログラムに適用して速度を比較してみたところ、1.18 倍から 2.98 倍の速度で動作した。」

となっている。これは x86、 PowerPC、ARM といった汎用プロセッサーの多くは、信号処理や数値計算処理など、大量のデーターを扱うプログラムの高速化を目的として、SIMD（Single Instruction Multiple Data）命令セットという命令セットを搭載している。ここで、SIMD 命令セットとは複数のデーターを一命令で演算可能な命令の集まりのことを言う。この命令セットを使ってプログラムを高速化するためには、プログラマーがアセンブラーや組み込み関数を使ってプログラミングする必要がある。しかし、それには SIMD 命令セットや SIMD 命令特有の問題を解決するための知識を要する上に、プログラミングにもデバッグにも時間がかかる。さらに、記述したプログラムは特定のプロセッサーに特化することになるため、移植性が低くなる。近年は解析手法に基づいて SIMD 命令を生成するための研究 がなされているが [文献 4]-[文献 7]、コンパイラーが SIMD 命令による最適化を実現するための制約は多く、まだ発展途上である。特に、条件分岐や、ループ実行前に繰り返し回数が定まらない WHILE 型ループなどがあり、制御構造が複雑なプログラムの最適化は、SIMD 命令の実行条件が制約となってほとんどなされていない。SIMD 命令の適用範囲を拡大し、SIMD 命令による最適化が自動化可能なプログラムを増やすためにも、制御構造を SIMD 並列化する的確な手法が要求される。

本論文では、条件分岐や WHILE 型ループなどの複雑な制御構造を持ったプログラムブロックの SIMD 並列化処理を自動化するための手法を提案する。そして流れは以下のようになっている。2 章で、SIMD 命令と、SIMD 並列化や提案手法に関わる研究動向についての紹介。3 章で、制御依存関係を明確にし、SIMD 並列化を容易にするために利用する制御フローグラフという解析手法についての説明。4 章で、本論文の提案手法である制御構造の SIMD 並列化を行うためのアルゴリズムの解説。5 章で、本論文で提案する手法を実装したいくつかのテストプログラムによる、SIMD 並列化することで提案手法の性能評価。6 章が全体のまとめとなっている。

従来は、SIMD 命令という複数のデーターを並列処理するための命令セットで、条件分岐や WHILE 型ループといった複雑な制御構造を並列処理するための手法はほとんど提案されていなかった。それらは（どういう訳か）自動 SIMD 並列化の対象からはずされている。本論文では、複雑な制御構造を SIMD 命令で並列実行可能な形に変換するための手法を提案し、提案した手法を COINS コンパイラーインフラストラクチャーで簡単な形で実現した。その実現したコンパイラーコンパイラーを使用し、テストプログラムをまがりなりにも SIMD 並列化して、SIMD 並列化前後で速度比較したところ、平均 1.18 倍から 2.98 倍の速度向上を得ている。今後の課題としては以下のような項目が挙げられる。

1. 今回扱った AltiVec という SIMD 命令セットには vselect, vcond any に相当する命令があったが、対応する命令が欠けている SIMD 命令セット（例えば、Intel x86 系の SSE 等）において、どれほどの速度向上が得られるものかの検証。

2. 提案手法がゼロ除算やオーバーフローなどの例外が発生し得るプログラムで、正しく動作させる為の方策、即ち例外の回避方法についての検討。

3. SIMD 命令による最適化のための手法として提案されている SLP 解析は、演算レベルで SIMD 並列化可能な部分を発見するための手法なのに対し、今回提案した手法は制御レベルの SIMD 並列化を行う手法である。そのため、SLP と提案手法を併用できるような SLP 解析の拡張。

以上が修士 2 年の学生が得た結果である。本研究の具体的実現方法等を考えてみると、COINS は実験（研究）用コンパイラーコンパイラーとしての意味合いが強く、現実の利用に耐え得る品質のコンパイルコード出すのはまだまだやり残された点が残されており、実験的に検討した解析手法並びにそれに基づく追加機能も修士論文から判断する限り限定的と思われる。従ってその適応範囲は広くは無く、そのままで実用に供することにはならないと考えている。(実用に耐え得るコンパイルコードを出すためには、長年に渡る地道な改善の努力と、利用者からのフィードバックに答える体制の整備が必須であるが、COINS プロジェクトの状況を考えると成長は止まったと考えて良いであろう。）本研究の学術的な意味は、これまで複雑な制御構造を有する計算プログラムの高速実行が、SIMD 命令により可能であることを示した点にある。今後、この結果に影響を受けたベンダー各社の提供するコンパイラーのプログラム解析能力と高品質なコンパイルコード生成能力の両方が向上し、多くの利用者に恩恵を与えることを期待している。

3.3 4 倍精度積和演算の高速化

本研究は本研究室の修士 1 年の永井貴博君が修士課程 1 年の間におこなった研究である。彼が現在投稿している論文 [発表 1] の日本語アブストラクトは、

「並列計算機の性能向上や数値計算法の進展は、大規模科学技術計算における大きな鍵となっている。特に浮動小数点数における演算においては，計算規模が増すにしたがってより多くの計算量を必要とし、計算誤差も増大する。そのために，倍精度演算より有効桁数が多い 4 倍精度演算の必要性が高まってきており注目されている。4 倍精度数の表現には倍精度浮動小数点数を 2 つ用いて表される 128 ビットデーター型があるが、SR11000 モデル J2 上の Hitachi 最適化コンパイラーにおいて、4 倍精度演算は 2 つの倍精度データ型を用いてソフトウェアーによって実現されており、倍精度演算に比べより多くの計算回数を必要とする。そこで本研究では、SR11000 モデル J2 上の Hitachi 最適化コンパイラーを用いて 4 倍精度演算を定量的に解析し、FMA 命令（Fused Multiply-Add） を用いて演算回数を削減することによって高速化を行い、最大で約 1.5 倍の高速な 4 倍精度積和演算を実現した。」

となっている。浮動小数点数における数値計算法の中には、計算規模が増すにしたがって計算誤差が増大し、より多くの計算量を必要としてしまうものもある。例えば、線形方程式 $Ax = b$ の解法であり、クリロフ部分空間を用いる CG 法などは誤差の影響を大きく受ける。また計算機上における浮動小数点数演算では、実数を一定の有限桁で近似してしまうため、丸め誤差が発生する。そのために、倍精度データー型から仮数部を増やした 4 倍精度データー型を用いた 4 倍精度演算の必要性が高まってきており注目されている。

4 倍精度数で表現された数には、倍精度浮動小数点数を 2 つ用いて表される 128 ビットデーター型があるが、ハードウェアーで実装されている倍精度演算に比べより多くの演算回数を必要とする。そこで本研究は、SR11000/J2 上の最適化コンパイラーを用いた 4 倍精度演算に焦点を当て、行列積に用いられる積和演算を定量的に解析し、演算回数を削減することによって高速化をおこなった。FMA 命令を用いて浮動小数点数の演算回数を削減し、ループ展開による演算命令のレイテンシー隠蔽など SR11000/J2 のアーキテクチャーを考慮した最適化を適用して高速な積和演算、行列積を実現した。その結果、最適化コンパイラーを用いた場合より約 1.5 倍の性能で演算を実現できることを確認した。

本条件付き採録となった投稿論文は第 2 章で 4 倍精度データー型、第 3 章で 4 倍精度演算を紹介し、第 4 章では利用環境である SR11000/J2 のアーキテクチャーについて触れ、第 5 章で 4 倍精度演算の

最適化について述べ、第 6 章で 4 倍精度における行列積の積和演算の性能評価を行っている。実験で使用したマシンは本センターのマシンであり、その上の最適化 C コンパイラー 01-03-/A を用いている。OS は IBM AIX バージョン 5.3 であり、ラージページの設定は行われてはいない。

本研究で提案する 4 倍精度数の一加算あたりのクロック数性能比較を実際の実験により得ている。この実験でおこなった各種条件は

1. L2 キャッシュの半分程度までメモリーを使用する演算の場合、
2. キャッシュメモリー全体にのる程度のデーターサイズの演算の場合、
3. L3 キャッシュの 1.5 倍のデーターサイズの演算の場合、

の 3 ケースである。これらの各ケースの実験により、アンローリング段数を増やすことによってパイプライン処理が効果を出すため、

1. 段数を増やすに従って 4 倍精度加算演算一回あたりに必要なクロックサイクル数の割合は小さくなっている。
2. 最適化を行うと L1 や L2 キャッシュなどを用いることで、データー転送の影響を受けない範囲では SR11000/J2 最適化コンパイラーより高い性能を示す。

ことが判明している。

次に本研究で提案する 4 倍精度数の一乗算あたりのクロック数性能比較について、同様のケースで実験をおこなったところ、乗算においてもアンローリング段数の増加によるクロック数の割合は小さくなっており、4 倍精度加算と同様の傾向が見られる。

提案する 4 倍精度数の一積和演算あたりのクロック数性能比較についても、同様に実験を行い、SR11000/J2 最適化コンパイラーと比較すると、提案手法ではデーターが L1 キャッシュに乗る場合、257.60/167.62 ≃ 約 1.5 倍の高い性能が出ていることが分かった。また、4 倍精度加算や乗算などそれぞれの演算だけでは性能が劣っている L3 キャッシュのデーターサイズでも、二つの演算をまとめた積和演算を最適化することによってデーターサイズに依存せず、アンローリング段数が 4 のときでも SR11000/J2 最適化コンパイラを上回り、高い性能が出ている。提案手法では L2 キャッシュと同程度のデーターサイズで演算性能が最大になった。

最後に提案する行列積における性能比較実験について、SR11000/J2 上での行列積と提案手法の行列積では、4 倍精度積和演算の性能がほぼ一定の割合を保ったまま推移している。これはデーター転送の割合が少ない 4 倍精度積和演算性能がそのまま表れた結果と思われる。本研究では議論しなかったキャッシュの考慮についての知見を得るため、ブロックサイズ 200 で提案手法に適用すると、行列サイズが大きくなっても性能の低下は見られなかった。さらなる最適化のためには、データー転送における考慮も必要である。

なお、今回報告しているのと同様な倍精度を二つ用いたアルゴリズムで Fortran、C++を用いて実装したライブラリーが公開されている [文献 2]。しかし倍精度数をさらに上位と下位部分に分割して計算する乗算アルゴリズムでは、今回用いた 4 倍精度乗算の約 2 倍の演算回数を必要としてしまう。積和演算を用いた乗算もあるが演算回数は多くなっている。Hida らは 4 倍精度演算のアルゴリズムを応用して倍精度数を 4 つ用いた 8 倍精度演算を提案している [文献 3]。本研究では IEEE754 の倍精度実数を 2 つ用いた 4 倍精度演算の性能を解析評価し、SR11000/J2 最適化コンパイラーにおいて積和演算及び行列積を実現、高速化した。SR11000/J2 最適化コンパイラーよりも浮動小数点演算数削減、さらに 4 倍精度における浮動小数点演算の計算量に着目したループアンローリング、浮動小数点命令のレイテンシー隠蔽などによって最大で約 1.5 倍の演算性能を実現している。今後はメモリー間のデーター転送を考慮した 4 倍精度演算の最適化についての検討を 2007 年度以降おこなうことになる。

3.4 pi.super-computing.org 上での π 並びに関連する情報公開

80 年代に入ってからこれまで 4 半世紀以上に渡り実施してきている円周率計算に関する学問的意義他については、一般向けの啓蒙・報告書 [文献 1] 並びに、その報告書中で引用している文献、また一般聴取者向けの講演で述べており繰り返しとなるので、本報告では省略することととする。

さて 1995 年 6 月末から、http://www.super-computing.org/上で円周率の値、並びに関連する（anonymous ftp による）情報の公開を継続しているが、情報公開開始以降これまでに 10 年が経過した。公開当初、一般公開中の物は 2 億桁までの π と 1/π、並びにそれらに関連する統計値であり、特別な使用同意契約を FAX でをやりとりして締結することにより、アクセス可能な期間は最大 1 ヶ月であったが、40 億桁の π 並びに 1/π がアクセス出来る様になっていた。しかしそれまで公開していた www.cc.u-tokyo.ac.jp（sun ws）マシンの不調が度重なった為、2000 年 9 月中旬にサーバーマシンを変更したついでに、40 億桁の π 並びに 1/π が自由にアクセス出来る様にした。

2006 年度中に転送された総データー量は、サーバー機器の老朽化もあってか、2005 年度実績値のほぼ 25% 減と 10 年前からの公開開始以降初めての減少を示した。しかし転送総データー量は減少したにもかかわらず、アクセスサイト数、アクセス回数、アクセス元のアドレスのそれぞれは増加している。アクセス元のトップアドレスは過去最大数を更新し前年度の 10% 増の 126 種類となった。各種マスメディアでの報道のとおり、2002 年度に新しい円周率計算の世界記録を樹立することが出来たので、その時に得られた数値を利用し 自由にアクセスできるデーター量及びデーター種の増加させる価値は十分にあったが、この種の計算桁数を Web で公開する為のハードウェア整備の為の競争的研究資金の獲得が困難という事情もあり、公開桁数の大幅な増加は 2006 年度も実施できなかった。しかし、別館の耐震基準不足に伴う部屋移動と大容量磁気ディスクが安価に購入できる様になったことがあり、大容量データーの Web 公開に関係するハードウェアの準備が年度末に行えたので、2007 年度中にはやっと公開桁数の大幅な増加が行える見込みとなった。

さて 2006 年度中の利用統計は 73,469 サイト（1995 年度は 748 サイト、1996 年度は 2,052 サイト、1997 年度は 5,604 サイト、1998 年度は 9,615 サイト、1999 年度は 16,396 サイト、2000 年度は 15,619 サイト、2001 年度は 15,354 サイト、2002 年度は 25,529 サイト、 2003 年度は 35,652 サイト、2004 年度は 60,728 サイト、2005 年度は 64,312 サイト）（DNS に名前が載っているサイトとしては 62,896 サイト（1995 年度は 641 サイト、1996 年度は 1,627 サイト、1997 年度は 4,654 サイト、1998 年度は 8,302 サイト、1999 年度は 14,092 サイト、2000 年度は 13,500 サイト、2001 年度は 13,518 サイト、2002 年度は 21,795 サイト、2003 年度は 29,503 サイト、2004 年度は 46,434 サイト、2005 年度は 54,459 サイト））から総計 129,497 回（1995 年度は 5,012 回、1996 年度は 9,296 回、1997 年度は 16,127 回、1998 年度は 24,495 回、1999 年度は 40,457 回、2000 年度は 41,191 回、2001 年度は 37,188 回、2002 年度は 58,243 回、2003 年度は 67,851 回、2004 年度は 108,379 回、2005 年度は 116,351 回）ファイルがアクセスされ、転送された総データー量は 126,747.6 MB（1996 年度は 6,918MB、1997 年度は 13,880MB、1998 年度は 14,455MB、1999 年度は 22,353MB、2000 年度は 55,392.6MB、2001 年度は 59,866.2MB、2002 年度は 111,082.7MB、2003 年度は 86,500.9MB、2004 年度は 120,464.2 MB、2005 年度は 167,199.9 MB）と初めて前年度を下回った。前年度を下回った理由は不明である。

これらのアクセス元のアドレス（u-tokyo.ac.jp の jp に対応する部分）種別は総計で 126 種類（ac ad ae al am ao ar arpa at au aw ba be bg biz bm bn bo br bt by ca cc ch cl cn co com cr cu cx cy cz de dk do ec edu ee eg es eu fi fm fr gh gl gov gr gt hk hr hu id ie il in info is it jp ke kg kr kz la lb li lk lt lu lv ma mc md mil mu mx my na nc net ni nl no np nu nz org pe pf ph pk pl pt py ro ru sa se sg sh si sk sm st su sy th to tr tt tv tw ua uk us uy uz ve vn ws yu za zm。なお 1996 年度は 50 種類、1997 年度は 59 種類、1998 年度は 65 種類、1999 年度は 78 種類、2000 年度は 68 種類、2001 年度は 76 種類、2002 年度は 82 種類、2003 年度は 89 種類、2004 年度は 104 種類、2005 年度は 115 種類）にも及び、今まで見たことのないアドレスが今年度も増加していることと、転送したデーター量がこれまでの最高値を記録している

こと、また電子メールによる問い合わせ数の増加等から、円周率関連のデーター入手に対する全世界からの根強い人気が伺える。

2006 年度末までの円周率計算の公表されている世界記録は 2002 年 11 月末に SR8000/MPP を利用して樹立した 10 進で 1 兆 2411 億桁である。なお、この時同時に 16 進で 1 兆強計算しており、この桁数も当然ながら新記録である。計算した全ての桁数の公開は近年の磁気ディスク記憶装置一台当たりの記憶容量の観点からは何ら問題は無いが、転送するデーター量を考えると、ネットワーク経由による最新の計算結果の公開はますます困難となっている。しかし Giga ビット回線が国内外で充実しつつあることと、高信頼性ディスク装置の容量当たりの単価が下がって行っており本章の初めの部分にも述べてあるように、公開用のハードウェアの準備が行えたので、2007 年度中には公開する予定である。

また 2006 年度末には 2004 年度末に導入した新しいスーパーコンピューター HITACHI SR11000/J1 の中間レベルアップが実施され、現在サービスしているスーパーコンピューターの 3 倍強の主記憶容量・総合演算性能を持つ新しいスーパーコンピューター HITACHI SR11000/J2 のサービスが開始される。新しいマシンはこれまでの 3 倍弱のノード数と総主記憶容量、3 倍強の性能と高信頼性・高速磁気ディスクを有しており、本センターでこれまで新機種の導入当初に行い、各種問題点を的確に検出してきた実績を有する、信頼性・可用性確認用の大規模計算を実施することになる。システムに高負荷をかけるこの計算の実施により、結果として計算桁数記録の更新が実現される可能性がある。

3.5 スーパーコンピューターを用いた大規模高速計算

3.5.1 並列実行性能の高い数値計算ライブラリー及びプログラム開発環境に関する研究開発

分散並列型計算機の性能を引き出せるプログラムを書くことは、職人芸的要素が多くまたなかなか評価を受けにくいという現状がある。従ってプログラミングを生業としている、あるいは計算科学を専門としない多くの研究者にとって、高性能プログラムの作成を困難にしているし、プログラム作成効率を高めるための努力や、プログラムそのものの実行効率を極限まで高めようとすることに興味を持ち、長期にわたり没頭することに喜びを感じる学生が生まれることは多くは無いという現状がある。しかし研究の最前線の進展速度を高める為には、多くの研究者にとって使いやすく、高い実行性能が出せ、ノード数に関するスケーラビリティーが高く、広範囲の機能を有する数値計算ライブラリーの提供が少なくとも必要である。その様な目的を持った数値計算ライブラリーが、一部ベンダーあるいは独立ベンダーから提供・販売されてはいるが、その導入コストが高かったり、機能的に不十分だったり、実現している機能間の性能にアンバランスがあったり、機種間の互換性に問題があるのが現状となっている。また将来的には、計算機を利用した研究・開発における研究環境のインフラ整備の一環として、高い実行効率を容易に出せたり、プログラムの保守管理が容易となる様なプログラム開発環境の開発や提供が今後必要となるであろう。

この様な理由により本研究に係わる目的として、分散メモリー型計算機および共有メモリー型計算機、また PC クラスターから、スカラー並列計算機あるいは（ハード及びソフト）ベクトル並列計算機までの広範囲のアーキテクチャーのどれに対しても高い並列実行性能を発揮でき、広範囲の機能をカバーし、自動チューニング機能等による使いやすい数値計算ライブラリーの研究開発、並びに簡単なプログラミングで実行効率が高いプログラム開発環境の研究開発、それぞれを目的としている。具体的には、東京大学新領域創成科学研究科基盤情報学専攻の学生諸君達とこの研究テーマで研究を行う予定である。2004 年度までにおこなった研究として、特に連立一次方程式の反復解法における階の直交性と収束特性に関する研究と、大きな素数の基底に対する高速フーリエ変換の高速化に関する研究を実施したが、2005・2006 両年度はこの種の数値計算に興味を持つ学生が全く存在しなかった為、本研究は一時停止状態になっている。もし 2007 年度に数値計算に興味を持つ修士課程の学生が進学して来たならば、2007 年度後半以降に研究活動を開始することになるであろう。なお関連して、「次世代スーパーコンピューティング応用言語研究会（Next Generation Computing Application Language

Consortium （略称 NSCAL)）」を昨年 11 月に立ち上げて、研究室外のメンバーと活動を開始した。今後地道な活動を継続することになろう。

これまでの年報等での報告にも繰り返し示しているように、従来型のライブラリーでは利用者が定義または指定しなければならないパラメーターが多く、利用する計算機の計算機アーキテクチャーや問題の性質の違いによる性能チューニングが困難で、計算機アーキテクチャーや数値計算を専門としない多くの利用者にとって使い易いものでは無い。しかし本研究に基づく数値計算ライブラリーでは、

1. 利用者が指定するパラメーターが少ない、
2. 演算カーネルに対する自動チューニング機能を有する、
3. プロセッサー間通信に関する自動チューニング機能を有する、
4. 使用する数値計算アルゴリズムに関する自動選択機能を有する、
5. 各種計算機アーキテクチャーにおいて、高い実行性能を有する、

という特徴を持つ。2004 年度に新しく追加した「プログラム開発環境の研究開発」により得られる開発関連パッケージにより、プログラムを必要悪として強要される研究者の研究効率を向上させることが今後可能となることが期待される。

本研究成果を広く利用できる様にするには、今後多く使用される様になると予想する、PC クラスター機から、各種対称型共有メモリー計算機、また（ハード及びソフト）ベクトルクラスター機に至る広範囲の計算機アーキテクチャー用の数値計算ライブラリーとそれらの環境で動作するプログラム開発環境の本格的提供が必要であり、その体制作りと実際の使いもになる数値計算ライブラリーの提供、またプログラム開発環境提供の準備を行う予定であるが、実際には修士課程の学生が複数人となり、博士進学者が出る 2007 年度以降に本格化することになる見込みである。

3.5.2 多項式級数に基づく数学関数及び数学定数の高速計算法に関する研究

多くの数学関数あるいは多くの数学定数は、有理数係数多項式級数や規則正しい連分数表示が可能である。このような関数や表示式を高速に計算できる方法をこれまでの研究で発見するとともに論文として発表した。この新方式による実際の高速計算アルゴリズム実現に関する研究を引き続き行っている。

具体的適応例としてπの計算を選び、プログラムの作成を開始し、並列計算機の並列性を生かした高速化を行うとともに、円周率計算桁数の世界記録、10 進で 1 兆 2411 億桁、16 進で 1 兆 307 億桁、の記録を 2002 年 11 月末に樹立し、同年 12 月 6 日に記者発表を行い、テレビ・ラジオ・新聞等のマスメディアにて当時大々的に報道されたが、2004〜2005 年度はさらにメモリー効率の高い多倍長乗算方法の実現方法を検討し、具体的なプログラム作業を開始したが、まだ完成には至ってはいない。これらプログラムは、センターが提供しているスーパーコンピューターの性能評価、可用性診断、バグ摘出等に非常に効果的であり、実際 2005 年度において成果を上げているが、本報告で詳細を記述するのは馴染まないので述べないこととする。

なお本研究は、元（株）日立製作所 後保 保範博士、現黒田 久泰助教らと一緒に長年に渡り実施してきているものである。

4 具体的成果

2006 年度は修士 2 年 2 名、修士 1 年 2 名の学生が研究室に存在したため、修士 2 年の修士論文を投稿論文とした論文 [査読付 1] を印刷・公表することができたので、これを具体的成果とする。もう一人の修士学生は博士に進学しており、2006 年度中に英語で国外の国際会議に投稿したが、2006 年度

中に採録報告が届かなかったので、2006 年度中の具体的成果とはしないこととする。来年度以降は、研究室の学生も充実し、博士課程学生も存在することとなり、今後はコンスタントに研究成果を得ることとなろう。

査読付論文リスト

[査読付 1] 廣松悠介、黒田久泰、金田康正：「複雑な制御構造を持つプログラムの SIMD 命令セットによる最適化」、情報処理学会論文誌：プログラミング、Vol.48、No.SIG 4(PRO 32)、pp.62-72、2007-3。

その他の発表論文リスト

[発表 1] 吉田仁：”Computation of $\pi(x)$”、東京大学新領域創成科学研究科基盤情報学専攻修士課程修士論文、2007 年 3 月。

[発表 2] 永井貴博、吉田仁、黒田久泰、金田康正：「SR11000 モデル J2 における 4 倍精度積和演算の高速化」、情報処理学会論文誌：ハイパフォーマンスコンピューティングシステム、2007 年 3 月（条件付き採録で、2006 年度末現在再審査中）。

[発表 3] H. Yoshida, H. Kuroda and Y. Kanada: ”Parameter Selection for Fast Computation of $\pi(x)$”, Submitted to WorldComp′07 （2007 年 3 月末現在査読中）。

[発表 4] T. Nagai, H. Yoshida, H. Kuroda and Y. Kanada: ”Fast Quadruple Precision Arithmetic for Multiply/Add Operations on SR11000/J2”, Submitted to WorldComp′07 （2007 年 3 月末現在査読中）。

引用文献関連

[文献 1] 金田康正:「計算機による円周率計算」数学文化、pp. 72–83、日本評論社、Vol. 1、No. 1、2003。

[文献 2] D.H. Bailey:A fortran-90 double-double library http://crd.lbl.gov/dhbailey/mpdist/mpdist.html.

[文献 3] Yozo Hida, Xiaoye S. Li, and D. H. Bailey: Algorithms for quad-double precision floating point arithmetic, Proceedings of the 15th Symposium on Computer Arithmetic, pp. 155 – 162, 2001.

[文献 4] Larsen, S. and Amarasinghe, S.: Exploiting Superword Level Parallelism with Multimedia Instruction Sets, Programming Language Design and Implementation, Vancouver, Canada, Proceedings of ACM SIGPLAN Conf, pp. 145 – 156 (2000).

[文献 5] Kudriavtsev, A. and Kogge, P.: Generation of Permutation for SIMD Processors, the 2005 ACM SIGLAN/SIGBED Conference on Languages, Compilers, and Tools for Embedded Systems. LCTES′05, ACM Press, pp. 147 – 156 (2005).

[文献 6] Alexandre E.Eichenberger, P.W. and O′Brien, K.: Vectorization for SIMD Architectures with Alignment Constraints, PLDI’04, Vol.9-11, pp. 82 – 93 (2004).

[文献 7] PengWu, A. E. E. and Wang, A.: Efficient SIMD Code Generation for Runtime Alignment and Length Conversion, Proceedings of the International Symposium on Code Generation and Optimization, San Jose, CGO′05, pp. 153 – 164 (2005).

次世代システムソフトウェアの研究開発
—オープンスパコン、グリッド、ディペンダブルシステム—

石川 裕

1 概要

平成18年度、次世代システムソフトウェアを実現するために、オープンスパコンアーキテクチャの仕様策定、グリッド環境下での高性能通信アルゴリズムの提案と評価、ディペンダブルシステムを支援するオペレーティングシステムの概念設計を行った。オープンスパコンアーキテクチャの仕様策定では、京都大学 学術情報メディアセンター、筑波大学 計算科学研究センターと共に共通基本仕様の検討を行なった。市販スーパーコンピュータ製品にとらわれず、コンピュータ市場を牽引している高性能プロセッサおよび通信技術によって構成されるスーパーコンピュータを導入し、大規模計算科学研究者がグランドチャレンジできる基盤を提供するとともに、従来のスパコン応用分野にとどまらず、大規模ゲノム情報処理、大規模データマイニングなど新興応用分野まで幅広く適応可能なスーパーコンピュータ構築を目指す。

グリッド環境でのコレクティブ通信アルゴリズムが提案されているが、これまで、広域ネットワークの通信バンド幅が小さいと仮定されていた。しかし近年、通信バンド幅は急速に向上し従来の仮定とミスマッチしてきた。そこで、高遅延かつ高バンド幅のネットワーク環境に対して、コレクティブ通信(Bcast、Allreduce、Gather)の新しいアルゴリズムを提案した。本アルゴリズムは、高バイセクション・バンド幅を持つネットワークで効率的な通信を実現している van de Geijn と Rabenseifner の各アルゴリズムを拡張し、クラスタ間ネットワークのバンド幅を有効利用できるようにしたものである。さらに、ネットワークの同時利用ノード数を制限することで広域ネットワークでの輻輳の抑制を行った。ネットワーク・エミュレーション環境を用いてアルゴリズムの実行性能評価を行い有効性を確認した。本研究は産業総合研究所グリッドセンターとの共同研究で行われた。

ディペンダブルシステムを支援するオペレーティングシステムを実現すべく概念設計を行った。本研究ではネットワークで接続された組込みコンピュータによるデータベースサーバや高性能計算システムを実現するための高信頼高性能組み込み並列分散オペレーティングシステムを実現する。処理能力に応じて、ディスクを搭載したコンピュータ群、マルチコア CPU とメモリだけを搭載したコンピュータ群が適時ネットワークでつながれるクラスタシステムを想定し、Linux を基に、高信頼単一システムイメージを提供する並列分散 OS を実現する。これにより、オペレーティングシステムレベルでハードウェアの耐故障機能および実時間性を実現する。

2 オープンスパコン

2.1 背景

科学技術の更なる発展のために、スーパーコンピュータによる問題解決やサービス提供が世界レベルで大きな関心になっていることは、言うまでもない。このような状況で、東京大学、京都大学、筑波大学の計算センターはそのサービスにおいて、応用領域の多様性に応え、広範なユーザニーズに

資するべきであるという理念で一致した。また、3 センターのスーパーコンピュータのリプレース時期が 2008 年 3 月と一致したこと、そして、3 大学のセンターに所属する教授陣が長年培ってきたコンピュータサイエンス分野における先端研究のノウハウ（PACS-CS、SCore、MegaProto 等）のシナジー効果により、より良いスーパーコンピュータ調達が可能になると考え、共同で調達の基本仕様を策定することになった。このための Feasibility Study を平成18年度行った。

2.2 内容

本仕様は、次の 3 つの理念に基づくことから、「オープンスパコン仕様」と名付けられた。

- 基本アーキテクチャのオープン性
  コンピュータ市場を牽引しているコモディティ高性能プロセッサを基本とする。コモディティ高性能プロセッサ開発の熾烈な競争により、最新の半導体技術を基にしたプロセッサが半年から 1 年という短いサイクルで市場に投入されている。最新技術を使用することにより、高性能かつ低消費電力を実現したシステムを導入することが可能となる。
- システムソフトウェアのオープン性
  オープンソースに基づく先端ソフトウェア技術を基本とする。プログラミング環境やオペレーティングシステムなどのシステムソフトウェアにおけるオープンソフトウェアの潮流は並列アプリケーションにも及んでいる。多くのユーザが使用するこれら資産をシームレスに利用できる環境を提供することにより、より多くのユーザが大規模並列処理環境へ移行することが促進できる。
- ユーザのニーズに対するオープン性
  従来スパコン応用分野に加えて、大規模ゲノム情報処理、大規模データマイニングなど多様なユーザニーズに応えるシステムを実現する。これら従来の計算センターユーザでないニーズに対して応えることは、計算センターの新しい使命であると認識している。

我々の基本仕様によって、以下のようなニーズに応えられると考えている。

- 大規模科学技術計算のニーズ
  素粒子・宇宙、物質・生命、地球環境などの科学分野、原子力、自動車・航空機などの工学分野では、より大規模な科学技術計算を必要としている。
- 新興応用分野のニーズ
  従来のスーパーコンピュータユーザに加えて、ゲノム情報処理、データマイニング、物理シミュレーションに基づくアニメーションなどの、新しいニーズが台頭している。アプリケーション利用技術としても、WEB ユーザインターフェイスやグリッド技術など、従来計算センターの運用では想定していない運用方法が求められるようになった。
- PC クラスタユーザのニーズ
  小・中規模の科学技術計算では計算センターのスーパーコンピュータだけではなく、ユーザ自身が設置した PC クラスタを利用することも増えている。PC クラスタはオープンソフトウェアとの親和性が高く、プログラミング環境もユーザに親しみやすいものとなっている。しかし PC クラスタは規模が大きくなるに連れて保守・運用が難しくなってきており、計算センターに対するユーザの期待が高まっている。
- 京速コンピュータへの橋渡し
  2012 年年度までに実現される京速コンピュータは、現在計算センターに設置されているスーパーコンピュータの約 1000 倍の能力となる。しかし、現状のアプリケーションプログラムを変更せずにマシン性能に見合った大規模問題を解くことは、必ずしも容易ではない。京速コンピュータの利用を想定したアプリケーションプログラム開発のための開発環境が望まれている。

2.3 具体的成果

「オープンスパコン仕様」は、従来の調達で行われてきた単なる市場調査の結果として定めたものではない。石川研究室で開発してきた MPI 通信ライブラリの一実装である YAMPI[公開 1]を使用してコモディティ部品（一般に市販されている部品）によるスーパーコンピュータの実現可能性をさぐった結果を反映している。中間報告として、2006年9月6日に「大学の知を支援するセンターマシンワークショップ」を開催するとともに報道機関への広報もおこなった[報道 1]。具体的な成果としては、情報処理学会 HOKKE2007 において4本のネットワークリンクを束ねたときの性能とその考察を行った[発表 2]。詳細は、松葉浩也の研究報告を参照のこと。

3 グリッド

3.1 背景

グリッド環境で複数のクラスタを接続して利用するための MPI システムとして MPICH-G2、MagPIe、PACX-MPI などが知られている。これらグリッド用 MPI システム用に、高遅延ネットワークに最適化されたコレクティブ通信アルゴリズムが提案されている。これらアルゴリズムでは、遅延が大きいことと同時に通信バンド幅が小さいことが想定されている。

しかし近年、広域ネットワークの通信バンド幅は広がり、10Gbps から 40Gbps といったバンド幅が提供されるようになってきた。このバンド幅は典型的なクラスタノードの NIC の持つ 1Gbps というバンド幅を越えている。この傾向は、クラスタノードの NIC にかけられるコストを勘案すると、しばらくは続くと考えられる。このようなネットワーク環境は、これまでのグリッド用 MPI における仮定とマッチしていない。

一方、クラスタ内の全ノードが一斉に通信を行うと、ネットワークが輻輳する。一般に、広域ネットワークでは通信に TCP/IP プロトコルが用いられる。TCP では輻輳により致命的に性能が低下することが知られている。バンド幅を有効利用するとともに、輻輳を避ける必要がある。そこで、コレクティブ通信のアルゴリズムを高遅延かつ高バンド幅なネットワークに最適化した。本研究は、産業総合研究所グリッドセンターとの共同研究で行われた。

3.2 内容

3.2.1 Bcast アルゴリズム

```
Bcast(void *buf, int siz)
{
    if (rank < (nprocs/2)) {
      for (i = 0; i < (log2(nprocs)-1); i++) {
        split_copy(buf, siz, i, hemisphere);
      }
    }
  counter_copy(buf, siz);
    for (i = 0; i < (log2(nprocs)-1); i++) {
      merge_copy(buf, siz, i, hemis-
phere);
    }
```

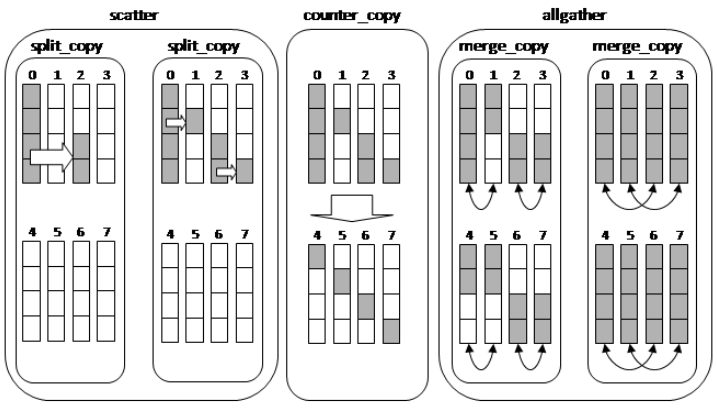


図 3.2.1 Bcast アルゴリズム

図3．2．1に van de Geijn の Bcast アルゴリズムをグリッド向けに拡張したアルゴリズムを示す。各ステップの処理は以下のようになる。
(1) split_copy のループは Scatter に相当する。i 番目の split_copy のステップで、各ノードはメッセージを半分に分け、その半分を 2 (nprocs/2-i) 離れたノードへ転送する。
(2) counter_copy は(1)のステップで Scatter されたデータをクラスタ間でコピーする。各ノードはそれぞれクラスタ間で 1 対 1 で対向した転送を行う。
(3) merge_copy のループは Allgather に相当する。i 番目の merge_copy のステップで、各ノードはメッセージを 2^i 離れたノードへ転送し自分の持つメッセージと転送されたメッセージをマージする。

counter_copy がクラスタ間通信のために追加されている。これを取り除くと van de Geijn の Bcast アルゴリズムそのものである。

counter_copy で全てのノードが一斉にクラスタ間通信を行うと、グリッド上のボトルネックになっているリンクが輻輳する。輻輳を避けるため、counter_copy では同時に通信するノード数を制限している。指定されたノードのみがクラスタ間通信を行い、それ以外のノードは指定されたノードへメッセージをフォワードする。同時に通信するノード数は、クラスタ間ネットワークに合わせて 1 から nprocs/2 の間で自由に選ぶことができる。

3.2.2 Allreduce

```
Allreduce(void *buf, void *tmp, int siz)
{
   for (i = 0; i < (log2(nprocs)-1); i++) {
   split_reduce(buf, tmp, siz, i, hemisphere);
 }
   counter_reduce(buf, tmp, siz);
 for (i = 0; i < (log2(nprocs)-1); i++) {
     merge_copy(buf, siz, i, hemisphere);
   }
}
```

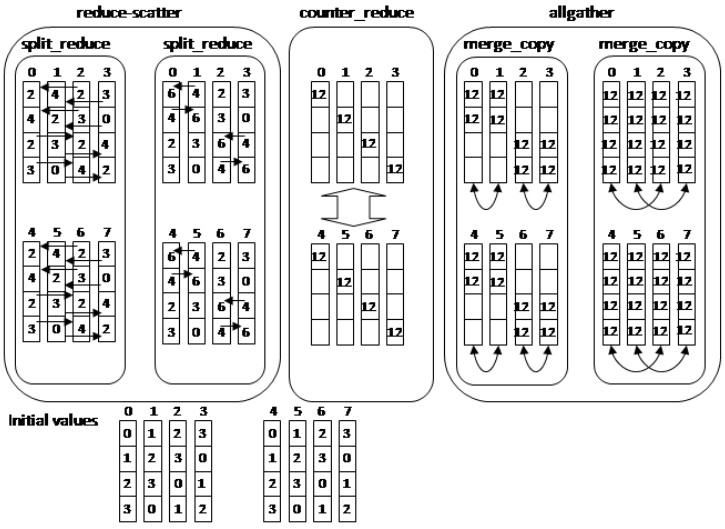


図 3.2.2 Allreduce アルゴリズム

図 3.2.2 に Rabenseifner の Allreduce アルゴリズムをグリッド向けに拡張したアルゴリズムを C 言語で示す。各ステップの処理は以下のようになる。
(1) split_reduce のループは Reduce-scatter に相当する。i 番目の split_reduce のステップで、各ノードはメッセージを半分に分け、その半分を 2(nprocs/2-i)離れたノードへ転送する。そこでリダクションを行う。
(2) counter_reduce は(1)のステップで Reduce-scatter されたデータをクラスタ間で交換しリダクションを行う。Bcast の counter_copy と同様であるが、双方向である点とリダクションを行う点が異なる。
(3) merge_copy のループは Allgather に相当する。これは Bcast の後半と同一である。

Rabenseifner との違いは、counter_reduce がクラスタ間通信のために追加されている点である。counter_reduce においても counter_copy 同様にクラスタ間の同時通信ノード数の制限を行っている。

3.2.3 Allgather

Allgather は van de Geijn の Bcast アルゴリズムの後半に現れる処理そのものである。まずクラスタ間で通信し、続いてクラスタ内で Allgather を行う。

3.2.4 Reduce

Reduce-scatter は Rabenseifner の Allreduce アルゴリズムの前半に現れる処理そのものである。クラスタ内で Reduce-scatter を行い、続いてクラスタ間で通信する。Reduce は Reduce-scatter+Gather(クラスタ内)に分解できる。

3.2.5 Gather

Gather は分散したデータを 1 つのルートノードに集約する通信である。Gather も Scatter と同様、ルートノードの通信能力から、各ノードに対して順次 1 対 1 通信を行う通信が最適である。

ただし、通信がルートノードに集中する通信であるので、輻輳が起る可能性がある。そこで、クラスタ間の通信に counter_copy を使用する。また、クラスタ内でも輻輳は起るため、クラスタ内の通信にはランデブープロトコルを使用する。ただし、クラスタ間とクラスタ内の通信はフェーズ分けされていないため輻輳は完全には抑制されていない。

3.3 具体的成果

これら提案したアルゴリズムおよび評価結果は、IEEE International Conference on Cluster Computing[査読付 1]および情報処理学会 研究報告 2006-HPC-87[発表 2]で発表した。ここでは評価結果として、Bcast と Allreduce の性能評価を示す。評価用に使用した計算機環境を図 3.3.1 に示す。

ノード	CPU	Opteron 2.0 GHz
	Memory	6 GB DDR333
	Network Card	Broadcom BCM5704
	OS	SuSE Enterprise Server 9 (Linux-2.6.5)
スイッチ		Huawei-3Com QUidway S5648

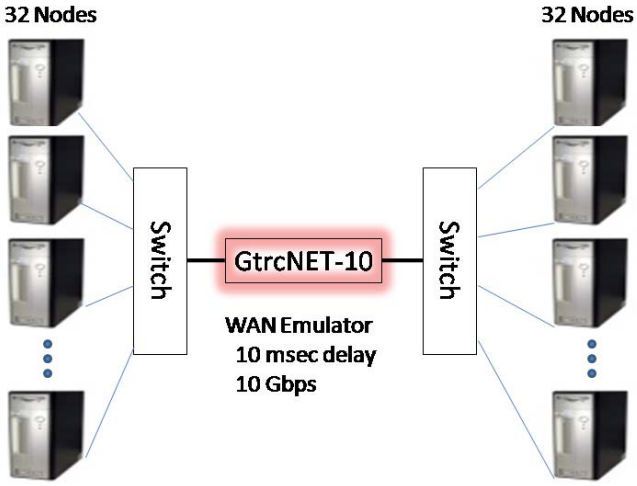


図 3.3.1 評価環境

ここで、GtrcNET-10 は産業総合研究所が開発した WAN 模擬装置で、10Gbps の性能を持つ。本装置を用いて 10 msec の通信遅延を生じさせている。図 3.3.2 に、Bcast アルゴリズムの評価結果を示す。ここで、n とは、クラスタ間通信における同時通信リンク数を示す。farfirst は、クラスタ間で、遅延のある通信を最初に行うというアルゴリズムで、既存のグリッド環境向け MPI 通信ライブラリで使用されているアルゴリズムの場合である。world は、van de Geijin アルゴリズムだけを使用した場合である。図からわかるとおり、我々が開発した新しい Bcast アルゴリズムは高性能な通信を達成している。

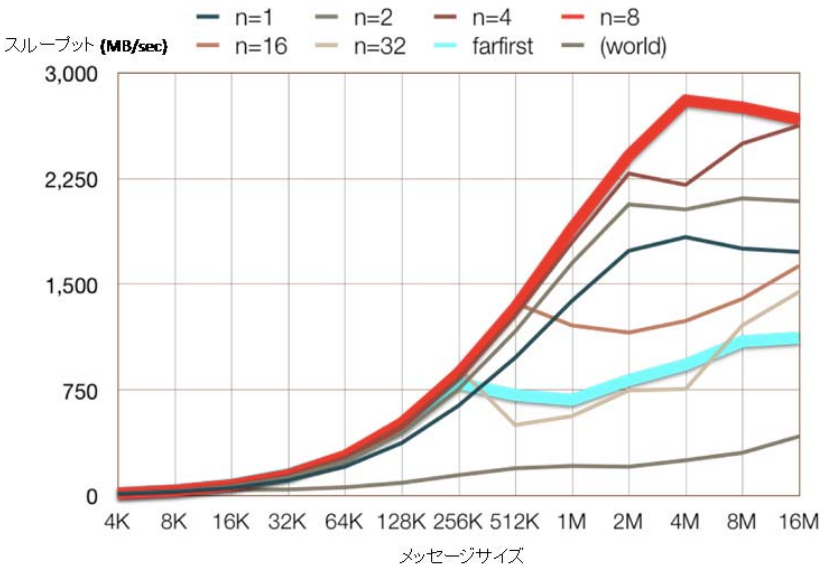


図 3.3.2 Bcast アルゴリズムの評価

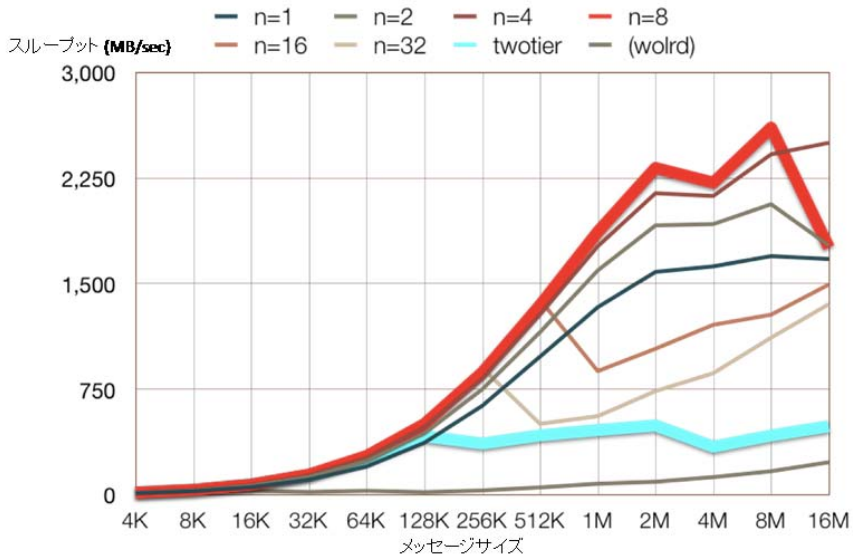


図 3.3.3 Allreduce アルゴリズムの評価

図 3.3.3 に、Allreduce アルゴリズムの評価結果を示す。Y 軸は、全体全で通信した場合の総転送データ量を通信時間で割った値である。ここで、n とは、クラスタ間通信における同時通信リンク数を

示す。Twotier は、クラスタ内でリダクションした値をクラスタ間で交換するアルゴリズムの場合で、既存のグリッド向け MPI で使われているアルゴリズムである。world は、Rabenseifiner アルゴリズムだけを使用した場合である。図からわかるとおり、我々が開発した新しい Allreduce アルゴリズムは高性能な通信を達成している。

本成果は、IEEE International Conference on Cluster Computing[査読付 1]および情報処理学会 SWOP2006[発表 1]で発表した。この他の結果に関しては、これら論文を参照のこと。

4 ディペンダブルシステム

4.1 背景

科学技術計算や基幹業務系で使われているスーパーコンピュータや大規模サーバシステムの故障により、数か月におよぶ計算結果や会計情報などの重要なデータを喪失し、社会に対して甚大な損失を与える。このため、このようなコンピュータシステム向けに、ハードウェア、ソフトウェアの双方で耐故障性を実現するための研究開発がおこなわれてきた。コンピュータ技術の進展に伴い、基幹業務系コンピュータに限らず、ホームサーバ、デジタル TV、組込み型高性能サーバ、車載制御装置、通信制御装置、ロボット、携帯機器、モバイル・ウェラブルコンピュータなど、社会の隅々にコンピュータが使われ、これらコンピュータがネットワークにつながり、相互連携しながら豊かな環境が構築されつつある。このようなシステムの故障は、社会に対して致命的な損失を与える。我々は、このような組み込み機器向けに対して、組込みディペンダブルシステムを支援するオペレーティングシステムの研究開発を科学技術振興機構が実施している戦略的創造研究推進事業 CREST研究領域「実用化を目指した組込みシステム用ディペンダブル・オペレーティングシステム」において、実施している。

4.2 内容

本年度は、プロジェクト開始の初年度であることから、研究領域「実用化を目指した組込みシステム用ディペンダブル・オペレーティングシステム」に参加している佐藤チーム、徳田チーム、中島チーム、前田チームと共に全体コンセプトおよびシステムアーキテクチャについて議論した。

4.2.1 全体コンセプト

コンピュータシステムがディペンダブルであるために、ソフトウェアの開発・配置・保守というライフサイクルに対して、我々は次のようなコンセプトを提案した(図 4.2.1.1 参照)。

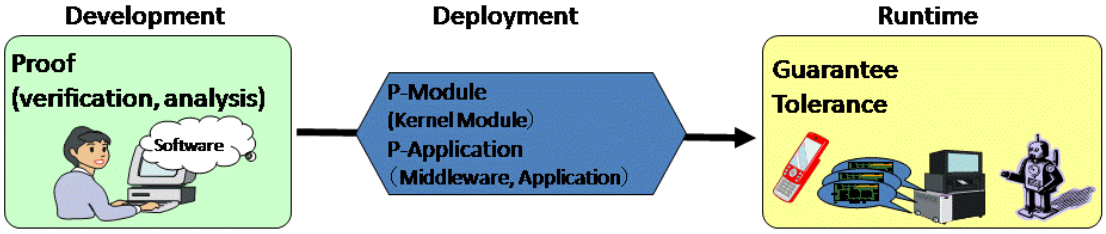


図 4.2.1.1 *-Proof 開発・配置・保守

- ソフトウェア開発時にソフトウェアの論理的機能、時間制約、脅威の可能性を検証あるいは解析する。
- 保証すべき性質を Failure-proof、Energy-waste-proof、Performance-degradation-proof、Threat-proof の4つに分類した。これらを総称して*-proof と呼んでいる。これらについては後述する。
- ソフトウェア開発時に検証あるいは解析したプログラムは、P-module (あるいは P-application)として提供される。
- 静的に検証・解析できない性質(予測不能性)は、実行時耐故障性機構により対処すると共にロギングする。

ソフトウェア開発時、実行時に保証する4つの性質についてその概要を示す。

- Failure-proof
  - ソフトウェア開発時:ソフトウェアの安全性、耐故障性検証
  - 実行時:故障を検知／予測し、故障回復、故障隔離保証
- Energy-waste-proof
  - ソフトウェア開発時:消費電力量解析、省電力復帰時間解析
  - 実行時:持続システム保証
- Performance-degradation-proof
  - ソフトウェア開発時:実時間性および性能解析
  - 実行時:実時間性および性能保証
- Threat-proof
  - ソフトウェア開発時:アタック回避検証
  - 実行時:アタック回避保証

アプリケーション分野によってこれら性質の実現方法は異なる。そこで、アプリケーション分野を携帯電話分野、OA・情報家電・センサー分野、サービスロボット・FA 分野に分け、各チームがどの分野においてどのような性質に対するディペンダビリティを実現するかを整理した(図 4.2.1.2 参照)。

図 4.2.1.2 *-proof とアプリケーション

4.2.2 システムアーキテクチャ

本領域で開発されるシステムの全体の概念レベルシステムアーキテクチャを図 4.2.2.1 に示す。以下の図で左側は開発時のツールを示している。静的プログラム安全検証ツールは前田チームが開発する。静的 WCET 検証ツールは石川研究室が CREST「ヒューマノイドのための実時間分散情報処理」(分担研究者、2008 年 9 月終了)において開発中のツールを発展させて実現する。静的 WCAT 検証ツールは徳田チームが開発する。

図 4.2.2.1 右側は*-proof を支援するためのオペレーティングシステム機構の概念図である。*-proof を実現するために、ハードウェアやソフトウェアの状態を検知する機構、検知したハードウェアあるいはソフトウェアの不具合を回復・回避する機構の大きく2つの機構を導入する。回復・回避機構は基本的な機構を提供するものであり、ポリシモジュールによって故障時の対応が決められる。P-

BUS は、検知機構、回復・回避機構、ポリシモジュール、OS カーネルとの間のインターフェイスおよびシステム全体の整合性をとるためのソフトウェアモジュールである。

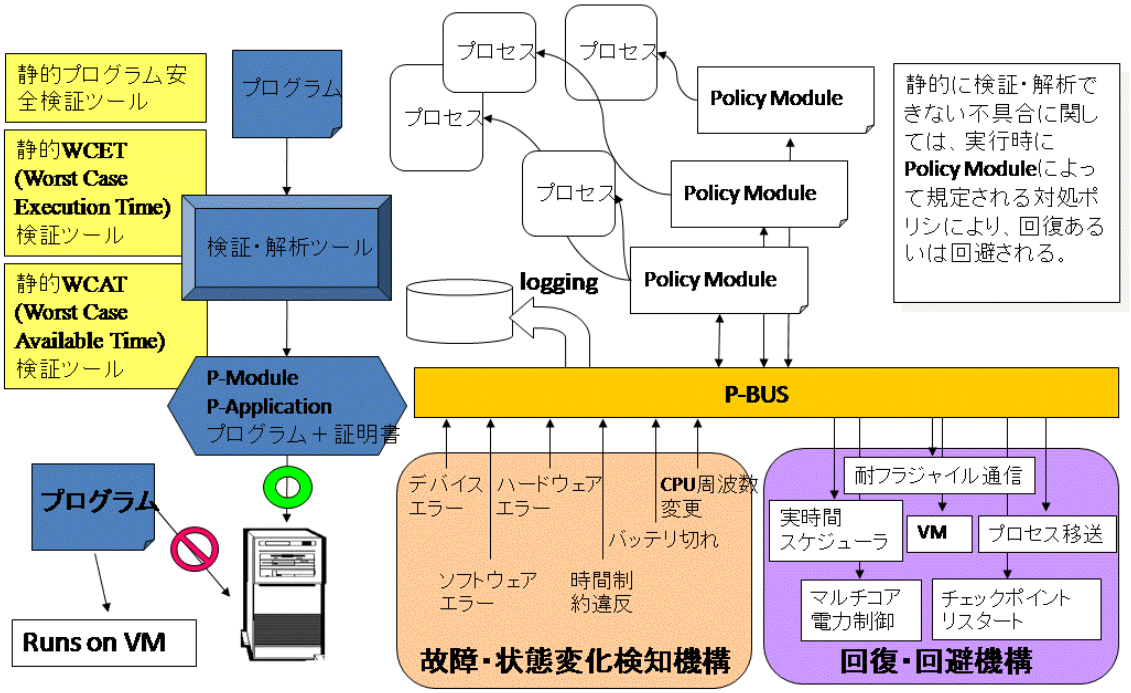


図 4.2.2.1 全体システム概念図

図 4.2.2.2 は、Linux カーネルにおけるシステムアーキテクチャである。図中、石川チームが開発を担当する部分を赤枠で示す。これらモジュールのプロトタイプを平成19年度開発する。

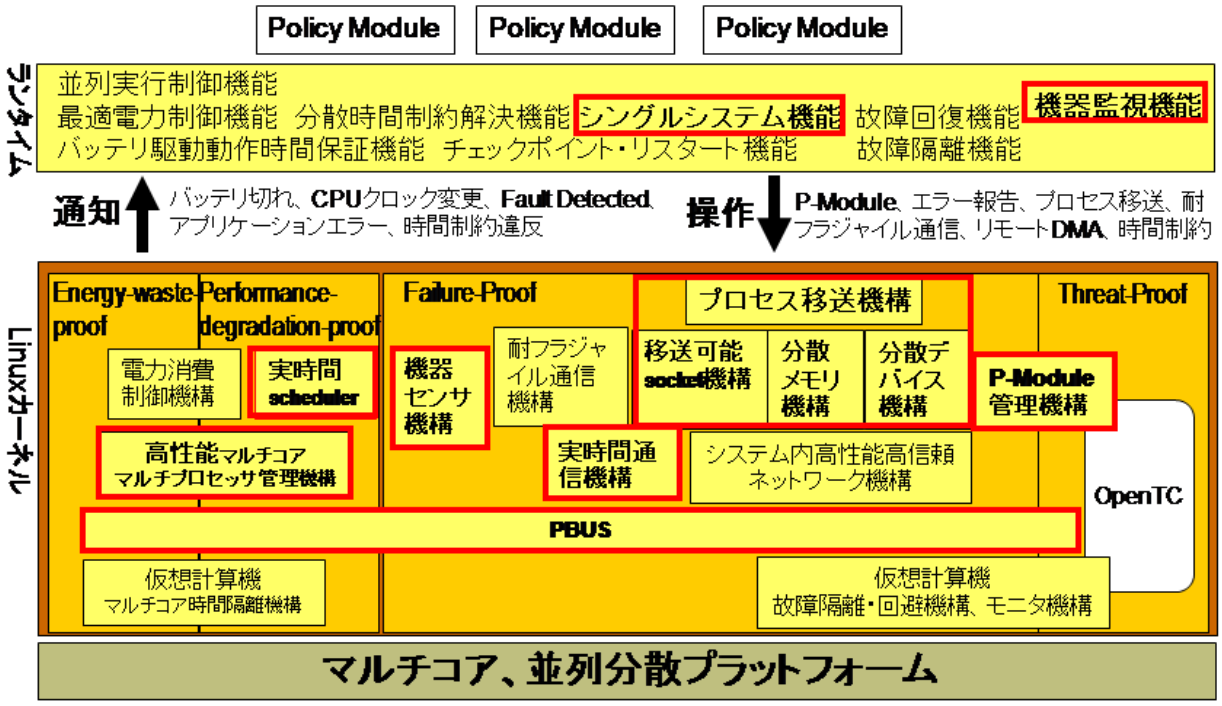


図 4.2.2.2 OS 概念図

4.3 具体的成果

本研究は、2006年10月から開始し、チーム間での連携方法、全体コンセプト作り、システムアーキテクチャの概念設計に焦点当ててきたために、外部に対する成果発表はない。

5 成果要覧

査読付論文リスト

[査読付 1] Motohiko Matsuda, Yutaka Ishikawa, Tomohiro Kudoh, Yuetsu Kodama, Ryousei Takano, "Efficient MPI Collective Operations for Clusters in Long-and-Fast Networks," IEEE International Conference on Cluster Computing, 2006. (Best Paper)

公開ソフトウェア

[公開 1] 石川裕、他：YAMPI、MPI 通信ライブラリ、http://www.il.is.s.u-tokyo.ac.jp/

その他の発表論文リスト

[発表 1] 松田、石川、工藤、児玉、高野、「グリッド上のコレクティブ通信アルゴリズム」、情報処理学会 研究報告 2006-HPC-87、pp. 257 - 262、2006.

[発表 2] 松葉、石川、「コモディティネットワークを用いた 5GByte/sec 通信の可能性」、情報処理学会 研究報告 2007-HPC-109, HOKKE 2007、pp.169 – 174,、2007.

報道関連

[報道 1] 記者会見「筑波大学、東京大学、京都大学　発　オープンスパコン仕様策定へ―大学の知を支援するスーパーコンピュータ実現を目指して―」 に対応して以下の報道があった。

毎日新聞「東大・京大・筑波大　共同で市販部品活用　超速スパコン」、9月６日、2006

朝日新聞「次期スパコン共同設計へ　東大など3大学」、9月６日、2006

読売新聞「スパコン開発　市販部品使い　東大・京大・筑波大」、9月６日、2006

日刊工業新聞「スパコン仕様共通化　東大・京大・筑波大　応用領域を拡大」、9月６日、2006

セキュリティを保証する言語処理系の研究

佐藤周行

1 概要

われわれは、「検証付最適化コンパイラ」のもと、最適化に際してセマンティクスを保証するための検証の枠組と、最適化の結果、それが性能を確実に上げていることを検証する枠組を研究してきた。

2006 年度は、その保証をセキュリティに対して行うための枠組の準備的な研究を行った。研究のプラットフォームは言語処理系に限らず、一般に言語で表現されているもの一般である。

さらに、コンパイラ技術を、最近重要度が急速に増している分散環境の上に適用することも重要であることを考え、分散環境下のコンパイル技術を研究することにした。特に、大規模分散環境下でのスケジューリングについて、研究をおこない、従来に対してデュアルな枠組が効果を持つことを実証した。

2 背景

コンパイラの研究の中心である最適化において、従来、研究の展開はアドホックなノウハウの蓄積によってなされてきた。しかし、副作用を許容する最適化、性能的にトレードオフを持つ最適化群の登場など、最適化が複雑・高度になった現在では、理論的基盤を持たなければ最適化適用の正しさの証明さえ危うくなってきている。事実 Gnu-C では、いまだに大量のバグレポートがなされている。

検証付きコンパイラ (Verifying Compiler) は、もともと Hoare の提唱したプログラムの正当性を併せて証明するコンパイラシステムとして 21 世紀の課題として提唱した概念である。本研究で目指すのは、コンパイラ最適化における上のような危機的状況に対応すべく最適化をその正しさの証明を込めて適用するコンパイラの構築であり、**最適化検証つきコンパイラ (Optimization Verifying Compiler)** と呼ぶべきものである。

従来の国内外における構成要素としての研究成果にはアセンブラの型付けによる動作正当性の証明 (Necula)、最適化の適用前後での意味保存の証明 (Necula)、最適化適用条件の時相論理等での形式化 (D Jones) を含む。さらに広義にはモデル検査の応用として動作正当性を証明することも含む。

現在、最適化検証付きコンパイラが検証することのひとつの柱は最適化適用の前後でプログラムの意味が保存されることである (Translation Validation)。証明を最適化の適用と同時に行うことを、アセンブラ言語に対する型理論として展開する。最適化適用はデータフロー解析を本質的に用いることから、データフロー理論を包含した型理論の構築につなげることができる。この最大の意義はデータフロー型の最適化の正当性を形式的に証明する体系を提示することである。また研究の基礎として、動作同一性証明に適したプログラム実行の適切な抽象化ができることになる。

また、最適化検証付きコンパイラはプログラム実行の抽象化を通して、最適化適用の最大の目

的である性能向上の理論的解析のベースになる。現実的に妥当な数理的性能モデルを仮定すると、その中で適用最適化が「真に」最適なコードを出すかどうかの検証ができる。(従来の最適化は、もとのコードと比較して、性能が良くなるかもしれないということを定性的にしか示せなかった。自動チューニング (Atlas 他) は、没理論でノウハウを適用することを目指しているものである。また、アーキテクチャ研究ではエミュレータが実質的なモデルの役割を果してきた。しかし、アンローリングなどのパラメトリックな最適化を解析するには数理的なモデルが必須である。) 性能モデルと、各種整数計画法の適用により、複数の最適化の組み合わせのどれが真に最適かを決定することができる。これはまさに性能面からみた検証であり、検証というプロセスが従来のアドホックな最適化研究に与えるインパクトの大きさを証明するものである。理論的な検証付きコンパイラが具体的な性能モデルの概念と結合して具体的な性能保証をすることは、理論と実践の両面での貢献が大きいものである。

さらに、これらコンパイラ技術が適用されるシーンについても注目している。特に、分散環境下でのプログラミングモデルには、スケジューリング、セキュリティなどの重要な視点が Web-service の中に次々と取り込まれている。これらの表現が今後重要になる。

また、要素技術としての CPU のアーキテクチャにも SIMD の再採用をはじめとして新しい波が押し寄せている。これらに対応するための言語技術も重要である。2006 年度は [編集 1] において、新しい波に対応するための言語側、特に Fortran の対応方針をまとめている。

3 内容

目的としてあげたゴールに対して 2006 年度、本研究では具体的に以下のようにテーマを設定した。

プログラムの分解と合成の高速化 プログラムの解析のために、プログラムから必要な部分を抽出する技術がスライシングである。さらに、スライシングされたプログラム（を最適化したもの）をいくつかあわせて、実際に動作するプログラムを合成する研究を進めている。

この研究は基盤情報学専攻修士 2 年の池西俊介との共同研究である。

分散環境でのスケジューリングの表現 大規模分散環境下でのスケジューリングについて、その最適化をおこなう。研究の枠組は、各計算資源が自律的に動作し、ジョブを自らの判断で取得して計算するものである。エージェント技術が背景にあり、したがって、ボトルネックになりやすい中央でのスケジューリングを廃止している。大規模分散環境が複雑系の挙動を示し始めると、機能することが期待される。

この研究は、基盤情報学専攻修士修了生の坂元賢太郎との共同研究である。

認証におけるセキュリティ表現 セキュリティを表現することが、分散環境下のサービス上でどのように応用されるか、の研究を行う。「セキュリティの表現」は、プログラミング言語を用いての表現にとどまらず、一般的にサービス記述の枠内で行うのが適当であると判断している。さらに、表現されたものから認証情報と認可情報をどう導出するかの研究を行う。

4 具体的成果

2006 年度は，それぞれについて以下の成果をあげた。

プログラムの分解と合成の高速化 検証付最適化コンパイラのための型システム [発表 2] を利用して、プログラムスライスを表現し、それをもとにプログラム合成の理論を構築した。この結果は現在ドラフトをまとめている段階である。

さらに、プログラムスライシングの効率化、特にダイナミックスライシングの効率化について、プログラム構造の利用を行うことによりよい結果を得ることができた。これは池西の修士論文としてまとめられている。

分散環境でのスケジューリングの表現 各計算資源が自律的に動作し、ジョブを自らの判断で取得して計算するジョブ処理モデルを作成し、そこで各計算資源が自らのポリシーにしたがってスケジューリングを行う枠組を提案した [査読付 1]。特に、計算資源の権限管理が複雑である現実社会における Grid 環境が上のモデルのよい例になっていると考えられる。実験は比較的小規模のもので行ったが、そこでも本方式のスケジューリングが効果を見せている。さらに大規模分散環境が複雑系の挙動を示し始めると、さらによく機能することが期待される。

認証におけるセキュリティ表現 強い認証のための基盤構築 [発表 4] と、その上での具体的なアプリケーション構築を行っている [発表 1, 発表 3]。これらの結果をもとにして、現実的な環境でのセキュリティモデルとその表現についての考察を行った。特に、データそのものが持つ認証情報と権限情報を計算環境下でどのように利用するかについて検討を行っている。

書誌編集リスト

[編集 1] Sato, H. Ed.: Proceedings of Fortran Day@東京大学, 2006. (http://www-sato.cc.u-tokyo.ac.jp/FortranDay/ProcFortranDay.html)

査読付論文リスト

[査読付 1] Sakamoto, K., Sato, H.: A Reource-Oriented Grid Meta-Scheduler based on Agents, Proc. 2007 Parallel and Distributed Computing and Networks, 109–114, 2007.

その他の発表論文リスト

[発表 1] Nishimura, T. and Sato, H.: Authentication with PKI – a Case Study in Information Technology Center in the University of Tokyo, Proc. Int'l Symp. Advanced ICT (AICT 2006), 251–256, 2006.

[発表 2] Sato, H.: "Program Analysis and Security Analysis in Optimizaton Verifying Compilers," Proc. COE Symposium on Advanced Electronics for Future Generations, –Secure Life Electronics for Quality Life and Society–, 271–275, 2007.

[発表 3] 西村, 佐藤："レガシー Web アプリケーションに対応する PKI を用いた簡易 Single Sign-On の実現," 情報処理学会 2006 年度第 4 回 QAI 研究会研究会報告, 2007.

[発表 4] 西村, 佐藤："自立的組織の集合体としての大学における PKI の運用," 情報処理学会第 70 回全国大会, 2007.

創薬・バイオの新基盤技術開発に向けた タンパク質反応全電子シミュレーションシステムの研究

佐藤　文俊

1　概要

当グループは、タンパク質をターゲットに量子論に基づく精密なシミュレーションシステムを開発している。ベースとなるソフトウエアは、大規模な系を取り扱うことができるよう独自に開発した、ガウス型関数を用いた密度汎関数法プログラム ProteinDF である。これを用いてこれまで 2000 年に 104 残基の金属タンパク質シトクロム *c*、「戦略的基盤ソフトウエアの開発」プロジェクトで 306 残基のインスリン 6 量体の全電子波動関数計算に世界で初めて成功した実績がある。本システムの研究開発の目的は ProteinDF を基に、タンパク質の新しい研究法、ならびに創薬やバイオ素子の設計などへの応用に必要な機能を追加・拡張し、インフラを整備し、創薬・バイオの新基盤技術開発に向けたタンパク質反応全電子シミュレーションシステムへと発展させ、これを公開することである。本システムは以下の 4 つの研究開発項目（サブグループ）、(1)密度汎関数法による超大規模タンパク質全電子計算エンジン、(2)タンパク質の構造最適化、分子動力学、自由エネルギー計算機能、(3)タンパク質波動関数 DB の大幅更新、(4)シミュレーションを支援し、物性を評価・解析できる高品位 GUI に関する諸研究開発成果を統括したシステムである。また、次年度から密度汎関数法－配置間相互作用法による電子励起・電子移動計算機能の研究開発を開始する。本システムは ProteinDF を含め計算単位が独立なオブジェクト指向言語 C++で構築されるため、各研究開発項目は独立かつ安全に開発することができる。今年度は、各サブグループの β 版の作成を行った。(1)ハイブリッド汎関数、新高速化手法、新初期波動関数作成法および新収束法の各開発とリファクタリング、(2)構造最適化機能の精密化、何種類かのインスリン 2 量体解離シミュレーションおよび自由エネルギー機能の設計、(3)20 種類のタンパク質波動関数の追加、相互作用解析法の開発(4)タンパク質構造モデリング機能の強化、反応シミュレーション解析のための新しい表現方法の開発を行った。

2　ProteinDF システムの開発

2.1　背景

創薬技術、バイオテクノロジーの発展は世界各国がしのぎを削る戦略目標である。その定番の手段となったゲノム情報の比較研究は、病原や薬剤に関与するタンパク質の同定などに威力を発揮するが、設計の決め手となる薬剤の効率やタンパク質の性能を調べることができない。そのため、タンパク質の反応シミュレーションは本来創薬に欠かすことができない基盤技術である。しかし、これまで実用的で信用に足る計算方法やそれに基づくソフトウエアは確立していなかった。タンパク質は電子移動、光励起、化学反応などの反応を制御して機能を発現するため、電子の働きが最も重要である。したがって、薬剤の設計には薬剤、基質、タンパク質を対等に取り扱う精密で大規模な量子化学計算によるシミュレーションが必須である。また、タンパク質は反応のプロセスで変形運動をする。この部分はニュートン方程式を用いて追跡するのが定石である。タンパク質どうしあるいはタンパク質と溶媒の相互作用には統計力学が必要である。このようにタンパク質の反応シミュレーションは必然的にマルチスケール・マルチフィジックスであり、その技術開発そのものも世界の戦略目標である。

我が国は文部科学省 IT プログラム「戦略的基盤ソフトウエアの開発」プロジェクト（前プロジェクト）などを通して、この分野のソフトウエア開発において世界をリードしている。その中で当グループは最先端の情報科学技術を駆使してソフトウエア ProteinDF を開発し、大規模タンパク質（306 残基のインスリン 6 量体）の全電子計算に成功している。ProteinDF がベースとする密度汎関数法はタンパク質のような金属とペプチド鎖を持つ分子の計算で電子相関を取り込んだ精密な計算ができるうえ、最近では、反応解析に必要な励起状態の計算も実験値を良く再現することが明らかになっている。間もなくペタフロップス超級の汎用スーパーコンピュータが登場する 21 世紀初頭の今、タンパク質の反応を全電子シミュレーションによって追跡できる環境が揃いつつある。

2.2 内容

本研究の目的は、ProteinDF を発展させて、タンパク質反応の全電子シミュレーションシステムを世界に先駆けて開発することである。この新しい手法は創薬の信頼性の高い基盤技術となる。実際に、超々速効性のインスリン設計などのシミュレーションを通して、その実用性も検証する。基礎研究のみならず、医薬品開発の高品位化、高効率化ならびに次世代の医薬品研究開発モデルの創成といった応用に貢献しうる本システムの開発はまさに急務の課題である。方法特許やビジネスモデル特許などを含め欧米に対するリードに寄与することも期待できる。触媒、分子素子、環境物質などへの応用も大いに期待してよいだろう。Fig. 1 に本プロジェクトで研究開発するタンパク質反応全電子シミュレーション・システム（ProteinDF システム）の概念図を示す。本システムは以下の(1)～(4)の研究開発項目、および次年度開発する密度汎関数法－配置間相互作用法を統括したシステムである。

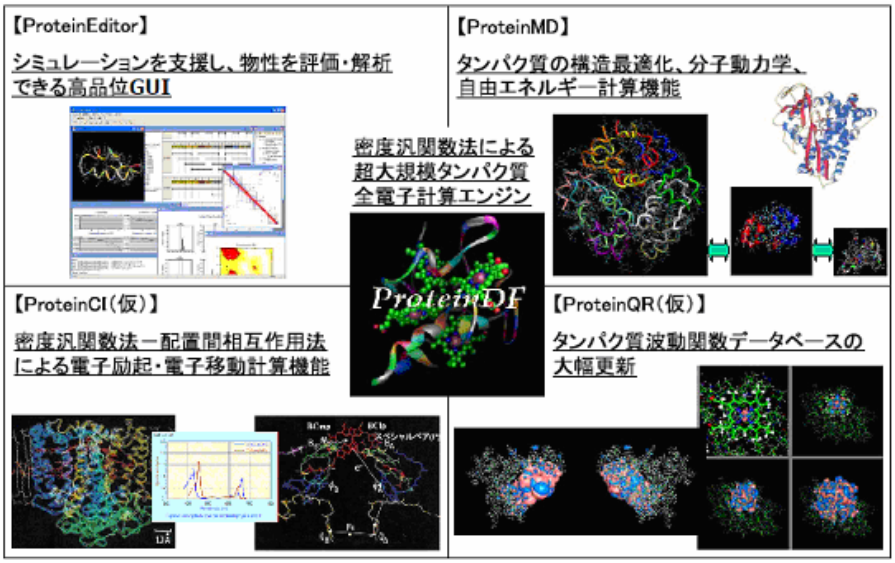


Fig. 1: Concept of ProteinDF system

2.3 具体的成果

(1) 密度汎関数法による超大規模タンパク質全電子計算エンジンの開発

プログラム名：ProteinDF, QCLO

a) ハイブリッド(混合)汎関数の開発

これまで ProteinDF は SVWN などの局所密度近似汎関数を中心に、タンパク質全電子カノニカル波動関数計算をサポートしてきた。近年、一般分子の密度汎関数計算では、ハートリー・フォック(HF)交換項や非局所密度近似補正も組み合わせたハイブリッド汎関数が使用されており、なかでも B3LYP ハイブリッド汎関数は数多く実験結果と比較が行われ、その高精度の結果から業界標準となっている。本年度はこのハイブリッド汎関数の実装を行った。ハイブリッド汎関数の一つである Becke3 パラメータによるハイブリッド汎関数は、以下の式により定義される。

$$E_{XC}^{B3LYP} = AE_X^S + (1-A)E_X^{HF} + BE_X^{B88} + CE_C^{LYP} + (1-C)E_C^{VWN} \quad (1)$$

ここで E_X^{HF} が HF の交換項である。HF の交換項を見積もるためには 4 中心 2 電子積分を計算する必要がある。ProteinDF では resolution of identity（RI）法を採用しており、3 中心 2 電子積分を高速チューニングしていた。RI 法を複数回使用して 4 中心積分計算を行うこともできるが、基本的に短距離力である交換項の見積もりでは 4 中心積分を直接計算しても、高速に計算できるため、これを Obara-Saika の積分公式を用いて実装した。

(i) コーン・シャム行列生成アルゴリズム

コーン・シャム行列生成アルゴリズムには、インテグラル・ドリブン、インデックス・ドリブンといった積分の対称性を利用したり、インデックスの独立性を利用したりするアルゴリズムがある。しかし、前者は通信量が多いため並列化の効率が出ず、後者は前者の 8 倍計算しなければならない。これに対し、最近、RT-並列アルゴリズムが提案された。インテグラル・ドリブンアルゴリズムに比べ 2 倍の積分計算が必要となるが、通信量はインデックス・ドリブンアルゴリズム並みに大幅に減少するため、大規模分子を並列計算機にて計算する場合に適している。これは後述のスクリーニングをうまく利用している。ProteinDF ではこのアルゴリズムを採用・実装した。

(ii) スクリーニング

計算すべき 2 電子積分の個数を少なくする手段として、積分値の極めて小さな積分計算を省略する方法(スクリーニング)は極めて効果的である。これにより、4 中心 2 電子積分の計算量をN^4からN^2~Nに減らすことができる。具体的には、シュワルツの不等式によるカットオフ、Direct SCF法と相性の良い密度行列によるカットオフ、長距離相互作用によるカットオフといった方法を組み合わせて、積分すべき個数を絞り込む。これらのスクリーニング手法は、前述の行列生成アルゴリズムとの適切な組み合わせにより、より効果的となる。これらのスクリーン手法も実装し、大規模分子系において、精度良くパフォーマンスを発揮できるスクリーン手法の組み合わせ・チューニングを行った。

以上の 4 中心 2 電子積分の開発・実装に基づき、数残基ペプチドなどでテスト計算を行った。今年度の実装段階では、実行速度や並列化の点でさらに改善できる余地は残されているものの、ProteinDF において HF 計算ならびにハイブリッド汎関数を利用した DF 計算が実現可能となった。

b) ProteinDF の改良・機能追加

(i) 初期波動関数作成法の追加

これまでの電子密度合成による初期値作成に加え、コアハミルトニアン、拡張ヒュッケル法による初期波動関数の作成を追加した。これまで通り QCLO 等で作成された波動関数も利用できる。

(ii) 収束法の追加

収束アルゴリズムに、アンダーソン法を追加した。この方法は、簡単で計算量が少ないながら効果が高い。インスリン 6 量体の SCF 計算は、このアンダーソン法を使って収束解を得た。

(iii) ソースコードのリファクタリング

ソースコードは ISO/IEC 14882:1998 準拠の C++で記述している。これまでオリジナルのコードで記述されている部分は、なるべく標準ライブラリを使用するように変更した。行列の線形代数演算については、LAPACK などのパッケージが有効利用できるよう改良した。

これらの成果により並列化率は 99％を超え、100CPU クラスの並列化はすでに十分となり、1,000CPU クラスの並列化が手に届くところとなった。このような高い並列化率をもつカノニカル分子軌道プログラムは大変稀であることを強調したい。

(2) タンパク質の構造最適化、分子動力学、自由エネルギー計算機能の開発

プログラム名 : ProteinMD

a) 構造最適化

タンパク質の構造最適化に必要な機能の実装とテスト計算を行った。特にサイズが大きい分子の電子構造はエネルギーギャップが小さく、SCF 計算が不安定になりやすい。このため構造最適化の

計算実行時に SCF 計算が不安定になり収束しなくなるという問題が発生する。SCF 計算が不安定になる原因は、コーン・シャム行列を計算するのに用いているアルゴリズムの精度があまりよくないことである。コーン・シャム行列の誤差 $\Delta \boldsymbol{F}$ に対する許容範囲は以下の式で与えられる。

$$\left| \boldsymbol{V}^{+} \Delta \boldsymbol{F} \boldsymbol{V} \right| < \Delta\varepsilon \left| \Lambda \right| \ . \qquad (2)$$

ここで $\Lambda, \boldsymbol{V}$ は重なり積分 $\boldsymbol{S}$ の固有値と固有ベクトルである。$\left| \boldsymbol{V} \right|$ のオーダは1、$\left| \Lambda \right|$ は 10^{-2}～10^{-3}程度、エネルギーギャップ $\Delta\varepsilon$ は 10^{-1}程度なので、コーン・シャム行列に要求される精度は 10^{-4}程度であることがわかる。交換相関項の計算アルゴリズムについて改良を行った。グリッド数値積分は、動径方向についてはLogスケーリング関数と組み合わせたEuler-Maclaurin求積法を用い、角度方向にはGauss-Legendre求積法を使用した。これにより、HeやNeなどの単原子分子の電子密度の積分は 10^{-10}程度と高精度であり、多原子分子においても 10^{-4}程度の精度を得ることが可能になった。テスト計算を行った結果を以下に示す。テスト分子にはACE-GLY-NME、交換相関汎関数にはSVWNとB3LYPを用い、構造最適化のアルゴリズムには準ニュートン法を使用した。エネルギー変化のグラフをFig. 2 に示す。ともに良好にエネルギーが減少しており、エネルギー極小化に成功した。

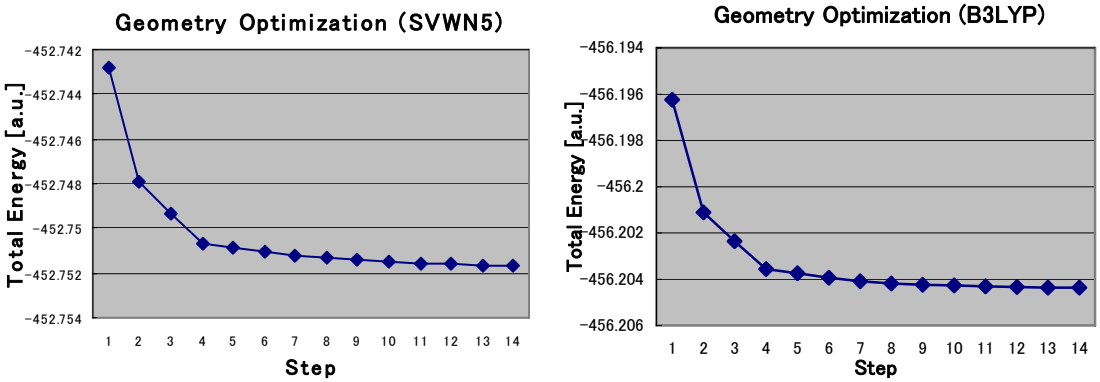


Fig. 2: Geometry optimization profile of ACE-GLY-NME
by using (left) SVWN and (right) B3LYP

b) 分子動力学シミュレーション

分子動力学シミュレーションの具体的な研究対象として、インスリン 2 量体の会合・解離をターゲットにした。インスリンは血糖値を降下させるホルモンタンパク質で、ヒト型インスリンには様々なアナログ製剤が開発されている。これらアナログは投与後効果が現れる速度が違い、これは単量体への解離速度が異なるからと考えられる。そこで、天然のヒト型インスリンとそのアナログであるリスプロ、アスパルト、グラルギンの4種類のインスリンそれぞれの 2 量体の分子動力学法シミュレーションを 10ns 実行して、これら 2 量体の安定性を解析した。この結果、2 量体の構成に重要と思われるβシート部位の構造の安定性に違いを見出した。今後、定量的な解析のために、自由エネルギー計算を実行する必要がある。十分なサンプリングを行うには上記シミュレーションの 10 倍以上のシミュレーション時間が必要である。AMBER では PentiumD3.0GHz を 8Core 使用した並列計算でも 8 ヶ月も必要となってしまう。この問題を解決すべく、クーロン相互作用高速専用計算ボード MDGRAPE-3 を利用する方針を取った。データ構造が全く異なる現在の ProteinMD でシミュレーションのパフォーマンスを落とさないよう MDGRAPE-3 専用の ProteinMD 特別版開発を開始した。特別版では原子の位置座標や力場パラメータなどのシミュレーションでのデータすべてを 1 つのコンポーネントが管理し、各計算ルーチンとそのデータコンポーネントとは STL スタイルに準拠して、イテレータを介したデータの遣り取りを行う。これにより、既存の ProteinMD とのスムースな統合を図る。シミュレーション実行中での不要な分岐をできるだけ排除し、計算パフォーマンスの向上を行っている。

(3) タンパク質波動関数データベースの大幅更新

プログラム名 : ProteinQR（仮）

a) 波動関数データベース

前プロジェクトにおいて構築したタンパク質波動関数データを質量ともに大幅に更新している。今年度は昨年度計算した 40 種類に新たに 20 種類を追加した。このペースを継続することにより、本プロジェクトの目標である 100 種程度の DB を完成させることが可能である。

b) QCLO と相互作用解析

タンパク質の密度汎関数法による全電子計算を達成するためには、精度の高い初期値が求められる。そのため、独自に擬カノニカル局在化軌道（QCLO）法を提案してきた。QCLO は一種の局在化軌道であり、その軌道は任意の単位では局在化しているものの、その単位内ではカノニカルな性質を持つ。これまでタンパク質の全電子計算では、タンパク質の一次構造に沿って各アミノ酸残基の QCLO を繋ぎ合わせることにより、大きなペプチド鎖分子の高精度の初期値を作成できた。タンパク質の相互作用には、静電相互作用、水素結合、疎水性相互作用などがあり、タンパク質の高次構造の保持に大きく寄与していることが知られている。これらの相互作用を取り込むことにより、さらに効果的に QCLO による初期値が作成できることが期待される。タンパク質の立体構造に基づいた新しい QCLO 法とともに、計算化学的手法を用いた相互作用の評価方法の開発を行った。

ここでは、グルカゴン(PDB-ID 1BH0、29 残基、原子数 470、基底関数 2,650、補助基底関数 4,776)の計算を例に報告する。8-21,23-24 番目の残基がαヘリックスを形成している。従来法(Case 1)では、最終ステップの1つ前の STEP 3 でおよそ 7 残基ずつ単純にタンパク質を分割して計算を行った。相互作用を考慮した QCLO 法(Case 2)では、αヘリックスを形成する 8-21,23-24 番目の残基をそれぞれ1つのフレーム分子として計算を行った。Fig. 3 に最終 STEP におけるマリケン電荷の初期値と収束値の差を示した。この差が小さいほど、初期値の精度が良い。Case 1 では比較的大きな(0.05 以上)差はαヘリックスを形成する部分に多く見られた。それらのほとんどの原子は、αヘリックスを形成する水素結合に関与する原子であった。一方で Case 2 では、8-21 番目の残基で形成されているαヘリックスの部分のマリケン電荷の差が改善されている。23-24 番目の残基近傍で見られる差は、αヘリックス内の原子とαヘリックス外の原子との水素結合と 2 つのαヘリックス間で形成された水素結合で、本ケースでは1つのフレーム計算からはずれたことが原因である。

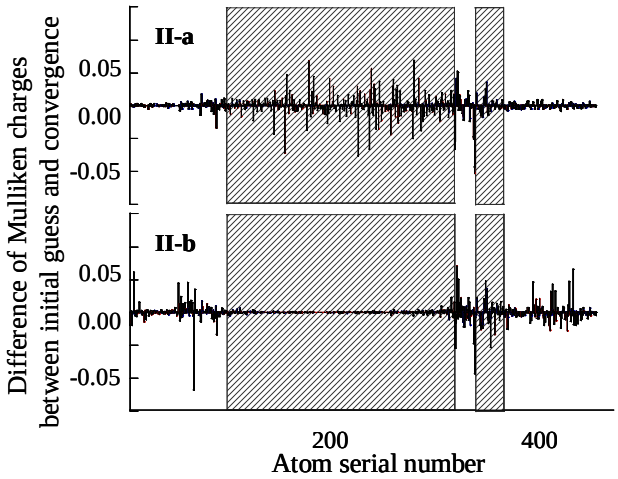


Fig. 3: Difference of Mulliken atomic charges between the initial guess and convergence in the STEP 4 of Glucagon. The shaded area represents Alpha-helix.

この結果を基に、相互作用を考慮した QCLO 法を、ステップ間のフレーム間相互作用エネルギーの差$|\Delta E_{INT}|$を基に評価した。STEP3における各ケースのフレーム間相互作用$|\Delta E_{INT}|$は Case

1,Case 2 でそれぞれ 60.338、45.037 kcal/mol であった。Case 2 ではフレーム分子の分割方法としてより理想的である。塩橋、β シートなどでも同様の結果を得た。これらの結果から、新しい QCLO 法は有効であると確認できた。この手法を利用すれば、あらかじめフレーム間相互作用が小さくなる理想的なフレームの分割方法を知ることができ、計算のトライ&エラーを減らすことが可能である。

(4) シミュレーションを支援し、物性を評価・解析できる高品位 GUI の開発

プログラム名 : ProteinEditor

今年度は計算規模の増大に伴う ProteinDF のリファクタリングを行なった。a)や b)の項目の機能強化は大規模計算に必要となるため、特に重点を置いた。

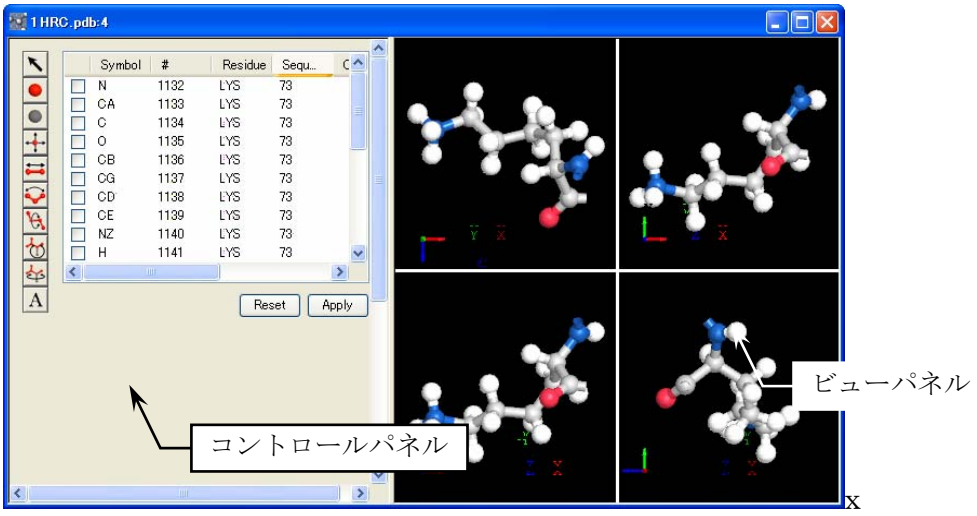


Fig. 4: A snap-shot of ModelingEditor（LYS73 of Cytochrome *c*: 1HRC）
Left pane (Control Panel): Operation functions for structural editing.
Right pane (View Panel): Display for structural editing

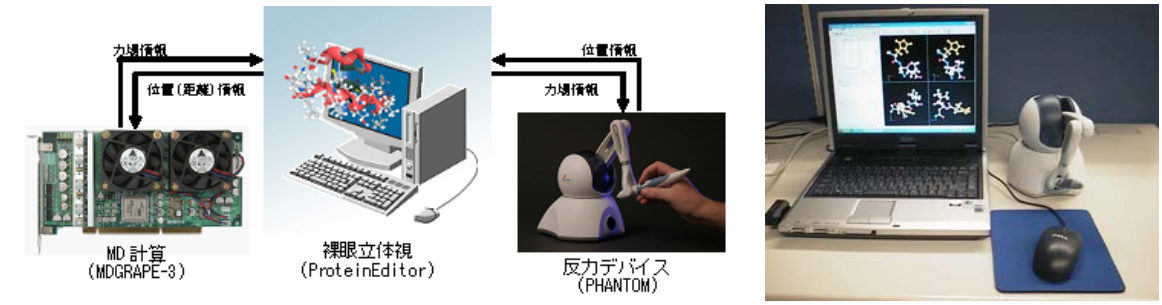


Fig. 5: A new protein modeling VR system based on ProteinEditor
Left: Concept of the VR system,
Right: Cooperation between Haptic-Device and PC comprising the solid naked eye view

a) タンパク質構造モデリング機能の強化

すでに実装されている水素付加やアミノ酸置換といった機能の高品位化とともに、シミュレーションに必須のモデリング機能を作成した。タンパク質計算構造決定は結果に影響を与えるため大変重要である。精細でフレキシブルな構造を作成するためには、分かりやすい操作で細部にわたって構造の編集ができることが必要がある。そこで、タンパク質構造のためのモデリング機能 ModelingEditor を実装した。Fig. 4 に操作画面を示す。分子グラフィックスで選択された構造に対して原子位置、結合距離、結合角、二面角などの変更ができる。分子動力学法や構造最適化法で局所解に落ち込み除去できなかった構造ひずみを対話的、かつ視覚的に編集することが可能になった。

これをベースに Virtual Reality（VR）技術を取り入れ、複雑なタンパク質立体構造の 3 次元空間における原子位置関係と原子間に生じる力の場を正確に把握し、対話的な構造編集を支援する機能として、一部生産技術研究の選定研究の補助も受け「タンパク質モデリング VR システム」の開発を行った。従来の分子モデリングでは、奥行き感があまり得られない 2 次元的な表示と平面的な動きしか表すことの出来ないマウスを用いることが多かったため、編集する構造の位置を直感的に決定することは困難であった。また数十原子以下の一般分子ではその必要性も低かった。Fig. 5 はタンパク質モデリング VR システムの構成と試験的な分子編集例である。反力デバイスを使用してタンパク質の異常構造を変異・修正させるが、その時に修正部位にかかる力を MD で計算し、デバイスにフィードバックさせる。分子グラフィックスの立体視表示とこの触感によって 3 次元構造と場を正確に認識し、直感的かつ精密に構造を修正することができる。対話的なモデリングを実現するにはリアルタイム性が重要となるため、タンパク質の力計算には MD シミュレーション専用計算機 MDGRAPE-3 の組み込みも行っている（H19 年度継続）。裸眼立体視の表示は、SHARP 製 3D 液晶を搭載したノート PC もしくは液晶ディスプレイで表示が行える。SGI 製の立体視ライブラリ ISL に対応したハードウェアにも対応できるように開発を進めている（H19 年度継続）。

b) 反応シミュレーション解析のための新しい表現方法

反応シミュレーションでは解析対象が動的になるため、データ量も物理量の種類も格段に増える。反応過程における分子軌道変化をストレスなく観察できる表現手法を設計した。本年度は分子構造変化表示は可能となったが、分子軌道変化を表示するにはさらに工夫が必要である。タンパク質表面のポケットの表示方法についても研究を行った。タンパク質（酵素）の表面や内部にはポケットと呼ばれるくぼみや空洞（鍵穴）があり、作用を受ける基質分子（鍵）はポケットに結合して適切な化学変化を受ける。一般に、この活性部位の立体構造は、鍵穴と鍵の関係のように特定の基質とぴったり合うようになっているため、ポケットの認識は非常に重要なステップである。本研究ではLIGSITEのアルゴリズムに基づくポケットの検出及び表示を開発した。このアルゴリズムではポケット検出が非常に高速に行うことが出来るが、反応中心となるポケット以外にも大小様々なくぼみも検出されてしまうため、その中から反応中心と考えられるものを選定する必要がある。また、ポケットの構造は複雑な上にタンパク質内部に隠れた状態になっていることがあり、構造を粗視化するような手法ではポケットと基質の結合状況、さらにはタンパク質立体構造との関係を正確に認識することが難しい。このような問題を解決するために没入型立体視システムによる表示を試みた。Fig. 6 にHoloStage™を用いて実験的にタンパク質のポケット表示を行った際の実験風景を示す。体験者の立体メガネに付いた位置センサにより、体験者の視点に同期して表示が移動・回転するため、体験者がシステム内を自由に移動し、視点方向を変えることによって表示されている構造の裏側もリアルタイムに確認できた。

Fig. 6: Display of a protein pocket by VR system: HoloStage™

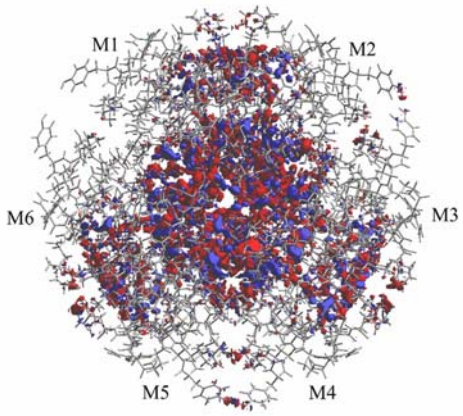


Fig. 7: Three-dimensional distribution of electron the density difference between insulin hexamer and its six monomers

(5) 基本的実証例題による機能の検証

a) ProteinDF の性能

RI法に基づく局所密度汎関数法の並列エンジンが完成した。計算律速は、分子積分演算、交換相関ポテンシャル係数を求める数値積分、そして行列演算である。同時にこれらの計算ルーチンは行列要素の保持に大量のメモリ領域を要する。例えば、インスリン 6 量体では 1 つの行列のサイズが倍精度実数で 5 GBにのぼる。分子積分演算は、全ての行列要素を各CPUに分散保存させながら、積分カットオフによる計算プレスクリーニングを組み合わせたダイレクト法によって、計算に要するワークエリアを圧縮してSPMDを達成するという独自のアルゴリズムで並列化した。交換相関ポテンシャル積分は多中心の数値積分であるが、一中心の数値積分の和に変換して計算するため並列化は容易で、原子毎にタスクを分割して計算する。行列演算は対角化、行列積、逆行列であり、これらの演算にはScaLAPACKの密行列用の関数を用いた。これも行列要素を分散保持するため、全ての演算は 1CPU毎に超巨大メモリを必要としないプログラム構造となった。306 残基、原子数 4,728、電子数 18,552、基底関数の数 26,790 のインスリン 6 量体の全電子状態計算を行なった。使用した計算機は 64 個のItanium2 プロセッサがNUMAlink^{TM}（3.2 GB/s）で結合されたSGI Altix3700 で、ピーク性能は 333 GFLOPSである。QCLOによる初期値作成法とアンダーソンの収束加速法により、26,790 次元のSCF計算を 17 回の繰り返しで無事に収束させるまで高速化に成功した。全計算時間は 65 時間で、82%の並列化効率が得られた。これは 99%以上の並列化率に相当する。

b) インスリンの電子構造

糖尿病の治療には注射によるインスリン投与がよく用いられるが、薬剤のような高濃度下では 6 量体がもっとも安定、続いて 2 量体が安定な構造となり、効くまでに時間がかかってしまう。タンパク質の自己集合には特に静電ポテンシャルが重要な働きをする。市販の即効型のインスリンは 2 量体界面の静電相互作用を小さくして、2 量体から単量体へ解離する過程が速くなるように設計されている。これまでタンパク質の理論設計には古典法しかなかったのだから、より正確な物理量が得られる量子化学計算を使用すれば、もっと高性能のインスリンが設計できないだろうか。Fig. 7 は 6 つのインスリン単量体が 6 量体を形成したときの電子分布の差である。タンパク質の電子は海のように豊潤で、会合によって波紋を作るように電子が再配置していた。このような結果は、当グループが全電子計算にこだわる理由を端的に表している。問題は、電子分布が変わってしまったため、クーロンエネルギーも大幅に変わってしまうことである。古典論の計算で用いられている力場は大変良くできているが、最も大きな割合を占めるクーロンエネルギーが変わってしまっては結果の見積もりを誤ってしまうだろう。実際、6 量体において古典論は解離エネルギーをだいぶ過小評価しており、これではインスリンの解離過程を正しく解析できない。さらにシミュレーションを積み重ねる必要があるが、量子化学計算によってそもそも単量体が安定な、投与すればすぐ効くインスリンが設計できる可能性がでてきた。

3 成果要覧

招待講演／招待論文

[招待 1] Toru Inaba, Fumitoshi Sato; “Development of Parallel Density Functional Program Using Distributed Matrix for Large Molecules”, 1st European Chemistry Congress, (2006).

[招待 2] 佐藤文俊; “タンパク質の量子化学シミュレーション”, ISM オープンフォーラム, (2006).

[招待 3] 佐藤文俊; “創薬の新基盤技術を目指したタンパク質量子化学シミュレーション：研究紹介と中期ビジョン”, 第 3 回 QuLiS シンポジウム, (2006).

[招待 4] 稲葉亨, 佐藤文俊; “密度汎関数法によるタンパク質電子状態解析プログラム ProteinDF の並列化方法とその性能”, 第 14 回 CCSE ワークショップ, (2006).

著書寄稿リスト

[著書寄稿 1] 佐藤文俊; “電子状態計算　electronic state calculation”, バイオインフォマティクス事典, （2006）626-628.

査読付論文リスト

[査読付 1] Toru Inaba, Fumitoshi Sato; “Development of Parallel Density Functional Program Using Distributed Matrix to Calculate All-Electron Canonical Wavefunction of Large Molecules”, Journal of Computational Chemistry, 28 (2007) 984-995.

[査読付 2] Toru Inaba, Naoki Tsunekawa, Toshiyuki Hirano, Tamotsu Yoshihiro, Hiroshi Kashiwagi, Fumitoshi Sato; “Density functional calculation of the electronic structure on insulin hexamer”, Chemical Physics Letters, 434 (2007) 331–335.

[査読付 3] 佐藤文俊, 稲葉亨; “タンパク質の全電子計算シミュレーションの発展”, 化学工業, 58（2007）19-23.

[査読付 4] 西村康幸, 吉廣保, 佐藤文俊; “タンパク質全電子計算のための GUI・ProteinEditor”, 生産研究, 59 (2007) 93-100.

公開ソフトウェア

[公開 1] 佐藤文俊他：ProteinDF システム、

http://www.rss21.iis.u-tokyo.ac.jp/theme/life/reaction/index.html

その他の発表論文リスト

[発表 1] Yasuyuki Nishimura, Tamotsu Yoshihiro, Noriko Nishino, Fumitoshi Sato; “ProteinEditor: Integrated Environment for Quantum Chemical Calculation System for Proteins”, International conference on Microtechnologies in Medicine and Biology, (2006).

[発表 2] Noriko Nishino, Toshiyuki Hirano, Tamio Nishimura, Soh Koike, Fumitoshi Sato: “Study of the Quasi-Canonical Localized Orbital (QCLO) Method Based on Protein Structures”, International conference on Microtechnologies in Medicine and Biology, (2006).

[発表 3] Naoki Tsunekawa, Fumitoshi Sato; “Development of a Molecular Dynamics Simulation Program for Quantum Chemical Calculations of Proteins”, XII-th International Congress of Quantum Chemistry, (2006).

[発表 4] Toshiyuki Hirano, Noriko Nishino, Fumitoshi Sato; “All-electron calculation of protein using the structure-based quasi-canonical localized orbital method”, XII-th International Congress of Quantum Chemistry, (2006).

[発表 5] Fumitoshi Sato, Toshiyuki Hirano, Naoki Ihara, Yasuyuki Nishimura, Noriko Nishino, Naoki Tunekawa, Tamotsu Yoshihiro, Hiroshi Kashiwagi, So Koike, Tamio Nishimura, Masahiro Sato, Hideaki Koike, Toru Inaba; “All-Electron Based Simulation System for Protein Reaction”, XII-th International Congress of Quantum Chemistry, (2006).

[発表 6] Tamotsu Yoshihiro, Yasuyuki Nishimura, Noriko Nishino, Fumitoshi Sato; “Development of an integrated environment of quantum chemical calculation system for proteins”, XII-th International Congress of Quantum Chemistry, (2006).

[発表 7] 恒川直樹, 伊藤宏比古, 佐藤文俊; “分子動力学法によるインスリン分子のダイナミックスの解析”, 分子構造総合討論会, (2006).

[発表 8] 西野典子, 平野敏行, 佐藤文俊; “擬カノニカル局在化軌道(QCLO)法によるタンパク質全電子計算と相互作用解析”, 分子構造総合討論会, (2006).

[発表 9] 井原直樹, 佐藤文俊; “擬カノニカル局在化軌道を用いた部分シミュレーション法の研究”, 分子構造総合討論会, (2006).

[発表 10] 西村康幸, 吉廣保, 石川寛人, 佐藤文俊; “VR 技術を取り入れたタンパク質量子化学計算システム統合環境の研究”, 分子構造総合討論会, (2006).

[発表 11] 佐藤文俊, 稲葉亨, 井原直樹, 恒川直樹, 西野典子, 西村康幸, 平野敏行, 吉廣保, 柏木浩; “RI 密度汎関数法の新しい並列化方法とその性能”, 分子構造総合討論会, (2006).

[発表 12] 西村康幸, 吉廣保, 佐藤文俊; “タンパク質量子化学計算システム統合環境 ProteinEditor の開発”, 第 23 回 CG・可視化研究会, (2006).

[発表 13] 佐藤文俊; “創薬・バイオ新基盤技術開発に向けた タンパク質反応全電子シミュレーション”, 文部科学省次世代 IT 基盤構築のための研究開発, 第 1 回「革新的シミュレーションソフトウェアの研究開発」シンポジウム, (2006).

[発表 14] 佐藤文俊; “生命現象シミュレーションへの全体取り組み方針”, 文部科学省次世代 IT 基盤構築のための研究開発, 第 1 回「革新的シミュレーションソフトウェアの研究開発」シンポジウム, (2006).

[発表 15] 佐藤文俊; “創薬・バイオ新基盤技術開発に向けた タンパク質反応全電子シミュレーション”, 文部科学省次世代 IT 基盤構築のための研究開発,「革新的シミュレーションソフトウェアの研究開発」ワークショップ(第 5 回) ―生命現象シミュレーション―, (2006).

[発表 16] 佐藤文俊, 稲葉亨, 井原直樹, 恒川直樹, 西野典子, 西村康幸, 平野敏行, 吉廣保, 柏木浩; “大規模密度汎関数法の並列化方法とインスリン 6 量体の電子構造”, 第 29 回情報化学討論会, (2006).

[発表 17] 恒川直樹, 伊藤宏比古, 佐藤文俊; “ヒトインスリン二量体の安定性に対する水溶媒の寄与”, 第 20 回分子シミュレーション討論会, (2006).

[発表 18] 伊藤宏比古, 恒川直樹, 佐藤文俊; “ヒトインスリンアナログのダイナミクスの解析”, 第 20 回分子シミュレーション討論会, (2006).

特記事項

[特記 1] 文部科学省次世代 IT 基盤構築のための研究開発:「革新的シミュレーションソフトウェアの研究開発」に参画。http://www.rss21.iis.u-tokyo.ac.jp/

報道関連

[報道 1] 日経産業新聞,「素材解析ソフトなど無料提供」, 2006 年 5 月 24 日.

[報道 2] 日刊工業新聞,「シミュレーションソフト 東大が 20 本公開」, 2006 年 5 月 24 日.

[報道 3] 化学工業日報社,「開発成果 20 本を公開へ」, 2006 年 5 月 24 日.

[報道 4] 科学新聞,「シミュレーションソフトウェア公開」, 2006 年 6 月 23 日.

数値計算の高速化に関する研究

黒田 久泰

1 概要

近年、あらゆる産業分野・学問分野で計算機の利用が急速に広まってきている。そして、大規模な問題を解くために並列計算機などの高性能計算機が利用される機会が増えつつある。しかし、ハードウェアの性能を向上させただけでは、それに見合ってプログラムの実行速度が向上するとは限らない。プログラムの速度を向上させるためには、ソフトウェアの面からのアプローチも必要であり、ハードウェアの性能を最大限に引き出すようなプログラム開発なども必要になってくる。そこで、ソフトウェアの面からプログラムの性能を向上させる方法について研究している。

- $\pi(x)$の計算の高速化

  現在、$\pi(x)$の値を早く求める方法として Gourdon の方法が知られているが、その計算方法の中には任意に決定することのできる 3 つのパラメータが存在する。その内の 1 つは利用可能なメモリ量から主に決定されるものであるが、残りの 2 つのパラメータの値は自由に設定できプログラムの実行時間に大きく影響する。そこで、具体的にパラメータの値をどのように設定すれば短時間で計算が行えるようになるのかについて研究を行った。

- 複雑な制御構造を持つプログラムの高速化

  多くの汎用プロセッサに標準で備わっている SIMD 命令を効果的に用いてプログラムを高速化する研究を行った。具体的には、条件分岐やループのような複雑な制御構造を持つプログラムに対して SIMD 命令による並列化を適用することを考え、実装を行った。その結果、プログラムによって SIMD 並列化による効果は大きく異なるものの 1.19 倍から 12.3 倍の速度向上を達成した。

2 $\pi(x)$の計算の高速化

2.1 背景

$\pi(x)$とは実数 x に対し x 以下の素数の個数を与える関数である。具体的な例として$\pi(10)$を考えると、10 以下の素数は 2, 3, 5, 7 の 4 つなので、$\pi(10)=4$ということになる。$\pi(x)$の値を求める方法としては古代ギリシアの時代から 18 世紀までは Eratosthenes の篩を用いることにより x 以下の全ての素数を列挙し、その個数を数える方法が一般的であった。その後、Legendre(1752～1833)により全ての素数を列挙することなく、大きな x について$\pi(x)$を計算する方法が提案された。

Legendre 自身、$\pi(10^6)$=78,526 と計算しているが、正確な値は 78,498 である。このように関数 $\pi(x)$は、関数自身の定義は単純であるが、複雑な計算が必要になっている。その後、1959 年に Meissel らによって新しい $\pi(x)$の計算方式が提案され、大きな x についての $\pi(x)$がコンピュータを使って求められるようになった。現在のところわかっている最大の $\pi(x)$の値は、2001 年に求められた $\pi(4\times10^{22})$である。

一方、$\pi(x)$に関連して素数定理がよく知られている。この素数定理は $\pi(x)$が Li(x)で近似できるというものであり、

$$\pi(x) \approx Li(x) = \int_2^x \frac{dt}{\log t}$$

という形で表わされる。しかし、現在のところこの近似式から $\pi(x)$の正確な値を求める方法は知られていないため、直接この式を利用して正確な $\pi(x)$を求めることはできない。

2.2 内容

$\pi(x)$を求める方法として、現在最も効率が良いとされているのは Gourdon の方法である。この方法で計算する場合、任意に決定することのできるパラメータ変数が 3 つある。ここではその 3 つのパラメータを k, y, z とする。k は使用メモリ量と計算量に関係したパラメータであり、利用可能なメモリ量によって主に決定される。y と z は $x^{1/3} \leqq y \leqq z < x^{1/2}$ を満たすように決定する。この 2 つのパラメータの値は計算時間を大きく左右する。

具体的な $\pi(x)$の計算方法は下記のようになる。ただし、Σ の中で変動する変数のうち p と q は素数だけを、n は自然数全体を動くものとする。

$$\pi(x) = A - B + C + D + \phi_0 + \sum_{i=0}^{6} \Sigma_i$$

$$x^* = \max(x^{1/4}, x/y^2)$$

$$A = \sum_{x^* < p \le x^{1/3}} \sum_{p < q \le (x/p)^{1/2}} \chi\left(\frac{x}{pq}\right)\pi\left(\frac{x}{pq}\right)、\ \chi\left(\frac{x}{pq}\right) = \begin{cases} 2 & (if\ \ x/pq < y) \\ 1 & (if\ \ x/pq \ge y) \end{cases}$$

$$B = \sum_{y < p \le x^{1/3}} \pi\left(\frac{x}{p}\right)$$

$$C = -\sum_{p_k < p \le x^*} \sum_{n: n \le z < pn,\ p_{\min}(n) > p,\ p_{\max}(n) \le y,\ n > x/p^3} \mu(n)\left(\pi\left(\frac{x}{pn}\right) - \pi(p) + 2\right)$$

$$D = -\sum_{p_k < p \le x^*} \sum_{n: n \le z < pn,\ p_{\min}(n) > p,\ p_{\max}(n) \le y,\ n \le x/p^3} \mu(n)\phi\left(\frac{x}{pn}, \pi(p) - 1\right)$$

$$\phi_0 = \sum_{n \le y,\, p_{\min}(n) > p_k} \mu(n)\phi\left(\frac{x}{n}, k\right)$$

$a = \pi(y)$、$b = \pi(x^{1/3})$、$c = \pi((x/y)^{1/2})$、$d = \pi(x^*)$

$$\Sigma_0 = a - 1 + \frac{\pi(x^{1/2})(\pi(x^{1/2}) - 1)}{2} - \frac{a(a-1)}{2}$$

$$\Sigma_1 = \frac{(a-b)(a-b-1)}{2}$$

$$\Sigma_2 = a\left[b - c - \frac{c(c-3)}{2} + \frac{d(d-3)}{2}\right]$$

$$\Sigma_3 = \frac{b(b-1)(2b-1)}{6} - b - \frac{d(d-1)(2d-1)}{6} + d$$

$$\Sigma_4 = \pi(y) \sum_{x^* < p \le (x/y)^{1/2}} \pi\left(\frac{x}{py}\right)$$

$$\Sigma_5 = \sum_{(x/y)^{1/2} < p \le x^{1/3}} \pi\left(\frac{x}{p^2}\right)$$

$$\Sigma_6 = -\sum_{x^* < p \le x^{1/3}} \pi\left(\frac{x^{1/2}}{p^{1/2}}\right)^2$$

Eratosthenes の篩を用いて$\pi(x)$を求める場合の計算時間オーダは$O(x \log\log x)$であるが、Gourdon の方法では、$O\left(\frac{x^{2/3}}{(\log x)^2}\right)$ となる。

2.3 具体的成果

プログラムは C 言語で記述し OpenMP を用いてノード内の並列化を行った。測定環境として SR11000/J1 の 1 ノード(8 プロセッサ)を利用した。x=10^{12}, 10^{13}, 10^{14}, 10^{15} において、y=$x^{i/300}$ (i は $100 \le i \le 150$ の範囲の 1 刻み)、z=$x^{i/300}$ (i は $100 \le i \le 150$ の範囲の 1 刻み)、$y \le z$ の範囲で測定した。また、k の値は使用メモリ量の関係から k=9 を利用した。y と z の値を変えて計算したときの実行時間は図 1 のようになった。

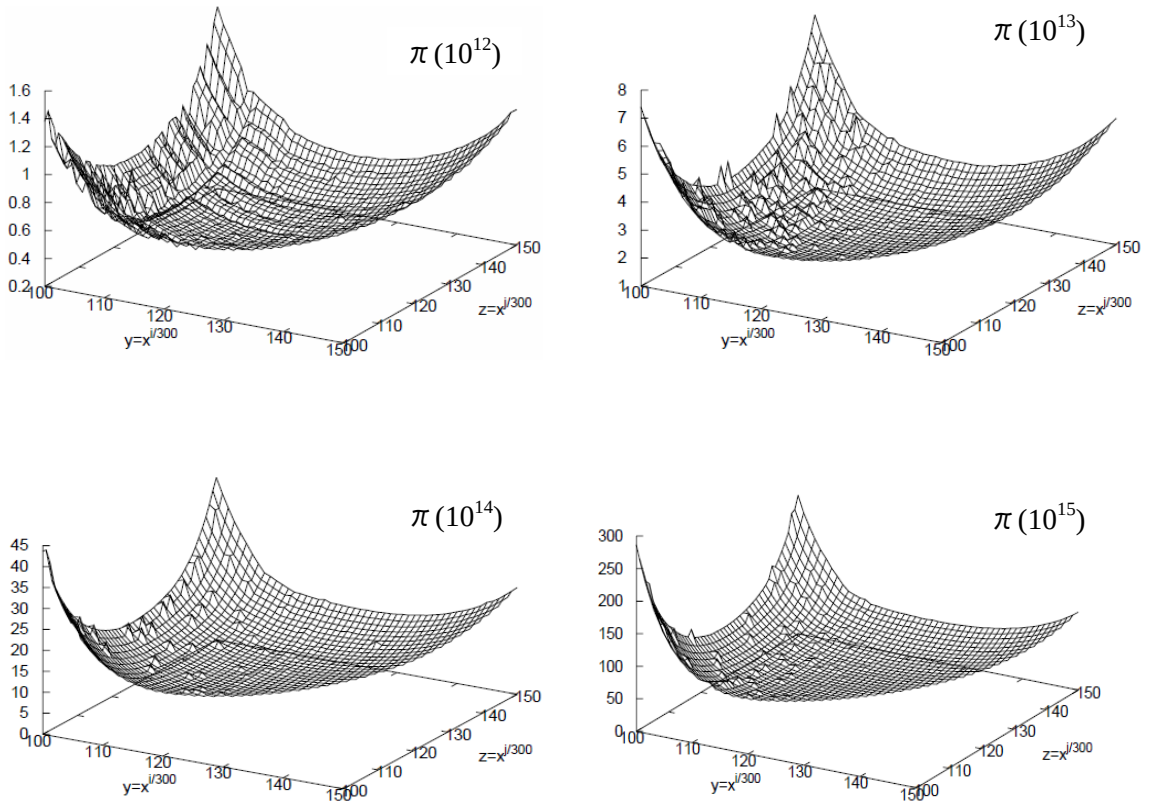


図 1: パラメータの値を変えたときの実行時間

図 1 からパラメータ y と z の値の取り方で大きく実行時間が異なることがわかる。x=10^{12}, 10^{13}, 10^{14}, 10^{15} で最適なパラメータ値を選択した場合と最悪なパラメータ値を選択した場合とでは、それぞれ 4.8 倍、6.1 倍、8.9 倍、14.5 倍の実行時間の開きが生じた。このことから x の値が大きくなるほど実行時間の差の割合も大きくなる傾向にあり、パラメータを正しく設定して計算することが重要であるということがわかる。

また、x の値が変わってもグラフの形状が似ていること、そして、実行時間が最短となるパラメータの値付近はグラフが滑らかになっておりパラメータの値が最適値から多少ずれていたとしても実行時間としてはあまり差が生じないであろうということもグラフからわかる。そのため、より大きな x についての $\pi(x)$を求める際には、それより小さな x の値における最適なパラメータの値が利用できるということになる。

具体的に x=10^{12}, 10^{13}, 10^{14}, 10^{15} について一番短時間で結果を得ることのできたパラメータの値を表 1 に示す。

表 1: 最適なパラメータの値

x	y	z
10^{12}	69183	173780
10^{12}	63095	173780
10^{13}	129817	318664
10^{13}	129817	261015
10^{13}	193493	318664
10^{13}	193493	288403
10^{13}	175119	352100
10^{13}	175119	318664
10^{13}	175119	288403
10^{13}	175119	261015

x	y	z
10^{13}	158489	261015
10^{13}	143438	318664
10^{13}	143438	288403
10^{13}	143438	261015
10^{14}	493552	844630
10^{14}	398107	844630
10^{14}	398107	758577
10^{14}	398107	681292
10^{15}	1258925	2238721
10^{15}	1122018	2238721

OpenMP を用いた並列化がどれくらい効果的であるかを調べるために、1 ノード内で使用するプロセッサを 1 個から 8 個に変えて実行時間を計測した。$\pi(x=10^{15})$の値について、パラメータを $y=x^{121/300}=1122018$, $z=x^{127/300}=2238721$ とした場合にかかった時間は図 2 のようになった。

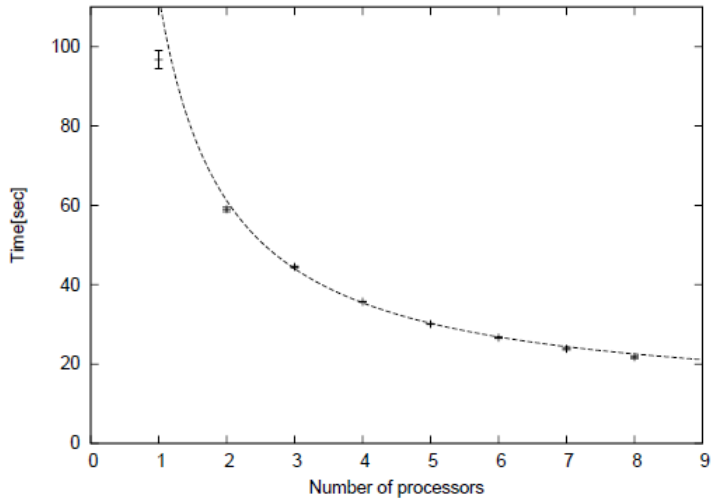


図 2: プロセッサ数を変えたときの実行時間

図 2 の破線は最小二乗法で近似した曲線を示したもので、下記の式のようになった。

$$Time\ (\text{sec}) = \frac{112.75}{0.0855 + \frac{0.9145}{N}} \qquad (N\text{はプロセッサ数}\ )$$

SR11000/J1 は物理的には 1 ノードあたり 16 プロセッサを備えているため、もし 1 ノード 16 プロセッサを使うとすると 1 プロセッサにおける実行時間と比べて約 7.0 倍の高速化が期待できることが見込まれる。

今後、これらの結果を元にしてより大きな x に対する $\pi(x)$を求める予定である。

3　複雑な制御構造を持つプログラムの高速化

3.1　背景

汎用プロセッサの多くは、複数のパックされたデータを一命令で演算可能な SIMD(Single Instruction Multiple Data)命令セットを搭載している。この命令セットはデータの並列性を利用して、大量のデータを通常の命令よりも高速に処理することが可能である。そのため、マルチメディア処理や数値計算処理の高速化に利用されている。これまで、自動解析によって SIMD 命令セットを使ったプログラムの並列化を実現するための研究が多くなされており、コンパイラによる SIMD 並列化も行われるようになりつつある。しかしながら、SIMD 命令はパックされた複数のデータに対して同じ演算しか実行できないため、要素ごとに演算内容が異なってくる条件分岐やループのような複雑な制御構造を持つプログラムは、あまり SIMD 並列化の対象として扱われなかった。そこで、そのような複雑な制御構造を持つプログラムについても SIMD 並列化の適用が可能となれば、より多くのプログラムが SIMD 命令の利用によって高速化が実現できる。

3.2　内容

SIMD 並列化を行う対象を条件分岐とループに分け、それぞれについて簡単に SIMD 並列化を適用するための手法を説明する。

条件分岐の SIMD 並列化

SIMD 命令はすべての要素に対して同じ演算を行うことしかできないので、条件成立の有無に関わらず成立時と不成立時の処理の両方を行うことで並列化を実現する。例として、次のようなプログラムを考える。

```
for(i = 0; i < n; i++){
   if(a[i] > 0)  a[i] = a[i] + 1;
   else  a[i] = -a[i];
}
```

ここで配列 a に対して、各要素の条件成立の有無に関係なくすべての要素に対して a[i]=a[i]+1 の計算と a[i]=-a[i]の計算を行う。最後に分岐条件を使って分岐内で起こった演算の結果を合成する。具体例として a[4]={5,-1,2,-3}の場合の例を図 3 に示す。a[i]=a[i]+1 の処理結果として{6, 0, 3, -2}の値を一時的に格納しておく。続いて、a[i]=-a[i]の処理結果として{-5,1,-2,3}の値も一時的に格納しておく。そして、最後に分岐条件を使って分岐内で起こった演算の結果を合成する。ただし、この合成の際にも SIMD 命令セットを効率良く適用させることで並列性を上げる。

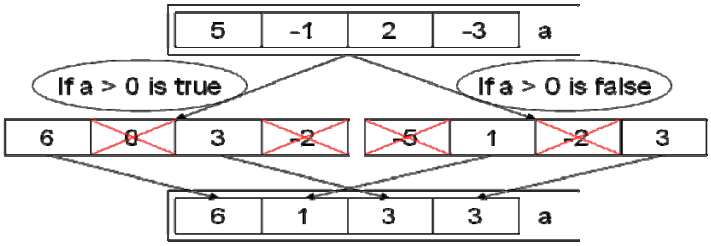


図 3: 条件分岐を含むコードの SIMD 並列化

ループの SIMD 並列化

ループについては、同時に処理される要素が全て終了条件を満たすまで、それら全ての要素に対して繰り返して処理を行うことにする。しかし、終了条件を満たした要素に対してループを繰り返すと余計な演算を行うことになり、ループ終了後に誤った結果を保持してしまうことになる。そこで、同時に処理される要素の個数に等しい真偽値の値を持ったループ条件変数を用意し、ループ継続中なら真、すでに終了条件を満たしていれば偽といったループ継続状態を保持させる。ループ中の処理は終了条件をすでに満たした要素に対しては反映させてはならないので、このループ条件変数で真の値を持っている要素に対してのみ演算を行った後の値を代入するようにする。

例として、次のようなプログラムを考える。このプログラムは、配列 a の各要素を 2 で割って 1 を引くという処理を繰り返して、0 以下になるまでの回数を配列 b の同じ添え字番号のところに格納する。

```
for(i = 0; i < n; i++){
   b[i] = 0;
   while(a[i] > 0){
      a[i] = a[i] / 2 - 1;
      b[i] = b[i] + 1;
   }
}
```

ここでは配列 a と b は 32bit 整数で SIMD 命令により 4 つずつ処理できるものとする。そして、a[4]={0, 1, 11, 8}の場合を考える。ループ条件変数は初期値{true, true, true, true}から始まり、a[i]>0 を満たさない要素のところは false に変わる。すべて false になったら while ループを抜ける。繰り返しの途中でループ条件変数が false になっている要素については、ループ中の計算結果は反映させないようにする。最終的には、b[4]={0, 1, 3, 2}及び a[4]={0,-1,-1,0}の値が格納される。この例の場合では、この 4 つの要素に対しては 3 回の反復が行われるが、各反復後における配列 a と b の値は図 4 に示すように推移する。

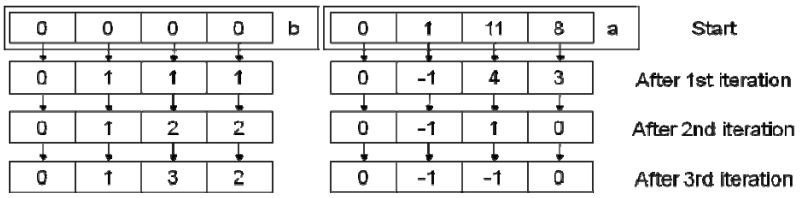


図 4: ループ部分のコードの SIMD 並列化

3.3 具体的成果

これらの手法を実装し、6 種類のテストプログラムを PowerPC の SIMD 命令セット向けに SIMD 並列化して実行時間を比較した。実験環境として PowerPC 970 1.8GHz を用いた。このプロセッサは SIMD 命令セットとして AltiVec(VMX)を備えている。テストプログラムの内容は次のとおりである。

ライフゲーム

ライフゲームを行うプログラムである。各セルにおいてその周りの生存しているセルの個数を調べて次の世代で誕生、維持、消滅のどの状態になるかを調べる。1 セルを 1 バイトで表現し、SIMD 並列化後に 16 セルの処理を同時に行えるようにした。ライフゲームの規模は 256x256 としてランダムな初期配置から 3200 世代進める、という処理を 10 回行う時間を比較した。

画像の二値化

与えられた画像の配色を閾値により二色化するプログラムである。256x256 のピクセル画像 12 枚を 1 枚に付き 2,000 回実行するのにかかった時間を比較した。

画像のエッジ検出

画像のエッジを検出するプログラムである。実験方法は画像の二値化と同様である。

画像の加算

2 つの画像の各画素に対して加算処理を行う。実験方法は画像の二値化と同様である。

ベータ分布関数

ベータ分布の計算に用いられる関数 $\beta(x, y)$の計算を並列化した。$\beta(x, y)$関数に毎回ランダムな値を渡して、合計 1000 万回実行するのにかかった時間を比較した。なお、計算アルゴリズムとして同時に計算することのできる並列数は 3 である。

マンデルブロ集合

マンデルブロ集合を計算して描画を行う。1024x1024 の領域の全体像の各画素を 128 段階調に分ける処理を 800 回行うのにかかった時間を比較した。

それぞれのプログラムについて、SIMD 並列化をしなかったものと SIMD 並列化をしたものとで時間計測を行った。その結果を表 2 に示す。

表 2: プログラムの実行時間と速度向上率

プログラム名	スカラ(sec)	SIMD(sec)	並列数	速度比率
ライフゲーム	20.78	1.70	16	12.3
画像の二値化	15.30	11.56	4	1.32
画像のエッジ検出	44.19	31.01	4	1.43
画像の加算	24.40	13.47	4	1.81
ベータ分布関数	6.89	5.78	3	1.19
マンデルブロ集合	32.63	10.95	4	2.98

ライフゲームでは 12.3 倍の速度向上が得られた。一方、ベータ分布関数では 1.19 倍といった速度向上しか得られなかった。プログラムによって速度向上には大きな違いがあるが、SIMD 並列化がうまく適用できるプログラムであれば大きな速度向上が期待できるということがわかる。

4 成果要覧

査読付論文リスト

[査読付 1] Ryosuke Yano, Kojiro Suzuki, and Hisayasu Kuroda: Formulation and numerical analysis of diatomic molecular dissociation using Boltzmann kinetic equation, American Institute of Physics, Physics of Fluids, Vol.19, No.1, pp.017103 1-13, 2007

[査読付 2] 廣松悠介，黒田久泰，金田康正：複雑な制御構造を持つプログラムの SIMD 命令セットによる最適化，情報処理学会論文誌:プログラミング，Vol.48, No.SIG 4(PRO32), pp.62-72, 2007

その他の発表論文リスト

[発表 1] Yasumasa Kanada and Hisayasu Kuroda: Towards Fast Computing Through Efficient Computation, Proceedings of COE Symposium on Advanced Electronics for Future Generations – “Secure-Life Electronics” for Quality Life and Society–, pp.257-262, 2007

[発表 2] 廣松悠介，黒田久泰，金田康正: 複雑な制御構造を持つプログラムの SIMD 命令セットによる最適化，情報処理学会研究報告 第 61 回プログラミング研究会，豊橋商工会議所，2006

[発表 3] 吉田仁，黒田久泰，金田康正： $\pi(x)$の計算におけるパラメータの選択に関する考察，情報処理学会研究報告 2004-HPC-108, pp.1-6, 2006

分散メモリ型並列計算機のための通信機構に関する研究

松葉　浩也

1　概要

近年のスーパーコンピュータの大部分は分散メモリ型の並列計算機である。この形式のスーパーコンピュータでは多くの使用用途において通信が必要となり、通信に関する機能や性能は計算機全体の使い勝手や性能に大きな影響を与える。この分散メモリ型並列計算機における通信に関して以下の２つの研究を行った。

- 複数の独立したノードから成るスーパーコンピュータやクラスタが単一の IP アドレスを持ち、インターネット上のクライアントに単一システムとしてサービスを提供するための新しい手法の研究
- 次期スーパーコンピュータで求めているノード間通信性能の実現可能性調査として、MPI レベルで 4GB/s 以上のバンド幅を得るための研究

2　単一 IP アドレスクラスタのためのプロトロルスタック

2.1　背景

近年のインターネットの普及により、情報発信を支える WWW などのサーバの負荷は高まる一方である。現在では複数のコンピュータに負荷を分散することで大量のアクセスに対応している Web サイトは珍しくない。また、サーバの負荷分散は文献検索やゲノムデータベース検索など研究分野における成果の共有システムでも重要であり、すでに数百台ものコンピュータから構成される分散環境でゲノム情報を公開している例もある。本センターにおいても、今後は大規模データを公開するためのプラットホームとして計算資源を提供することを考えていく必要があり、そのためには情報発信サーバの負荷分散技術は必須である。

複数の分散したコンピュータでサービスを提供する場合、ユーザーに単一のシステムとしての操作性を提供する方法が必要となる。そのための方法として Virtual Server と呼ばれる技術が存在するが、ネットワークの性能が生かせない問題がある。この問題を解決し高性能な負荷分散システムを構築することは一般的な大規模サーバにとっても、今後の本センターのサービスにとっても重要な課題である。

2.2　内容

本研究ではクラスタが単一のシステムとしてサービスを提供することを想定し、クラスタ全体が単一の IP アドレスを持つことを可能としつつネットワーク性能を十分に生かすことのできる、高性能オペレーティングシステムを提案した。

クラスタに単一の IP アドレスを与える機能自体は Virtual Server と呼ばれる手法が提案されており、すでに広く使われている。この手法はクラスタとインターネットの間に Load Balancer と呼ばれるコンピュータを配置し、Load Balancer がユーザーからのリクエストをクラスタの各ノードに分散して転

送する。この手法では Load Balancer がクラスタを代表する IP アドレスを持つので、ユーザーに単一 IP アドレスを提供することができる。一方で図 1 のように複数のサーバが同時に返信データを送信した際には、クラスタ内のスイッチでパケットが混雑し一部のパケットが失われる。この場合 TCP プロトコルの輻輳回避アルゴリズムによりサーバからの転送速度は低く抑えられる。このためパケットの消失を繰り返すとサーバからの送信速度は低いままとなり、ネットワーク本来のバンド幅を生かすことができない。

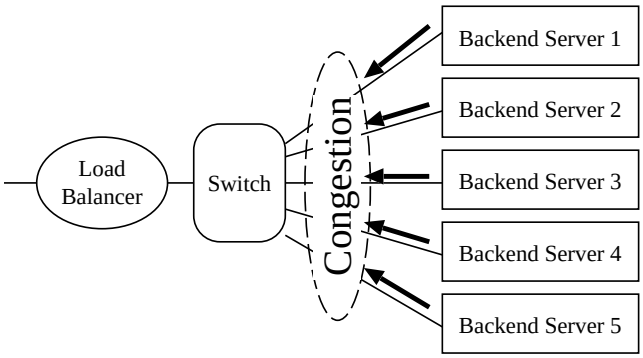


図 1: クラスタにおけるパケットの混雑

このパケット混雑による性能低下の問題を解決するため、クラスタ内では TCP/IP を用いずに通信を行う新たな方式 SAPS (Single Address Protocol Stack) を提案した。本方式では Virtual Server における Load Balancer に代わって I/O サーバと呼ばれるコンピュータを配置する。I/O サーバは Load Balancer のように単にパケットの転送を行うのではなく、TCP/IP のプロトコルハンドリングを行う。そして、サービスを提供するアプリケーションが動くクラスタノード群 (アプリケーションノードと呼ぶ) とはクラスタ専用の高速インターコネクト (転送用ネットワークと呼ぶ) を使用して通信する。本方式はクラスタ内で TCP/IP を使用しないため、クラスタ内で頻繁に発生するパケットロスが TCP/IP によって観測されることはない。そのため、TCP の輻輳回避アルゴリズムが過度に転送速度を落とす Virtual Server の問題が解決される。

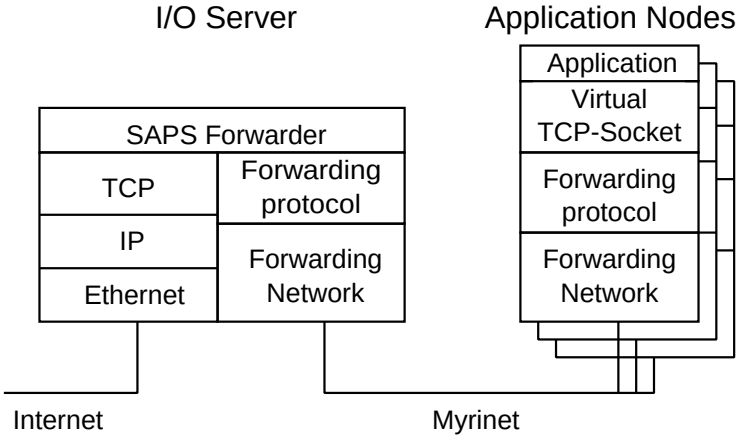


図 2: SAPS の設計

SAPS の具体的な設計を図 2 に示す。図の左側が I/O サーバであり、インターネットへの接続に使用する Ethernet デバイスとその上の TCP/IP プロトコルスタックを持つ。一方、右側のアプリケーションノードではサービスを提供するサーバプロセスが動作し、そのプロセスはソケットインタフェースを

用いて通信を行う。このソケットインタフェースを Virtual TCP Socket と呼ぶ。そして、I/O サーバ, アプリケーションノード共に転送用ネットワークとそれを扱うためのプロトコルスタックを持つ。転送用ネットワークは Myrinet や Infiniband のようなクラスタインターコネクトが想定されている。

インターネットからデータを受信した場合、まず I/O サーバが TCP/IP のプロトコル処理を行う。途中でパケットが失われていた際の再送要求などもすべて I/O サーバが行い、送信された完全なデータ列が復元される。受信されたデータは転送用ネットワークを用いて アプリケーションノードに転送される。そしてこのデータは Virtual TCP Socket を通じてアプリケーションに届けられる。アプリケーションがデータを送信する際はまず送信されるデータが転送用ネットワーク経由で I/O サーバに届けられる。I/O サーバは TCP/IP のパケットを組み立てクライアントに送信する。サーバとして動作するため TCP の状態を変更するシステムコール (listen() など) はアプリケーションノードで発行され、その要求が I/O サーバに伝えられる。I/O サーバはクライアントからの接続要求を受け取ると TCP/IP の接続を確立した上で、受信可能状態のアプリケーションノードを一個選択しサーバプロセスに新たなクライアントの接続を知らせる。

実装は Linux 2.6 を用いて行った。すべてカーネルモジュールとして実装し、カーネルパッチやユーザーレベルのデーモンなどは必要ない。上で述べたように、クライアントからのデータ受信の際、I/O サーバは TCP/IP パケットを組み立ててクライアントが送信したデータ列を復元してからアプリケーションノードに転送する設計である。しかし実装上は I/O サーバは TCP/IP のパケット単位で転送処理を行っている。TCP プロトコルスタックは順序が入れ替わったり欠落したパケットの並べ替えや再送要求を終えた後、クライアントが送信した順番にパケットを受信キューに挿入する。I/O サーバはその受信キューの先頭から順にパケットを取り出し、TCP/IP のヘッダが書かれていた部分に転送用ネットワークのヘッダを上書きした上で転送用ネットワークのデバイスに送信要求を出す。アプリケーションからの送信データも同様であり、I/O サーバは転送用ネットワークのヘッダを TCP/IP のヘッダに書き換えてクライアントに送信する。この実装では I/O サーバは各パケットのヘッダ部分を上書きするのみであり、オーバヘッドの大きい送信データ自体のコピーが発生しない。

アプリケーションノードはソケットインタフェースを提供する。既存の TCP/IP 用のソケットは、同じコンピュータに TCP や IP のプロトコルスタックが存在することを仮定して実装されているため、SAPS では使用できない。そこで新たに SAPS 用のソケットインタフェース (Virtual TCP Socket) を作り直している。このソケットは既存の TCP/IP と完全な互換性を持つ。したがって既存の TCP ソケットの実装を Virtual TCP Socket で置き換えることにより、アプリケーションを一切変更することなく SAPS を利用可能である。

転送用ネットワークには代表的なクラスタインターコネクトである Myrinet XP を用いている。Myrinet のための通信ライブラリには PM を使用し、カーネル内部から PM を使用するためのインタフェースを SAPS 用に作成した。

2.3 具体的成果

SAPS を実装し、ネットワーク性能を測定したところ、SAPS はギガビットイーサネットの帯域を使い切ることができることが示された。また、Web サーバとして使用した場合の性能を SPEC Web 2005 ベンチマークを使用して測定したところ、既存手法の Virtual Server と比較して 7.9%高速であることが示された。以下、この性能評価の詳細について述べる。

ベンチマークは 1 台の I/O サーバ、4 台のアプリケーションノード、9 台のクライアントを用いて行った。評価に使用したコンピュータの仕様を表 1 に示す。性能評価では、まず基本的なネットワーク性能を測定するためにバンド幅を測定するプログラムを作成し、サーバからクライアントに送信する際のバンド幅を測定した。次にアプリケーションを使用した際の全体の性能を測定するため、SPEC Web 2005 ベンチマークを用いて Web サーバとしての性能を測定した。結果は Virtual Server と比較した。

バンド幅測定の結果を図 3 に示す。最も負荷が高くなる大きなサイズのメッセージにおいて、SAPS

表 1: Specifications of cluster nodes

	I/O server	Applications Node
CPU	AMD Opteron 248 (2.2GHz) × 2	Xeon 2.8GHz × 2
PCI	PCI-X (64bit 133MHz)	PCI-X (64bit 100MHz)
Ethernet	Broadcom BCM5703X GE	Intel Pro/1000 Server
Myrinet	Myrinet XP	Myrinet XP

は Virtual Server と同等の性能を示している。この性能はほぼギガビットイーサネットの限界値であり、I/O サーバの負荷は TCP/IP と転送用ネットワークの双方を処理しても 1Gbps を処理できる程度に抑えられていることがわかる。

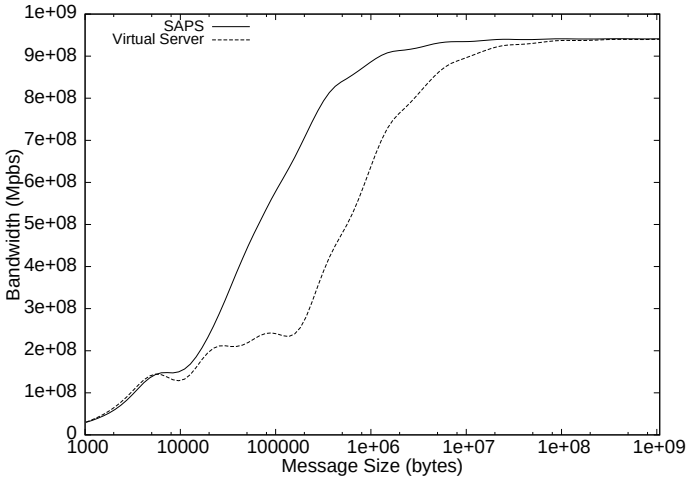


図 3: Point-to-Point Bandwidth

次にアプリケーションベンチマークである SPEC Web 2005 の結果を表 2 に示す。本ベンチマークは Web サーバの性能を測定するものであり、Web サイトとしてコンピューターメーカーのユーザーサポートのページを模擬している。ユーザーは自身が所有するコンピュータのページを検索し、修正ソフトなどをダウンロードする。表中の Simultanious Sessions は同時にこのサイトを使用しているユーザー数、Requests は 5 分間に発生したユーザーからのアクセス数、TimeGood は 3 秒以内に応答された数、TimeTolerable は 5 秒以内に応答された数である。表 2 の結果をまとめると、SAPS は Virtual Server と比較して、最大 30%多くのアクセスを 3 秒以内、13%多くのアクセスを 5 秒以内に処理し、全体で 7.9%多くのアクセスを受け入れた。図 3 はベンチマーク中のネットワークの状態である。Retransmission は TCP が観測したパケットロスの数であり、SAPS は Virtual Server の 100 分の 1 程度であることがわかる。Bandwidth はサーバからクライアントに送信されたデータのバンド幅の平均であり、SAPS は Virtual Server よりも有効にネットワークを使用していることがわかる。これらのネットワーク性能差がアプリケーション性能の差の原因である。

本研究の貢献は、クライアント、サーバ双方のアプリケーションの互換性を保ちつつ、クラスタ内の通信に TCP を使わない方式を提案したことである。これにより、クラスタ内で発生しがちなパケット混雑の影響を低減することに成功した。

表 2: SPEC Web Results

	Simultanious Sessions	Run	Requests	TimeGood	TimeTolerable	Error
SAPS	2800	1	75569	88.5%	97.6%	0
		2	75777	88.8%	97.7%	0
		3	75307	87.8%	97.7%	0
Virtual Server	2800	1	70227	73.6%	93.0%	13
		2	Failed	-	-	-
		3	70641	74.8%	93.3%	2

表 3: Detailed Statistics

	Retransmission (packets/s)	Bandwidth (Mpbs)
SAPS	103	930.0
Virtual Server	11984	917.8

3 コモディティハードウェアによる 4GB/s 通信の可能性の検討

3.1 背景

スーパーコンピューティング部門では 2008 年にスーパーコンピュータの更新を控えている。次世代スーパーコンピュータは「オープンスーパーコンピュータ」と命名されており、現在のユーザーにとどまることなく、より多くの人に利用してもらうことを目標としている。ユーザーの利便性向上のため、オープンスーパーコンピュータの CPU アーキテクチャは広く普及しているパーソナルコンピュータのアーキテクチャと同じものを採用することが決定している。一方で、PC クラスタにとどまらないスーパーコンピュータとして十分な性能を確保するため、ノード間を接続するネットワークには物理性能で 5GB/s 以上、MPI レベルで 4GB/s 以上という高い性能を求めている。

現在開発されている最高性能のネットワークを使用しても、単一リンクでは物理性能で 5GB/s 以上、MPI レベルで 4GB/s 以上を実現することはできない。そこでこの要求を満たすためには複数のネットワークリンクを束ねて使用するトランキングが必須となる。このクラスの性能を達成するトランキングは世界的に見ても例がないため、トランキングを実現する MPI ライブラリを実際に作成し、要求している仕様を満たすことが可能であることを確認する。

3.2 内容

本研究ではノード間接続ネットワークの例として Infiniband DDR を使用する。これは一本当たり 16Gbps のネットワーク（実際には PCI-Express のボトルネックで 13.4Gbps）であり、これを 4 本使用することでオープンスーパーコンピュータの仕様を満たすことを試みる。

3.2.1 性能測定の方法

MPI によるトランキングの実験を行うため、YAMPI と呼ばれる MPI ライブラリに Infiniband 用の通信ドライバを実装し、そこにトランキング機能を持たせる。トランキングは大きなバンド幅を必要とする際に用いられるため、本実験で作成した MPI ライブラリでも大きなメッセージを送受信する場合に限ってトランキング機能を提供するようにした。

送受信のメッセージサイズが大きい場合、MPI では一般的に「ランデブー通信」と呼ばれるプロト

コルが使用される。これは、送信側が実際のデータ転送に先立ち受信側の準備状況を確認するプロトコルであり、Infiniband のリモートメモリアクセス機構 (RDMA) を使用できる利点がある。具体的には以下の手順で通信を行う。

1. メッセージの送信要求が出されると、MPI ライブラリが受信側に対しメッセージのサイズとマッチング情報 (MPI における tag, communicator, 送信者のランク) を送信する
2. 上記のメッセージを送信した後、送信バッファの仮想アドレスと物理アドレスを固定する操作をカーネルに要求し、ネットワークカード内にページテーブルを準備する（メモリレジストレーションと呼ばれる）
3. 受信側はアプリケーションによりマッチする受信命令が発行され次第、受信バッファをメモリレジストレーションし、送信側に準備完了を伝える
4. 送信側は RDMA によってデータ送信を行い、完了後に RDMA 完了メッセージを送信する
5. 受信側が RDMA 完了メッセージを受信し次第、すべての操作の終了を上位層に報告する
6. メモリレジストレーションを解除する

送信メッセージサイズが大きい場合、上記手順にかかるほとんどの時間は手順４で占められる。そこで、手順４において RDMA 通信を行う際に通信領域をネットワークの本数で等分し、複数のネットワークを用いて送信する。メッセージサイズが十分に大きく、上記手順４以外に要する時間が無視できる程度であれば、この方法によるトランキングでネットワークリンクの数に比例して性能が向上する。

バンド幅測定実験は Opteron プロセッサを持つコンピュータを用いて行った。測定用コンピュータの仕様を表 4、コンピュータ内部も含めた接続図を図 4 に示す。

表 4: 測定用コンピュータの仕様

CPU	Opteron 2212 (Dual Core, 2.0GHz) x 2
Chipset	nVIDIA nForce Pro 3600 + 3050
Main Board	Tyan Thunder n6650W (S2915)
	BIOS version: 1.01D beta
Memory	DDR2-667 Dual Channel (2GB)
PCI	PCI-Express x8 + x8 from 3600
	PCI-Express x8 + x8 from 3050
Network	Mellanox InfiniHost III Lx (x4 DDR)
	Memfree(NIC にメモリを持たない)
OS	Fedora Core 6 (Linux 2.6.18)

実験環境で注意すべき点は、すべてのメモリが同じ速度でアクセスできるとは限らないことである。図 4 中、CPU(0) と Memory(0) は直結されているが、CPU(0) から Memory(1) にアクセスする場合はハイパートランスポート (HT) を経由する必要がある。これらには性能差が生じる。この負均質性はネットワークから見たメモリの距離にも存在し、NIC(0) から見たとき、Memory(0) は最短距離にあるが Memory(1) は CPU 間を接続するハイパートランスポートを経由する必要がある。この後の性能評価実験では NIC(0) から見た Memory(0) のように最短距離にあるメモリを「近いメモリ」、プロセッサ間のハイパートランスポートを経由する位置にあるメモリを「遠いメモリ」と呼ぶ。

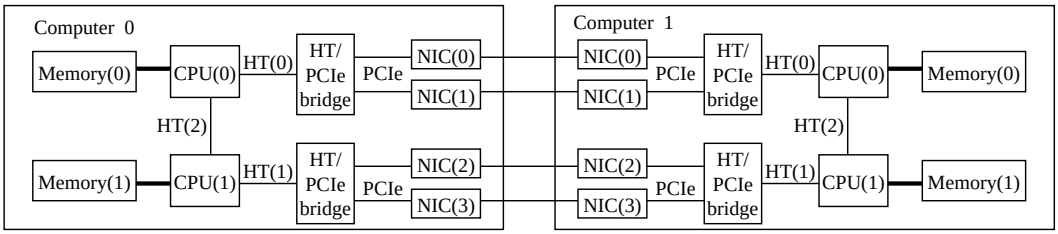


図 4: 接続図

```
benchmark(cpunode, memnode[],
          size, nr_nic){
 /* プロセスを特定の CPU に固定 */
 numa_run_on_node_mask(cpunode);
 buf = malloc(size);
 /* NIC 一枚当たりのサイズ */
 chunk = size / nr_nic;
 /* 物理アドレスが割り付けられる前に
    メモリの場所を指定する */
 for(i = 0; i < nr_nic; i++){
  numa_tonodemask_memory
    (buf+chunk*i, chunk, memnode[i]);
 }
 /* 物理メモリを割り付ける */
 memset(buf, 0, size);
 /* 測定開始 */
 st = MPI_Wtime();
 if (node==0){
   for(i = 0; i < niter; i++) {
     MPI_Send(buf, size, ...);
     MPI_Recv(buf, 4, ...);
   }
 } else {
   for(i = 0; i < niter; i++) {
     MPI_Recv(buf, size, ...);
     MPI_Send(buf, 4, ...);
   }
 }
 ed = MPI_Wtime();
 /* C は前もって測定した片道遅延 */
 time = (ed - st)/ niter - C;
 bandwidth = size / time;
}
```

図 5: 測定プログラムの概要

メモリの距離と通信性能との関係を明らかにするために、通信性能測定プログラムは、使用するメモリの物理的配置を固定している。図 5 に測定プログラムの概略を示す。

結果を示す前にこのシステムにおける理論性能を検討する。検討対象は、1) メモリバンド幅、2)HyperTransport のバンド幅、3)PCI-Express のバンド幅、4)Infiniband DDR のバンド幅である。1) のメモリは DDR2 667MHz のメモリを Dual Channel で使用しているため理論性能は 10.6GB/s である。2) の HyperTransport の理論性能は 6.4GB/s である。3) の PCI-Express については、実験に使用したネットワークカードが 8 レーンの PCI-Express 接続であるため、理論性能は 2Gbps×8 = 16Gbps = 2GB/s である。4) の Infiniband DDR の理論性能は 16Gbps である。これらを総合すると 1 リンク当たりの理論性能は 2GB/s となる。ただし PCI-Express に関しては、実装によってはパケットヘッダのオーバヘッドが無視できないほどに大きくなる。今回実験に使用したチップセットの場合、PCI-Express パケットのデータ部分の最大サイズは 128byte でありヘッダは最大 24byte なので、実際に得られる性能は 2GB/s × 128/(128+24) = 1.68GB/s となる。今回の測定環境では、この性能が全体のボトルネックであり、これを基に計算すると 2 リンク、4 リンクの最大バンド幅はそれぞれ 3.37GB/s、6.74GB/s となる。

3.2.2 性能測定結果

まず基本性能測定のため、４本ある Infiniband のうちの１本のみを使用して性能を測定する。測定は送受信共にネットワークカードから近いメモリを使用する場合と遠いメモリを使用する場合について行った。また、今回作成した YAMPI の OpenIB 層に性能バグが存在しないことを確認するため、結果を広く使われている MPI 実装のひとつである OpenMPI における測定結果と比較する。図 6 に結果を示す。

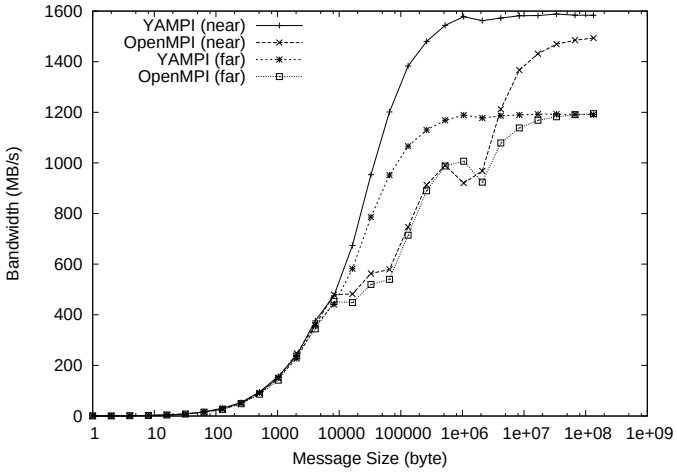


図 6: １リンク使用時のバンド幅

遠いメモリ、近いメモリ、それぞれにおいて YAMPI は OpenIB よりも高速であり、YAMPI の実装に問題がないことが確認された。また、近いメモリを使用した場合 YAMPI は最大 1.6GB/s を達成しており、理論性能の 95%の性能が得られている。

注目すべき点は近いメモリと遠いメモリの性能差である。HyperTransport の理論バンド幅に問題がないにもかかわらず、遠いメモリを使用した場合には大きな性能低下が見られる。原因を調査するため、送受信の一方のみを遠いメモリにして性能を測定した結果を図 7 に示す。送受信とも近いメモリまたは遠いメモリの結果も併せてプロットされているが、これは先に示した図 6 の YAMPI の結果の再掲である。

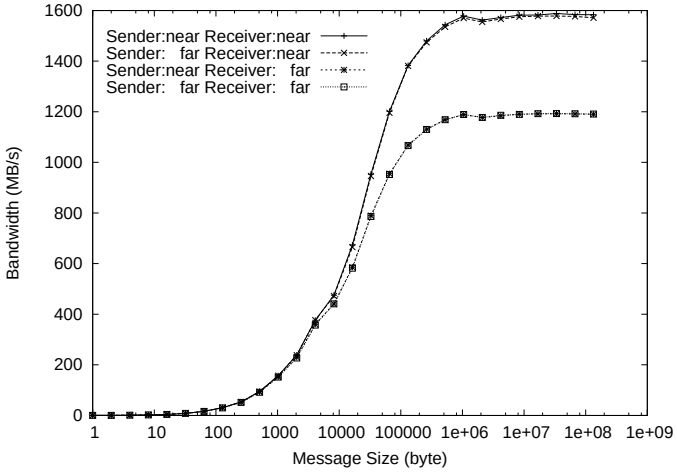


図 7: メモリとネットワークの距離と性能

図 7 の結果から、受信側におけるメモリとネットワークカードの距離が性能に影響しており、送信

側の距離はほとんど問題にならないことがわかる。受信側においてのみ、メモリとネットワークカードとの距離が問題になる理由は現時点では不明である。

2 リンク使用時の性能

本節では 4 本のうちの 2 本を使用した場合の性能を測定する。高い性能が期待できる近いメモリを使用した場合について性能を測定する。2 本のリンクの選び方として、以下の 3 通りがある。

1. nForceProfessional（以下、NFP）3600 に接続された 2 本のリンクを使う
2. NFP3050 に接続された 2 本のリンクを使う
3. NFP3600 から 1 本、NFP3050 から 1 本を使う

これらについて測定結果を図 8 に示す。比較のため、NFP3050 から 2 本を使用する場合については、2007 年 1 月時点での最新 BIOS であるバージョン 1.01Dβ を使用した場合と、その前バージョンになるバージョン 1.01 を使用した場合の二通りについて測定を行う。なお、先に述べたようにトランキングが有効となるのは大きなメッセージを通信するときのみであり、本測定ではメッセージサイズが 32KB 以上のときのみである。

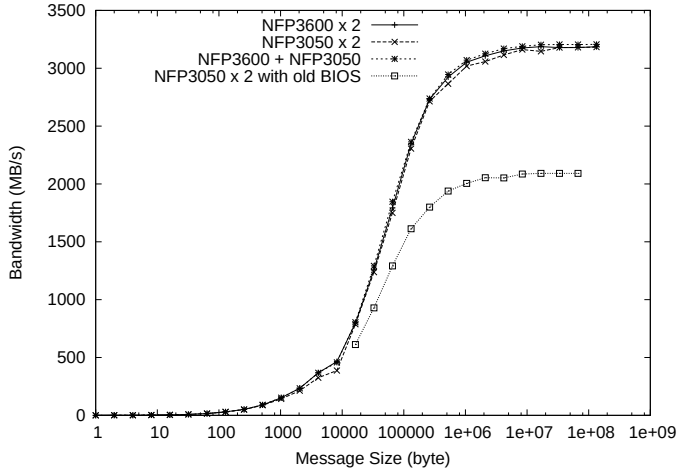


図 8: 2 リンク使用時のバンド幅

結果からわかるように、いずれの場合もほぼ同じ性能を示しており、ピークで 3.0GB/s である。これは先に検討した理論性能の 89%である。また、NFP3600 と NFP3050 の性能差はない。ただし、この結果が得られたのは BIOS のバージョンが 1.01Dβ の時のみである。グラフ中の一番下の線に示されているように、バージョン 1.01 の BIOS では NFP3050 の性能が十分でなく、2 枚のネットワークを接続するのに十分な性能は得られていなかった。その時の性能は 2.1GB/s なので、BIOS のバージョンアップにより 42%性能が改善したことになる。バージョン 1.01 の BIOS はチップセットを適切に設定していなかったものと推測される。

4 リンクの性能

最後に 4 リンクすべてを使用した場合の性能を図 9 に示す。”Near”はすべてが近いメモリになるように配置した場合、”Far” は NFP3050 を使用する 2 リンクのうち 1 リンクの受信側のみを遠いメモリにした場合である。すべて近いメモリを使用した場合のバンド幅は最大で 6.17GB/s である。これは理論性能の 91.5%であり、ほぼ期待される性能である。また、オープンスーパーコンピュータ要求要件である 4GB/s の 1.54 倍であり、余裕を持って要件を満たしていると言える。一方、1 リンクの受信側

のみで遠いメモリを使用した場合、性能は安定せず最大でも 3.7GB/s にとどまっている。この結果から、最大限に性能を発揮させるためには NUMA を考慮し、近いメモリのみを使用して通信を行うことが必要不可欠であることがわかる。

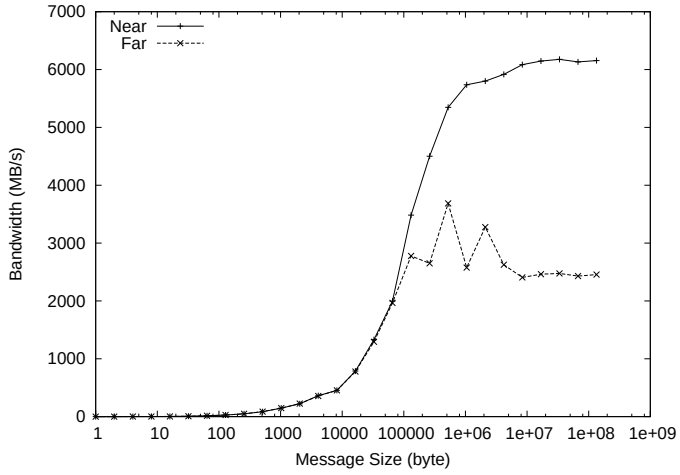


図 9: 4 リンク使用時のバンド幅

3.3 具体的成果

現在入手可能な高速ネットワークを用いることで、特殊設計されたハードウェアを用いることなく MPI レベルで 6.17GB/s の性能が実現可能であることを示した。また、この性能を実現するためには NUMA アーキテクチャを意識したメモリ配置を行うこと、またハードウェアの設定を適切に行う BIOS を用いるなど、ソフトウェアの最適化が必要であることを示した。

4 成果要覧

受賞関連

[受賞 1] 松葉 浩也: コンピュータサイエンス領域奨励賞, 情報処理学会, 2006 年 6 月

[受賞 2] 松葉 浩也: 山下記念研究賞, 情報処理学会, 2007 年 3 月

査読付論文リスト

[査読付 1] Hiroya Matsuba and Yutaka Ishikawa: Single IP address cluster for internet servers, 21st IEEE International Parallel and Distributed Processing Symposium (IPDPS2007) CD-ROM, March 2007.

その他の発表論文リスト

[発表 1] 松葉 浩也、石川 裕: コモディティネットワークを用いた 5GByte/sec 通信の可能性, 情報処理学会研究報告 HPC-109, 札幌,2007 年 3 月

PKI プロジェクト

PKI と既存システムとの連携を実現するサポートシステム

西村 健

1 概要

PKI（Public Key Infrastructure, 公開鍵認証基盤）は技術的には成熟した分野であるが、現状では実用面において普及しているとは言い難い。これはシステムが統一的でない大学のような組織で顕著であると思われるが、大小様々な形態で存在している既存システムとの連携が容易ではなく PKI 導入から得られるメリットが制限される、また連携を行ないそのメリットを享受するためには多大なコストが必要であるためと考えられる。例えば東京大学においても全学的な PKI の導入に向けた取り組みを行なっている[発表 3]が、既存のシステムと PKI の連携については最良の解を見出せていない。

本年度は手始めとして、ID/パスワードを用いる Web アプリケーションに PKI 認証を適用するためのシステムを提案している[発表 1,2]が、今後はより適用範囲を拡大して様々なアプリケーションに PKI 対応を可能にするサポートシステムの構築を行ない、各方面での PKI の普及を図る。

2 レガシー Web アプリケーションに対応する PKI を用いた簡易 Single Sign-On の実現

2.1 背景

今日我々は Web を通して数々の情報サービスを受けられるようになっている。Web は情報提供の場として重要度を増していくとともに、別の側面で特定のコミュニティにおける情報共有の場としても利用されてきている。多くの大学の Web ページが各学部からの広報の場として利用されている。その利便性から多くの情報が Web 上に掲載されるようになってきた。学部、学科など特定の単位のコミュニティの人々がこの Web によってつながりを持っている。この場面でのアクセス制御の一つの解が PKI であり、PKI は適切に運用されるという前提の下で最も厳格なアクセス制御を行なうことができる。

PKI アプリケーションの普及において重要なのはサーバが PKI 認証を利用することである。例えば東京大学ではすでに全学レベルから研究室レベルまで多くのサーバが稼動しており、多くの情報が Web を通じて提供されている。しかし多くのサーバで認証は旧来の ID およびパスワードによるものである。財務部門やオンラインジャーナルなど機密情報を扱う必要があるサーバはアクセスを大学内からのみに制限し安全性を保っている。つまり利用者は大学内にある端末からそのようなサーバにアクセスしなければならない。PKI 認証を採用すればこの制限を緩和できる可能性がある。

しかしながらサーバには既製品のソフトウェアパッケージにより構築されたものも多い。つまりそのパッケージが提供する（多くの場合 ID およびパスワード方式の）特定の認証方式に制限されるということである。そのようなシステムで PKI 認証を組み入れることは難しく、そのようなサーバ群をカバーするのは大きな問題になっている。本研究ではこの問題を、大学内でよく使われていて認証方式に ID/パスワードを採用しているグループウェアの名称をとって「サイボウズ問題」と呼んでいる。

2.2 内容

本研究では、既製の SSL-VPN 製品によって通信の安全性を担保しつつ認証において PKI 認証を用いたゲートウェイシステムを提案する。このゲートウェイがクライアントとサーバ群との仲介者となり、

2.2.1 経路上での盗聴

SSL-VPN 本来の機能である SSL による暗号化によって、通信経路上での盗聴対策はなされている。またクライアント端末－password pushout server 間についても、PKI 認証すなわち SSL クライアント認証が行なわれるため同様に SSL による暗号化がなされている。

2.2.2 キーロガーによるパスワードの窃盗

私有鍵の PIN 以外にパスワードを入力することはなく、その意味では安全であるといえる。ただしクライアント端末上で任意の悪意あるプログラムが実行できる場合には、password pushout server から送信された暗号化されたパスワードを復号するか、コンテンツサーバへ送信する平文のパスワードを盗聴することによりコンテンツサーバに対するパスワードの窃盗は可能である。

2.2.3 password pushout server 管理者による不正

password pushout server には各種のサーバに対する ID/パスワードの組が保管されるため、この部分におけるセキュリティは特に重要となる。前述のとおり password pushout server に保管されるパスワードは利用者の公開鍵により暗号化されており、その点では安全であるといえる。ただしクライアント端末は password pushout server から送信されたスクリプトに従いパスワードの復号等を行なうため、そのスクリプトに悪意のあるコードが含まれていた場合平文のパスワードの奪取が可能になる。

つまり安全性は利用者がクライアント端末で実行するスクリプトが信頼できるものであるかどうかの一点に集約されることになる。

2.2.4 本システムの位置付け

以上の考察から、本システムは、(1)クライアント端末に悪意あるプログラムが実行される可能性がなく、(2) password pushout server から送信されるスクリプトが第三者による精査で問題ないと判断されそれがデジタル署名等で確認できる場合に有効なシステムと位置付けることができる。

2.3 具体的成果

2.3.1 システムの実装

実際に password pushout server システムを構築し、F5 ネットワークス社の SSL-VPN アプライアンス FirePass とともにクライアント上のブラウザ Firefox により動作試験を行ない、本システムが現時点のインフラで実現可能であることを確認している。

ただしクライアント上で暗号化パスワードの復号などの複雑な処理を行なわなければならずそれらの処理系はブラウザ毎に異なるため、各ブラウザでの実装は課題として残されている。また実運用での適用および検証は今後の課題である。そしてそれらの課題を解決した後に成果として公開を考えている。

2.3.2 発表

本研究の成果については[発表 1,2]にて発表を行なっている。またそこへ至る背景としての PKI プロジェクトでの取り組みは[発表 3]にて発表を行なっている。

3 成果要覧

その他の発表論文リスト

[発表 1] 西村 健, 佐藤 周行, “レガシー Web アプリケーションに対応する PKI を用いた簡易 Single Sign-On の実現,” インターネットアーキテクチャ研究会, 電子情報通信学会技術研究報告 Vol.106 No.465, pp. 61-66, 2007 年 1 月

クライアントからの PKI 認証の後にゲートウェイが各サーバに対する ID およびパスワードを送出する。このシステムによってサーバを PKI 認証のために改変することなく、PKI による安全性を享受することができる。

SSL-VPN の形態として機能がシンプルで設定も容易なリバースプロキシ型を想定すると、PKI 認証による SSL-VPN の一つの問題は、VPN サーバとコンテンツサーバで認証情報を共有できないことである。コンテンツサーバ側での認証は VPN サーバでの PKI 認証とは独立に用意しなければならない。旧来の ID/パスワード式の認証の場合は VPN 接続後にそのログイン画面が表示され、利用者は ID およびパスワードをそれぞれのサーバに対して入力する必要がある。

本システムの構成を図 1 に示す。ここで認証の仲介を行なうゲートウェイを password pushout server と呼んでいる。password pushout server は(証明書の DN, server ID, userID, 暗号化されたパスワード)の組をそれぞれの利用者証明書、コンテンツサーバに対して保管しておく。SSL-VPN に対する認証が成功し、利用者が特定のコンテンツサーバへ接続を指示すると、SSL-VPN はその要求をフックし password pushout server へ接続するように要求する。クライアントが password pushout server へ接続すると、そのクライアント証明書情報から対応する userID およびパスワードをクライアントへ返す。クライアントは取得した userID およびパスワードを SSL-VPN 経由でコンテンツサーバへ送信し正しく認証を行なうことができ、以降コンテンツサーバからサービスを受けられる。

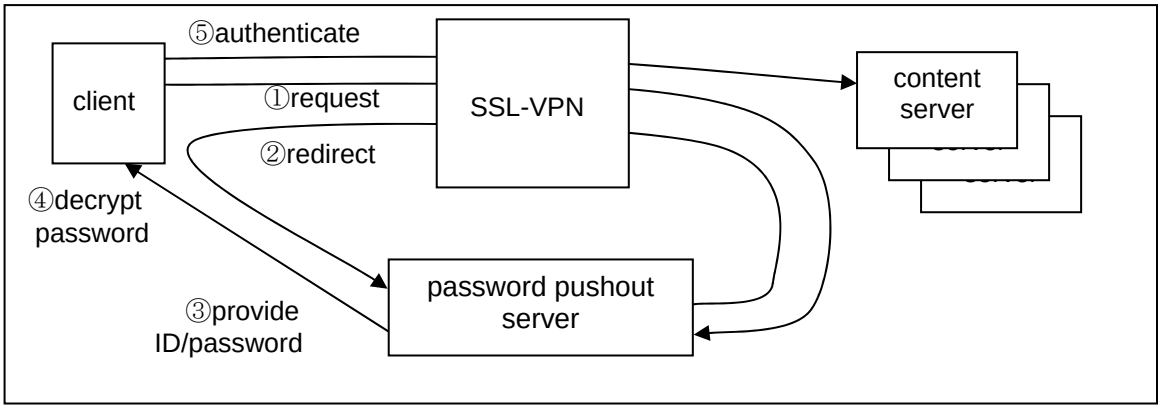


図 1: password pushout server およびその他のサーバの構成

ここでのポイントは password pushout server に保管されるパスワードは利用者の公開鍵で暗号化されているということである。Password pushout server は暗号化されたパスワードのみを扱い、コンテンツサーバへの送信の直前にクライアント側でブラウザ上のスクリプトによって復号を行なう。これによって平文のパスワードの漏洩先が最小限に制限される。

技術的にはこのシステムはサーバ側でサインオンを行なわないため厳密な意味での SSO (Single Sign-On) システムではない。しかし利用者の視点ではゲートウェイでのみ認証を行なっているように見えるため、利用者の立場では SSO システムとして動作していると言える。このシステムの利点はサーバを改変することなく行なえることであり、SAML ベースの SSO を実現するのに必要な莫大なコストを抑えることができる。小規模の組織に対するセキュリティの改善案としては安価なコストでの現実的な解になりうると考える。

本システムについて、利用者に対して想定される脅威には以下のものが挙げられる。

- 経路上での盗聴
- キーロガーによるパスワードの窃盗
- password pushout server 管理者による不正

以下、それぞれの脅威に対して考察を行ない、本システムが満たすセキュリティ上の位置付けを明確にする。

[発表 2] T. Nishimura, H. Sato, "Authentication with PKI – a Case Study in Information Technology Center in The University of Tokyo," International Symposium on Advanced ICT, pp. 251-255, 2006.

[発表 3] 西村　健, 佐藤　周行, “自律的組織の集合体としての大学における PKI の運用,” 情報処理学会全国大会 講演論文集(分冊 4), pp. 327-328, 2007 年 3 月

PART 3

教育・サービス活動

（学内共同利用）

情報メディア教育部門

図書館電子化部門

キャンパスネットワーキング部門

学内共同利用３部門共通　講習会等

（全国共同利用）

スーパーコンピューティング部門

PKI プロジェクト

情報メディア教育部門

情報メディア教育部門

概要

山口和紀　　平野光敏

情報化の進展を受け、教育の様々な場面でコンピュータやネットワークが必要とされる機会が増えてきている。そのなかで部局や学科、個人では対応することが難しく、情報基盤センターが大学全体にサービスを提供することで、集約効果があり、ノウハウが蓄積されるものを当部門では担当している。サービスには、システムを運用し利用に供するものだけではなく、ノウハウを提供するものなど、必要に応じて最も適切なことを行っている。サービスを提供する上での経費負担に関しては、利用者に課す負担金により利用が阻害される事なく、かつ、無駄に使われないような負担体系を構築し、資源の効率的な利用を行っている。

具体的なサービスとしては、教育で必要となる端末とサーバ群の総称である教育用計算機システム(ECCS)がある。ECCS のサービスでは、不特定多数が使用する端末を安全に使えるようにすること、運用コストをできるだけ下げることが大きな課題であるが、それだけでなく、利用形態の多様化により、新たな課題が日々発生してくる。例えば、学外者を含む講習会などでライセンスの制約のため特殊設定が必要になるというようなことがある。このような課題を解決することも重要な仕事である。また、ECCS では、3 万人分にアカウントを発行しているが、IC カード学生証を利用した即時発行システムを構築したことでアカウント配布の利便性の向上及び効率化を達成した。プリンタについてもプリペイドカード方式と運用の一部を外注する事で、利便性を高めつつランニングコストを下げた。更に、IC カード学生証を利用したポストペイ方式を開発したことで、一層の利便性の向上を達成した。また、大学としての責任を果たすために、新規利用者講習会を開催し、基礎的な情報倫理教育を行っている。利用方法の周知については、利用の手引き、ウェブ広報、窓口対応と電子メール対応、電話対応を行っている。更に、学生からの質問を学生の目線で対応できるように、学生を教育して、相談員として質問にあたらせている。

ECCS が安全かつ効率的に運用できていることから、それを発展的に利用したいという要望が多く寄せられている。直ぐにサービスとして提供するのが難しいものは、実験サービスという形でサービスの可能性を探り、できるものからサービスとして広く提供するようにしている。例えば、ECCS のアカウントを部局で導入したシステムで利用したいという要望を受け、教養学部や当センターPKI プロジェクトの VPN ゲートウェイ等に提供している。持ち込み PC のインターネット接続環境のための認証ゲートウェイ機能も ECCS のものを法学政治学研究科などに提供している。 ECCS のシステムのメールサーバ、ウェブサーバ、ネットワークストレージなどは、その有用性が教育に限定されるものではなく、要望も多いので、独立したサービスとして研究・事務にも提供している。

最近、遠隔講義が増えてきているが、設備を用意することが難しかったり、設備を用意できても利用や先方との調整が難しいことがある。そこで、テレビ会議システムを備えた遠隔講義室や貸し出し用システムを用意している。先方との接続性などについてはキャンパスネットワーク部門とも協力しながら利用の支援を行っている。また、個別に持つことが難しい多地点接続装置を用意して、利用に供している。

映像配信の利用も増えてきているが、映像配信のストリーミングサーバを利用ごとに持つのではなく、複数の利用をまとめれば、同じコストで同時接続数の上限を実質的にあげることができて効率的である。そこで当部門ではストリーミングサーバの提供をしている。これは、UTOCW, TODAI TV, 卒

業式 Live 中継、英語 I のリスニング、スペイン語部会の映像教材、情報倫理ビデオ、セミナー・講習会の映像配信等に利用されている。ライセンスや内容に応じて配布範囲を学内に限定したり、パスワードをつけるなど要望に応じて開発・調整を行っている。

ECCS では、講義用に教員が使えるウェブサーバを運用しているが、教育の IT 支援の進展に従い、単なるウェブサーバではなく、Wiki や学習支援システムが要望されるようになってきた。これを受けて、当部門では学習支援システム CFIVE を開発し、運用している。更に、出欠を簡単に取りたいという要望を受け、IC カードの学生証を利用した出欠管理システムを構築した。これは、学術俯瞰講義などの授業で利用されている。また、PowerPoint によるプレゼンテーションをできるだけ簡単に記録して配信したいという要望に対して、自動的に収録し配信するシステムを開発し、運用を始めている。

教育用計算機システム運用報告

1 経過報告

1.1 Mac(iMac, Xserve)

教育用計算機システム全体では標準端末として iMac 端末(1,149 台)を設置している。iMac 端末はハードディスクを搭載せず、ブートサーバ(Apple Xserve)から OS(Mac OS X)のイメージを読込んで起動する。1 台のブートサーバは iMac 端末 20 数台を制御し、49 台(駒場 34 台、本郷 15 台)のサーバを設置している。また、RDC(Remote Desktop Connection)を使用して、Windows Server 2003 を利用することもできる。

システムイメージの更新は OS のバージョンアップ、セキュリティアップデート、アプリケーションのインストールやアップデートに伴って行われる。
2006 年度の主な更新は以下の通りである。なお、不具合の修正や広報の表示の際にもイメージの更新を行うことがある。

2006 年 3 月 27 日 - 4 月 4 日 Unix サーバの MacOS X 10.4(Tiger)バージョンアップ
-Unix サーバ:ux101, ux102, ux103, ux104
2006 年 4 月 2 日 iMac 端末の OS バージョンアップ
-Mac OS X 10.3(Panther)から 10.4(Tiger)へのバージョンアップ
2006 年 5 月 1 日 iMac 端末のシステム更新
-Mac OS X 10.4.6
-MATLAB 7.1
2006 年 5 月 26 日 iMac 端末のシステム更新
-Security Update 2006-003 (Mac OS X 10.4.6)
-JMP6 ライセンス更新
-QuickTime 7.1
-Adobe Reader 7.0.5 管理者パスワードが要求される不具合の修正
2006 年 6 月 12 日 iMac 端末のシステム更新
-Xcode2.3
-Firefox 1.5.0.4
-Thunderbird 1.5.0.4
2006 年 10 月 3 日 iMac 端末のシステム更新
-Mac OS X 10.4.7
-Security Update 2006-004 (Mac OS X 10.4.7)
-Microsoft Office 2004 for Mac 11.2.6
-Adobe Reader 7.0.8
-R 2.3.1
-自動ログアウトする機能を追加(アイドル状態が 60 分続いた時)
2006 年 12 月 14 日 iMac 端末のシステム更新
-Mac OS X 10.4.8
-Security Update 2006-007 (Mac OS X 10.4.8)
-Microsoft Office 2004 for Mac 11.3.0
-Opera 9.0.2
-Firefox 2
-Xcode 2.4
-iTunes 7.0.3

-Quicktime 7.1.3
-Acrobat 6.0.5
-VLC 0.8.5

1.2 VID(Windows, Linux)

VID 端末は、Windows や Linux でしか動作しないアプリケーションを必要とする講義などで用いることを想定したディスクレス端末である。また、CAD などを使用するために大きなメモリ容量や高精細な画面が必要なアプリケーションを動かすための VID+端末もあり、本郷・駒場・柏に合計約 250 台設置している。
Windows において、ディスクイメージファイルの更新時には OS のセキュリティアップデートとウイルス対策ソフトのパターンファイル更新を必須に行い、その他にソフトウエアの更新などを行った。
2006 年度の主な更新は以下の通りである。

2006 年 4 月 14 日 VID 端末(Windows)のソフトウェア更新
-Microsoft WindowsXP、Microsoft Office アップデート
-MATLAB 7.1 バージョンアップ
-TSP 5.0 バージョンアップ
2006 年 5 月 18 日 VID 端末(Windows)のソフトウェア更新
-Microsoft WindowsXP アップデート
-Mozilla Suite 1.7.13 バージョンアップ
2006 年 6 月 19 日 VID 端末(Windows)のソフトウェア更新
-Microsoft WindowsXP アップデート
-Firefox1.5.0.4 インストール
-Thunderbird1.5.0.4 インストール
-JMP6 インストール
-StarSuite 8 バージョンアップ
2006 年 7 月 21 日 VID 端末(Windows)のソフトウェア更新
-Microsoft WindowsXP、Microsoft Office アップデート
-Flash Player 9.0.16.0 インストール
-J2SE Runtime Environment Update インストール
-Mozilla Suite アンインストール
2006 年 8 月 24 日 VID 端末(Windows)のソフトウェア更新
-Microsoft WindowsXP、Microsoft Office アップデート
-Adobe Acrobat 6.0 Standard を 6.0.5 バージョンアップ
-Firefox 1.5.0.6 バージョンアップ
-Thunderbird 1.5.0.5 バージョンアップ
-Opera9.01 インストール
2006 年 9 月 21 日 VID 端末(Windows)のソフトウェア更新
-Microsoft WindowsXP、Microsoft Office アップデート
-Firefox 1.5.0.7 バージョンアップ
-Thunderbird 1.5.0.7 バージョンアップ
2006 年 12 月 14 日 VID 端末(Windows)のソフトウェア更新
-Microsoft WindowsXP、Microsoft Office アップデート
-Firefox 1.5.0.8 バージョンアップ
-Thunderbird 1.5.0.8 バージョンアップ
-Opera9.02
-Flash Player 9.0.28.0

-J2SE Runtime Environment 5.0 アップデート
2007 年 3 月 27 日 VID 端末(Windows)のソフトウェア更新
-Microsoft WindowsXP、Microsoft Office アップデート
-Stata/SE 9.2
-MATLAB 7.3.0(R2006b)
-Adobe Acrobat 6.0 Standard 6.0.6
-VIRUS BUSTER 2006
-Firefox 1.5.0.11
-Thunderbird 1.5.0.10
-Opera9.10
-J2SE Runtime Environment 5.0 アップデート

2 サービス関係

2.1 ポストペイ方式プリンティングサービスの開始について

2006 年 9 月 1 日より、ECCS(教育用計算機システム)でのプリンタ出力として新たにポストペイ方式プリンティングサービスを開始した。

従来のプリペイドカード方式では、事前に購入した専用プリペイドカードを用いて印刷しましたが、新サービスではポストペイ用として使用可能な IC カードを用いる。印刷は IC カードを IC カードリーダにタッチすると印刷枚数分の精算、印刷ができる。

なお、ポストペイ方式のサービスを利用するには、IC カード学生証(キャンパスカード[学生用クレジットカード]契約者のみの IC カード学生証)、IC カード教職員証(統合型教職員証のみ)が必要となり、他の IC カードは利用できません。

ポストペイ用プリンタの設置場所は以下の通りである。
<本郷地区>
・総合図書館 2F メディアプラザ2
・情報基盤センター 1F 自習室前
<駒場地区>
・駒場図書館 2F
・情報教育棟 1F 自習室(相談員室)

2.2 コミュニケーションプラザ(北館/南館)への有線 LAN 無線 LAN サービスの提供について

駒場キャンパスの駒場寮跡地にコミュニケーションプラザ(北館)という学生サービスを目的とした建物の建築中に, 2006 年 4 月のサービス提供開始にあたり、有線／無線 LAN サービスのテストとして「テスト用アカウント」を発行・提供し、他の目的には使わないように覚書を交わした。

また、2006 年 10 月 1 日、コミュニケーションプラザ(南館)のオープンに伴い、南館に 8 台の無線 LAN ステーションの接続サービスを開始したため、覚書を交わした。

なお、4 月からのコミュニケーションプラザの有線／無線 LAN の環境は以下の通りである。

品目	個数	設置場所
有線 LAN	27 口	北館 2 階
	10 口	北館 3 階
	12 口	和館 1 階
	1 口	南館 3 階

無線 LAN	5 ステーション	北館 2 階
	8 ステーション	南館 1～3 階

表 1. コミュニケーションプラザ 有線／無線 LAN(2006 年 10 月)

2.3 赤門総合研究棟への有線 LAN サービスの提供について

大学院公共政策学教育部が赤門総合研究棟に用意する有線 LAN に対して、運用代行の依頼があったことから、2006 年 4 月に教育用計算機システムの有線 LAN 接続サービス提供の覚書を交わした上でこれを実施した。

設置場所は以下の通り。

品目	個数	設置場所
有線 LAN	32 口	3 階 338 室
	110 口	地下第 8 教室

表 2. 赤門総合研究棟 有線 LAN(2006 年 4 月)

2.4 大学院工学系研究科・工学部への無線 LAN サービスの提供について

大学院工学系研究科・工学部が工学部 2 号館、8 号館、11 号館に用意する無線 LAN に対して、運用代行の依頼があったことから、2006 年 4 月に教育用計算機システムの無線 LAN 接続サービス提供の覚書を交わした上でこれを実施した。

設置場所は以下の通り。

品目	個数	設置場所
無線 LAN	1 ステーション	工学部 2 号館 2 階
	〃	工学部 2 号館 2 階
	〃	工学部 2 号館 5 階
	〃	工学部 8 号館 1 階
	〃	工学部 11 号館 2 階

表 3. 大学院工学系研究科・工学部 無線 LAN(2006 年 4 月)

3 稼働状況

3.1 部局別実利用者数

教育用計算機システムを一度でも利用したことのある学部学生、大学院生、研究生、教職員の月毎累積の部局別実利用者数である。比較のため、2006 年度(各月右側)と 2005 年度(同左側)を掲せてある。

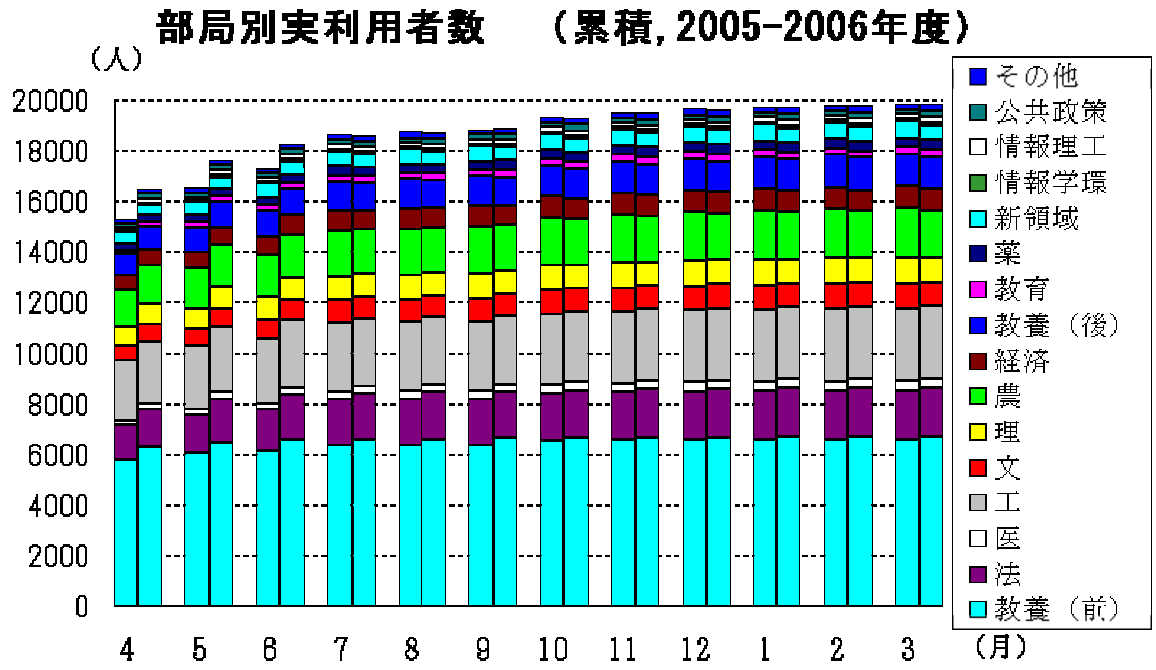


図 1. 部局別実利用者数

3.2　端末利用状況

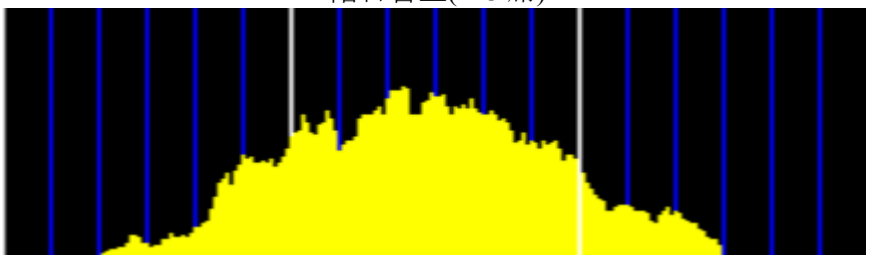
情報メディア教育部門では、教育用計算機システムを管理しており、その端末等を本郷キャンパスと駒場キャンパスに配置している。本郷では、情報基盤センター(浅野地区)の演習室に端末等を集中配置すると共に、総合図書館、留学生センター、各学部／研究科等の 30 数箇所にも端末を分散配置している。駒場では、教養学部情報教育棟に端末等を集中配置すると共に教養学部図書館、数理科学研究科にも端末を分散配置している。

これらの配置場所のうち、利用の多い本郷の総合図書館と駒場の情報教育棟と駒場の教養学部図書館の端末利用状況は図2の通りである。(下から、総合図書館(2 階メディアプラザ 2 と同 3 階メディアプラザ 3 の合計(51 台))、駒場教養学部図書館 2 階メディアパークプラザ(35 台)、情報教育棟 3 階大演習室 2(195 台)、同 2 階大演習室 1(195 台)、1 階自習室(123 台)の 2007 年 1 月 31 日におけるものである。)

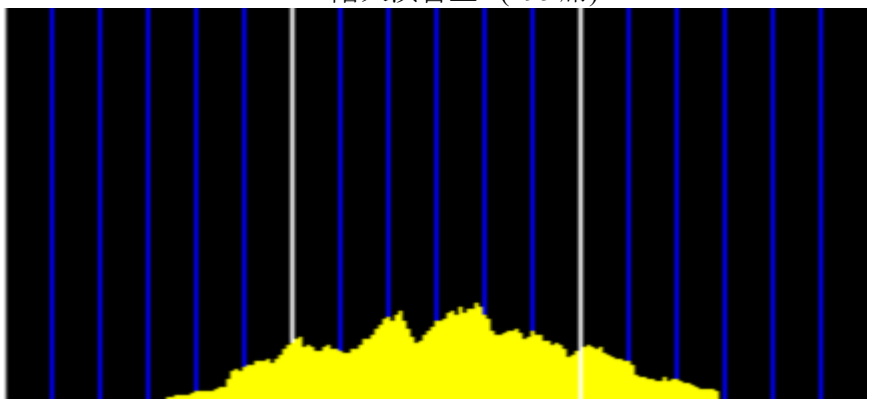
2007 年 1 月 31 日　iMac 端末利用状況
各部屋の時間帯ごとの端末利用率のグラフである。
縦軸は端末数で横軸は時刻(6 時から 24 時まで)である。

情報教育棟

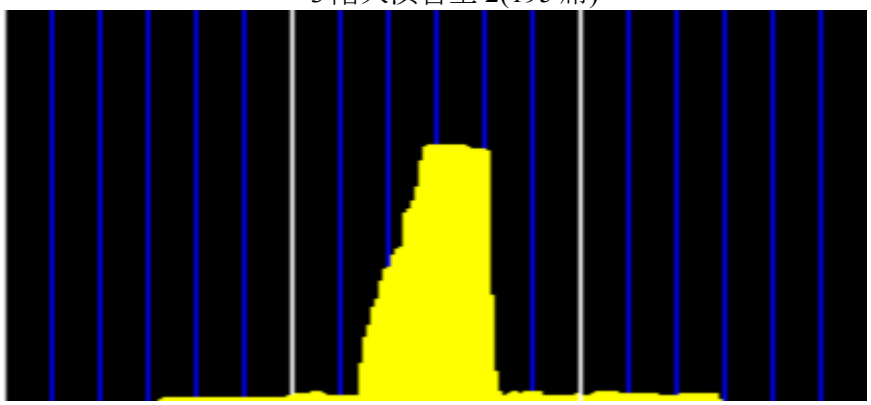
・1 階自習室(123 席)



・2 階大演習室 1(195 席)



・3 階大演習室 2(195 席)



駒場図書館

・2F-メディアパークプラザ (35 席)

総合図書館

・2F-メディアプラザ 2, 3F-メディアプラザ 3(51 席)

図 2. 本郷総合図書館、駒場図書館、駒場情報教育棟の端末利用状況

4　次期システムについて

情報基盤センターでは、2008 年 3 月に教育用計算機システム(ECCS:2004 年 3 月導入)の更新を予定している。現在、仕様策定委員会の設置に向けての準備を進めている。導入手続き日程としては、市場調査手続きから仕様書作成手続きの段階に入ったところである。

ここに、システム更新に向けて行なった手続き等を掲載する。

4.1 アンケート

情報メディア教育専門委員会(2006/6/27)で、現行システムの利用状況、次期システムへの要望などを伺うためのアンケート実施について説明があった。アンケートは、「ECCS2004 に関するアンケート(教員向け)」、「ECCS2004 に関するアンケート(全ユーザ向け)」の 2 種類の内容を用意した。回答方法は、教育用計算機システムのアカウントを用いたユーザ認証(CFIVE 利用)により、1 人 1 回のみの回答で、回答期限は 7/3-10/31 とした。

調査項目は、以下に示すものを設定した。

[教員向けアンケート]

1 講義でお使いのソフトウェアについて
2 ECCS で提供しているソフトウェアの利用料金について
3 ECCS に必要で提供されるとよいソフトウェアについて

[全ユーザ向けアンケート]

1 ECCS を利用する主な目的について
2 ECCS で利用する主な環境やサービスについて
3 ECCS で利用する主なソフトウェアについて
4 ECCS を利用する頻度について
5 ECCS で満足している点について(ネットワーク、OS、ソフトウェア、端末など)
6 ECCS で不満な点について(レスポンスタイム、端末数、容量/利用制限、印刷など)
7 新 ECCS に必要なものについて(環境、サービス、ソフトウェア、その他)

4.2 公聴会

教育用計算機システムを用いて授業を担当される教職員を対象に次期システムの基本システム案を説明し、要望などを伺った。また、アンケートの中間報告(7/3-9/27)も併せて行った。公聴会は、駒場地区(9/28)、本郷地区(9/29)で開催した。

4.3 教育用計算機システム(ECCS)導入作業部会、導入説明会

情報メディア教育専門委員会(10/2)で、同専門委員会の下に「ECCS 導入作業部会」を組織することとした。導入作業部会は各学部からの代表等により構成される。

ECCS 導入作業部会は、2 回(12/13、1/17)開催され、アンケート結果・公聴会での意見等を参考に、次期システム案の大枠についての検討を行った。その後、導入説明書案を作成し導入説明会(2/28)を開催した。

4.4 今後のスケジュール

仕様書作成手続きに入るため、仕様策定委員会を設置して各社からの提案書・資料をもとに仕様書案の作成、仕様書案説明会と続く。その後、入札手続きの段階に入り、入札説明会の開催、応札締切日となり、開札日に業者が決定する。そして、2008 年 3 月に新システムによるサービス運用の開始となる。

5 タッチパネル式 WWW キオスク端末の増設について

2005 年度に設置したタッチパネル式 WWW キオスク端末の利用回数のログを集計してみると、全体のアクセス数は多くなく、平日のほうが利用されている傾向にある。

当初は、窓口の閉まっているときや休日での利用を想定していたが、窓口の横に設置されているためか、現在は窓口での業務軽減として役立っている。しかし、図書館や自習室に設置すると当初の

目的に利用されるものと考えられるため、3/15～28 にかけて本郷地区(基盤センター1F 受付前、総合図書館)、駒場地区(情報教育棟自習室、駒場図書館)に計 4 台設置した。

6　柏出張講習会

学部学生、大学院生および研究生が教育用計算機システムの利用を新規に申し込む際には、新規利用者向け講習会を受講する必要がある。しかし、柏キャンパスに所属する学部学生、大学院生については、本郷への移動は時間がかかるため、受講者の負担が大きく、また、遠隔キャンパスとしては所属人数が多いという事情を考慮して、年に 1 回程度の頻度で出張講習会を開催している。

2006 年度の実施状況は以下の通りである。
・5 月 12 日(金) 講習会開催および利用申込書受付
新規利用申込者 31 名
継続利用申込者 110 名
・5 月 19 日(金) パスワード配付

平成18年度情報教育研究集会

1　概要

広島大学の主催で平成 18 年度情報教育研究集会が、平成 18 年 11 月 24 日(金)から平成 18 年 11 月 25 日(土)に行なわれ、当部門からも教職員が参加し発表等を行った。この集会の目的は、国公私立の大学・短期大学・高等専門学校における情報教育(情報を専門とする学科の専門科目の授業を除く)を担当する教職員が今日の情報化社会の急速な進展に対応した情報教育の授業を実施するために必要な教育の理念・内容・方法等について討議するものである。

2　プログラムの内容

○11 月 24 日(金)

広島大学教育学部 K201 教室、L205 教室に於いてチュートリアルが、また広島大学サタケメモリアルホールに於いて、開会挨拶、特別講演、パネルディスカッションが行なわれた。

・チュートリアル

「教育における著作権」

尾崎史郎氏 (メディア教育開発センター)

「インストラクショナル・デザイン入門」

内田実氏 (メディア教育開発センター)

・開会挨拶

牟田泰三氏 (広島大学　学長)

福島哉史氏 (文部科学省　専門教育課)

・特別講演1

「大学入試と教科『情報』」

中森眞理雄氏(情報処理学会コンピュータと教育研究会主査、東京農工大学工学部)

・特別講演2

「e-Learning2.0 を目指した高等教育の試み-テクノロジーは授業を「楽」にする-」

安武公一氏(広島大学大学院社会科学研究所)

・パネルディスカッション「教育における著作権」

尾崎史郎氏 (メディア教育開発センター)

坂井素思氏 (放送大学教養学部)

植村八潮氏 (東京電気大学出版局)

司会:中村純氏 (広島大学情報メディア教育研究センター)

・特別講演3

「大学教育におけるユニバーサルアクセス

佐野(藤田)眞理子氏 (広島大学大学院総合科学研究科)

○11 月 25 日(土)

広島大学総合科学部講義棟 K 棟、L 棟に於いて、46 セッション(ポスターセッションを含む)、合わせて 246 件の論文発表等が行なわれた。()内の件数は論文発表等の件数を表す。

・e ラーニング(1) (5 件)

・e ラーニング(2) (4 件)

・e ラーニング(3) (4 件)

・e ラーニング(4) （5 件）
・現代 GP セッション(1) （7 件）
・現代 GP セッション(2) （7 件）
・情報リテラシ教育(1) （4 件）
・情報リテラシ教育(2) （4 件）
・情報リテラシ教育(3) （5 件）
・情報リテラシ教育(4) （5 件）
・情報リテラシ教育(5) （6 件）
・情報リテラシ教育(6) （6 件）
・情報リテラシ教育(7) （5 件）
・情報リテラシ教育(8) （5 件）
・情報リテラシ教育(9) （5 件）
・情報リテラシ教育(10) （6 件）
・情報教育のあり方(1) （6 件）
・情報教育のあり方(2) （5 件）
・情報教育のあり方(3) （5 件）
・情報教育のあり方(4) （5 件）
・情報教育のあり方(5) （6 件）
・教育手法と評価(1) （5 件）
・教育手法と評価(2) （5 件）
・情報教育のあり方(6) （5 件）
・情報教育のあり方(7) （6 件）
・教育用情報基盤システム(1) （5 件）
・教育用情報基盤システム(2) （6 件）
・教育用情報基盤システム(3) （5 件）
・教育用情報基盤システム(4) （5 件）
・教育用情報基盤システム(5) （7 件）
・ユニバーサルデザイン （6 件）
・教育学習支援システム(1) （5 件）
・教育学習支援システム(2) （5 件）
・教育学習支援システム(3) （4 件）
・教育学習支援システム(4) （4 件）
・デジタルコンテンツ(1) （4 件）
・デジタルコンテンツ(2) （4 件）
・デジタルコンテンツ(3) （5 件）
・デジタルコンテンツ(4) （5 件）
・デジタルコンテンツ(5) （4 件）
・情報メディア教育(1) （4 件）
・情報メディア教育(2) （6 件）
・情報メディア教育(3) （5 件）
・情報メディア教育(4) （5 件）
・情報メディア教育(5) （5 件）
・ポスターセッション （16 件）

3　付録論文

当部門からは、「IC カード機能付学生証を用いた出欠管理システムの構築と運用」（発表者、佐藤）、「IC カードを使用したパスワード通知書発行システム」（小川）、「オープンソースを利用したウェブキオスク端末の構築」（前田）、「東京大学 ECCS プリンタ課金システムにおけるポストペイ決済の導入」（安東）の口頭発表を行った。

平成18年度情報教育研究集会

IC カード機能付学生証を用いた出欠管理システムの構築と運用

佐藤 孝明　寺脇 由紀　関谷 貴之　尾上 能之　山口 和紀
東京大学 情報基盤センター
{takaaki,terawaki,sekiya,onoue}@ecc.u-tokyo.ac.jp,yamaguch@mail.itc.u-tokyo.ac.jp

1. はじめに

大学において、学生の授業の出欠を簡単に採取・管理したいとの教員からの要望がある。一方、受講者の人数や授業が行われる場所は授業によって様々である。そこで、東京大学情報基盤センターは柔軟かつ効率的に運用が可能な IC カード機能付学生証を用いた出欠管理システムを構築し、2005 年 10 月より運用を開始した。本論文は本システムの構成、システムの運用を通じて得られた知見について報告する。

2. 出欠管理システム

図 2 に全体構成を示す本システムは、ハンディターミナルで学生証を読み込んで得られた出席情報を、学習管理システム CFIVE[5] で管理するものである。以下、システムの詳細を説明する。

2.1 IC カード機能付き学生証、教職員証

本学の教職員、学生は東京大学ユーティリティカード (以降、UTCard と記述する) と呼ぶ IC カード機能付教職員証、学生証を所持する。この UTCard は、身分証明証はもちろん、図書館の入館ゲート、生協における買い物に利用されており、特に本センターの教育用計算機システム (ECCS) においてはパスワード通知書発行システム [4]、プリンタでのポストペイシステム [3] に利用されている。出欠管理システムでは、UTCard 上にある共通 ID とカードの再発行回数を読み取る。この共通 ID が本学の全構成員を識別する ID である。

2.2 ハンディターミナル

図 1 は UTCard を読み取るハンディターミナルと、ハンディターミナルで読み込んだ出席データをサーバに送信するための通信ユニットである。据置型のカードリーダと比較してハンディターミナルは、工事が不要で屋外でも利用できる利点があるが、採取したデータを一時本体に蓄積する事から、送信前に故障するとデータが失われる欠点がある。このハンディターミナルは汎用的な製品のため多数のボタンが配置されているが、利用する教員が操作に困らないよう、次のようにインタフェースを設計した。

(1) 出欠を採取する講義を決める「時間割番号入力画面」、UTCard を読み取る「学生証読取画面」、採取した出席データを送信する「サーバに送信画面」に、常に決まったボタンで移動可能とした。
(2) 時間割番号が入力済みの時は、最も利用頻度の高い学生証読取画面を電源投入時に開くようにした。
(3) 誤って出席データを削除しないように、削除用のボタンは設けず、メモリが不足する時に送信済みのデータを削除することとした。

図 1 ハンディターミナルと通信ユニット

さらに、時間割番号入力画面では、数字ボタンによる番号入力と対応するバーコードの読込みという 2 種類の方法で時間割番号を入力可能である。

ハンディターミナルは次の 3 種類のファイルを格納する。

(1) 授業情報マスタ: 時間割番号、講義情報など
(2) 制御情報マスタ: 出席、欠席、遅刻、早退など処理内容とそれを表すコード
(3) 出席記録 IC カードファイル (出席データ): 時間割番号、共通 ID、再発行回数、読み取り時刻など

(1) は時間割番号入力画面で番号を入力した際に、対応する講義の名称や担当教員等を表示するために用いる。(2) は出欠以外の情報を記録するために用いる。ハンディターミナルで読み込んだ UTCard の共通 ID や読み取った日時は (3) に蓄積される。(3) は講義を識別する時間割番号毎に異なるファイルとなる。(1) ～ (3) のファイルはハンディターミナルを通信ユニットに置いて必要な操作を行う事で、出欠管理サーバとの間で送受信を行う。

2.3 出欠管理サーバ

ハンディターミナルから送信される出席データを CFIVE サーバに登録できる形に変換する役割を担っている。ただし、利用者が操作するのはハンディターミナルと CFIVE であり、出欠管理サーバを利用者が操作するケースはない。このサーバには次のファイルを保持する。

(1) IC カードファイル: 共通 ID、学生証番号、所属、再発行回数など
(2) ユーザ名プレフィックスファイル: 所属部局を表すコード
(3) 更新用授業情報マスタ、制御情報マスタ

また、主に次の役割を担う。

A. ハンディターミナルから送られた出席データ上の共通 ID と再発行回数の情報を、(1)、(2) に基づいて個人を特定し、CFIVE 上のユーザ名に変換する。
B. A. の結果を授業の出席情報 (出席記録ファイル) として、CFIVE サーバに送信する。

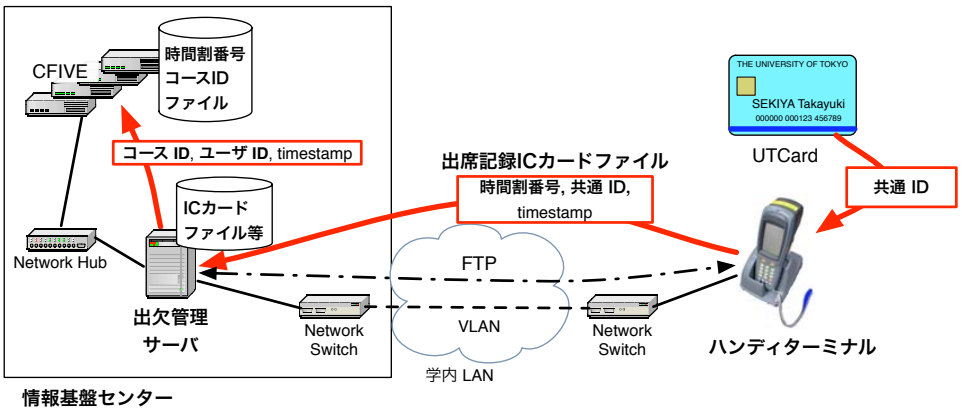


図 2　出欠管理システムの仕組み。学習管理システム CFIVE の継続的な開発と運用の評価 [2] より。

C.　必要に応じて (3) をハンディターミナルが受信可能とする。

ハンディターミナルからの出席データの送信を、5 分毎に調べ、新たなファイルが存在すれば自動的に A.、B. を行う仕組みとなっている。

なお、当初個人の特定に当たっては、学生証上の共通 ID と再発行回数の両方が一致する事を条件としていた。しかし、実際に運用を開始すると、再発行回数が一致しない学生証を保持する学生が居る事が分かった。これには、紛失等で再発行された UTCard の情報が本センターに伝達されず、再発行回数の値がファイル (1) と学生が保持する UTCard とで異なってしまった等の原因があった。結局再発行回数によらず CFIVE に登録するように変更したが、ログや元データは残すことで、後にチェックすることを可能にしている。

2.4　CFIVE

ユーザが出欠情報を管理するインタフェースは CFIVE を用いている。CFIVE は、2004 年 4 月より学内向けのサービスとして本センターによって運用が始まり、2006 年 9 月末までに計算機関連・図形科学実習・語学・物理・化学などの 80 以上の講義 (コース) で利用されている。

2.4.1　CFIVE サーバ上での処理

CFIVE サーバは、時間割番号と CFIVE 上のコースの ID との対応関係表となる時間割番号コース ID ファイルを保持する。出欠管理サーバから送信された出席記録ファイルの時間割番号に基づいて CFIVE 上の対応するコースを特定し、UTCard を読み取った時刻に該当する授業に出欠情報を登録する。当該時刻に開かれている授業のデータが CFIVE 上に無い場合、当該”日時”の授業のデータを自動的に作成した後の登録となる。

一方、学期の初めの授業では履修登録が完了しておらず、CFIVE 上に受講者が未登録の場合がある。しかし、本システムでは UTCard を読み取った学生が、CFIVE 上の当該コースに登録済か否かに関わらず、その学生の出欠情報を登録する。このように本システムでは、事前に学生の CFIVE への登録といった準備をしなくても出席情報を採取できるよう工夫している。

図 3　CFIVE における出席一覧の画面

2.4.2　CFIVE における出欠管理

教員や TA は、CFIVE にログインし、「授業」機能を利用する事で、授業毎の学生の出欠情報を管理する事ができる。行える操作は以下の 2 点である。

- CFIVE で出欠情報を修正できる。
  UTCard を学生が忘れたり、UTCard がハンディターミナルで読み込めなかった場合に、CFIVE を通じて教員や TA が出欠情報を修正できる。
- CFIVE から出欠情報をファイルとしてダウンロードし、成績の編集に利用できる。
  表計算アプリケーション等で処理できるように、授業毎の出欠情報を CSV 形式と HTML の表のファイルとしてダウンロードできる。

このファイルに記述されているデータは、
受講者 ID, 受講者名, 状態, 出席状況, 登録日, 更新日
となっている。ここで、「状態」とは、「出席状況」が出席の場合”0”、欠席の場合”1”であり、さらに、出席を取った時間を確認できるように登録日の項目が設けられている。また、これらのファイルは文字コードに UTF-8 を使用しているが、UTF-8 で記述されたテキストファイルを Microsoft Excel で読み込むと文字化けする現象 [7] がある。一方で、Excel の仕様で、UTF-8 の TABLE タグで構成さ

れた HTML ファイルの読み込みでは文字化けが起こらない。よって、ダウンロードした出欠情報の操作にスクリプトなどを作成、表計算ソフトを利用など、ユーザの環境の違いによる対応ができるように 2 種類のファイルをダウンロードできるようにしてある。

また、学生は CFIVE の「授業」機能を用いて、受講する講義の各授業における自分の出席状況を確認できる。

このように、採取した出席データを教員や TA、受講者である学生といった利用者が管理したり、閲覧したりするにあたって、CFIVE の 1 機能を利用することで別のシステムの操作方法を覚えることがなくなり便利になった。

2.5 出欠管理システムの構成

システム構成を図 2 としたのは以下の理由による。

(1) サーバ、ハンディターミナル間で送受信されるファイルを覗かれることを防ぐ。

(2) 通信ユニットでのファイルの送受信は FTP のみに対応する。

ハンディターミナルは先生が立ち寄り易い事務室に置き、出欠管理サーバは安全に管理するためにサーバ室に設置するため、両者は物理的に離れている。更にセキュリティとコストのバランスに配慮した上で、上記条件を満たす手段として、サーバとハンディターミナルで構成されるネットワークを、VLAN を使って学内の一般のネットワークから分離した。

3. 実際の運用とユーザの反応

ここでは、システムが利用されるまでの流れについて説明する。また、現在は教養学部前期課程の授業でのみ出欠管理システムが利用されているが、これらの授業を通して収集できたユーザの反応についても述べる。

3.1 実際の運用形態

教員は出欠管理システムの利用に際し、学期の最初に次の 2 点を行う。

(1) ハンディターミナルを借りる曜日や時限を事務室に連絡

(2) CFIVE の利用登録申請

次に、毎回の授業における教員或いは TA の出欠管理システムの利用手順は以下の通りである。

(1) 事務の窓口でハンディターミナルを貸りる。

(2) ハンディターミナルに講義の時間割番号を入力し、表示される講義名を確認する。

(3) 授業が行われる教室において、受講者である学生が所持する UTCard を読み込む。
このとき、UTCard を忘れたり、読み込みができなかった UTCard を所持する学生がいた場合はその学籍番号をメモする。

(4) 授業終了後、窓口に設置されている通信ユニットにハンディターミナルを置き、採取した出席データをサーバに送信、返却する。

(5) 必要があれば、CFIVE にログインし出欠情報の修正を行う。

学生は授業終了後に CFIVE にログインすることで、自分の出席状況の確認が行える。

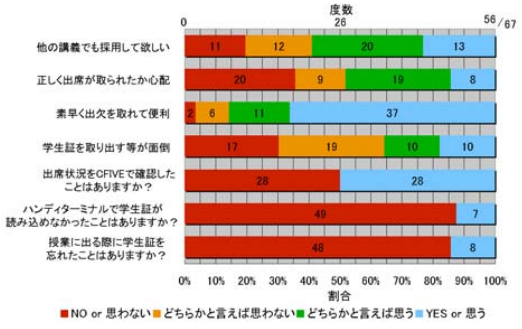


図 4 アンケート結果

本センターは、故障などの突発的な事態への対処や日頃のメンテナンス以外に、以下の作業を行う。

- 年度始めに、大学本部から学生の所持する UTCard に関するデータの提供を受け、出欠管理サーバの IC カードファイルを更新する。
- 学期 (半年) ごとに、教養学部教務課から授業情報が記載されたファイルの提供を受け、出欠管理サーバ内の授業情報マスタを更新する。
- ハンディターミナル中の授業情報マスタの更新作業を各事務の担当者に行ってもらう。

各事務の担当者はハンディターミナルを操作し、授業情報の更新を行うが、その作業はパスワード認証の後に開く管理メニューにおいて、授業情報マスタの受信画面を呼び出し、通信ユニットに置くだけと単純な操作となっている。

2005 年 10 月以降、約 1 年間の運用において IC カードファイルなどのサーバ上の各ファイルの更新を容易にするスクリプトを作成したり、cron を使用して、時間割番号コース ID ファイルといった定期的な更新を要するファイルの自動更新を行う仕組みを作成したりといった努力を続けてきた。他にも、教養学部教務課から提供される授業情報マスタの提供をオンラインで行える仕組みを利用した。結果、システム自体については管理者が時間を割く必要がない環境になりつつある。

3.2 ユーザの反応

3.2.1 学生の反応

2005 年 10 月に運用が開始された後、約半年間、履修者 400 名の規模の講義において試験運用が行われた。この講義では、教室の入り口にハンディターミナルを所持した TA を数名配置し出欠を取るという形式を採っている。この運用の中で UTCard の不良が原因と思われるカードを所持する学生がおり、出欠の採取が行えないケースが約 20 件ほど発生したが、授業終了後に CFIVE において出欠情報の修正が行えたために、該当者の氏名、学生番号を紙に控えるのみの対応で済み、問題とはならなかった。

この講義の最後の授業にアンケートを行ったところ、結果は図 4 の示すものとなっている。「素早く出欠を取れて便利」だと思う学生が 85 % に上った。意見には JR 東日本の Suica 利用時の自動改札のような印象を持っている傾向が感じられた。ハンディターミナルそのものに対して

も、出席を採取するにあたって今回の製品の選択がベストのものか疑問に持たれているようだ。また、「正しく出席が取られたか心配」の「はい」と「いいえ」が約半々になっていることは、学生への出欠管理システムについての説明が十分と言えず、CFIVE により出欠の確認が行えることが周知されていなかったのが原因と考えられる。また別の講義にて収集された意見の中には、ハンディターミナルにおいては読み取りに成功した際に「ピッ」という電子音が鳴り振動するが、これでは成功なのかどうかわかりにくい、さらに、ボタンがたくさんあって使い方がわかりにくいという声もあった。

3.2.2 教員の反応

2006 年 4 月より、8 つの講義で出欠管理システムが利用された。システムを利用した教員からはメールやアンケートにより、以下の意見が得られた。

本システムに関わる CFIVE の「授業」機能に対する意見

(1) 各回の授業ごとではなく、縦にアカウント名、横に授業名がある表がダウンロードできるといい。
(2) 出席一覧画面に授業を判別できる情報が欲しい。
(3) 出席一覧のファイルの拡張子は csv がよい。
(4) 出席一覧のファイルの名前に授業の情報が欲しい。

受講者の人数やハンディターミナルの利用方法、利用状況

(5) 約 60 ～ 130 人程度の受講者がいる授業で用いられた。(カードのトラブルで読み取れない学生の数は 1, 2 人程度の割合)
(6) 1, 2 台程度のハンディターミナルが授業で使われる。
(7) 授業中でのハンディターミナルの運用形態は、先生が机の間を回る、TA に任せる (机の間を回る or 教室の入り口)、学生に本体を回す、と様々な方法が取られていた。

それ以外の本システムに対する意見

(8) ハンディターミナルに入力する時間割番号を間違えると特殊な処理が必要になってしまう。
(9) 期末試験のときは便利かもしれない。

(1) の要望は出席の状況が一目で把握できる表が欲しいという要望であり、現状では各回の授業ごとの出欠情報が記録されたファイルから、表計算ソフトなどを使用して自分で作成しなければ同等のものは得られない。教員にとってこの作業は面倒であると受け取られているようだ。(2) ～ (4) は概ね現在の CFIVE における出欠管理システムが係わるインタフェースの不備を指摘したものである。いずれももっともであり、今後の CFIVE の改修に合わせて修正される予定である。

さらに、(5) ～ (7) によって、実際の授業でどのように用いられているのかがわかった。特にハンディターミナルの運用方法は利用されている講義が少ないにも係わらず様々な方法が取られる傾向があるようだ。

また、ハンディターミナル上で間違えて入力した時間割番号を修正する方法がなく、間違えた時間割番号のまま読み取りを行ってしまった場合はもう一度採取をやり直すか、データを送信してもらい担当者がデータの修正作業を行うしかないという不備があること、まだ要望はないが一回だけ使いたいといったケースに対し、CFIVE の利用登録申請など面倒な点があるのではないのか考える必要があることが明らかになった。

4. まとめと考察

約 1 年間の運用において見る限り、ユーザからは概ね良い評価を得られた。特に、

(1) 講義の履修者が確定していなくても出欠が取れる。
(2) CFIVE の 1 機能として利用できる
(3) 授業にてハンディターミナルの運用が柔軟に行える。

から、以下のような運用が可能になった。

(1) 1 回目の授業からシステムの利用が可能。(2.4.1 参照)
(2) ユーザは多くの操作を習得する必要がない。
(3) 各教員の方針に従ったシステムの利用ができる。

このような柔軟さが好印象が得られた原因だろうと考える。しかしながら、ユーザからの声によって以下の点の改善が必要なことも分かった。

- ハンディターミナルにおいて UTCard の読み取りが成功したかが分かりにくい。
- 誤ってハンディターミナルで設定した時間割番号による出席情報の修正が簡単にできない。

しかし、ハンディターミナル上のソフトウェアの改良には、ROM を修正せねばならず本センターでは作業ができない。そこで本センターで改良が可能な CFIVE や出欠管理サーバ上のソフトウェアの改良で対処したい。

謝　　辞

本論文のテーマである出欠管理システムの共同開発者である NTT コミュニケーションズと DENSO の方々に深く感謝致します。この場をかりてお礼を申し上げます。

参 考 文 献

1) 関谷 貴之, 寺脇 由紀, 尾上 能之, 山口 和紀, 学習管理システムの開発と運用 : 情報教育シンポジウム SSS2004, pp. 141 - 144 (2004)
2) 関谷 貴之, 佐藤 孝明, 寺脇 由紀, 尾上 能之, 山口 和紀, 学習管理システム CFIVE の継続的な開発と運用の評価 : 情報教育シンポジウム SSS2006, (2006)
3) 安東 孝二, 山口 和紀, 関谷 貴之, 東京大学 ECCS プリンタ課金システムにおけるポストペイ決済の導入 : 平成 18 年度情報教育研究集会, F4-5 (2006)
4) 宮嵜 洋, 小川 大典, IC カードを使用したパスワード通知書発行システム : 平成 18 年度情報教育研究集会, F2-4 (2006)
5) CFIVE Project (2006), http://cfive.itc.u-tokyo.ac.jp/
6) UTcard を用いた出欠管理システム (2005), http://cfive.ecc.u-tokyo.ac.jp/DigitalLife_2005_11/iccard.html
7) UTF-8 形式のテキストファイルが文字化けする, http://support.microsoft.com/default.aspx?scid=kb;ja;821863
8) 東京大学情報基盤センター (2005), IC カードによる出欠管理システム 操作説明書, http://cfive.itc.u-tokyo.ac.jp/iccard/manual.zip

IC カードを使用したパスワード通知書発行システム

宮嵜 洋、小川 大典
東京大学情報基盤センター

東京大学情報基盤センター教育用計算機システムでは毎年、前年度からの継続ユーザに対してアカウントの更新を行っている。従来はパスワードを印刷した通知書を事前に封緘しておき、配布時に学生証を確認して本人に封筒を手渡す手続きを行っていたが、本学学生証の IC カード化を機に IC カードリーダによる読み取りとその場で印刷を行う通知書発行システムを開発した。ここでは、本システムの概要と構成について述べる。

1 経緯

教育用計算機システム(Educational Campuswide Computing System, 以下 ECCS)は、本学の学生や教職員が教育や研究のために使用する計算機システムである。ECCS ではアカウントの新規発行、パスワード再発行、次年度への継続発行といった契機にユーザへパスワード通知書を配布している。これまで通知書の配布は、学生証を確認しつつ、学生証番号に対応するユーザの通知書（封筒又は用紙）をストックから探し出し手渡すという方法で行ってきた。特にパスワードの変更を伴う年度更新は前年度から継続するユーザ全てが対象となるため、その数は約 2 万人に及び、通知書の準備とその配布にかなりのコストを費やしていた。

その中、2005 年 4 月に本学では学生証及び教職員証の IC カード化が実施され、全学共通 ID（以下共通 ID）と呼ばれるユニークな番号が記録された IC カードが全学生及び教職員（一部を除く）に配られた。

共通 ID は学生証の IC カード化以前から事務処理等に利用されており、ECCS でもユーザを特定する情報の一つとして扱ってきた。共通 ID からユーザ名への変換が容易であったことから、窓口で学生証の提示を求めるパスワード再発行手続きに利用できると考え、2005 年 10 月に IC カード読み込みプログラムの開発に着手した。IC カードに関する規格、仕様の調査から、カードリーダの入手、動作確認を経て 2006 年 1 月下旬にはパスワード再発行システムが完成した。

さらに 2006 年 5 月には次年度への継続ユーザのアカウント更新に対応すべく拡張を行った。本システムにより、これまで毎年行ってきた事前に封緘、配布時に封筒を検索するという方法から、配布時に学生証を読み取り、その場で印刷する方法へと手続きの形態が変わった。

2 システム構成

2.1 概要

本システムは IC カードリーダ、プリンタが接続された PC とユーザ情報を管理するデータベースサーバ（以下 DB サーバ）で構成される。

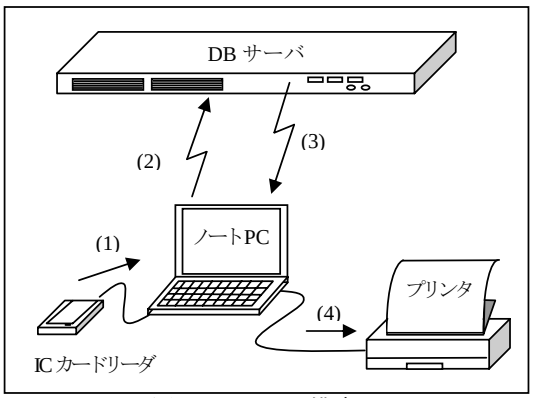


図 1 システム構成

動作は次のようになる。

(1) PC で IC カードの共通 ID を読み取る
(2)サーバにユーザ名を問い合わせる

(3)ユーザに対応した印刷ジョブを PC に送る
(4) PC は通知書を印刷する

2.2 IC カードとカードリーダ

本学で学生証・教職員証に採用されたICカードはISO/IEC 14443 TypeB（非接触）、及びISO 7816（接触）という国際標準規格に相当するものである。また、インターフェース規格はPC/SC(Personal Computers/SmartCard)[1]に準拠しており、仕様が公開されているため我々でも開発が可能であった。また、ICカードに搭載されている共通IDアプリケーションのインターフェース仕様の著作権は本学が所有しており、それを使用することができた。

カードリーダには非接触型と接触型がある。ICカードを軽くタッチするだけでよいという操作性を考慮して非接触型が候補であったが、カード損傷時など非接触で読み込めない場合でも対応できるよう接触型としての機能もあるSCR331DI-NTTCom（NTTコミュニケーションズ）[2]を採用した。本製品はPCとUSBによる接続が可能であり、Windows、Mac OS X（PPC、Intel）に対応したドライバが用意されている。

2.3 サーバワークステーション

DB サーバは本システムのために特別に用意した物ではなく、ユーザ管理用に以前から使用しているワークステーション（OS は Linux）である。元々パスワードの発行や通知書の印刷は、このサーバ上のスクリプト（Perl）を実行し、データベースを操作していたが、今回はクライアント PC からの指令を受けてサーバプログラムがスクリプトを起動する仕組みとした。本プログラムは Perl で記述しており、クライアントとの通信には socket を使用している。

2.4 クライアント PC とプリンタ

IC カードから共通 ID を読み込むための PC（OS は WindowsXP）で動作するプログラムは主に IC カードリーダを制御するものであるが、取り扱うデータが個人情報を含むため、PC 上にはデータを保持しないよう設計した。

また、このPCはプリントサーバとしての機能も果たしている。ICカードを読み取ったその場で印刷する必要があるため、プリンタはUSBで接続可能なレーザプリンタ（20ppm）とした。ただし、DBサーバから送られるPostscript形式のデータを扱うため、Redmon(Redirection Port Monitor)[3] 及びGhostscript[4] を用いてPC上で変換し印刷するという方法をとっている。

なお、本 PC はユーザインターフェースも兼ねているため、受付窓口やロビーなどでも使用できるよう、運びやすいノート PC を採用した。

3 プログラム概要

3.1 ユーザインターフェース

実際の使用例を簡単に解説する。まず PC 上で実行ファイルを起動するとコマンドプロンプトが開き、次のメニューが表示される。

```
[メインメニュー]
  1. 窓口受付（パスワード再発行など）
  2. 継続発行（期間中のみ）
  3. 保守メニュー
  q. 終了
  number>
```

通常は 1. の窓口受付を選択し、次の画面でカードの読み込みを待ち受ける。

```
[窓口受付]
----------------------------------------
カードを読み込んでください。[or q(quit)]
----------------------------------------
```

窓口に来たユーザからカードを受け取り、カードリーダにタッチすると、カード認証とともに共通 ID を読み込む。正当なカードと認識すると DB サーバに問い合わせて共通 ID に該当するユーザ名及びユーザの情報を取得し、画面に表示する。

```
>>> カード認証 OK(共通 ID: 00000000000000001)
  ユーザ名(uid):  xxxxxxxx
  氏名(fullname): ○○○○
```

このときユーザが既に通知書を受け取っていれば、

```
>>> 今年度の継続通知書（又は新規）は受け渡し
    済です。
  パスワード再発行をしますか？(y/n)
```

となり、受け取っていなければ、

```
>>> 今年度の継続通知書（又は新規）を受け取っ
    ていません。
  継続通知書を発行しますか？ (y/n)
```

と、通知書発行の履歴により次の処理が変わる。いずれの場合も処理を実行するならば「y」、そうでなければ「n」と答える。「y」の場合は通知書の

印刷が行われ、その後「カードを読み込んでください」の表示に戻り、再び待ち受け状態となる。

以上が一連の動作例である。処理の結果はサーバ上の履歴ファイルに記録され、次回以降、同じカードが読み込まれたときの処理に使用される。

3.2 カード制御

カード制御は本システムの主要な部分であるが、これにはリソースマネージャと呼ばれるプログラムが介在する。リソースマネージャにはICカード（リーダ）を制御する関数[5]が用意されており、次の関数を(1)～(6)の順に呼び出すことでICカードとの通信ができる。

(1) SCardEstablishContext()
　リソースマネージャとの通信を確立。
(2) SCardListReaders()
　接続されているカードリーダを確認。
(3) SCardConnect()
　カードと接続し、カードの状態を確認。
　（戻り値をチェックすることによってカードが置かれたかどうかなどがわかる。）
(4) SCardTransmit()
　データを送出し、結果を受信。
　（カードアプリケーション固有の命令を送出する。戻り値にカードからの情報が含まれる。）
(5) SCardDisconnect()
　カードとの接続を切断。
(6) SCardReleaseContext()
　リソースマネージャとの通信を終了。

これらの関数を使用するために、本プログラムはC 言語で記述した。Windows の場合リソースマネージャは OS の機能に含まれており、ライブラリ ”winscard.lib” をリンクすることで実行ファイルを作成できる。今回はその環境が整っているVisual C++（Microsoft Visual Studio 6.0）を使用した。なお、他のOSにおいては、例えばMUSCLE（Movement for the Use of Smart Cards in a Linux Environment）[6]の提供するライブラリが同等の機能を備えており、MacOS X（Xcodeに含まれる）やLinuxでの利用も可能である。

3.3 サーバプログラム

DB サーバは主に共通 ID とユーザ名の対応、パスワードの変更処理、履歴の管理、通知書の作成と印刷の指令等を担っている。クライアント PC からの通信を待ち受け、指示に従って DB サーバ上のスクリプトを実行するために、inetd 起動によるサーバプログラムを作成した。例えば以下のようなコマンドを用意している。

(1) challenge
　カード認証のための乱数リクエスト。
(2) auth　<カードの応答>
　カードの正当性をチェック。（カードの応答とサーバの計算の比較による確認。）
(3) user　<共通 ID>
　カードの共通 ID からユーザ名の検索。
(4) info　<ユーザ名>
　ユーザ情報の取得と表示。
(5) request　<ユーザ名>
　パスワード再発行リクエスト。
(6) print　<ユーザ名>　<通知書の種類>
　通知書の印刷リクエスト。

これらのコマンドは従来から ECCS で使用している継続通知書やパスワード再発行通知書などの発行用のスクリプトを呼び出す仕組みとなっている。それと同時に履歴を保存し、過去にユーザが行った処理（再発行や更新など）に応じて次回の処理が異なるようにプログラムしている。

複数の PC から同時にリクエストが来る可能性があるため、サーバプログラムが使用するファイルは排他制御を行っている。また、印刷処理中に作成される一時ファイルにはプロセス ID 番号を含んだファイル名を付けるようにし、ファイル名が同一にならないよう工夫している。

3.4 データの流れ

カードリーダに IC カードがタッチされてから、印刷までのデータの流れを示す。基本的に PC がカードやサーバに指示やデータを送り、結果を受け取る形態である。そのうちカードの認証を行っている部分について解説すると、

(A) サーバに乱数列の生成を要求
(B) カードに乱数列を送り、暗号化させる
(C) サーバでも同様の暗号化を行う
(D) (B)の結果と(C)の値を比較する

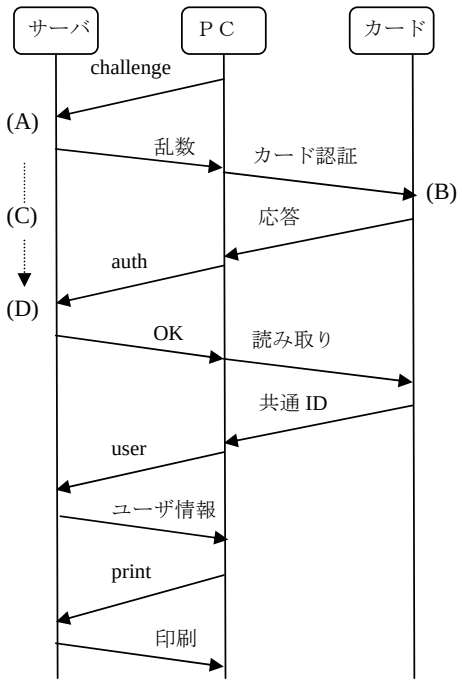


図 2　データの流れ

という手順である。(B)はカードのアプリケーションが持つ認証機能であり、本システムでは(D)の結果が正しければ偽造カードではないとしている。

カード認証を行った後、引数の与え方次第では別のユーザの情報も取得できてしまうが、サーバへ接続可能な PC を IP アドレスで制限することでセキュリティを確保している。

4　結果

本システムはまずパスワード再発行システムとして、2006 年 1 月下旬に試験的に窓口に 1 セット（PC、カードリーダ、プリンタ）設置した。年度末を迎え、パスワード再発行の手続きの件数は少なかったので、この間に動作不良の修正やメッセージの変更、機能の追加等を行うことができた。その後、新年度のアカウント継続手続きに向けて機能を拡張し、2006 年 5 月 9 日～5 月 18 日（駒場）、5 月 19 日～6 月 8 日（本郷）の配布期間までに、本郷、駒場に 3 セットずつ合計 6 セットを揃えた。継続手続き期間後は、1 台もしくは 2 台を窓口に常置して期間中に受け取れなかった学生や、パスワードの再発行に引き続き使用している。

動作に関しては、IC カードの読み込みから印刷完了までは約 20 秒程度であり、待ち時間を感じさせない。また、1 台が紙切れや紙詰まりで停止しても、他でカバーすることで滞りなく配布することができた。今年度の継続手続き期間には約 10,000 件の処理をおこなったが大きな支障はなく、1 時間に最大 150 件（同時に複数台使用）を処理することもあり、予想以上に順調であった。このため従来の手作業での封筒検索と比べて省力化が実現でき、配布担当者には特に好評であった。

また、従来は学生の所属によって本郷・駒場キャンパスのいずれかに封筒がストックされていたため、封筒が無い方に学生が来たときはその場で手渡せなかったが、本システムによってどちらのキャンパスでも更新手続きができるようになった。

5　おわりに

IC カードは急速に普及しており、本学でも既に学生証・教職員証が出欠管理、プリンタ課金、図書館入館、健康診断等に利用されている。そのような最中、本システムの開発する機会を得られたのは幸いであった。今回の構築の過程で分かったことは、IC カードは規格に則っているため、ごく一般的なツールで開発可能なこと、プログラミングが容易であること、OS やアプリケーションに制約をあまり受けないことである。他のシステムへの組み込みも容易とあってか、本システムの副産物である共通 ID を読み出すプログラムをツールとして学内のサービス（図書貸し出し、バス予約システム）に提供する機会も得た。このことはコスト軽減のみならず、身近なサービスへの幅広い応用の可能性を示すことができたといえる。

今後の課題として、窓口の手続きやユーザ登録の簡略化など、他の業務へさらに IC カードが応用できないか、検討してみたい。

参考 URL

[1] http://www.pcscworkgroup.com/
[2] http://www.ntt.com/jpki/
[3] http://www.cs.wisc.edu/~ghost/redmon/
[4] http://www.cs.wisc.edu/~ghost/
[5]http://msdn.microsoft.com/library/default.asp?url=/library/en-us/secauthn/security/scardtransmit.asp
[6] http://www.linuxnet.com/

オープンソースを利用したウェブキオスク端末の構築

前田 光教

東京大学 情報基盤センター
maeda@ecc.u-tokyo.ac.jp

概要：学生窓口において教育用計算機システムの広報や FAQ を確認することを目的としたタッチパネル式のウェブキオスク端末を、試験的に 1 台導入した。キオスク端末専用機ではなく、汎用の端末やオープンソースソフトウェアの利用を中心とし、入力方法にはタッチパネル式のモニタとした。本稿ではその端末の構築方法を報告する。

1 導入の経緯

情報基盤センターでは学生や教職員に教育や研究のための教育用情報基盤を提供するシステムとして教育用計算機システム(Educational Campuswide Computing System, ECCS)を運用している。学部 1 年の必須講義である「情報処理」を中心として講義の支援を行っているが、情報とは関連のない講義にまで ECCS 端末が利用されることが増えている。そのため、講義の形態が多様化してその講義を支えるシステムも多様化せざるを得ずシステムが複雑になっている。

システムが多様化し複雑になるとユーザの個人環境も多様化して発生する問題も多様化する。初出の障害の復旧に全力で当たるのは当然であるが、既出の定型的な問題も多くあり、そうした問題には自己解決できるようにウェブによる広報や FAQ を充実する努力を行っている。それでも気付かずに窓口に来ることは多い。しかし、窓口は対応できる人材も対応できる時間も限られているために、時間帯によっては、次の日まで問題を解決できる手段がなくなってしまう。窓口には掲示板を設けてウェブの広報などを印刷して張り出してあるがその情報は限られている。紙に印刷されている方が分かり易いという見解はあるものの、ウェブでの情報量とでは比較にならない。特に簡単な内容の確認にはウェブの方が優れていると考える。そこで、窓口が開いていない時間にも、ある程度の問題解決ができるようにウェブキオスク端末を設置することとした。

図 1　ウェブキオスク端末

図 2　操作画面

2 端末の構成

構築する前提として OS には Linux を利用することを決めていたこともあり、端末とモニタは Linux での動作保証か実績があることが決め手となった。端末には「プロサイド Mecolo a121」[1]、モニタには「NDS LCD タッチモニタ LC-19NA-USB」[2] を採用した。どちらも「Redhat9」ではあるが Linux の動作を保証している。それをルータの NAT(Network Address Translation)を利用してインターネットと接続した。端末は特別なものではなく通常の PC で、その選択肢は無限にあると言っていい。タッチパネ

ルはドライバが必要でLinuxのドライバを提供するメーカは少ない。しかし、表向きにサポートしていなくとも直接電話で問い合わせれば導入した実績のあるメーカもあった。

OSには「Redhat9」と構成が似ている「VineLinux3.2」[3]を選択した。「Redhat9」は現在もFTP版と呼ばれる無料版が配布されていて利用できるものの、有志が行ってはいるが、正式なサポートが行われていないために選択対象としなかった。「VineLinux3.2」は最新のアップデートを行うことができる。

3 端末の構築

本章ではタッチパネル式ウェブキオスク端末の構築方法を説明する。

3.1 PCのセットアップ

- インストール時にX環境と「GNOME」デスクトップのカテゴリを選択
  ブラウザを利用するので最低限「X-Window(X.Org)」を選択しなければならない。特定ユーザで自動ログインさせたいのでその設定の容易な「GNOME」を選択した。インストールの途中で要求されるので一般ユーザとしてキオスク用ユーザを作成する。
- インストール終了後に各パッケージのアップデート(apt-getコマンドを使用)
- 「iptables」のインストール
  「iptables」はLinuxで汎用的に利用されているIPパケットフィルタリングソフトウェアである。
  ルータでもフィルタリングが可能な製品は多いので、どちらかに一本化した方が管理しやすいだろう。
- 「IPAフォント」のインストール
  Linuxはフォントが奇麗とは言えない。そこで情報処理推進機構(IPA)が地理情報システム(GIS)「GRASS」の一部として公開しているTrueTypeフォント「IPAフォント」[4]を利用することとした。
- ログイン画面「gdm」を制限
  「gdm」は「GNOME」で利用されているログインウインドウシステムである。
  「gdm」のログイン画面からログインウインドウの設定ができないようにする。キオスク用ユーザで自動ログインするようにする。時間待ちログインも設定する。リモートログインできないようにする。
- 自動起動サービスの見直し
  「kudzu」「murasaki」「iptables」「network」「syslog」「random」「sshd」「rawdevices」「crond」以外は起動しないようにした。「sshd」を起動するようにしたのは、本報告ではキーボードやマウスでの操作を制限するため、日々の運用はリモートで行えるようにするためである。
- ブートローダ「lilo」の設定(図3)
  ブート時に起動オプション(シングルユーザ起動など)を付けた場合はパスワードを入力しないと起動しないようにする。

```
password="********" <-- パスワード
restricted
```

図3　ブートローダの設定に追加

3.2 タッチパネルの設定

- タッチパネルの設定
  付属CD-ROMのLinux用インストール説明書を読んで設定する。必ず「root」でログインして作業すること(管理者権限のないユーザでは設定に失敗する)。
- ログイン画面を制限
  「gdm」のログインウインドウから「root」でログインできないようにする。

3.3 端末のBIOS設定

- キーボードが接続されていなくても起動するようにする。
- ハードディスクからだけしかブートしないようにする。
- BIOSのパスワードを設定する。

3.4 Operaのインストールと設定

- Operaのホームページよりパッケージをダウ

ンロード
「 opera-8.51-20051114.1-static-qt.i386-en.rpm」(2006 年 1 月 16 日現在の最新版)
日本語ファイルのダウンロード。「ouw850_ja.lng」(Linux 版は「言語ファイル」が必要)

- Opera のインストール(rpm コマンドを使用)
- 作成したキオスク用ユーザでログインしなおし、Opera を起動
- Opera の設定で JAVA などが動かないように設定
  「一般設定」の「起動」を「前回終了時の場所から続行する」(「前回終了時の場所から続行する」にしたのは、それ以外ではキオスクモードでフルスクリーンにならないため)にする。「ホームページ」を「http://www.ecc.u-tokyo.ac.jp/」にする。
- パネルの整理と配置
  「外観の設定」の「メインバー」にチェックして、配置を「下に表示」にする。
  button.html(図 4)をもとに、「下へスクロール」「上へスクロール」「FAQ」ボタンを作成し、表示した「メインバー」に「ホームページ」「更新」「戻る」「進む」「ページを閉じる」「下へスクロール」「上へスクロール」「FAQ」を配置する。
  「ページを閉じる」ボタンを作成したのは、html の「a」タグに指定できるターゲットに「_blank」とあると新規ページで開いてしまうことで、フルスクリーンモードではそのページを閉じない限り元のページに戻ることができなくなってしまうからである。
- フォントの変更
  すべてのフォントを「IPA ゴシック」か「IPA 明朝」にし、フォントサイズをモニタに合わせて大きくする。
- Opera がクラッシュしても再起動するようにシェルスクリプトを作成(図 5)

```
<a href='opera:/button/Scroll down,,,, Sort descending & Scroll down & Scroll down & Scroll down & Scroll down' title="下へスクロール">Scroll Down Button</a><br>
<a href='opera:/button/Scroll up,,,, Sort ascending & Scroll up & Scroll up & Scroll up & Scroll up' title="上へスクロール">Scroll Up Button</a><br>
<a href='opera:/button/Go to page , "http://www.ecc.u-tokyo.ac.jp/faq/" ,,, Panel Info' title="FAQ">FAQ Button</a><br>
<a href='opera:/button/Close page,,,, Blocked' title="ページを閉じる">Close current page</a>
```

図 4　button.html

```
#!/bin/sh
while true
do
        /usr/bin/opera -kioskmode -kioskbuttons -nocontextmenu -nodownload -noexit -nohotlist -nokeys -nomaillinks -nomenu -noprint -nosave -resetonexit www.ecc.u-tokyo.ac.jp
done
```

図 5　openopera.sh

- 作成した「openopera.sh」がログイン時に自動起動するように「GNOME」のログインウインドウシステムに設定
- すべてのパネルとデスクトップアイコンを削除
- スクリーンセーバーの設定
- 同時に、モニタのパワーマネジメントを有効にして「電源 OFF」の時間を設定する。

本章では書ききれなかった、より詳細な、コマンドレベルから記述したウェブページを用意したので参考にしていただきたい。
http://user.ecc.u-tokyo.ac.jp/~mmbit/webkiosk.html

4 Opera

ブラウザに「Opera」を選択したのは、キオスクモードが標準で実装してあり、ボタンの作成・

配置が自由に構成できる柔軟さと豊富な起動オプションの存在であった。「Opera」はオープンソースではなく実行ファイルでの配布となっていて 2005 年 9 月 20 日からは通常の PC での利用は無料となり、ライセンス料を支払わなければ利用できなかったキオスクモードが利用可能となった。一部の起動オプション[5]を図 6 に示す。

kioskmode **か** k	Opera **をキオスクモードで起動する。自動的に有効となるオプションがある。**
kioskbuttons	**メインバーとメインバーに配置されたボタンが表示される。**
nocontextmenu	**コンテキストメニュー(マウスの右クリック)を無効にする。**
nodownload	**ファイルのダウンロード無効にする。**
noexit	Opera **の終了を無効にする。**
nokeys	**キーボードショートカットを無効にする。テキストボックスなどへの入力は有効である。**
nomenu	**メインメニューを無効にする。**

図 6　「Opera」の起動オプション

5 評価

導入した 2006 年 1 月から 9 月までのアクセスデータをもとに、月別(図 7)、曜日別(図 8)で集計した。

本端末は、建物の出入口からも演習室からも離れていることもあり、窓口の開いていない時よりも開いている時に多く利用された傾向が現れた。当初の目的とは違っているものの、窓口の近くに設置したことで窓口業務の軽減につながっていることが確認できた。

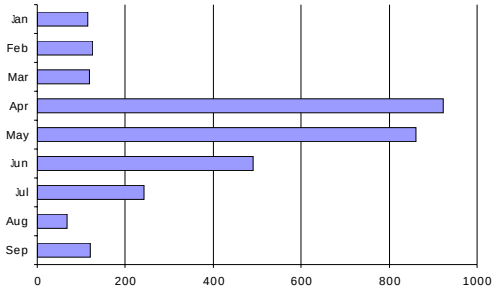


図 7　月別アクセス数

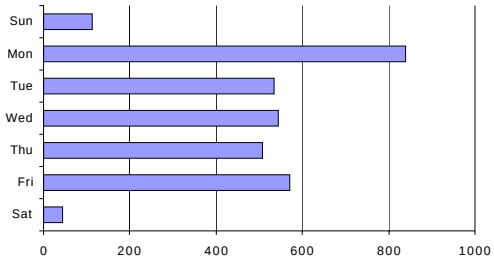


図 8　曜日別アクセス数

6 今後の応用

本報告では入力インターフェイスにタッチパネルを採用したが、インターフェイスをマウスとして、古いハードウェアを再利用することで費用を掛けずに構築することができる。キーボードとマウスを接続しても「Opera」にはショートカットを無視させることや「Opera」を終了できなくすることができるため、他のソフトウェアの利用制限を維持できる。メールクライアントをウェブメールで提供するシステムが増えていて、業務などに必要な作業がブラウザだけで完了するようなウェブアプリケーションが充実すれば、情報基盤システムの一つとして提供することもできるだろう。Linux に明るくなく Windows で構築したい場合にも、PC の購入時に付属する OEM 版の Windows の利用を前提としても同様に構築できる。Windows であれば PC もタッチパネルもさらに選択肢が増えることになる。

学生証の IC カード化が東京大学だけでなく他大学でも進んでいる。ウェブキオスク端末と IC カード読み取り装置とを連携することでパスワード再発行などの手続きを簡略化することもできる。

ECCS 以外でも、すでにウェブによる情報提供を行っているならば、本報告の設定のほとんどを変更することなく設置することができる。設置箇所によっては大きな効果が期待できる。

参考文献/URL

[1]http://www.proside.co.jp/product/detail.php?id=57378
[2]http://www.nds-jpn.com/Products/MagicTouch/LCD-TM/LC-19N.html
[3]http://www.vinelinux.org/
[4]http://www.grass-japan.org/FOSS4G/readme-grass-i18n-ipafonts.eucjp.htm
[5]http://www.jp.opera.com/docs/switches/#kiosk

東京大学 ECCS プリンタ課金システムにおけるポストペイ決済の導入

安東 孝二, 山口 和紀, 関谷 貴之

東京大学 情報基盤センター

ando@itc.u-tokyo.ac.jp,yamaguch@itc.u-tokyo.ac.jp,sekiya@itc.u-tokyo.ac.jp

概要：東京大学では学生証および職員証を IC カード化するとともに、その機能を用いて大学生協などにおいてポストペイ決済サービスを行っている。今回、教育用計算機システム（ECCS）のプリント課金システムに従来のプリペイドカード方式に加え、学生証や職員証を用いたポストペイ方式を導入したので、これを報告する。

1 背景

東京大学情報基盤センターでは、以前より情報基盤として教育用計算機システム(Educational Campuswide Computing System、以下 ECCS)を開発・管理・運用している。総発行アカウント数が３万を超えるシステムであるため、日々の運用コストが非常に大きい。ECCS は現在、東京大学の本郷キャンパス・駒場キャンパスそして新たに柏キャンパスにデスクトップ機を配置しているが、それを維持する技術職員は本郷キャンパスに２名、駒場キャンパスに３名しかいない。安定運用のためにはあらゆる努力が求められる。

運用コストの中でも特にプリンタに関するコストは非常に大きく、常に業務を圧迫してきていた。1995 年導入のシステムではプリント枚数の管理すること自体のコストを嫌い、プリント枚数のコントロールに多くのリソースを割くことはなかったが、WWW の爆発的な流行によりプリントシステムの問題と対峙せざるを得なくなった。

このころからプリント要求に従来に見られない変化が見られるようになった。WWW 自体の普及と使いやすい GUI の Web ブラウザの出現によって、プリントという行為が、能動的なものから受動的なものへと変化してきたことが最も顕著な変化である。従来はプリントするという行動は、自らが書いたソースコードをデバグするために出力したり、自分で編集したワープロの文書をプリントするという生産性を伴う能動的な行動であったものが、Web ブラウザを眺めて、リンクをクリックして、さらにプリントボタンをクリックするという、まさに TV 番組を眺めるのと同じような受動的行動へと変化したのである。その結果、ECCS では（教育的に必要と思えない）プリンタ出力の激増という現実に直面した。そのため、1999 年導入のシステム（以下 ECCS1999）では上限値管理を中心とするシステムを構築した。これを原型として、（株）リコーより Ridoc IO Gate が開発され、現在では全国約１００校で導入されており、現場での支持が得られている。また、2005 年には桝田らが Linux を用いたアプローチを発表しており（参考文献１）、プリントシステムの管理は大規模システムにおいて解決しなければならない問題の一つであることは明らかである。

図 1 教育用計算機システム自習室

2004 年 3 月に東京大学情報基盤センターは ECCS2004 へ全面的にリプレースした。(図 1) 全ての端末を diskless netboot で行うことや、端末の大半がアップル社の G4 になったことが大きく注目されることとなったが，その他にも大学の計算機センターによく見られる様々な問題にもチャレンジしている。従来からのプリントシステムの改善もその一つである。ECCS2004 のプリントシステムでは以下の点が改善した。

① 管理コストの低減
② 本当に必要なプリントを促進
③ カラープリンタの導入
④ 持ち込み PC への対応

図 2 ECCS2004 プリントシステムの制御端末とプリペイドカード読み取り機

問題解決に大きな役割を果たしたのはプリントを有料化である。プリペイドカードを販売しプリントに課金をすることで管理の外注化と無駄なプリントアウトの抑制を実現した。(図 2) それによりカラープリンタの導入も可能になった。持ち込み PC への対応も USB メモリと PDF を利用することで可能となった。(参考文献 3)

このような状況の一方で、東京大学では「多機能 IC カード導入プロジェクト」(cf. http://www.u-tokyo.ac.jp/public/documents/1613.pdf)が本格化し、学内ではポストペイによる IC カードでの小額決済が普及してきた。東京大学生協で利用できるなど学内での利便性は非常に高い。そこで、将来へ向けてのアプローチとして ECCS2004 のプリント課金システムについても、多機能 IC カードでのポストペイ小額決済への対応を図ることとなった。

図 3 統合型 IC カード職員証サンプル

2 IC カードと小額決済機能

東京大学では教職員および学生が利用する身分証明書の IC カード化を進めている。(図 3) 各建物の入退館システムへの対応、健康診断などでの利用、出席管理システムでの利用など大学システムの IT 化への一端を担っている。ECCS でもパスワード再発行の煩雑さを解消するため IC カード身分証を利用したシステムを開発、利用している。

これらの個人を特定する機能に加え、クレジ

ット機能を申し込むとポストペイ小額決済が可能となる。ポストペイは私的な決済に限られるため必ずクレジットの申し込みが必要である。月毎に与信枠内（３万円）を上限とした小額決済が可能で、後日まとめてクレジットカードへ請求される。毎回のチャージにかかる手数料を節約できるため小額決済が可能となっている。IC カード自体は接触、非接触のインターフェイスを持ったハイブリッドカードであるが、ポストペイの際は非接触インターフェイスが使われることが多い。

3　IC カード対応プリンタシステム

IC カード対応プリンタシステムは、ECCS2004 におけるプリントシステム（参考文献３）と以下の点で異なっている。

① IC カードによるポストペイ決済のみ可能な点
② USB メモリからの出力は PDF だけでなく MS Word などの比較的使われているフォーマットに対応した点
③ プリンタ制御端末が液晶一体型である点
④ ユーザ ID の入力を必要としない点
⑤ 課金の際にデポジットを必要とする点

①は運用および開発の都合上の理由でポストペイ専用とした。複数の決済を混在させることによるユーザの混乱を避けるためと、複数の決済手段への技術的対応によるコスト増を避けたためである。
②は(株)リコーでの商品化の際に付け加えられた機能である。
③は省電力と省スペースを鑑みて、液晶一体型の汎用 PC を制御用端末に採用したものである。そのためタッチパネル機能を得るため、サードパーティーのセンサーを後から取り付ける必要があった。
④は IC カードの基本的な機能を使ってユーザを特定することが可能にしたものであるが、身分証の本人の ID は東京大学で共通 ID と呼ぶ一意の ID であり、ECCS のアカウントとの関連はない。そのため、双方を結びつける DB を維持・管理する必要がある。今回は、PostgreSQL を用いた専用変換サーバを用意することで共通 ID と ECCS アカウントとの変換を実現した。
⑤は IC カードリーダに非接触型を選択したために必要となった。従来のプリペイドカード方式では、フロントローディング方式の磁気カードリーダを利用しており（図２）、一度プリペイドカードを挿入してしまえば、決済が終了するか決済を中止するまでユーザがプリペイドカードを持ち去ることは出来ない。しかし、今回の IC カードリーダには利便性のため非接触式を採用したためユーザは決済のタイミングで IC カードを読み取らせないことが可能である。そのため、今回のシステムではデポジットを取ることとし、システムの予想した金額をデポジットし、最後に精算する仕組みとした。ユーザは印刷の際に、ID 確認のため、デポジット支払いのため、精算のためと最大で計３回非接触式 IC カードリーダにカードをかざすこととなる。通常はデポジットのための予想金額と実際の課金が等しいので２回の操作で終わるはずである。

図 4　IC カードポストペイ対応プリントシステム

4　印刷の流れ

印刷の流れを制御端末の画面と共に説明する。

① IC カードをリーダにタッチしてログイン（図５左上）

② 印刷するジョブを選択し、印刷ボタンを押す（図５右上）

③ IC カードをリーダにタッチして、デポジット後、印刷を行う（図５下）

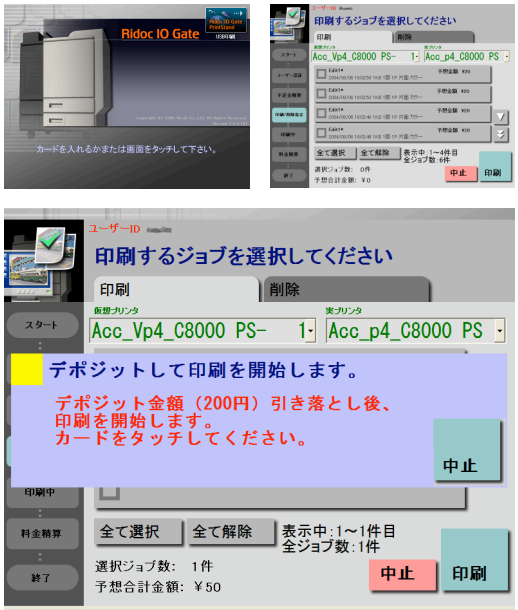


図 5　通常印刷の流れ

ただし、デポジット金額と実際の課金に不整合があった場合は図６のように再度 IC カードをリーダにタッチして精算のための手続きが必要となる。

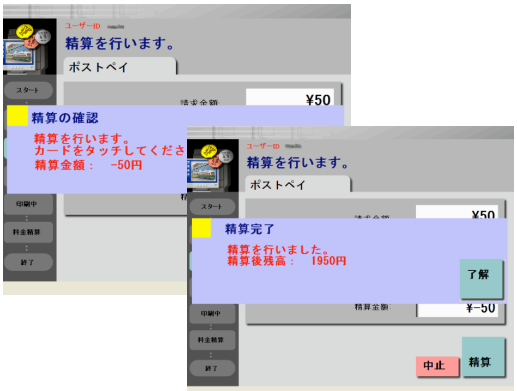


図 6　精算のための手続き

5　まとめ

2006 年９月より本郷キャンパスと駒場キャンパスにおいて各 2 台のポストペイプリンタの運用を開始している。全ての学生が利用できる訳ではないなど、制度上も含めて問題点はあるが、学内での小額決済の利便性を示すことは出来た。

謝辞

（株）リコー、NTT コミュニケーションズ（株）、（株）DC カードの方々に多大なご協力を賜りました。

参考文献

[1] 桝田秀夫、小川剛史、斎藤明紀、中村匡秀、近藤弘一、中西通雄、「教育用計算機システムにおけるプリンタシステムに求められる要求とその実装」、情報処理学会論文誌、Vol.46 No.4、pp.930—939、2005.

[2] 吉岡顕、田中哲朗、安東孝二、「低 TCO を目指した大規模教育用システムの設計」,情報処理学会分散システム/インターネット運用技術シンポジウム 2000、Vol.2000、pp.1-6, 2000.

[3] 安東孝二、関谷貴之、「大学における印刷管理の新しい試み」、情報処理学会分散システム/インターネット運用技術研究報告、No.2005-DSM-038、pp.13-16、2005.

平成 18 年度名古屋大学総合技術研究会

1　概要

平成 18 年度名古屋大学総合技術研究会が平成 19 年 3 月 1 日(木)から平成 18 年 3 月 2 日(金)に名古屋大学の主催で行われ、当部門からも職員が参加し発表を行った。

この研究会は大学共同利用機関および各大学・高等専門学校の技術者が、日常業務で携わっている実験装置の開発、維持管理の話題から改善、改良の話題に及ぶ広範囲な技術的研究支援活動について発表する研究会である。

2　プログラムの内容

名古屋大学経済学研究科カンファレンスホールに於いて、開会式、特別講演が行われ、名古屋大学工学研究科 IB 電子情報館中棟に於いて、口頭発表、ポスター発表が行われた。

○3 月 1 日(木)
・開会挨拶
　主催者挨拶
・特別講演
　「ものづくりと材料イノベーションに想う」
　　名古屋大学総長　平野眞一氏
・各分野口頭発表
・ポスター発表
・情報交換会
○3 月 2 日(金)
・各分野口頭発表
・ポスター発表
・見学会

(研究会の分野)

1. 機械・ガラス工作技術研究会
2. 装置技術研究会
3. 回路・計測・制御技術研究会
4. 極低温技術研究会
5. 情報ネットワーク技術研究会
6. 生物化学研究会
7. 分析・環境技術研究会
8. 実験・実習技術研究会

3　付録論文

当部門からは、5.情報ネットワーク技術研究会に参加し、「IC カードを使用したパスワード通知書発行システム」(前田、2006 情報処理教育研究集会における小川の発表と同等)の口頭発表を行った。

学内組織向けメールサーバ(MAILHOSTING)運用報告

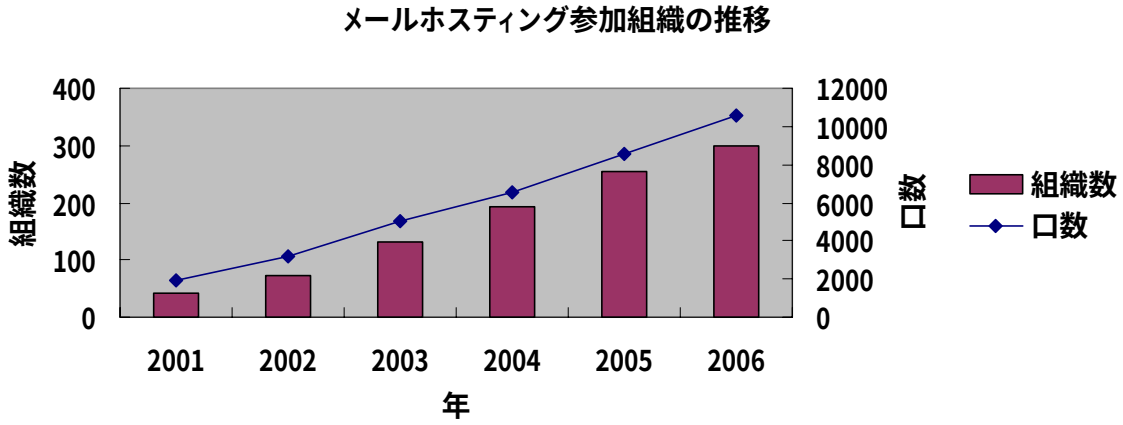
1　経過

2006 年6月27日　第24回情報メディア教育専門委員会において、利用負担金免除制度の適用継続を希望していた1組織に対し、制度の適用継続が認められた。

2006 年10月29日10:00～12:00　計画停電のためサービスを休止した。

2　参加組織

平成 18 年度末現在、本サービスを利用している組織は 300 組織であり、それらの申込み口数の合計は、10565 口である。平成 17 年度末に比べ、46 組織、2035 口増加している。



3　広報

本年度に行った主な広報は、以下のとおりである。

- xGate のパスワードの変更について　更新:2006 年 4 月 11 日
- メールサーバの緊急メンテナンス(終了)　更新:2006 年 5 月 15 日
- xGate 設定(Outbox)に関する注意　更新:2006 年 5 月 15 日
- 迷惑メールに関する一部メールの受信拒否について　更新:2006 年 6 月 29 日
- 学内組織向けメールサーバの利用負担金について　更新:2006 年 7 月 19 日
- メンテナンスによるメールサーバ停止について(終了)　更新:2006 年 8 月 18 日
- 8/24(Thu) システム障害について(終了)　更新:2006 年 8 月 24 日
- 9/13(水) UTnet 基幹工事による教育用計算機システムのサービス停止について(終了)　更新:2006 年 9 月 13 日
- Mac OS X 上での Eudora 6.0J の設定について　更新:2006 年 10 月 18 日
- 10/29(日)10:00～12:00 計画停電に伴うサービス停止について　更新:2006 年 10 月 24 日
- 11/15 メールサーバへの接続障害について　更新:2006 年 11 月 15 日
- 10/29(日) 事故停電　更新:2006 年 11 月 21 日
- Windows 版 Eudora(version6 および 7)で SSL を使用するための注意事項　更新:2007 年 1 月 11 日
- 1/22 メールサーバの障害について　更新:2007 年 1 月 24 日

・ 2/21 メンテナンスによるメールサーバ停止について(終了)　更新:2007 年 2 月 21 日
・ 2/13(火) 一部メールサーバの障害について　更新:2007 年 3 月 28 日

WEB PARK サービス運用報告（2006 年 4 月ー2007 年 3 月）

1　経過

2006/06/27（火） 第 24 回情報メディア教育専門委員会において、利用負担金免除制度の適用継続を希望していた 5 組織に対し、制度の適用継続が認められた。
2006/07　サービス利用者ガイド（web 版）を大幅に整備した。
2006/10/29（日）10:00～12:00、17:00～18:50　電源設備の定期点検のためサービスを休止した。

2　利用組織数

2006 年度に利用を開始した組織数は 67、利用を中止した組織数は 13 であった。
2007 年 3 月末現在の総利用組織数は 434 である。

3　WEB PARK サービスを利用している学内組織一覧（2007 年 3 月末現在）

（各カテゴリーにおいては申し込み順、各組織の名称は 2007 年 3 月末現在）

センター、委員会等 : 76 組織

アジア生物資源環境研究センター（6 組織）
- アジア生物資源環境研究センター　(anesc)
- 地域資源評価研究分野　(horiken)
- 生物環境評価大部門 土地環境評価研究分野 福代研究室　(dinos)
- 共生機能開発研究分野 樹木生理学・熱帯造林学研究室　(tree)
- 耐性機能開発研究分野 白子研究室　(anesctmv)
- 土地環境評価研究分野 環境材料設計学研究室　(smd)

遺伝子実験施設　(mgrl)
総合文化研究科／教養学部（14 組織）
- AIKOM委員会　(aikom)
- 進学情報センター　(agc)
- 学生相談所　(soudanjo-komaba)
- 情報教育棟　(i-edu)
- 駒場図書館　(lib-c)
- ドイツ・ヨーロッパ研究センター　(desk)
- 文化施設運営委員会オルガン委員会　(organ)
- 「人間の安全保障プログラム」運営委員会　(human_security)
- 教養教育開発室　(komed)
- 広域科学専攻 広報委員会　(kouiki)
- 社会連携委員会　(friday-lecture)
- アメリカ太平洋地域研究センター　(cpas)
- 美術博物館　(museum)
- 生命科学構造化センター　(csls)

先端科学技術研究センター（13 組織）
- フォトニクス材料分野(宮野研究室)　(myn)
- 知的財産権大部門　(ip)

- 知的財産マネジメント研究会 (smips)
- 先端まちづくり研究ラボ (planning)
- 藤井研究室 (findata)
- 御厨貴研究室 (mikuriya)
- バリアフリー分野 (bfp)
- 馬場研究室 (zzz)
- 研究戦略・社会システム大部門 資源・エネルギー・環境政策 (wlab)
- 化学認識機能材料分野 (crm)
- 菅研究室 (cbl)
- 渡辺研究室 (fennelweb)
- 藤井研究室 (finlab)

海洋研究所（17 組織）

- 海洋研究所 (ori)
- 海洋生命科学部門 行動生態計測 (fishecol)
- 新プログラム「海洋生命系のダイナミクス」 (dobis)
- 海洋生態系動態部門 (ori-ecosystem)
- 海洋物理学部門 (ori-dpo)
- 先端海洋システム研究センター 海洋システム計測分野 (ori-cer)
- 海洋化学部門 (ori-co)
- 観測研究企画室 (ori-tso)
- 海洋生物資源部門 (ori-lmr)
- 海洋生命科学部門 (ori-bio)
- 国際沿岸海洋研究センター (icrc)
- 先端海洋システム研究センター (ori-camr)
- 海洋生命科学部門 生理学分野 (physiol)
- 海洋研究所図書室 (ori-library)
- 海洋科学国際共同研究センター 企画情報分野 (ori-cicplan)
- 先端海洋システム研究センター 海洋システム解析分野 (ori-msa)
- 海洋研究連携分野「生物圏環境学」 (ori-mbe)

低温センター (crc)

情報基盤センター（4 組織）

- 情報メディア教育部門 (itc-media)
- キャンパスネットワーキング研究部門 若原研究室 (wakahara)
- 情報メディア教育部門 電子教材係 (elearn)
- 図書館電子化部門 (dl-itc)

東京大学バリアフリー支援室 (spds)

生物生産工学研究センター（2 組織）

- 生物生産工学研究センター (biotec-res-ctr)
- 植物機能工学研究室 (ppk)

情報倫理審査専門委員会 (cie)

産学連携本部 (ducr)

埋蔵文化財調査室 (aru)

保健センター (health)

UT-CERT (ut-cert)

医療ナノテクノロジー人材養成ユニット (NBEP)

留学生センター (ic)

社会科学研究所（1 組織）

- 図書室 (iss-lib)

国際・産学共同研究センター（1 組織）
- 香川研究室 (cm)

大学総合教育研究センター (jun)
ナノバイオインテグレーション研究拠点 (CNBI)
生命科学教育支援ネットワーク (lse)
総括プロジェクト機構 ジェロントロジー寄付研究部門 (gero)
北京代表所 (beijing-office)
柏地区事務部 渉外・広報室 (utkk)
サステイナビリティ学連携研究機構 (ir3s)
情報の知識化・価値化技術協議会 (chishiki)

学部／研究科, 学科／専攻, 研究室等：352 組織

法学系研究科／法学部（1 組織）
- 久保研究室 (kubo)

医学系研究科／医学部（25 組織）
- 機能生物学専攻 統合生理学研究室 (iphysiol)
- 病因病理学専攻 免疫学研究室 (immunol)
- 健康科学・看護学専攻 成人看護学／ターミナルケア看護学分野 (adng)
- 国際保健学専攻 (sih)
- 健康科学・看護学専攻 生物統計学／疫学・予防保健学 (epistat)
- 健康科学・看護学専攻 地域看護学分野 (chn)
- 健康科学・看護学科 (hn)
- 国際地域保健学教室 (ich)
- 病因病理学専攻 微生物学講座 (microbiology)
- 健康科学・看護学専攻 健康社会学分野 (hlthsoc)
- 健康科学・看護学専攻 健康学習・教育学分野 (sg)
- 健康科学・看護学専攻 母性看護学・助産学分野 (midwifery)
- COE 機能生物学・細胞分子生物学 (coem)
- 公衆衛生学教室 (public health)
- 健康科学・看護学専攻 看護体系・機能学分野／看護管理学分野（基礎看護学教室） (nurs-adm)
- 国際保健学専攻 人類生態学分野 (humeco)
- 健康科学・看護学専攻 老年看護学／創傷看護学分野 (rounenkango)
- 機能生物学専攻 細胞分子生理学分野 (cmphysiology)
- 健康科学・看護学専攻 家族看護学分野 (fn)
- 分子予防医学教室 (macrophage)
- 分子神経生物学 (mnp)
- 国際保健学専攻 人類遺伝学教室 (humgenet)
- 疾患生命工学センター 健康・環境医工学部門 (env-health)
- 神経生化学 (neurochemistry)
- 特定領域研究「感染現象のマトリックス」事務局 (matrix)

工学系研究科／工学部（110 組織）
- 航空宇宙工学専攻 (aero)
- 航空宇宙工学専攻 塩谷研究室 (shioya_lab)
- 航空宇宙工学専攻 森下研究室 (tmorisi)

- 航空宇宙工学専攻 ジェットエンジン研究室 (jetlab)
- 超電導工学専攻 鹿野田研究室 (kanoda_lab)
- マテリアル工学系専攻 (material)
- 材料学専攻 先端デバイス工学研究室 (adam)
- 物理工学専攻 教務室 (ap_office)
- 物理工学専攻 藤原研究室 (fujiwaralab)
- 物理工学専攻 古澤研究室 (furusawa)
- 物理工学専攻 五神研究室 (gono)
- 物理工学専攻 十倉研究室 (cmr)
- 物理工学専攻 初貝研究室 (ap_hatsugai)
- 建築学専攻 (arch)
- 機械系専攻 (microsystemlab)
- 地球システム工学専攻 (geosys)
- 地球システム工学専攻 物探・油層研究室 (gpre)
- 計数工学科 (keisu)
- 技術経営戦略学専攻 (knowledge)
- 工学・情報理工学図書館 (cllib)
- 附属総合研究機構 (sogo)
- 電気工学科 山地・藤井研究室 (yamaji)
- 電気工学専攻 横山研究室 (syl)
- マテリアル工学系専攻 材料システム研究室 (mse)
- 電子工学専攻 田中雅明研究室 (cryst)
- 都市工学科 都市デザイン研究室 (ud)
- マテリアル工学専攻 プラズマ工学研究室 (plasma)
- 社会基盤工学専攻 建設マネジメント・開発システム研究室 (cmis)
- マテリアル工学専攻 COE 推進委員会 (coe-material)
- 社会基盤学専攻 海岸・沿岸環境研究室 (coastal)
- 社会基盤工学専攻 交通・都市基盤計画研究室 (trip)
- 物理工学専攻 為ヶ井研究室 (tamegai_lab)
- 社会基盤学専攻 河川／流域環境研究室 (reel)
- 航空宇宙工学専攻 青木研究室 (aokilab)
- 社会基盤工学専攻 応用力学／岩盤力学研究室 (ohriki)
- 量子相エレクトロニクス研究センター 染谷研究室 (someya)
- 地球システム工学専攻 加藤研究室 (katolab)
- 量子相エレクトロニクス研究センター (qpec)
- 環境海洋工学専攻 (naoe)
- マテリアル工学科 山口研究室 (ssi-lab)
- 工学教育推進機構 国際化推進室(GWP) (gwp)
- 建築学専攻 環境系 (envarch)
- 環境海洋工学専攻 海洋機器研究室 (gisolab)
- 船舶海洋工学専攻 構造システム研究室 (nasl)
- 21 世紀 COE 強相関物理工学 (coe-ap)
- 物理工学専攻 (ap)
- マテリアル工学専攻 (metall)
- 航空宇宙工学専攻 河野－津江研究室 (kono-tsue_lab)
- 環境海洋工学専攻 実現化工学講座 湯原研究室 (yuhalab)
- マテリアル工学専攻 マイクロフォトニクス研究室 (microphotonics)

- 物理工学科 香取研究室 (amo)
- 先端学際工学専攻 元橋研究室 (mo)
- 環境海洋工学専攻 生産システム工学研究室 (msel)
- 物理工学科 樽茶研究室 (tarucha-lab)
- 物理工学専攻 永長研 CREST (nagaosa_crest)
- 環境海洋工学専攻 海洋空間計画研究室 (ospl)
- 化学システム工学科 山下研究室 (tcl)
- 精密機械工学専攻 (pe)
- 技術経営戦略学専攻 (biz-model)
- 精密機械工学専攻 小林研究室 (ein)
- 機械工学専攻 牛田・古川研究室 (tissue)
- 技術部情報センター (ttc)
- 社会基盤学専攻 コンクリート研究室 (concrete)
- 化学生命工学専攻 (chembio)
- 総合研究機構 俯瞰工学 (metatechnica)
- 総合研究機構 強力 X 線実験室 (hpx)
- 精密機械工学専攻 新井・横井・太田研究室 (arai-lab)
- 社会基盤学専攻 上田研究室 (infra_economics)
- 環境海洋工学専攻 輸送・環境システム計画研究室 (suiso)
- 電子工学専攻 大津研究室 (uuu)
- 航空宇宙工学専攻 西成研究室 (tknishi)
- 機械工学専攻 ホロニックエネルギーシステム学講座 (hes)
- 総合研究機構 井上研究室 (ultracold)
- 総合研究機構 触媒反応工学領域 (nishiba)
- 地球システム工学専攻 藤田研究室 (tfujita-lab)
- 建築学専攻 岸田研究室 (design)
- システム創成学科 (si)
- 都市工学専攻 国際都市計画・地域計画研究室 (onishiken)
- 建築学専攻 難波研究室 (namba)
- マテリアル工学専攻 阿部研究室 (nanostructures)
- 社会基盤学専攻 橋梁研究室 (bridge)
- 応用化学専攻 平尾研究室 (qcl)
- 総合研究機構 石原研究室 (ishihara)
- 技術経営戦略学専攻 (TMI)
- システム創成学科 知能社会システムコース (psi)
- バイオエンジニアリング専攻 (bioeng)
- 応用化学専攻 橋本研究室 (hashimoto_lab)
- マテリアル工学科 片岡研究室 (kklab)
- マテリアル工学専攻 吉田亮研究室 (yoshida-lab)
- 電子工学専攻 保立・山下・何研究室 (sagnac)
- 物理工学科 今田研究室 (imada)
- 都市工学専攻 関澤研究室 (fse)
- システム量子工学専攻 古田研究室 (cse)
- 物理工学専攻 求研究室 (motomelab)
- 航空宇宙工学専攻 岩崎研究室 (sal)
- 社会基盤学専攻 英語コミュニケーション教室 (english)
- 建築学専攻 野口研究室 (bmat)

- 都市工学科 都市交通研究室 (ut)
- 環境海洋工学専攻 マリタイム・イノベーション（常石造船）寄附講座 (mil_tsuneishi)
- 建築学専攻 伊藤研究室 (itolaboratory)
- 環境海洋工学専攻 設計研究室 (nakl)
- 1st Asian Biomaterials Congress (abc1st)
- 第 34 回 日本臨床バイオメカニクス学会 (biomech34)
- 地球システム工学専攻 登坂研究室 (tosaka-lab)
- 機械工学専攻 加藤研究室 (sstl)
- 化学システム工学専攻 土橋研究室 (dobashi)
- マテリアル工学専攻 複層鋼板プロジェクト (lism)
- 化学システム工学専攻 堂免・久保田研究室 (domen-lab)
- 化学システム工学専攻 大久保研究室 (zeolite)
- 化学システム工学専攻 (chemsys)

人文社会系研究科／文学部（1 組織）

- 言語動態学研究室 (dootai)

理学系研究科／理学部（3 組織）

- 技術部 (s-tech)
- 物理学専攻 島野研究室 (shimanolab)
- 生物情報科学学部教育特別プログラム (upbsb)

農学生命科学研究科／農学部（85 組織）

- 生産・環境生物学専攻 応用昆虫学研究室 (applent)
- 応用生命化学専攻 分析化学研究室 (bunseki)
- 応用生命化学専攻 生物機能開発化学研究室 (biofunc)
- 応用生命化学専攻 生物有機化学研究室 (seiyu)
- 応用生命化学専攻 土壌圏科学研究室 (soil-cosmology)
- 応用生命化学専攻 植物分子生理学研究室 (pmp)
- 応用生命化学専攻 植物栄養・肥料学研究室 (syokuei)
- 応用生命化学専攻 放射線植物生理学研究室 (radio-plantphys)
- 応用生命工学専攻 分子生命工学研究室 (molbiotech)
- 生物・環境工学専攻 生物機械工学研究室 (bme)
- 生物材料科学専攻 生物材料物理学研究室 (bmp)
- 森林科学専攻 (forest)
- 森林科学専攻 森林利用学研究室 (foresteng)
- 生圏システム学専攻 (es)
- 農業・資源経済学専攻 (arec)
- 水圏生物科学専攻 (fs)
- 附属牧場(高等動物教育研究センター) (bokujo)
- 放射性同位元素施設 (kanri)
- 農学生命科学図書館 (aglib)
- 獣医学専攻 (vetmed)
- 生物・環境工学専攻 農地環境工学研究室 (land)
- 水圏生物科学専攻 水産資源研究室 (shigen)
- 水圏生物科学専攻 水圏生物工学研究室 (suikou)
- 農業・資源経済学専攻 農業史研究室 (aghist)
- 森林科学専攻 森林風致計画学研究室 (fuuchi)
- 獣医学専攻 動物育種繁殖学研究室 (ikushu)
- 獣医学専攻 獣医解剖学研究室 (kaibo)

- 獣医学専攻 獣医生理学研究室 (seiri)
- 獣医学専攻 獣医薬理学研究室 (yakuri)
- 獣医学専攻 獣医微生物学研究室 (microbio)
- 獣医学専攻 獣医公衆衛生学研究室 (eisei)
- 獣医学専攻 比較病態生理学研究室 (hibyou)
- 獣医学専攻 獣医病理学研究室 (byouri)
- 獣医学専攻 獣医内科学研究室 (naika)
- 獣医学専攻 獣医外科学研究室 (geka)
- 獣医学専攻 実験動物学研究室 (jitsudo)
- 獣医学専攻 獣医臨床病理学研究室 (vcpb)
- 応用動物科学専攻 応用免疫学研究室 (immune)
- 応用動物科学専攻 応用遺伝学研究室 (iden)
- 応用動物科学専攻 動物細胞制御学研究室 (saibo)
- 応用動物科学専攻 細胞生化学研究室 (seika)
- 応用動物科学専攻 獣医動物行動学研究室 (koudou)
- 応用動物科学専攻 高度医療科学研究室 (koui)
- 生物･環境工学専攻 生物環境情報工学研究室 (beie)
- 森林科学専攻 森林植物学研究室 (fb)
- 応用生命化学専攻 有機化学研究室 (org-chem)
- 応用生命化学専攻 食品生化学研究室 (food-biochem)
- 生物･環境工学専攻 環境地水学研究室 (chisui)
- 応用生命工学専攻 細胞遺伝学研究室 (cell-gene)
- 水圏生物科学専攻 水圏生物環境学研究室 (fol)
- 応用生命工学専攻 醗酵学研究室 (hakko)
- バイオトロン研究室 (biotron)
- 農業･資源経済学専攻 経済学研究室 (agriecon)
- 森林科学専攻 造林学研究室 (silviculture)
- 生物･環境工学専攻 生物環境情報工学研究室 (joho)
- 応用生命工学専攻 微生物学研究室 (Lab_Microbiology)
- 食シグナル･生体統御系間相互作用（明治乳業）寄付講座 (shoku-signal)
- 応用生命化学専攻 食糧化学研究室 (foodchem)
- 応用生命化学専攻 生物化学 (inositol)
- 生物･環境工学専攻 (bee)
- 21 世紀 COE プログラム「生物多様性･生態系再生」研究拠点 (coe-ber)
- 生圏システム学専攻 保全生態学研究室 (coneco)
- 生産･環境生物学専攻 (aeb)
- 生物･環境工学専攻 生物環境情報工学研究室 (agrinfo)
- 生産･環境生物学専攻 植物分子遺伝学研究室 (pmg)
- 付属家畜病院 (vmc)
- 生物･環境工学専攻 生物プロセス工学研究室 (bipren)
- 附属水産実験所 (suijitsu)
- 生産･環境生物学専攻 園芸学研究室 (engei)
- 生産･環境生物学専攻 植物分子遺伝学研究室 (ICPMB2007)
- 電脳土壌センター (soilDB)
- 水圏生物科学専攻 水圏生物工学研究室 (motil)
- 農学国際専攻 (ga)
- 農業･資源経済学専攻 食料･資源経済学研究室 (frec)

- 産学官民連携型農学生命科学インキュベータ機構 (agri-cocoon)
- 農学国際専攻 国際情報農学研究室 (iai)
- 農学国際専攻 新機能植物開発学研究室 (pbt)
- 森林科学専攻 森林経理学研究室 (forestmanag)
- 農学国際専攻 国際森林環境学研究室 (gfes)
- 森林科学専攻 森林動物学研究室 (zoology)
- 応用生命化学専攻 栄養化学研究室 (ayo)
- 農学国際専攻 国際水産開発学研究室 (lgfs)
- バイオミネラルゼミ (biomineral)
- 食の安全研究センター (frc)
- 水圏生物科学専攻 魚病学研究室 (fishparasite)

経済学研究科／経済学部（1 組織）

- 都市経済学 金本良嗣研究室 (urban-eco)

総合文化研究科／教養学部（49 組織）

- 広域科学専攻 広域システム科学系 嶋田研究室 (shimada-lab)
- 教務課 (kyomu)
- 広域科学専攻 生命環境科学系 丹野研究室 (tanno)
- 国際社会科学専攻 (kiss)
- 国際社会科学専攻 相関社会科学講座 (kiss-sr)
- 国際社会科学専攻 国際関係論講座 (kiss-ir)
- 国際社会科学専攻 山影進研究室 (yamakage-ken)
- 国際社会科学専攻 山本泰研究室 (sociology-komaba)
- 化学部会 (chemistry)
- 広域科学専攻 生命環境科学系 安田研究室 (bio)
- 地域文化研究学科 ドイツ分科 (doitsuka)
- 駒場 21 世紀 COE プロジェクト「融合科学創成ステーション」 (rcis)
- 総合文化研究科・教養学部 (c)
- 教養学部等事務部経理課 (keirika)
- 地域文化研究専攻 (ask)
- 地域文化研究専攻 イギリス科 (british section)
- 超域文化科学専攻 表象文化論 (repre)
- 地域文化研究専攻 若林研究室 (wakasemi)
- 地域文化研究専攻「ジェノサイド研究の展開」 (cgs)
- 広域科学専攻 生命環境科学系 (bio-komaba)
- 国際社会科学専攻 廣松研究室 (stat-komaba)
- 教養学部等事務部総務課 広報・情報企画係 (safe-office)
- 英語I 運営班 (eigoichi)
- 超域文化科学専攻 比較文学比較文化コース (hikaku)
- 広域科学専攻 相関基礎科学系 吉岡研究室 (yoshioka-lab)
- 広域科学専攻 相関基礎科学系 (dbs)
- 国際社会科学専攻 内山研究室 (politics-komaba)
- 文系三学科 地域文化研究科 ラテンアメリカ分科 (latinamerica)
- 広域科学専攻 生命環境科学系 陶山研究室 (suyamalab)
- 広域科学専攻 広域システム系 人文地理学 (humgeo)
- 広域科学専攻 佐々研究室 (sasa)
- 広域システム科学系 (system)
- 基礎科学科 (kisokagakuka)

- 広域科学専攻 広域システム科学系 増田研究室 (masuda_lab)
- 広域科学専攻 染田研究室 (somedalab)
- 地域文化研究学科 フランス分科 (sfr)
- 文学･芸術の社会的統合機能の研究 (lac)
- 広域科学専攻 相関基礎科学系 科学史･科学哲学研究室 (hps)
- 地域文化研究専攻 (古田元夫ゼミ) (furuta-semi)
- 広域科学専攻 生命環境科学系 栗栖研究室 (xtalstructure)
- 広域科学専攻 和田研究室 (lipid)
- 広域科学専攻 相関基礎科学系 加藤雄介研究室 (kato-yusuke-lab)
- 科学技術インタープリター養成プログラム (STITP)
- 地域文化研究専攻 カンパニア地方の都市とヴィッラ集落をめぐる社会史的研究 (campania)
- 超域文化科学専攻 文化人類学コース (bunjin)
- 超域文化科学専攻／超域文化科学科 (choiki)
- フランス語部会 (langue_fr)
- 教養教育開発機構寄付研究部門「教養教育への囲碁の活用研究部門」 (igo)
- 情報システム研究グループ (graco)

教育学研究科／教育学部（2 組織）

- 教育学研究科･教育学部図書室 (ikuto)
- 附属中等教育学校 (hs)

新領域創成科学研究科（59 組織）

- 先端生命科学専攻 分子認識化学分野 (molecular-recognition)
- 国際協力学専攻 佐藤研究室 (satoj)
- 国際協力学専攻 (cois)
- 国際協力学専攻 湊研究室 (minato)
- 国際協力学専攻 農業環境学研究室 (aee-labo)
- 環境学研究系 自然環境学専攻 (nenv)
- 環境学研究系 自然環境学専攻 生物圏情報学分野 (bis)
- 自然環境学専攻 自然環境形成学分野 (nef)
- 環境学研究系 社会文化環境学専攻 沿岸域環境研究室 (coastal-env)
- 物質系専攻 木村研究室 (phys)
- 環境学専攻 自然環境評価学研究室 (nee)
- 自然環境学専攻 自然環境構造学分野 (nes)
- 環境学専攻 久田研究室 (sml)
- 複雑理工学専攻 齊木研究室 (yukimuki)
- 人間環境学専攻 鳥居研究室 (torii-lab)
- 環境学専攻 環境健康システム学分野研究室 (envhlth)
- 環境学専攻 自然環境学コース 生物圏機能学分野 (lbf)
- 複雑理工学専攻 佐々木研究室 (sas)
- 国際協力学専攻 柳田研究室 (yanagita)
- 環境学専攻 社会文化環境学大講座 (sbk)
- 環境学専攻 徳永研究室 (tokunaga)
- 複雑理工学専攻 村重研究室 (murashige-lab)
- 複雑理工学専攻 藤森研究室 (atsushi)
- 環境学専攻 環境システムコース 島田研究室 (shimada-labo)
- 人間環境学専攻 人間エネルギー環境学大講座 (hee)
- 先端生命科学専攻 河村正二研究室 (kawamura-lab)

- 物質系専攻 川合研究室 (surfchem)
- 先端生命科学専攻 動物生殖システム分野 (K-medaka)
- 国際協力学専攻 佐藤研究室 (rdm)
- 物質系専攻 和田研究室 (wadalab)
- 国際協力学専攻 中山幹康研究室 (nakayama)
- 国際協力学専攻 中山研究室「水のグローバルガバナンス」プロジェクト (ggwater)
- 複雑理工学専攻 鳥海研究室 (gaea)
- 環境システム学専攻 松橋・吉田研究室 (globalenv)
- 先端生命科学専攻 久恒辰博研究室 (hisatsune-lab)
- 環境学専攻 環境モデリング統合研究室 (lemons)
- 先端生命科学専攻 資源生物制御学分野 (seigyo)
- 環境学専攻 保坂研究室 (ems)
- 環境学系 岡本研究室 (vis)
- 環境学専攻 環境システムコース (envsys)
- 先端生命科学専攻 医薬デザイン工学分野 (iyaku)
- 人間環境学大講座 メディア環境学研究室 (mesh)
- 人間環境学専攻 (h-k)
- 先端エネルギー工学専攻 藤井研究室 (fujiilabkashiwa)
- 環境学専攻 国際環境協力コース 國島研究室 (kunishimalab)
- 先端生命科学専攻 (ib)
- 先端生命科学専攻 人類進化システム分野 (jinrui)
- 先端エネルギー工学専攻 鈴木研究室 (aesuzuki)
- 社会文化環境学専攻 環境音響学研究室 (env-acoust)
- 社会文化環境学専攻 味埜研究室 (minolab)
- 環境システム学専攻 地球環境工学研究室 (geelhome)
- 環境システム学専攻 大島研 (oshima-lab)
- 自然環境学専攻 自然環境構造学分野 春山研究室 (nesh)
- 先端生命科学専攻 生命応答システム分野 (outou)
- 人間環境学専攻 鳥居研究室 (droptech)
- 自然環境学専攻 自然環境変動学分野 (changes)
- 生涯スポーツ健康科学研究センター (hss)
- メディカルゲノム専攻 生命分子解析学分野 (phys-biochem)
- 基盤情報学専攻 高木研究室 (takagilab)

情報学環・学際情報学府（1 組織）

- 佐倉研究室 (sakuralab)

情報理工学系研究科（14 組織）

- システム情報学専攻 第七研究室 (mels)
- 知能機械情報学専攻 下山・松本・星野研究室 (leopard)
- 21 世紀 COE「情報科学技術戦略コア」実世界情報システムプロジェクト (coe-rwisp)
- 知能機械情報学専攻 (ist)
- システム情報学専攻 第5研究室 (cyb)
- 数理情報学専攻 武市研究室 (ipl)
- 数理情報学専攻 (mist)
- 国際交流室 (oir)
- システム情報学専攻 (ipc)
- 21 世紀 COE「情報科学技術戦略コア」 (coe-inside)
- 情報理工学系研究科 (ist-inside)

- 知能機械情報学専攻 CREST 移動体プロジェクト (crestidotai)
- 創造情報学専攻 石塚研究室 (ishizuka-lab)
- 知能機械情報学専攻 (irt)

公共政策大学院 (graspp)

留学生会、同窓会、クラブ等：6 組織

東京大学中国留学生学友会 (acsut)
東京大学台湾留学生会 (todai-tsa)
ThaiTodai (thaitodai)
PAKISTAN STUDENTS CULTURAL SOCIETY (pakistan)
Tokyo University Islamic Cultural Society (tuics)
農学生命科学研究科 生物・環境工学専攻 同窓会 (shikou)

関連 URL：http://park.itc.u-tokyo.ac.jp/

遠隔講義支援サービス運用報告（2006.4 – 2007.3）

1　IP 遠隔会議システム

1.1　会議等利用状況（学外接続利用のみ抜粋）

１）本郷遠隔講義室

2006/04/06(木) 12:30-18:00
　情報基盤センター　中山助教授　会議　E-SIDE と接続
2006/04/11(火) 10:00-12:00
　新領域創成科学研究科　山崎助手　学外との会議　名古屋大学、大阪大学、京都大学と接続
2006/05/11(木) 13:00-15:00
　情報基盤センター　中山助教授　LIVE E!打合せ　奈良先端科学技術大学院大学、大阪大学と接続
2006/05/17(水) 10:00-18:00
　情報基盤センター　中山助教授　WIDE board 会議　慶應義塾大学と接続
2006/07/12(水) 13:00-16:00
　情報基盤センター　中山助教授　LIVE E!打合せ　奈良先端科学技術大学院大学、アジア工科大学院、広島大学と接続
2006/07/15(土) 11:00-17:00
　情報基盤センター　安東助手　会議　佐賀大学、広島市立大学と接続
2006/11/01(水) 10:00-12:00
　情報理工学系研究科　山崎講師　学外との会議　名古屋大学、大阪大学、京都大学と接続
2006/11/08(水) 10:00-12:00
　情報基盤センター　加藤助教授　コンピュータネットワーク（理学部 情報科学）　WIDE San Francisco Studio と接続
2006/11/16(木) 11:00-13:30
　情報基盤センター　中山助教授　DUMBO 国際ワークショップ打合せ　奈良先端科学技術大学院大学、アジア工科大学院と接続
2006/11/21(火) 10:00-11:30
　情報理工学系研究科　山崎講師　学外との会議　名古屋大学、大阪大学、京都大学と接続
2006/12/20(水) 13:00-15:30
　情報基盤センター　関谷助手　協同研究先との打合せ　慶應義塾大学と接続
2007/01/05(金) 10:00-12:00
　情報理工学系研究科　山崎講師　学外との会議　京都大学と接続
2007/02/21(水) 13:00-19:00
　情報基盤センター　関谷助手　協同研究打合せ　慶應義塾大学と接続
2007/03/15(木) 10:30-12:00
　情報理工学系研究科　山崎講師　学外との会議　大阪大学、京都大学と接続

２）駒場遠隔講義室

2006/04/26(水) 10:00-12:00
　総合文化研究科　植田助教授　ビジネスミーティング　京都大学、国立情報学研究所と接続
2006/04/28(金) 13:00-14:30
　総合文化研究科　佐藤教授　平和構築研究会 研究連絡会　名古屋大学と接続

2006/05/09(火) 18:00-22:00
　生産技術研究所　 西川技術官　 接続テスト　 ヨーク大学と接続
2006/05/17(水) 10:00-12:00
　総合文化研究科　 植田助教授　 ビジネスミーティング　 京都大学、国立情報学研究所と接続
2006/05/31(水) 10:00-12:30
　総合文化研究科　 植田助教授　 会議　 京都大学、国立情報学研究所と接続
2006/06/14(水) 10:30-12:30
　総合文化研究科　 佐藤教授　 公開セミナー　 名古屋大学と接続
2006/06/21(水) 10:10-12:30
　総合文化研究科　 植田助教授　 会議　 京都大学、国立情報学研究所と接続
2006/06/30(火) 17:30-19:30
　総合文化研究科　 佐藤教授　 公開研究会、研究連絡会　 名古屋大学と接続
2006/07/18(火) 10:30-12:30
　総合文化研究科　 植田助教授　 会議　 京都大学、国立情報学研究所と接続
2006/08/29(火) 17:00-22:00
　生産技術研究所　 西川技術官　 会議（UK-JP DTA ワークショップ） ヨーク大学と接続
2006/09/21(木) 10:30-12:00
　総合文化研究科　 植田助教授　 会議　 京都大学と接続
2006/10/17(火) 13:30-15:30
　総合文化研究科　 植田助教授　 会議　 京都大学と接続
2006/11/07(火) 10:00-12:00
　総合文化研究科　 植田助教授　 会議　 京都大学と接続
2006/11/17(金) 11:00-14:40
　総合文化研究科　 佐藤教授　 平和構築研究会　 名古屋大学と接続
2006/11/27(月) 18:00-21:00
　総合文化研究科　 金井助教授　 INSA de Lyon での Ph.D 学生の審査　 INSA de Lyon と接続
2007/01/16(火) 10:00-12:00
　総合文化研究科　 植田助教授　 会議　 京都大学と接続
2007/02/21(水) 13:30-17:00
　総合文化研究科　 植田助教授　 会議　 京都大学と接続

3）その他

2006/11/11(土) 10:00-13:00
　総務部　 第 5 回東京大学ホームカミングデイ　 安田講堂と駒場キャンパス数理科学研究科を接続

1.2　講義等利用状況

1）本郷遠隔講義室

2006/04/17-2007/01/29　 毎週月曜日　 13:00-14:45
　新領域創成科学研究科　 先端エネルギー工学輪講　 柏キャンパスと接続
2006/04/17-2006/07/10　 毎週月曜日　 14:45-16:25
　新領域創成科学研究科　 プラズマ核融合学講義　 柏キャンパスと接続
2006/04/13-2006/07/20　 毎週木曜日　 10:00-12:00
　新領域創成科学研究科　 ネットワークコンピューティング講義　 柏キャンパスと接続
2006/04/07-2007/01/19　 毎週金曜日　 10:15-11:55
　新領域創成科学研究科　 平成 18 年度基盤情報学専攻輪講　 柏キャンパスと接続

2006/10/05-2007/01/25　毎週木曜日　16:20-18:00
　数理科学研究科　情報理論講演会　駒場キャンパスと接続

２）駒場遠隔講義室

2006/04/10-2006/07/03　毎週月曜日　17:05-18:45
　情報学環　学際理数情報学基礎 I　情報学環と接続
2007/10/04-2007/01/31　毎週水曜日　14:30-16:30
　情報学環　学際理数情報学基礎 VII　情報学環と接続

2　MCU サービス（試行）

（MCU：テレビ会議システム多地点接続装置 2005 年度より導入）

2.1　会議等利用状況

2006/04/10(月) 09:30-12:30
　情報基盤センター　中山助教授　9 地点同時接続

2006/04/13(木), 20(木), 27(木), 05/11(木), 18(木), 25(木), 06/08(木), 15(木), 22(木),
　07/13(木), 14(金), 27(木), 08/24(木), 09/05(火), 07(木), 14(木), 21(木), 10/12(木), 19(木), 26(木),
　11/09(木), 16(木), 22(水), 12/07(木), 14(木), 21(木),
　2007/01/11(木), 18(木), 25(木), 02/08(木) 12:00-15:00
　人工物工学研究センター　淺間教授　3 地点同時接続

2006/04/17(月) 17:30-21:00
　工学系研究科　山本助教授　2 地点同時接続

2006/05/30(火) 15:00-17:00
　物性研究所　矢田技術職員　3 地点同時接続

2006/06/06(火) 15:00-18:00
　情報基盤センター　中山助教授　6 地点同時接続

2006/06/16(金) 15:00-17:00
　情報基盤センター　中山助教授　3 地点同時接続

2006/07/13(木) 09:30-12:00
　情報基盤センター　中山助教授　3 地点同時接続

2006/11/13(月), 20(月), 27(月), 12 月 04(月), 11(月) 16:30-18:20
　新領域創成科学研究科　江尻助教授　3 地点同時接続

2006/11/22(水) 17:00-21:00
　工学系研究科　中須賀教授　5 地点同時接続

2006/12/08(金) 15:45-19:00
　人工物工学研究センター　大武助教授　3 地点同時接続

2006/12/08(金) 17:30-21:30
　新領域創成科学研究科　小野教授　6 地点同時接続

2006/12/11(月) 08:00-10:00
　新領域創成科学研究科　小野教授　6 地点同時接続

2006/02/15 08:30-12:00
　新領域創成科学研究科　小野教授　7 地点同時接続

2.2　講義等利用状況

2006/10/16-2007/01/29　毎週月曜日　15:20-18:20
　教養教育開発機構　2006 年度冬学期 学術俯瞰講義「生命の科学」　4 地点同時接続

3　インターネットライブ中継

2006/10/16-2007/01/29　毎週月曜日　16:20-17:50（学内のみ）
　大学総合教育研究センター　2006 年度冬学期 学術俯瞰講義「生命の科学」

2007/01/23(火) 11:00-12:00（学内のみ）
　総務部　アクション・プラン全学説明会

2007/02/03(土) 14:00-17:00
　理学系研究科　第 1 回高校生のための「サイエンスカフェ本郷」

2007/02/06(火) 14:00-15:30（学内のみ）
　情報基盤センター　第 55 回コンピュータ・ネットワーク利用セミナー
　「IIJ の迷惑メール対策 ~ここまでやる現場での対策~」

2007/03/23(金) 09:00-12:00
　総務部　平成 18 年度東京大学卒業式

関連 URL　http://elearn.itc.u-tokyo.ac.jp/

CFIVE 運用報告（2006 年 4 月ー2007 年 3 月）

1　経過

CFIVE はプログラムソースを公開している学習管理システムであり、2004 年 4 月より東京大学でのサービスを開始した。サービス開始以来 CFIVE を利用する講義数は順調に増加しており、2005 年度は 35 講義であったが、2006 年度は 63 講義で利用された。

以下に利用された講義名等のデータ、改良された機能、広報等を示す。

2　夏学期に CFIVE を利用した講義

夏学期（2006 年 4 月ー9 月）に CFIVE を利用した講義は以下の通りである。

項番	開講組織	講義名	担当教員（敬称略）
1	工学部	自然共生システムのデザイン学（金 2）	佐藤愼司
2	教養学部(前期課程)	図形科学実習（金 3,4）	加藤道夫、堤江美子
3	教養学部(前期課程)	図形科学実習(木 3,4)	安達裕之、深野暁雄
4	教養学部(前期課程)	図形科学実習(水 3,4)	鈴木賢次郎、近藤邦雄
5	教養学部(前期課程)	図形科学実習(火 3,4)	田中一郎、深野暁雄
6	教養学部(前期課程)	図形科学実習(月 3,4)	鈴木賢次郎、金井崇
7	教養学部(前期課程)	図形科学実習(共通)	鈴木賢次郎ほか
8	教養学部(前期課程)	学術俯瞰講義「社会の形成」(木 5)	山本泰、森田朗
9	教養学部(前期課程)	心に挑む-東大の心理学(木 4)	丹野義彦
10	教養学部(前期課程)	宇宙科学実習 II(火 4,5)	鈴木建
11	教養学部(前期課程)	宇宙科学実習 II(月 4,5)	鈴木建
12	教養学部(前期課程)	情報(金 4)	中村政隆、尾上能之
13	教養学部(前期課程)	情報(金 3)	中村政隆
14	教養学部(前期課程)	英語 2 列(2)R(金 2)	荒木純子
15	教養学部(前期課程)	中級英語(LS)(金 1)、	荒木純子
16	教養学部(前期課程)	プログラム構成論(木 5)	山口和紀
17	教養学部(前期課程)	図形科学(木 4)	山口泰
18	教養学部(前期課程)	情報(木 3)	辰己丈夫
19	教養学部(前期課程)	物質化学(無機系化合物)(木 3)	高橋正
20	教養学部(前期課程)	数学1演習(1) (木 2)	堤正義
21	教養学部(前期課程)	情報 (木 2)	田中哲朗
22	教養学部(前期課程)	英語2列(1)R(木 2)	アルヴィなほ子
23	教養学部(前期課程)	記号論(水 5)	石田英敬
24	教養学部(前期課程)	情報(水 4)	山口和紀
25	教養学部(前期課程)	情報(水 3)	中村政隆
26	教養学部(前期課程)	情報(水 2)	玉井哲雄
27	教養学部(前期課程)	基礎現代化学(水 1)	永田敬
28	教養学部(前期課程)	情報(火 4)	美馬秀樹
29	教養学部(前期課程)	情報(火 3)	玉井哲雄
30	教養学部(前期課程)	基礎現代化学(火 1)	永田敬
31	教養学部(前期課程)	政治経済学	丸山真人
32	教養学部(前期課程)	数学1(1) (月 1)	堤正義
33	教養学部(前期課程)	情報(月 1)	増原英彦

3　冬学期に CFIVE を利用した講義

冬学期（2006 年 10 月－2007 年 3 月）に CFIVE を利用した講義は以下の通りである。

項番	開講組織	講義名	担当教員（敬称略）
1	教養学部(後期課程)	情報システム科学実習 I	山口泰
2	教養学部(後期課程)	情報システム科学 V	山口泰
3	教養学部(後期課程)	システム基礎数学 3	川合慧、山口和紀
4	情報基盤センター	技術職員研修	尾上能之
5	経済学部	経営科学	安田雪
6	農学部	情報工学	中村典裕
7	工学部システム創成学科	WWW アプリケーション入門	大和裕幸、浅間 一
8	工学部	コンピュータ及び演習(火 1)	阿久津好明
9	教養学部(前期課程)	学術俯瞰講義「学問と人間」(木 5)	小林康夫
10	教養学部(前期課程)	全学自由研究ゼミナール(木 4)	アルヴィなほ子
11	教養学部(前期課程)	英語 2 列(PO)(金 1)	河合祥一郎
12	教養学部(前期課程)	学術俯瞰講義「生命の科学」(月 5)	浅島誠
13	教養学部(前期課程)	英語 2 列(PO)(金 2)	荒木純子
14	教養学部(前期課程)	英語 2 列(PW)(金 1)	荒木純子
15	教養学部(前期課程)	英語 2 列(PO)(水 4)	河合祥一郎
16	教養学部(前期課程)	英語演習 2(水 2)	河合祥一郎
17	教養学部(前期課程)	情報科学(金 4)	山口和紀
18	教養学部(前期課程)	図形科学 I(金 3)	鈴木賢次郎
19	教養学部(前期課程)	情報科学(金 3)	上田哲郎
20	教養学部(前期課程)	数学1演習(2)(木 2)	堤正義
21	教養学部(前期課程)	英語2列(P1)(木 2)	アルヴィなほ子
22	教養学部(前期課程)	応用倫理学(水 5)	廣野喜幸
23	教養学部(前期課程)	図形科学 I(水 4)	鈴木賢次郎
24	教養学部(前期課程)	情報科学(水 4)	田中哲朗
25	教養学部(前期課程)	英語 2 列(P1)(水 4)	アルヴィなほ子
26	教養学部(前期課程)	情報科学(水 3)	玉井哲雄
27	教養学部(前期課程)	情報科学(月 4)	立堀道昭
28	教養学部(前期課程)	数学1(2)(月 1)	堤正義
29	教養学部(前期課程)	スペイン語 1 列	網野徹哉

4　組織毎の利用状況

組織毎にまとめた 2006 年度の利用状況は、以下の通りである。

開講組織	講義数
教養学部(前期課程)	54
教養学部(後期課程)	3
工学部	3
農学部	1
経済学部	1
情報基盤センター	1
計	**63**

5　新たに追加された機能

2006 年度に新たに CFIVE に追加された機能は、以下の通りである。

- (管理機能)コース、ユーザ管理作業を, WWW ブラウザ経由で実行できる機能
- (課題機能)コメント編集画面の文字数制限の緩和
- (課題機能)レポート一覧のソート機能
- (課題機能)課題 NO. が変更しないようにする機能
- (課題機能)レポート提出、管理における WebDAV 機能
- (ユーザ機能)学生が複数のレポートを提出する際, 複数ある電子ファイルをフォルダに整理でき、移動、アップロードができる機能
- (ユーザ機能) コース内の学生個人へレポート(電子ファイル)を返却できる機能
- (ユーザ機能)講師が, 自分自身が担当するコースのユーザの追加や削除を可能とする機能
- (インストール支援機能) インストーラ
- (テスト機能)穴埋め問題の文字マッチングの改善
- (テスト機能)テスト公開に関するルールの改善
- (テスト機能)テストの一覧表示のソート方法の改善
- (テスト機能)テスト問題作成の文字数制限を大幅に超えるとシステムダウンする原因の改善
- (掲示板機能)掲示板の投稿状況の確認機能
- (掲示板機能)掲示板の編集機能とサブメニュー等のインターフェイスの改善
- (授業機能)出席一覧のダウンロード機能
- (全般)受講者が多いコースにおける問題の調査と改善
- (全般)テキストフィルタ機能
- (全般)エラーに応じたエラーメッセージの表示をするよう改善
- (全般)ダイレクト URL 経由での CFIVE の利用に関するバグの改善

6　広報

2006 年度に行った主な広報は、以下の通りである。

利用者への案内

- 平成 18 年度夏学期の CFIVE のご利用について(教職員向け)
- 平成 18 年度冬学期の CFIVE のご利用について(教職員向け)
- Utask-Web のアカウントでの CFIVE の利用について
- IC カードによる出欠管理システムについて(教職員向け)
- 学生自身による CFIVE のコースへのユーザ登録(学生向け)
- c5-support への問い合わせについて

お知らせ

- CFIVE サーバのメンテナンスのお知らせ(2007/03/14)　　更新 2007/3/14
- 2007/02/01 の 12:15~13:10 頃に発生した CFIVE にアクセスできなかった現象について　更新 2007/2/1
- CFIVE サーバのメンテナンスのお知らせ(2006/11/24)(終了)　更新 2006/11/24

- CFIVE 開始のお知らせ　更新 2006/11/17
- 課題機能における不具合について(2006/10/30)(終了)　更新 2006/10/31
- Utnet 基幹ネットワーク停止による CFIVE への影響について(2006/10/29)(終了)
  更新 2006/10/30
- CFIVE の「課題」機能をご利用の際の注意点　更新 2006/9/19
- PukiWiki 整形ルールについて(2006/05/24)　更新 2006/9/19
- UTnet メンテナンスによる CFIVE への影響について(2006/09/13)(終了)　更新 2006/9/13
- 2006/08/24 に発生の障害について(終了)　更新 2006/8/24
- CFIVE サポート休業について(2006/08/14,08/15)　更新 2006/8/9
- 2006/07/27 の 12:15~12:50 頃に発生した CFIVE にアクセスできなかった現象について
  更新 2006/7/27
- 「課題」機能にて登録するレポートファイルのファイル名(学生向け 2006/06/19)　更新 2006/6/19
- システム障害に関するお詫び　更新 2006/6/16
- テスト機能の不具合について(2006/06/15)　更新 2006/6/15
- テストの提出ができないときは　更新 2006/6/4
- サービス停止のお知らせ (終了)　更新 2006/6/4
- 課題機能における不具合について(20060509)(終了)　更新 2006/5/11
- 課題登録の課題名(教員, TA 向け)(終了)　更新 2006/5/11
- バージョンアップによるシステム停止のお知らせ (終了)　更新 2006/4/3

以上

図書館電子化部門

図書館電子化部門

概要

図書館情報係　本多　玄

図書館電子化業務部門は、図書館情報係、デジタル・ライブラリ係、学術情報リテラシー係の 3 係からなる。

図書館電子化業務部門は附属図書館と連携し、本学における学術情報リソースの提供とその利便性を高めることで本学の学術研究活動を支援する一方、本学の研究成果を社会に向けて発信する役割も担っている。

図書館情報係は主に全学の図書館室で使用される図書館システム（「附属図書館学術情報システム」）の運用・管理を行なっている。平成 18 年度は前年度末にシステムのリプレイスがあったことから、さらなる安定稼動と本学向けカスタマイズ部分のシステム適用について、メーカー側との連絡・調整にあたった。

デジタル・ライブラリ係は本学で生産される学術情報を電子化・発信することがその主な任務であるが、昨年度は本学の学術機関リポジトリ（「UT Repository」）の本格サービス開始により、広報活動や著作権処理、そしてさらなる機能高度化に取り組んだ。

学術情報リテラシー係は、本学に必要と思われる商業学術データベースの選定と契約を行なうとともに、学内で利用可能な各種データベースの利用に関する講習会やパンフレットの作成を通じて学内教職員・学生の学術情報リテラシーを高める活動を本年度も継続的に行なった。

デジタルコンテンツサービス

デジタル･ライブラリ係　　鴫　邦宏

1　運用報告

1.1　東京大学学術機関リポジトリの構築

東京大学学術機関機関リポジトリ～UT Repository～とは、東京大学で生産されたさまざまな研究成果を電子的な形態で集中的に蓄積･保存し、学内外に公開することを目的としたインターネット上の発信拠点である。平成 16 年度から附属図書館と連携し学術機関リポジトリの構築を行い、平成 18 年 4 月 1 日から「東京大学学術機関リポジトリ(UT Repojsitory)」としてサービスを開始した。
(http://repository.dl.itc.u-tokyo.ac.jp/)
平成 18 年度は、国立情報学研究所の最先端学術情報基盤(CSI)構築事業による外部資金を調達･活用し、附属図書館と連携協力して次の業務を行った。

1.1.1　広報活動

各部局教授会でのプレゼンテーション資料及び、学内の研究者に向けた説明資料と著作権ガイドを1つにしたパンフレット「知っていますか？UT Repository」等を作成。以下の各部局教授会にて附属図書館長が UT Repository のプレゼンテーションを行う際に資料を配布した。
パンフレット「知っていますか？UT Repository」は、別途、学内全図書館･室へ必要部数を配付した。

UT Repository の部局教授会でのプレゼンテーション

部局	月日	資料配布部数
文学部・人文社会系研究科	10/18	100
情報理工系研究科	11/2	30
先端科学技術研究センター	11/8	70
工学部・工学系研究科	11/9	75
理学部・理学系研究科	11/15	90
海洋研究所	12/20	100
生産技術研究所	12/20	100
宇宙線研究所	1/17	35
物性研究所	1/18	60
農学部・農学生命科学研究科	1/25	200

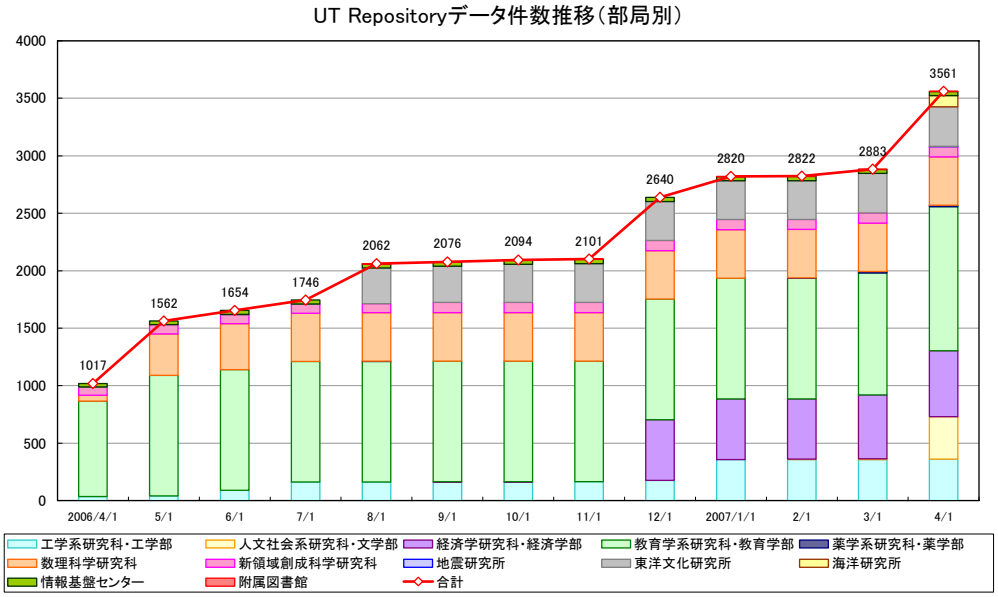
図書館広報誌「図書館の窓」(2006.10)で特集記事を、情報基盤センター広報誌「Digital Life Vol.8」に紹介記事を、学内広報誌「学内広報」No.1354（2007.3.14）の特集記事を掲載し、学内広報の拡大を図った。

また、教職員情報基盤センターの学内教職員セミナー「情報基盤センターのサービス紹介」において、UT Repository の紹介を行い、附属図書館職員研修プログラムにおいてリポジトリ構築についての発表を行った。

1. 1. 2 コンテンツ作成と著作権許諾処理

紀要、学位論文、学術雑誌論文を中心としてコンテンツの収集と作成を行い、3 月末には 3,561 件のデータを公開した。

機関リポジトリでは本文そのものを PDF 等で公開するため、単に電子化作業を行うだけでは無く、著作権の許諾処理を必須として作業を行っている。著者本人や共著者の許諾はもちろんであるが、学術雑誌掲載論文では出版者の許諾も必要であるため、機関リポジトリへの掲載許諾情報を調査し、必要に応じて学協会等へ個々に許諾の確認を取る作業を行った。また、紀要及び COE リサーチアシスタント報告書等のバックナンバーの電子化にあたり、著者への許諾を電子メールと郵送により実施した。



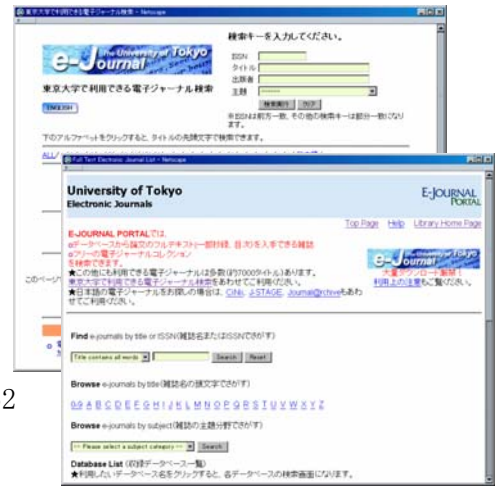
1.1.3 システム開発

前述のとおり、著作権許諾処理は著者本人、共著者、出版者の許諾が必要であり、学生の場合は指導教員の許諾が必要となる場合もあるため、これらの許諾処理を同一のデータベースで一元管理し、全ての許諾処理が完了しないコンテンツは公開しないように制御するデータベースとシステムの開発を行い、著作権許諾管理とリポジトリへの登録管理業務の効率化を図る。平成 18 年度は根幹のデータベース設計とデータベースの実装及び最小限の機能実装を行った。今後、データベース等の評価を行い、データベースの改善と機関リポジトリシステムと連携する機能の実装を行う予定である。

1.1.4　学内システム連携

今後、標準実績データベース等の学内関連システムとの連携を図り、研究業績等を One Writing で登録できると共に、容易に各種研究成果リストの出力が可能となるようなシステムの構築を推進する予定である。

1.2　電子ジャーナルのゲートウェイサービス



電子ジャーナルの提供を円滑に行うための電子ジャーナルリンク集データベースのメンテナンス及び利用統計採取のためのプロキシサーバの維持・管理と、本学が契約するデータベースを経由して利用できる電子ジャーナルとフリーの電子ジャーナル（約 27,000 タイトル）についての管理ソフトを導入した E-Journal Portal の学内へのサービスを行っている。
また、Elsevier 社の電子ジャーナルサービス ScienceDirect を学外から利用できるリモートアクセス用 ID を、教員・大学院生に対し発行する業務を行っており、平成 18 年度は 1,252 件発行した。
(http://ejournal.dl.itc.u-tokyo.ac.jp/)

1.3　インターネット学術情報インデックス（IRI）

インターネット学術情報インデックス（IRI）はインターネット上に存在するリソースの中から教育・研究上有用な情報源に限定して収集し、効率的な検索サービスを提供するサブジェクトゲートウェイであり、学内外サイトからのリンクも多い。
平成 18 年度は 177 件のリソース情報を追加した。
(http://resource.lib.u-tokyo.ac.jp/iri/url_search.cgi)

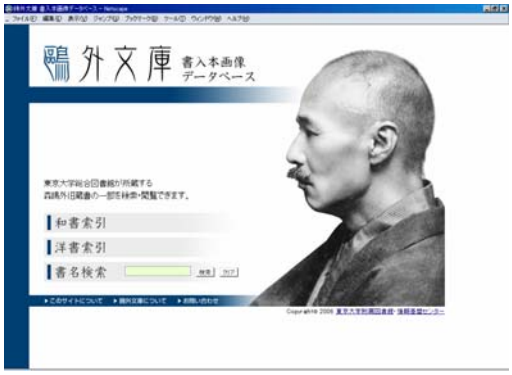
1.4　東京大学学位論文要旨データベースの構築

学位論文データベースは従来冊子体で刊行されていた「東京大学博士学位論文の内容の要旨と審査結果の要旨」を電子化して公開しているものである。
平成 18 年度は平成 16 年度論文要旨 1,265 件の電子化を行った。昨年度の提出電子ファイルの種別調査を踏まえ、提出電子ファイルを全面的に利用し, データベース作成の効率化と経費の削減を図った。
(http://gakui.dl.itc.u-tokyo.ac.jp/)

1.5　貴重書のデータベース化支援

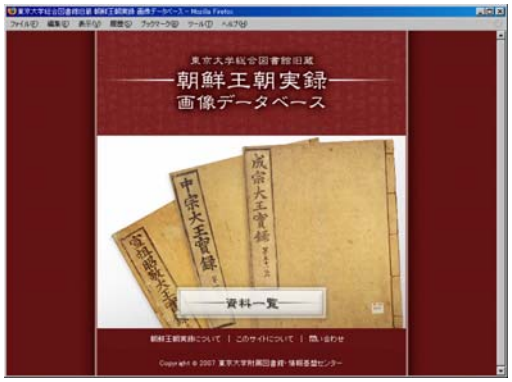
1.5.1　鴎外文庫書入本画像データベース

平成 17 年度に引き続き、附属図書館が平成 18 年度科学研究費補助金（研究成果公開促進費　課題番号: 188017）の交付を受けた鴎外文庫書入本画像データベースへ作成協力を行い、コンテンツの拡充を行った。
(http://rarebook.dl.itc.u-tokyo.ac.jp/ogai/)



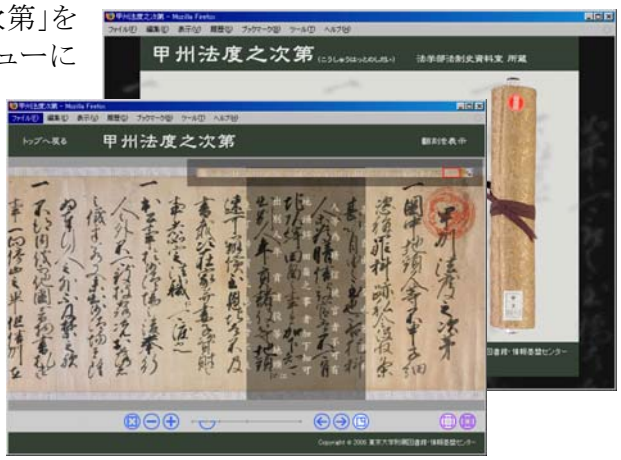
1.5.2　朝鮮王朝実録

総合図書館旧蔵の朝鮮王朝実録について、附属図書館と協力し、本文画像を公開するためのシステムの作成を行った。



1.5.3　甲州法度之次第

法学部法制史資料室所蔵本の「甲州法度之次第」を電子化し公開システムの作成を行った。ガラスビューにより翻刻を透過表示する機能を有している。公開は平成 19 年度予定である。



1.6　貴重書展示会の支援

附属図書館展示委員会に協力し、通常展示および特別展示への協力を行っている。平成 18 年度の特別展示会「知の職人たち―南葵文庫に見る江戸のモノづくり―」では、パンフレット及びポスターのデザイン作成を行い、電子展示版（ホームページ）を作成・公開した。
(http://www.lib.u-tokyo.ac.jp/tenjikai/tenjikai2006/index.html)

学術情報リテラシー支援

学術情報リテラシー係　　吉野　明美

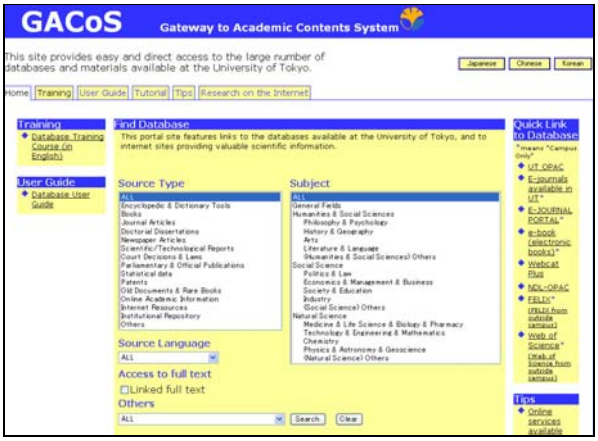
1　運用報告

1.1　情報探索ポータルサイト GACoS(Gateway to Academic Contents System)

デジタル学術情報にアクセスするためのポータルサイトとして GACoS：Gateway to Academic Contents System を構築、運用しているが、平成 17 年度に CGI によるデータベース検索機能を加え、デザインを一新した日本語版 GACoS に続き、英語版、中国語版、韓国語版の GACoS もリニューアルして公開した。

また、日本語版についても、検索条件に「その他の条件」として「学外のみ/学外 OK/学外 OK（手続き必要）など」を追加するとともに、東京大学 OPAC など基本的なデータベースは検索せずにアクセスできるよう「定番データベース」としてクイックリンクを設けるなど、機能の改善を行った。

日本語版 GACoS のトップページ
(http://www.dl.itc.u-tokyo.ac.jp/gacos/)



英語版 GACoS のトップページ
(http://www.dl.itc.u-tokyo.ac.jp/gacos/e/index-e.html)

1.2　リテラシー支援資料の作成･更新

1.2.1　「図書館利用ガイド」

附属図書館情報リテラシー教育部会編集の東京大学の図書館を初めて利用する学生等を対象とした「図書館利用ガイド 2007」及び、英語版「Guide to UT Libraries 2007」の作成に協力した。

1.2.2　「ネットでアカデミック」

東京大学における文献調査のガイドブックとして発行している「ネットでアカデミック」(日本語版、英語版、中国語版、韓国語版の 4 ヶ国語版)の増刷を行い、学内各図書館･室に配布した。後述の情報探索ガイダンスのテキストとしても活用した。

1.2.3 データベースリーフレット

東京大学で利用できる以下のデータベースについて、それぞれの特徴や検索方法を紹介したリーフレット 7 種(日本語版、英語版の 2 ヶ国語版)を作成し、学内各図書館・室に配布した。今年度は、機能の更新などにあわせて、「GACoS」と「東京大学 OPAC」について改訂を行った。

リーフレット 7 種類は次のとおり。

- GACoS / IRI : Index to Resources on Internet
- 東京大学 OPAC
- FELIX : ジャーナルコンテンツ・データベース
- 東京大学学位論文データベース / ProQuest Dissertations and Theses (PQDT)
- 電子版貴重書コレクション / 電子版自然真営道
- 東京大学で利用できる電子ジャーナル検索 / E-JOURNAL PORTAL
- JapanKnowledge

1.3 Litetopi メールマガジン

平成 17 年 12 月に開始した東京大学所属者に各種データベースサービスや、情報探索ガイダンスの案内などを送る「Litetopi(リテトピ)メールマガジン」の発行を継続して行った。平成 18 年度は、No.7 から No.27 まで 21 回発行した。登録者も約 500 人に増加した。

バックナンバーは、GACoS(http://www.dl.itc.u-tokyo.ac.jp/gacos/literacy.html)から参照できるようにしている。

1.4 学術データベースの拡充

1.4.1 Web of Science バックファイルの導入

主要な学術雑誌(約 9,000 誌)に掲載された論文の引用索引データベース「Web of Science」のバックファイルを購入し、検索可能な年代の範囲を 1970 年以降から次のように拡大した。

- Science Citation Index Expanded (SCI-EXPANDED): 自然科学系→1945 年以降
- Social Sciences Citation Index (SSCI): 社会科学系→1956 年以降

1.4.2 日本歴史地名大系オンライン版の導入

『日本歴史地名大系』(平凡社刊)全 51 冊を完全データ化した「日本歴史地名大系オンライン版」(JapanKnowledge JK セレクトシリーズ)のトライアルサービスを実施し、平成 19 年度からの正式導入を決定した。

2 講習会・研究会開催報告

2.1 情報探索ガイダンスの実施概要

データベースをはじめとした電子的学術資料を学習・研究に効率的に利用してもらうことを目的とする情報探索ガイダンスを年間を通じて開催した。1 年間で 111 回開催し、合計で 1416 人の参加があった。平成 17 年度に比べると、開催回数は減少したが、参加人数は 234 人増加した。

特に会場の記載がないコースは、総合図書館 1 階メディアプラザ I の講習会コーナーが会場である。

2.2 平成 18 年度情報探索ガイダンス開催実績

2.2.1 テーマ別ガイダンス

1 つのテーマについて、検索実習を交えながら解説した。コースの目的をわかりやすくするため、データベースユーザトレーニングの「FELIX と CiNii」コースを 10 月から「日本の論文を探すには?」

コースに改編した。また、学外から利用できるデータベースについての問い合わせが多いため、7 月から「自宅から検索するには?」コースを新設した。このコースでは、附属図書館相互利用係の協力を受け、帰省先などで東京大学以外の図書館を利用する場合の手続きや注意点も説明している。

また、10 月には、文献管理ソフト EndNote について、販売元から講師を招き、講習を行った。このコースと連動し、EndNote ヘルプデスクを講習会コーナーに設けた。

コース名	月日	回数	人数	内容
図書館の蔵書を検索するには?	10/3,3/8	2	3	OPAC の使い方を中心に、図書館の蔵書を検索できるデータベースを紹介。
日本の論文を探すには?	10/11, 1/26,3/12	3	13	FELIX、CiNii の使い方を中心に、日本の雑誌論文を検索できるデータベースを紹介
電子ジャーナルを利用するには?	5/18,7/18, 10/19, 3/14	4	14	電子ジャーナルにアクセスする方法や、代表的な出版社の電子ジャーナルサイトを例にとって実際の利用方法を解説。
自宅から検索するには?	7/6,8/29, 1/31,3/2	4	19	自宅や帰省先、海外からも利用できるデータベース、電子ジャーナルを紹介。
効率的な文献管理をするには? -EndNote 利用法	10/25	2	34	代表的な文献管理ソフト「EndNote」を例に各種データベースの文献情報の取り込みや引用文献リストの自動作成方法を解説。
EndNote ヘルプデスク	10/31	1	6	ヘルプデスクを設け、個人対応も行った。
合計		16 回	89 名	

2.2.2 データベースユーザトレーニング

特定のデータベースの使い方について、検索実習を交えながら解説した。「FELIX と CiNii」コースは、平成 17 年度までの定期講習会「FELIX 編」を拡充したものだが、2.2.1 に記載のとおり、さらに「日本の論文を探すには?」コースに改編した。

コース名	月日	回数	人数	内容
FELIX と CiNii	5/24,7/21	2	12	FELIXに収録されている各データベース（雑誌記事索引SwetScan、PCI）の検索方法とCiNiiの利用方法を解説。
Web of Science	8/2,8/23, 9/21,2/14, 3/20	5	19	Web of Science の特徴を解説し、キーワード検索や引用文献をキーにした検索方法を紹介。
合計		7 回	31 名	

2.2.3 分野別データベースユーザトレーニング

平成 17 年度の「分野別データベース講習会」（自然科学系分野と人文社会科学系分野）に「総合分野」を加え、「分野別データベースユーザトレーニング」とし、シリーズ化して各分野の主要なデータベースごとの講習会を実施した。データベースの提供元から講師を招き、実習形式で行った。

データベースによっては、利用者が参加しやすい会場を考慮し、駒場図書館や農学生命科学図書館などでも開催した。

・総合分野

コース名	月日	時間	人数	会場
JapanKnowledge	5/29	15:00-16:30	12	総合図書館 講習会コーナー
LexisNexis Academic	6/23	13:30-14:30	9	
MC-統計	6/28	15:00-16:30	11	

・人文社会科学系分野

コース名	月日	時間	人数	会場
Literature Resource Center with MLA International Bibliography	6/1	13:30-14:30	4	総合図書館 講習会コーナー
InfoTrac Custom	6/1	14:45-15:45	8	
ASSIA・LLBA	6/7	15:00-16:30	7	
LLBA	10/30	15:00-16:00	13	駒場図書館会議室
Business Source Elite	6/12	13:15-14:45	5	総合図書館
PsycINFO・PsycARTICLES	6/12	15:00-16:30	10	講習会コーナー
Web of Science　人文社会系	6/19	10:30-12:00	8	情報基盤センター 5 階大演習室 2
	6/27	13:15-14:45	11	駒場図書館会議室
Lexis.com	6/23	14:45-16:15	8	総合図書館
JSTOR	6/26	15:00-16:00	5	講習会コーナー
eol ESPer	6/29	15:30-16:30	5	

・自然科学系分野

コース名	月日	時間	人数	会場
SciFinder Scholar	6/16	14:45-16:15	5	総合図書館 講習会コーナー
	7/4	15:00-16:30	8	駒場図書館会議室
Web of Science　自然科学系	6/19	13:15-14:45	16	情報基盤センター 5 階大演習室 2
	6/27	15:00-16:30	16	駒場図書館会議室
BIOSIS Previews	6/19	15:00-16:30	6	情報基盤センター 5 階大演習室 2
AGRICOLA・ProQuest Agriculture Journals	6/30	13:30-15:00	15	農学生命科学図書館
合計	20 回		182 名	

2.2.4　総合図書館オリエンテーション

総合図書館オリエンテーションに協力し、春季は「OPAC 入門」、秋季は「データベース講習会」として、各 30 分のコースを担当した。

コース名	月日	回数	人数	内容
OPAC 入門	4/6～4/19	14	71	OPAC の使い方と GACoS の紹介
データベース講習会	10/11,10/13	2	27	OPAC、FELIX、電子ジャーナルの利用方法
合計		16 回	98 名	

2.2.5　留学生向け情報探索ガイダンス

2.2.4 記載の総合図書館オリエンテーションの英語コースに協力するとともに、データベースユーザトレーニング、テーマ別ガイダンスの英語編を開催した。11 月には、「留学生向け情報探索ガイダンス」を 4 ヶ国語で開催した。日本語コースを除き、外国人講師によって講習を行った。

コース名		月日	回数	人数
総合図書館オリエンテーション	OPAC 入門英語編	4/11,4/14	2	5
	データベース講習会英語編	10/12,10/18, 10/20	3	29
Web of Science 英語編		6/20,9/15,2/20	3	9
Electronic Journals 英語編		5/22,7/10,8/21	3	15

留学生向け情報探索ガイダンス	中国語コース	11/14	1	11
	韓国語コース	11/15	1	6
	日本語コース	11/17	1	1
	英語コース	11/21	1	6
合計			15 回	82 名

2.2.6 事務職員向け情報探索ガイダンス

事務職員が業務をする上で役立つデータベースの紹介を行った。この実績を基に、平成 19 年度新規採用職員研修のプログラムの 1 コマに組み込まれることが決定した。

月日	時間	人数	内容
12/11	16:00-17:00	9	辞書・事典ツール、新聞・雑誌記事、官報、統計、企業情報の検索方法や図書館サービスを紹介。
12/19	11:00-12:00	12	
合計	2 回	21 名	

2.2.7 出張講習会

利用者(教員、学生等)からの依頼に合わせて行う出張講習会を実施した。データベースの検索で得た情報から文献入手までの手順がわかるよう、関係する部局図書館から図書館(室)のサービス案内を行うなど、附属図書館との連携を強化した。今年度は、駒場図書館、大学院人文社会系・文学部図書室、理学部生物学図書室、経済学部図書館、医学図書館、総合図書館と協力して行った。

教養学部の基礎科目「基礎演習」の授業の 1 コマでの講習も出張講習会の一環として行った。さらに、教養学部の「英語」と「会計」のクラスでも教員からの要望があり、今年度から授業に出向いての講習を実施し、教員と連携した授業組み込み型の講習会が増えつつある。

また、医学部附属病院看護部からの依頼により、看護師の臨床実践や研究活動に役立てるための講習会も始めることになった。

いくつかの講習会では、受講者に対して、担当教員等の協力の下、事前アンケートや 2～6 ヵ月後の追跡アンケートを実施し、利用者のニーズや講習会の効果を調査した。事前アンケートの結果は、講習会の内容に反映させ、講習の効果を高めるように努めた。

研究科・研究室名等	月日	時間	人数	内容	会場
教養学部 「基礎演習」	4/18	13:00-14:30	22	・情報検索の基礎の解説と実習	駒場 情報教育棟 3 階中演習室 3
	4/24	14:40-16:10	41	・ JapanKnowledge 、OPAC、Webcat Plus、Book Contents 、FELIX、CiNii 等の利用方法	駒場 情報教育棟 4 階中演習室 2
	4/25	10:40-12:10	22		
	4/26	14:40-16:10	24		
	4/27	13:00-14:30	20		
	4/28	9:00-10:30	24		
	4/28	10:40-12:10	42		
学際情報学府 山内研究室	5/22	13:00-14:00	4	・ JapanKnowledge 、OPAC 、Web of Science、CiNii、電子ジャーナル等の利用方法	情報学環アネックス 山内研究室
人文社会系研究科 “Academic Writing”	5/30	16:50-18:30	10	・ JapanKnowledge 、OPAC、CiNii、Web of Science 、電子ジャーナル等の利用方法 ・文学部図書室案内	総合図書館 講習会コーナー
文学部 “Academic Writing”	6/1	17:00-18:00	13	・ JapanKnowledge 、OPAC、Webcat、CiNii	総合図書館 講習会コーナー
	12/21	17:00-18:00	9		

				等の利用方法 ・文学部図書室案内	
理学部 生物学科 動物学コース・植物学コース	6/21	13:30-15:00	13	・OPAC、Web of Science、電子ジャーナル、BIOSIS、PubMed 等の利用方法 ・生物学図書室案内	総合図書館 講習会コーナー
総合文化研究科 三谷研究室	6/27	10:00-11:30	7	・OPAC、Webcat、CiNii、JSTOR 等の利用方法	総合図書館 講習会コーナー
医学部附属病院 看護部	10/24	17:30-18:30	19	・OPAC、電子ジャーナル、CiNii、医中誌 Web 等の利用方法 ・医学図書館案内	医学部附属病院 中央診療棟 2 7 階会議室
	11/16	17:30-18:30	18		
	12/5	17:30-18:30	20		
教養学部「会計」	10/27	13:00-14:30	350	eol ESPer の利用方法	駒場キャンパス 13 号館 1323 教室
教養学部「英語」	11/6	13:30-14:30	15	・OPAC、Webcat、FELIX、電子ジャーナル等の利用方法	駒場図書館 会議室
	11/6	14:40-15:40	20		
文学部社会心理学研究室「社会心理学調査実習」	1/18	10:30-11:15	15	・OPAC、Webcat、電子ジャーナル、CiNii、PsycINFO、Web of Science 等の利用方法 ・図書館ツアー	総合図書館 講習会コーナー
	1/18	11:15-12:00	14		
合計	21 回		722 名		

2.2.8 個人講習会

平成 16 年度より開始している個人の希望に対応した個人講習会を今年度も継続して行った。

月日	時間	所属	身分	人数	内容
4/20	11:00-12:00		教員	1	GACoS 及び東京大学で利用できるデータベースの紹介と投稿雑誌の探し方
6/15	13:00-14:00	新領域	大学院生	1	OPAC、FELIX、CiNii、電子ジャーナル、LexisNexis Academic、JSTOR の検索
7/6	16:00-16:30	新領域	大学院生	1	JSTOR の検索
9/1	13:00-14:30	文学部	学部生	1	雑誌記事の検索
9/19	14:00-15:30	総合文化	大学院生	1	雑誌論文の検索
10/17	14:00-16:00	新領域	大学院生	1	OPAC、FELIX、CiNii、自宅からの検索のための手続き、各種図書館サービス
12/21	14:00-15:00	工学系	職員	1	OPAC、CiNii、Web of Science
2/15	14:00-15:00	工学系	職員	1	朝日新聞オンライン記事データベース「聞蔵」、LexisNexis Academic
2/23	13:00-14:00	医学系	大学院生	1	CINAHL Plus with Full Text
合計				9 名	

2.2.9　セミナー・説明会

6 月 5 日から 10 月 6 日までトライアルサービスを実施した遺伝子配列事典「AssEST」の説明会を実習形式で行った。2 月には、文献管理・論文執筆支援ツール「EndNote Web」や著者識別システム「DAIS」など新機能が追加された「Web of Science」についての説明会を開催した。3 月には EndNote（デスクトップ版）の初心者向けの説明会を駒場図書館、柏図書館と協力し、3 キャンパスで開催した。各回ともデータベースの提供元から講師を招いて行った。

また、総合図書館大会議室を会場に開催した「Web of Science 新機能説明会」と「EndNote 入門セミナー」では、電子教材係の協力の下、テレビ会議システムを用いて、柏図書館メディアホールに中継を行い、柏キャンパスの利用者が本郷キャンパスまで出向かなくても同じ内容を視聴できるようにした。

講習会名	月日	時間	人数	会場
AssEST セミナー	9/5	13:30-15:00	5	総合図書館講習会コーナー
小計	1回		5	
Web of Science 新機能説明会	2/8	15:00-16:30	55	総合図書館 3 階大会議室
			5	柏図書館メディアホール (総合図書館からの中継)
小計	1回		60	
EndNote 入門セミナー	3/8	15:00-16:30	27	駒場図書館会議室
	3/14	15:00-16:30	27	柏図書館メディアホール
	3/27	15:00-16:30	58	総合図書館 3 階大会議室
			5	柏図書館メディアホール (総合図書館からの中継)
小計	3 回		117	
合計	5 回		182 名	

図書館関係システム運用・管理

図書館情報係　　本多　玄

1　運用報告

本章では図書館情報係管轄のサーバ群の運用・管理について記述する。

1.1　附属図書館学術情報システム

本システムは OPAC、MyLibrary、Web 画面からの図書購入依頼等の「利用者サービス」面と図書・雑誌の受入、目録作成等の「図書館業務システム」面からなるシステムである。

平成 17 年度末に現在のシステムにリプレイスされたが、本学向けカスタマイズ案件の多くの部分および新規サービスである MyLibrary 機能のリリースは平成 18 年度に持ち越されることとなった。

上記を考慮して評価用途と図書職員のシステム操作練習用途を兼ねた「評価／講習会用サーバ」を新たに構築し、「メーカー開発＝＞評価環境適用＝＞東大側での検証＝＞問題なければ本番環境適用」という流れで改修を進めることができた。

平成 18 年度はこれらの評価環境での動作確認や、すでに本番稼動している部分の障害・不具合に関するメーカーとの連絡・調整に業務の大部分が充てられたといってよい。

また、ハードウェアの初期不良、大規模大学である本学での使用を考慮した場合にソフトウェア設定で不備な部分もあり、障害発生の都度、その原因調査を通じて対応策を講じ安定稼動に向かっているところである。

なお、学術雑誌購入経費の共通経費化が決定したことに伴い、図書館業務面でも対応した業務が可能なよう新規のシステム改修を行なった。

1.2　情報サービス用サーバ関連

図書館システムリプレイスにともない、旧図書館システムで使用していたサーバ数台が他用途に転用可能となったため、以下のサーバについて機種更新とサーバOSの変更、セキュリティの強化を行なった。

- 附属図書館 Web サーバ
- 図書職員用 Wiki サーバ
- バックアップ用サーバ

また、ブックコンテンツデータベースについては機器類更新のみでなく、システム設計を全面的に変更することとし、附属図書館学術情報システムとの完全統合を実現するためメーカーと仕様打ち合わせ、評価環境でのテスト等を行なった。リリースは平成 19 年 5 月中の予定である。

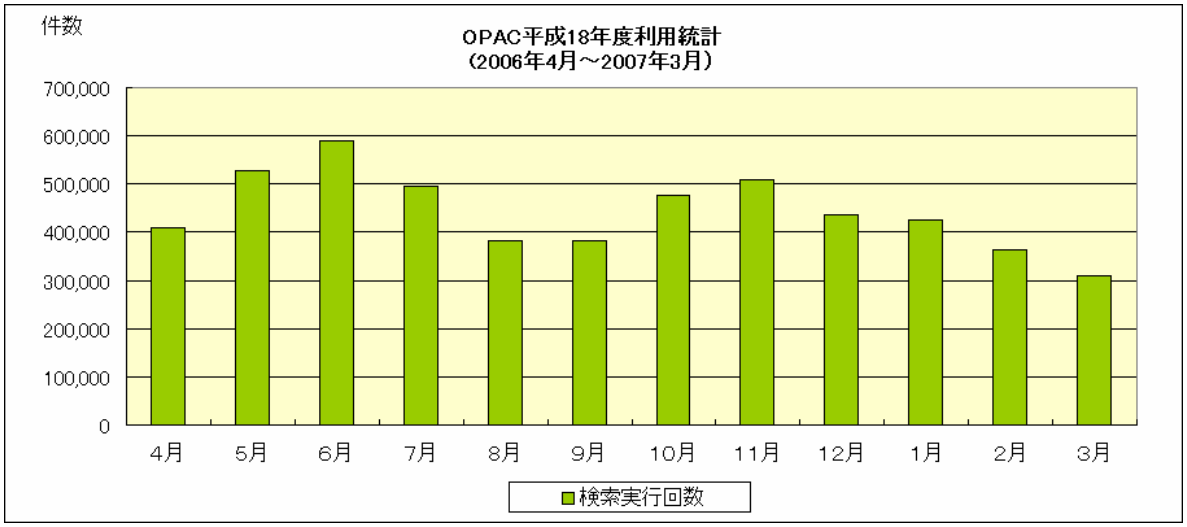
2　講習会・研究会開催報告

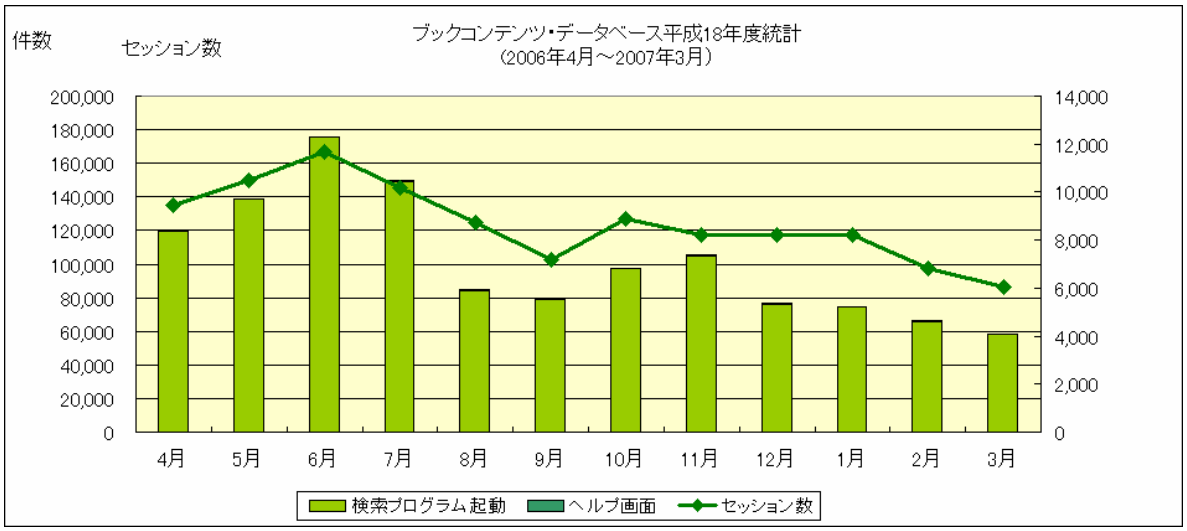
7 月の職員人事異動に対応し、7 月 13 日から 21 日の間、新規採用および異動によりシステム上での担当業務未経験者を対象にシステム操作の講習会を行なった（附属図書館主催）。

図書館情報係は講習会用端末からサーバへのアクセス権設定、ネットワークセグメントの変更（講習会期間中は端末を図書館 VLAN に一時的に移行）等で支援した。

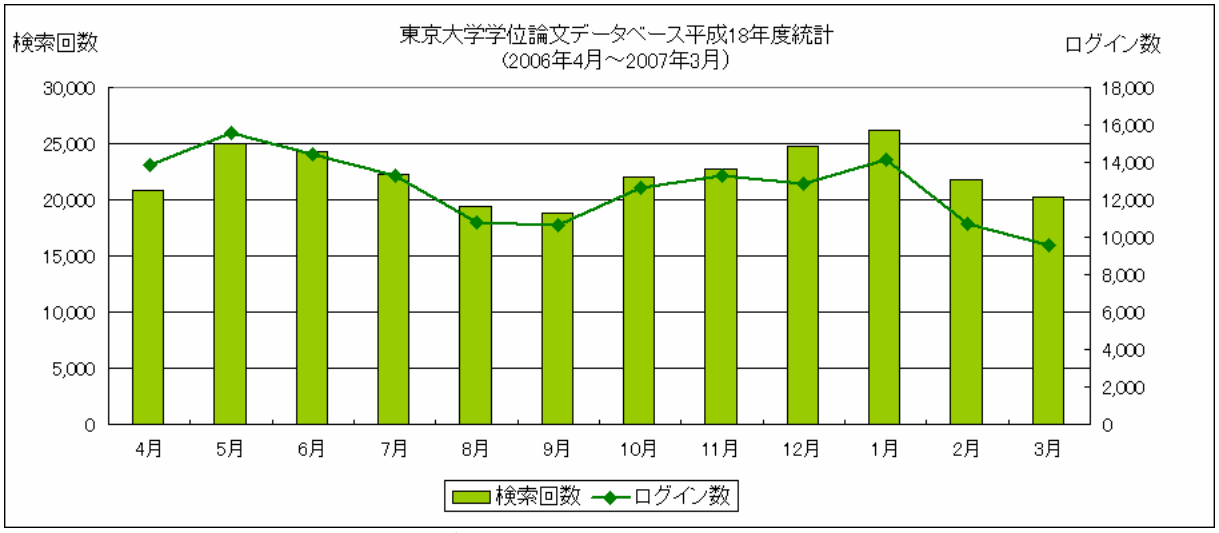
3　サービス統計

OPAC 平成 18 年度利用統計

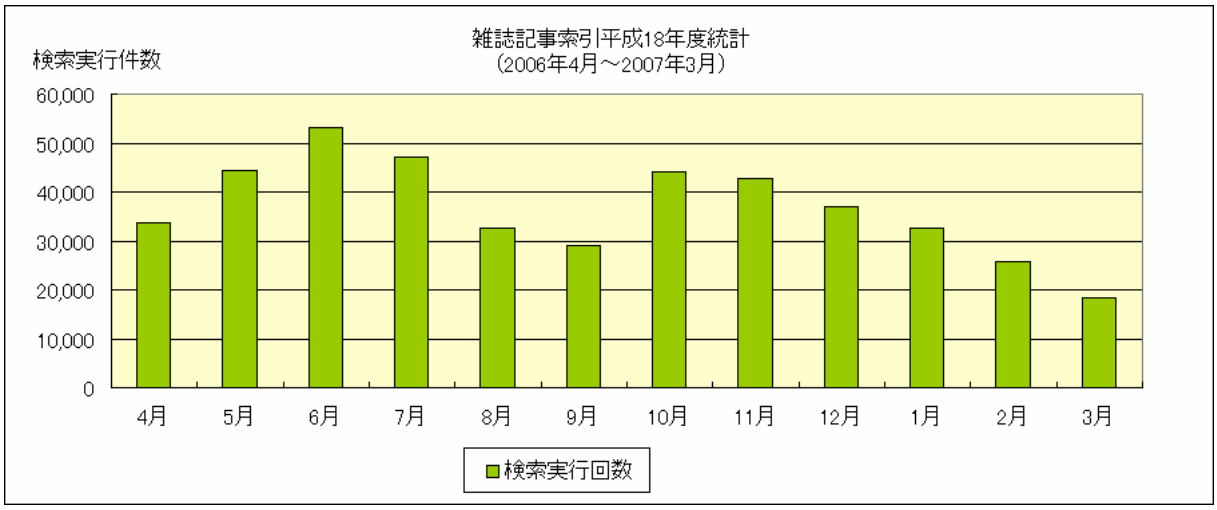




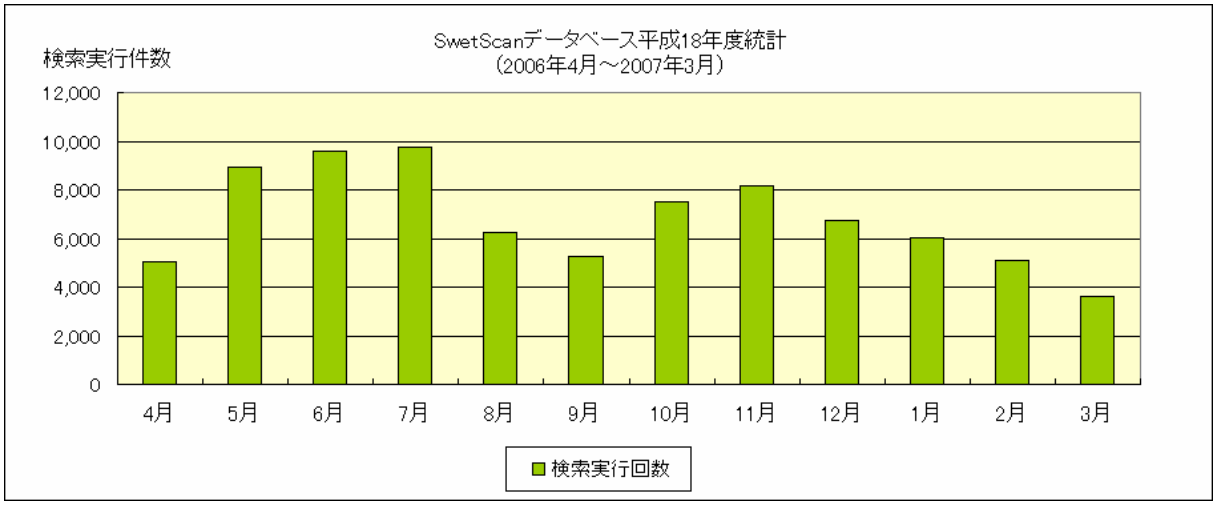
ブックコンテンツデータベース平成 18 年度利用統計



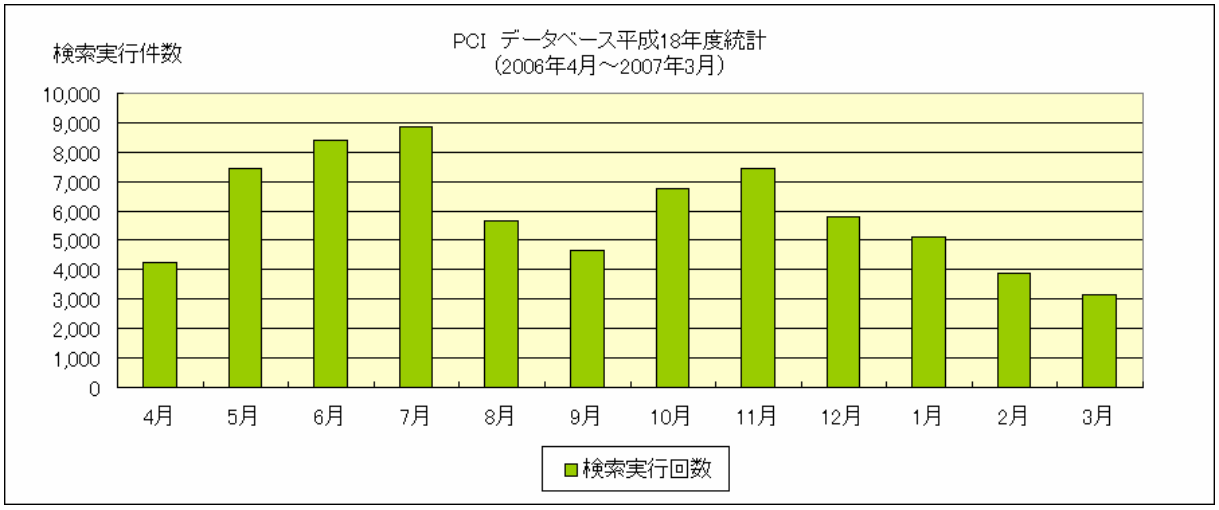
学位論文データベース平成 18 年度利用統計



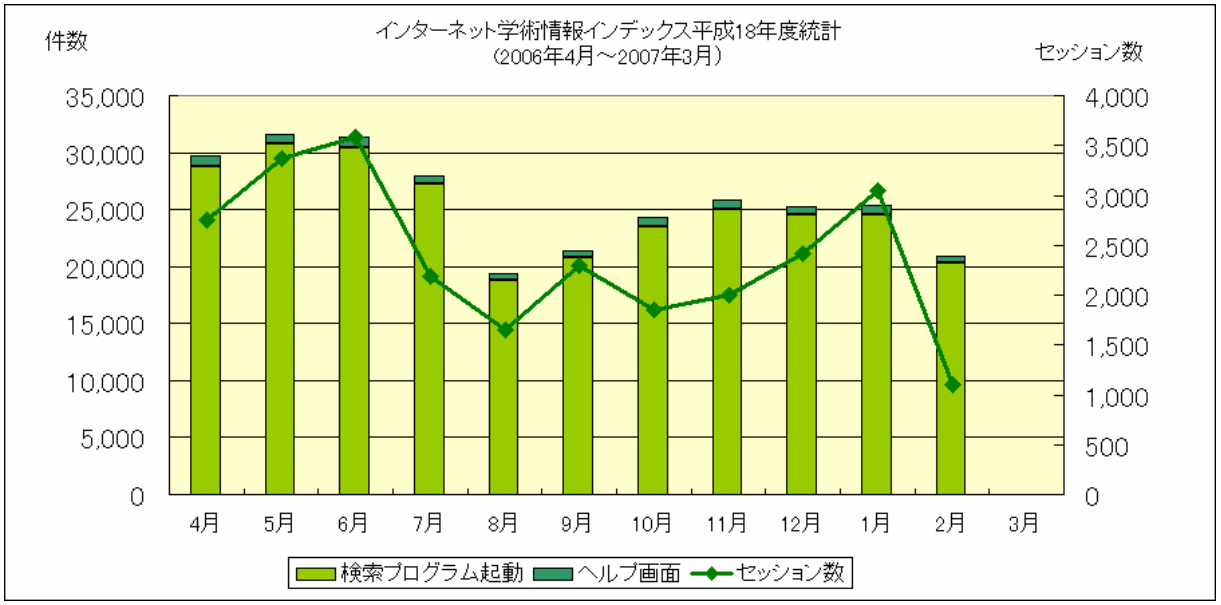
雑誌記事索引平成 18 年度利用統計



SwetScan 平成 18 年度利用統計



PCI 平成 18 年度利用統計



IRI インターネット学術情報インデックス

キャンパスネットワーキング部門

キャンパスネットワーキング部門

概要

早野裕士

■ 東京大学キャンパスネットワークシステム（UTnet3）の運用管理

本学のキャンパスネットワークシステム UTnet3(University of Tokyo network system 3)は、各建物内の支線ネットワーク(支線)及び建物間接続や学外との接続のための基幹ネットワーク（基幹）とから構成されている。情報基盤センターの本キャンパスネットワーキング部門は、以下に示す通り、基幹の運用管理に加え、これに関連する業務も含めて取り組んだ。

1. 基幹ネットワークの運用管理
基幹の中核を占めるネットワーク機器は、本郷地区、駒場 I 地区、駒場 II 地区、柏地区、白金地区、中野地区の各ハブサイトに設置したギガビットスイッチ(L3)であり、基幹の基本トポロジーは情報基盤センターを中心にしたスター型になっている。基幹に対するトラヒック需要の増加に対応するため、ギガビットスイッチの転送処理ボードの増強、回線速度の高速化等を実施した。また遠隔研究施設については本郷地区との接続性を確保しているが、同様にトラヒック需要に応じた増強を実施した。

2. 基幹ネットワークと支線ネットワークとの接続
各建物にレイヤー2 スイッチを設置し、基幹のギガビットスイッチ(L3)と支線を接続しているが、ネットワーク利用の増大への対応と高機能化のため、2005 年度に引き続き、レイヤー2 スイッチの順次更新を行った。

3. VLAN 対応
UTnet3 では仮想 LAN(VLAN: Virtual LAN)に対応している。VLAN によって、部局や研究室が複数の建物やキャンパスに分散して配置されている場合でも部局や研究室のまとまりごとに同一のサブネットに収容することを可能としている。本学では、建物の新設や組織変更等が少なくなく、それに伴った VLAN に関する要求は非常に多く、それら要求に応じてネットワーク機器の設定変更や増設等を実施した。

4. キャンパス間接続、及び対外ネットワーク接続
キャンパス間接続、及び学外への接続については、ギガビットスイッチを設置し、Super-SINET をはじめとした広帯域回線による接続を行っている。2006 年度は、接続回線速度の高速化や高機能化を実施した。

5. その他関連業務
UTnet3 基幹ネットワークの運用管理に関連する業務として、ネットワーク機器監視業務の改善、ドメイン名の割り当て、DNS(Domain Name System)のサービス、UTnet 光ファイバケーブル専用利用、及び無線 LAN サービスの各業務に取り組んだ。

■ セキュリティ対応

最近のネットワークおいてはセキュリティへの対応が必須となっており、本学においてもセキュリティインシデントは日常茶飯事と言っても過言ではないほどに頻繁に発生している。特にコンピュータへの不正侵入の試み、電子メールや Web ページを通してのウィルス感染等が問題となっている。そこで、本学のセキュリティの維持確保に向け、以下の通り、セキュリティ対応業務を実施した。

1. 不正侵入の監視
学外と学内との間の通信に関し、基幹部分において不正侵入の試みを検知し事前に予防する仕組みとして、UTnet3 では不正侵入検知装置 (IDS: Intrusion Detection System)と異常トラフィック監視システムを導入して監視し、その結果に応じて必要な対処を実施した。

2. ウィルス感染への対応
メールや Web ページ等を通してのウィルス感染に関する対策として、情報基盤センターでは、適切なウィルス対策ソフトウェア製品を学内利用者の希望に応じて配布し、予防に役立てている。2006 年度は、配布製品を拡大するとともに、利用者の増大を図って、ウィルス対策を積極的に進めた。

■ 東京大学情報システム緊急対応チーム(UT-CERT)との連携
UTnet3 ではセキュリティに関する対応として、基幹部分のギガビットスイッチ(L3)で、全学的に問題となる事象や各部局の個別の事象についてのフィルタを設定している。東京大学情報システム緊急対応チーム(UT-CERT)から委託業務を受け、このフィルタを活用することによって、全学的なセキュリティ対策を行った。

■ 学内ソフトウェアライセンス
学内に利用者の多いソフトウェアについて、全学サイトライセンスを取得し、そのライセンス管理と配布サービスを行った。

東京大学キャンパスネットワークシステム（UTnet3）の運用管理

中村昇平　　佐島浩之　　佐々木馨
松本浩一

1　運用報告

本学のキャンパスネットワークシステムは UTnet(University of Tokyo network system)と称し、これまでの更新経緯に応じて現在は UTnet3 と呼ばれている。UTnet3 は支線ネットワーク(支線)と基幹ネットワーク（基幹）とから構成されている。支線は、各建物内に設置されたネットワークで、当該部局によって運用管理されている。基幹は、支線の相互接続及び学外との接続のために設置されたネットワークで、情報基盤センターの本キャンパスネットワーキング部門が運用管理している。本部門では、以下に示す通り、基幹ネットワークの運用管理に加え、これに関連する業務も含めて取り組んだ。

1.1 基幹ネットワークの運用管理

本郷地区の基幹には、 5 箇所のハブサイト（情報基盤センター、電話庁舎、附属図書館、工学部 8 号館、農学部 3 号館）があり、各ハブサイトにはギガビットスイッチ(L3)を設置し、基幹は情報基盤センターを中心にしたスター型の構成としている。駒場 I 地区、駒場Ⅱ地区、柏地区、白金地区、中野地区の各郊外地区についても、各ハブサイトにギガビットスイッチ(L3)を設置し、専用回線等を介して本郷地区と接続している。このような基幹の運用管理について、主に以下の取り組みを行った。

－Super-SINET によって接続されている柏－本郷両地区間の接続の EoMPLS (Ethernet over Multi-Protocol Label Switching)方式への変更
－本郷地区内の 5 台のギガビットスイッチ(L3)の CPU ボードの増強
－柏に設置しているギガビットスイッチ(L3)の CPU ボードの増強

以上の変更・増強等の実施によって、柏と本郷の両地区にまたがる VLAN の設定を可能とする等の機能の向上と、データ転送能力に係わる性能の向上を達成し、基幹に対する学内からの要求に応えることによって、本学の研究・教育の一層の円滑な推進を図った。

基幹のネットワーク構成を図 1 に示す。また、本郷地区と、駒場 I 地区、駒場Ⅱ地区、柏地区、白金地区、中野地区の各キャンパスの間のトラフィック量の推移を図 2 に示す。

1.2 基幹ネットワークと支線ネットワークとの接続

基幹と各支線の接続については、各建物にレイヤー2 スイッチを設置し、このスイッチとハブサイト内のギガビットスイッチ(L3)との接続で実現している。 UTnet3 導入当初にはレイヤー2 スイッチを約 200 台設置したが、ネットワーク利用の増大への対応と高機能化のため 2005 年度より順次更新作業を開始した。これまでに約 90 台の更新を行い、2006 年度には、更に 74 台のレイヤー2 スイッチの更新を行った。

1.3 VLAN 対応

本学では、部局や研究室が、複数の建物やキャンパスに分散して配置されている場合及び他部局の建物の一部に入室する場合が増えており、これに伴って、部局や研究室のまとまりご

とに同一のサブネットを利用できる建物間仮想LAN（Virtual LAN：VLAN）の申請が増えている。VLANに関して2006年度にあった申請に伴って実施した、基幹ネットワークの構成変更およびサブネットワークの割当は表１の通りである。

1.4 キャンパス間接続、及び対外ネットワーク接続

以下に示す通り、学内キャンパス間接続及び対外ネットワークとの接続に関して、変更作業を実施し、これによって、各接続に関して高機能化・高速化を達成した。

－2006年9月に、超高速・高機能研究開発テストベッドネットワークであるJGN2(Japan Giga Network 2) との接続を、1Gb/sから10Gb/s に変更した。

－2006 年10 月に、Super-SINET によって接続されている柏－本郷間をEoMPLS接続に変更した。（1.1 参照）

－当初ADSL回線で接続を開始していた柏Ⅱキャンパスとの接続を光ファイバー(FTTH：Fiber To The Home)に変更した。

－医科学研究所奄美研究施設、文学部附属北海文化研究常呂実習施設を、ADSL回線によるVPN(Virtual Private Network)接続に変更した。

－情報理工学系研究科の秋葉原研究室の設置に伴い、Super-SINET経由でのtunnel設定による接続を行った。

キャンパス間回線のトラフィック量の推移は図 2 の通りであった。

1.5 その他関連業務

(1)ネットワーク機器監視業務の改善

UTnetのスイッチ監視業務に利用しているシステムを、センター開発のスクリプトから、システムやネットワーク監視機能を持つフリーソフトウェアのNagios への利用に移行した。これによって、以下の通りネットワーク機器監視に関する機能を改善した。

・ICMP(Internet Control Message Protocol)のみによる監視から、SNMP(Simple Network Management Protocol)併用による監視への変更になり、詳細な状態の監視が可能になり、障害等の状況の把握と対応が容易となった。

・旧システムでは、障害の発生時及び回復時にメールが関係者に送られるだけだったが、Nagios システム導入後は、Web ブラウザによる視覚的な監視も可能になり、監視業務の効率化が図れた。

・サーバ監視機能があるため、これまでは対象外だったDNSやWWWサーバのデーモン監視が可能になった。

(2)DNSのサービス

PC等のDNSクライアントからDNSサービスを利用可能とするために、２台のDNSサーバ（ns.nc.u-tokyo.ac.jp(133.11.124.164)及びns2.nc.u-tokyo.ac.jp(133.11.123.45)）を運用した。また、別の３台の DNS サーバ（dns1.nc.u-tokyo.ac.jp(133.11.0.1)， dns2.nc.u-tokyo.ac.jp(130.69.0.1) 及び dns3.nc.u-tokyo.ac.jp(157.82.0.1)）を運用し、u-tokyo.ac.jp直下のサブドメインの正引きとIPアドレスの逆引きの権限委譲、及び、利用者からの希望に応じたセカンダリサーバ運用サービスを実施した。

なお、ns/ns2は、キャッシュサーバ(Recursive Server)、dns1/dns2/dns3はコンテンツサーバ(Authoritative Server)として機能を分離させ、信頼性の向上と性能の向上を図った。

(3)ドメイン名の割り当て

2006 年度に申請があり、キャンパスネットワーク専門委員会で承認されたドメイン名の割当を行った。内訳は、新規が5件、変更が1件で、詳細を表2に示す。

(4)UTnet 光ファイバケーブル専用利用

本学では建物間で光ファイバを専用利用することが可能で、2006 年度に行ったUTnet 光ファイバケーブル専用利用の変更及び割り当てを表3 に示す。

(5)無線LAN サービス

2005年3月より開始した山上会館での無線LANサービスに引き続き、浅野地区に建てられた武田先端知ビル内の武田ホール、及び弥生地区に建てられた弥生講堂・一条ホールにおいて無線LANを含めたネットワークサービスを、それぞれ2006年6月及び2006年7月に開始した。これらに対する2006年度の各部局別の申請利用状況を表4に示す。

2　講習会・研究会開催報告

[第4回UTnet Meeting]
最近のネットワークの動向や管理の問題点に関して情報交換を行うための集まりとして開催しているUTnet Meeting を、2006年9月27日に理学部1号館の小柴ホールで行った。
プログラムは次の通りであり、参加者は、情報基盤センターの関係者を含め 109 名であった。

1. SINET3 について(国立情報学研究所)　漆谷 重雄
2. UTnet3 の拡充計画等　佐々木 馨
3. UTnet3 の無線LAN スポットについて　佐島 浩之
4. UT-CERT の今の体制　中山 雅哉
5. レポートシステム　坂井 朱美
6. 迷惑メール対策　安東 孝二
5. ユーザからの事例報告(工学部)　山崎 一彦
6. フリーディスカッション

[平成18 年度東京大学技術職員研修(コンピュータ関係)]
本学職員のうち、研究室やセンター等においてコンピュータやコンピュータ・ネットワークの初心者レベルの利用者、及びそれらの運用管理業務やセキュリティ対策業務に携わる初心者レベルの管理者を対象に、セキュリティ対応を含むコンピュータやコンピュータ・ネットワークの利用に関する研修を、以下の通り実施した。参加者は約20名であった。

・期間
平成18年11月28日(火)～平成18年11月30日(木)(3日間)
・場所
平成18年11月28日(火):情報基盤センター4階講義室
平成18年11月29日(水):情報基盤センター4階講義室、同センター5階演習室
平成18年11月30日(木):情報基盤センター4階講義室、同センター5階演習室

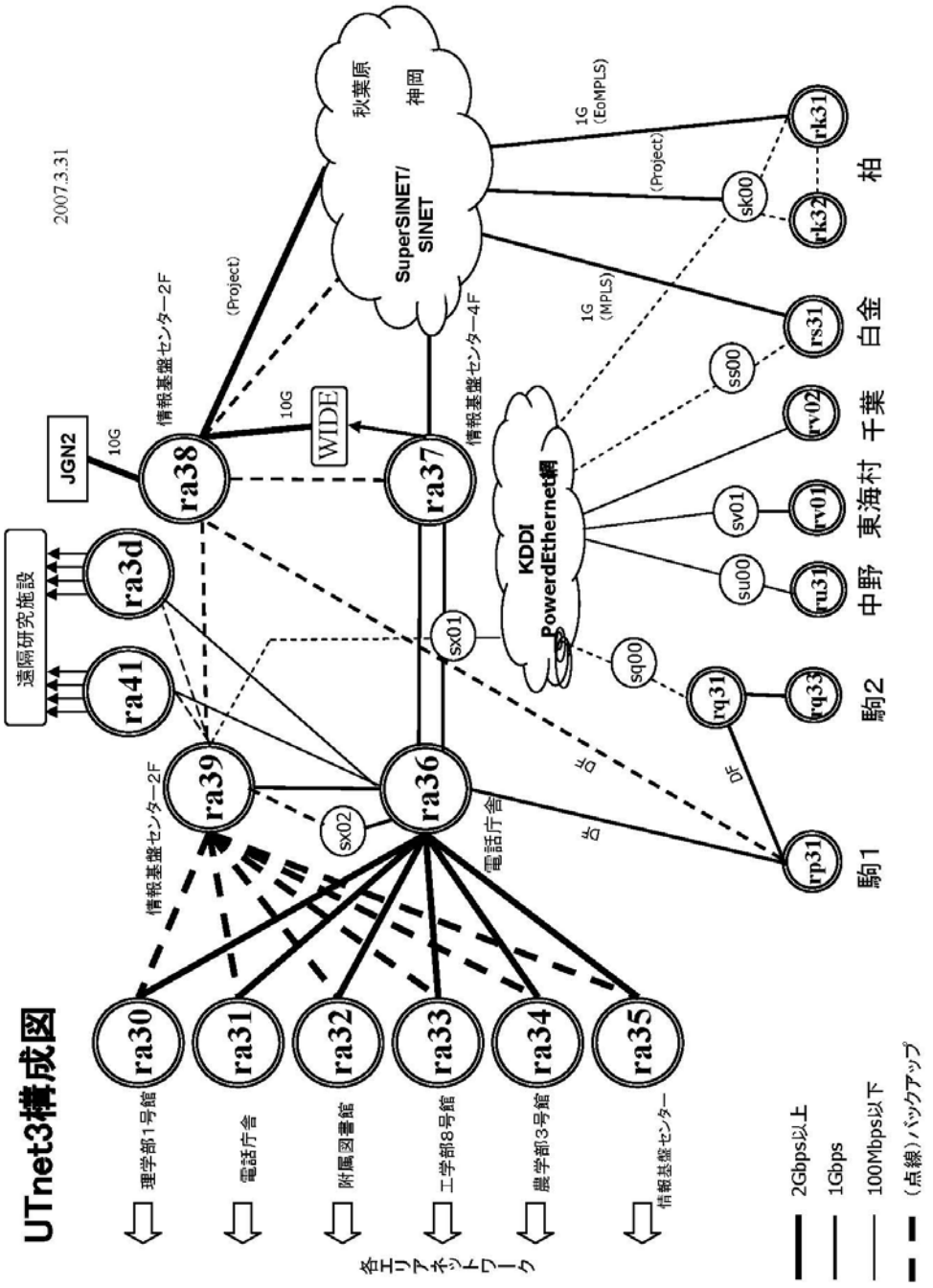


図1　Utnet基幹構成図

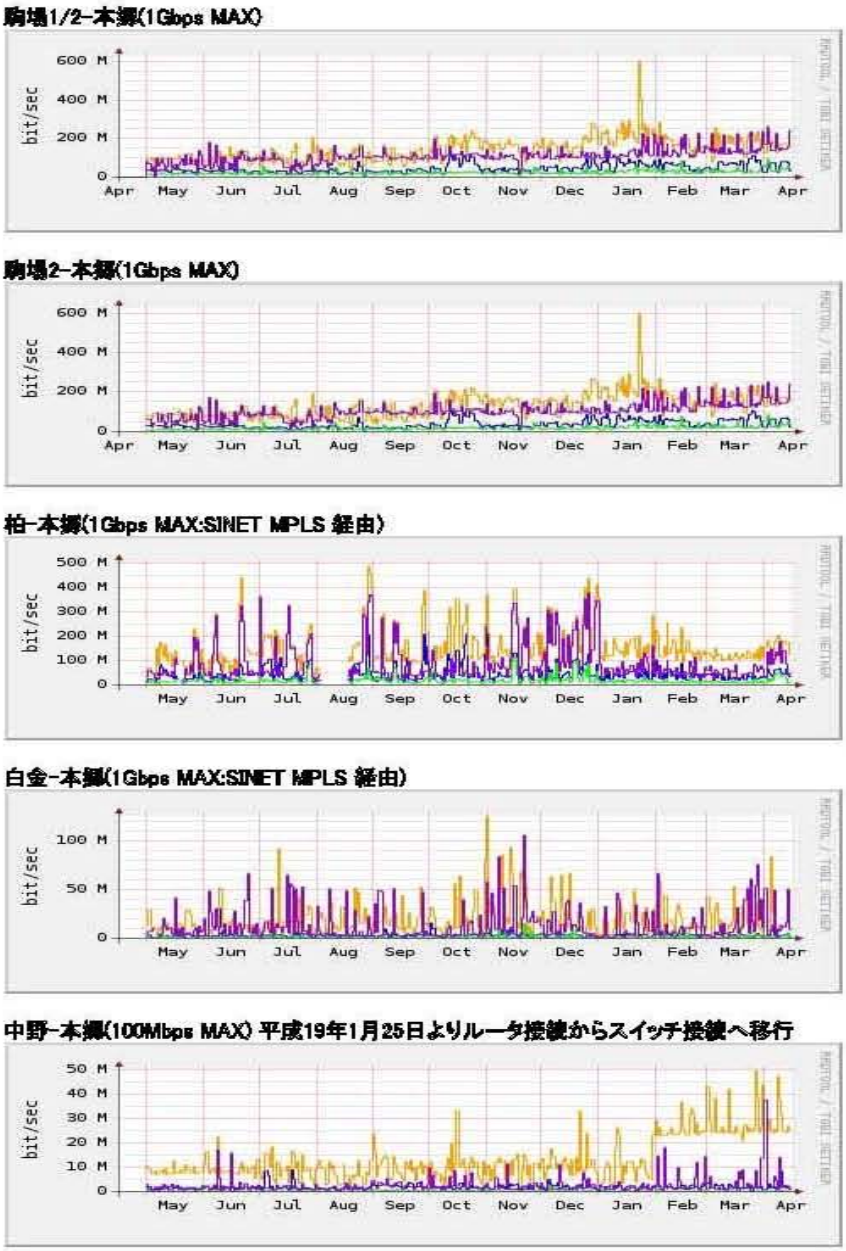


図2 UTnetのキャンパス間トラフィック

表 1 基幹ネットワークの構成変更およびサブネットワークの割当

項番	申請部局名	設置場所	変更内容
1	本部事務組織	御殿下記念館	建物間 VLANの申請
2	大学院理学系研究科	理学部1号館	サブネットアドレスの申請
3	情報基盤センター	工学部2、8号館	建物間 VLANの申請
4	本部事務組織	本部庁舎	サブネットアドレスの申請
5	情報基盤センター	情報基盤センター	基幹ネットワークの構成変更
6	本部事務組織	本部庁舎	サブネットアドレスの申請
7	情報基盤センター	情報基盤センター	基幹ネットワークの構成変更
8	情報理工学系研究科	工学部新2号館	サブネットアドレスの申請
9	情報基盤センター	弥生講堂	建物間 VLANの申請
10	大学院総合文化研究科	教養101号館	建物間 VLANの申請
11	情報理工学系研究科	工学部14、新2号館	建物間 VLANの申請
12	気候システム研究センター	気候システム研究センター	サブネットアドレスの返却
13	本部事務組織	地震研、地震研1号館、列品館	建物間 VLANの申請
14	情報基盤センター	情報教育棟	建物間 VLANの申請
15	情報理工学系研究科	工学部新2号館	建物間 VLANの申請
16	大学院総合文化研究科	教養900番教室	建物間 VLANの申請
17	情報基盤センター	医学部1号館	建物間 VLANの申請
18	本部事務組織	情報基盤センター	建物間 VLANの申請
19	情報学環	第 2本部棟、法学1号館、附属図書館	建物間 VLANの申請
20	人文科学研究科	文学部アネックス	建物間 VLANの申請
21	東洋文化研究所	東洋文化研究所	建物間 VLANの申請
22	大学院農学系研究科	弥生講堂	基幹ネットワークの構成変更
23	本部事務組織	工学部8号館、東洋文化研究所	建物間 VLANの申請
24	情報基盤センター	本部庁舎、情報基盤センター	建物間 VLANの申請
25	大学院工学系研究科	工学部列品館	建物間 VLANの申請
26	情報基盤センター	工学部11号館	建物間 VLANの申請
27	大学院工学系研究科	工学部3号館	基幹ネットワークの構成変更
28	大学院工学系研究科	タンデム加速器研究棟	建物間 VLANの申請
29	情報学環	第 2本部棟、法学1号館、附属図書館	建物間 VLANの申請
30	大学院農学系研究科	農学部7号館B	建物間 VLANの申請
31	大学院理学系研究科	三鷹	サブネットアドレスの申請
32	情報学環	工学部新2号館	建物間 VLANの申請
33	大学院人文科学研究科	常呂実習施設	サブネットアドレスの申請
34	新領域創成科学研究科	柏	サブネットアドレスの申請
35	大学院工学系研究科	工学部新2号館	基幹ネットワークの構成変更及びサブネットアドレスの申請
36	IML	IML棟	基幹ネットワークの構成変更、建物間 VLANの変更
37	情報基盤センター	情報基盤センター、情報学環	建物間 VLANの申請
38	情報基盤センター	情報基盤センター、情報学環	サブネットアドレスの返却
39	本部事務組織	情報基盤センター、本部庁舎	基幹ネットワークの構成変更及びサブネットアドレスの申請、建物間 VLANの変更
40	情報基盤センター	薬学部総合研究棟、薬学部本館	建物間 VLANの申請
41	本部事務組織	本部庁舎	建物間 VLANの申請
42	医科学研究所	医科研4号館	建物間 VLANの申請
43	理学部	理学部1号館	サブネットアドレスの申請
44	情報基盤センター	情報基盤センター、理学部7号館	建物間 VLANの申請
45	情報基盤センター	情報基盤センター、柏図書館	建物間 VLANの申請
46	情報基盤センター	情報基盤センター	建物間 VLANの申請
47	情報理工学系研究科		サブネットアドレスの申請

48	情報基盤センター	情報基盤センター	サブネットアドレスの申請
49	大学院総合文化研究科	教養学部8号館	基幹ネットワークの構成変更
50	本部事務組織	本部庁舎	基幹ネットワークの構成変更及び サブネットアドレスの申請
51	人工物工学研究センター		サブネットアドレスの返却
52	情報理工学系研究科	工学部新2号館	基幹ネットワークの構成変更及び サブネットアドレスの申請
53	情報基盤センター	情報基盤センター	サブネットアドレスの申請
54	大学院農学系研究科	農学部2号館別館	サブネットアドレスの申請、返却
55	大学院工学系研究科		サブネットアドレスの返却
56	気候システム研究センター	柏総合研究棟	サブネットアドレスの申請
57	東洋文化研究所	東洋文化研究所	建物間 VLANの申請
58	本部事務組織	情報基盤センター	建物間 VLANの申請
59	新領域創成科学研究科	柏総合研究棟	サブネットアドレスの返却

表 2 ドメイン名の割当

項番	ドメイン名	申請部局名	備考
1	kashiwa.u-tokyo.ac.jp	柏地区事務部	柏キャンパス共同学術経営委員会
2	kj.u-tokyo.ac.jp	柏地区事務部	kashiwa-jimu(1 2月末まで併用)から変更。
3	editoria.u-tokyo.ac.jp	本部事務組織	地球観測データ統融合連携研究機構
4	iaeste.u-tokyo.ac.jp	大学院工学系研究科	東京大学イアエステ委員会
5	kyodo-sankaku.u-tokyo.ac.jp	本部事務組織	男女共同参画室
6	mail.u-tokyo.ac.jp	本部事務組織	東京大学事務系職員公式メールアドレス

表 3 UTnet光ファイバケーブル専用利用の変更及び割当

項番	申請部局	利用区間	種類	
1	情報理工学系研究科	工学部新2号館－情報基盤センター	SM 2芯	新規
2	大学院理学系研究科	理学部1号館－理学部化学館	SM 2芯	廃止
3	大学院理学系研究科	理学部1号館－理学部4号館	SM 2芯	廃止
4	新領域創成科学研究科	工学部2号館旧－工学部8号館	SM 2芯	新規
5	情報理工学系研究科	理学部1号館－情報基盤センター	SM 2芯	新規
6	大学院総合文化研究科	教養16号館－アドバンスドラボ棟	GI 2芯	廃止
7	情報理工学系研究科	工学部3号館－工学部10号館	SM 2芯	廃止
8	情報理工学系研究科	工学部10号館－工学部13号館	SM 2芯	廃止
9	生産技術研究所	生研C棟－先端研4号館(駒場オープンラボ)	SM 2芯	新規
10	大学院工学系研究科	工学部2号館－情報基盤センター	SM 2芯	新規

表4　ネットワーク利用の申請部局・組織別申請件数

項番	申請部局・組織	山上会館	武田ホール	弥生講堂
1	21COE機械システム・イノベーション		1	
2	医学系研究科	1		
3	医学部付属病院	1		
4	海洋研究所	1		
5	経済学研究科	1		
6	研究協力部	1		
7	工学系研究科	6	8	2
8	産学連携本部			1
9	史料編纂所	1		
10	情報学環	1	2	1
11	情報基盤センター		1	1
12	情報理工学系研究科	2	1	1
13	新領域創成科学研究科	2		
14	素粒子物理国際研究センター	1		
15	大学総合教育研究センター			1
16	東洋文化研究所	1		
17	農学生命科学研究科		1	5
18	法学政治学研究科	1		
19	理学系研究科	4		
20	青山学院大学			1

セキュリティ対応

早野裕士　　佐野徹　　坂井朱美

1　運用報告

最近のネットワークおいては、高速な接続性の実現とともに、セキュリティへの対応も重要となっている。本学においてもセキュリティインシデントは日常茶飯事と言っても過言ではないほどに頻繁に発生している。特にコンピュータへの不正侵入の試み、電子メールや Web ページを通じてのウィルス感染は問題で、感染を受けた側の被害はもとより、学内や学外への不正アクセスの踏台となることも珍しくない。しかし、一方で、セキュリティを厳重にすることは利用者の利便性の低下にもつながる。一般論としていえば、セキュリティ対策の要点は安全性と利便性のトレードオフといえるが、本学のように多種多様な部局が存在している環境では、このトレードオフのバランス点を一つに収束させることは容易ではない。このような配慮のもと、本学のセキュリティの維持確保に向け、以下の通り、セキュリティ対応業務を実施した。

1.1 不正侵入の監視

本学内のコンピュータに対して不正なアクセスを試みている可能性を検出し、必要ならば当該コンピュータが収容されているネットワークの管理者と連絡をとって不正侵入を防ぐべく対応を行った。このような不正アクセスの検知には、以下の2システムを活用した。

・不正侵入検知システム (IDS :Intrusion Detection System)
　これは、通常のネットワークアクセスでは発生しないようなアクセスパターン (例えば多数のホストを順にスキャンしていく動作など)を対象として検知し、警告を発するシステムである。実際には、マカフィー社の IntruShield 4000を導入し、学内と学外との通信をタップしIDSに接続して監視した。

・異常トラフィック監視システム
　これは、サービス妨害攻撃またはサービス不能攻撃などと呼ばれる、インターネット経由での不正アクセスの 1 つで、大量のデータや不正パケットを送りつける行為などの不正な攻撃を検知できるシステムである。実際には、沖電気工業社製 Secure Traffic Probe(ソフトウェア)を導入し、学内と学外との通信をタップし異常トラフィック監視システムに接続して、ポート単位での異常なトラフィックの変化について監視した。

1.2 ウィルス対策

Windows パソコンやサーバマシン等における、ウィルスの脅威と感染被害を未然に防ぐ有効な手段として、コンピュータウィルス対策ソフトウェアがある。情報基盤センターでは、昨年度に引き続きトレンドマイクロ社の製品についてサイトライセンス契約を拡大継続することによって、学内への安価な配布サービスを推し進めた。また、Mac OS(Intel 版 Mac)対応として Sophos Anti-Virus for Mac、及び多言語対応(基本的には留学生向け)として、 Windows OS の簡体中国語と繁体中国語用の Sophos Anti-Virus for Windows の試行運用をあらたに開始した。コンピュータウィルス対策ソフトウェア配付の 2006 年度の申請状況は表1の通りである。

表1　サイトライセンスソフトウェアの申請状況（2006年度）

ソフトウェア名	申請件数	部局数	申請台数
ウイルスバスターコーポレートエディション	１３	８	６７２
ウイルスバスター2006（日本語版）	６２６	４０	２２，３５５
ウイルスバスター2006（英語版）	７０	２２	６９３
Server Protect	５３	２０	２３９
InterScan VirusWall	１５	１１	４２
InterScan VirusWallエンタープライズエディション	４	３	３０

2　講習会・研究会開催報告

（東京大学キャンパスネットワークシステム(UTnet3)の運用管理の 2. を参照）

東京大学情報システム緊急対応チーム（UT-CERT）との連携

早野裕士　　佐野徹　　坂井朱美

1　運用報告

情報基盤センターでは、UTnet3 の基幹ネットワークでセキュリティ対策を実施していることから、ネットワークを介したセキュリティインシデントについて、東京大学情報システム緊急対応チームUT-CERT(University of Tokyo-Computer Emergency Response Team)から以下の委託業務を受け、UT-CERT と連携協力しながら、全学的なセキュリティ対策を実施した。

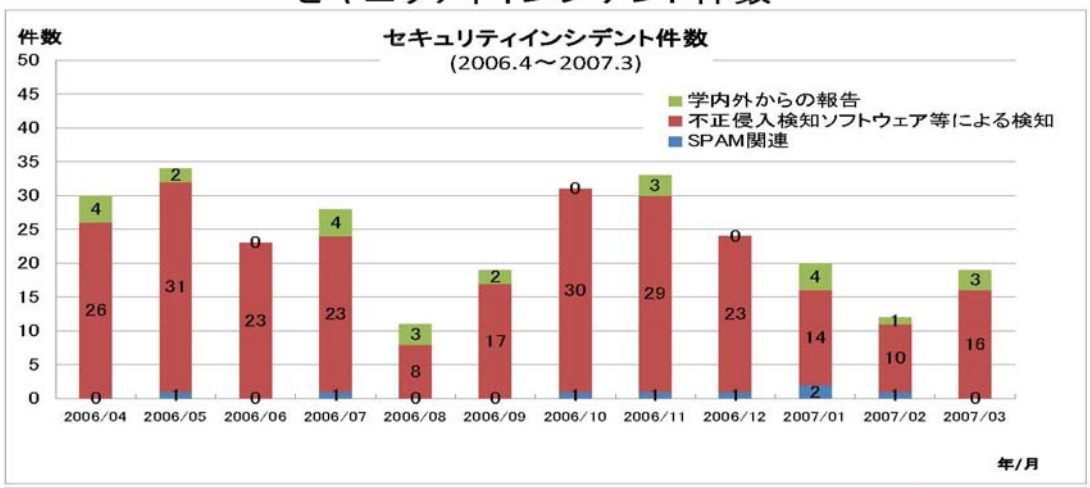
・セキュリティ関連情報の収集と学内への注意喚起等
　メーカやセキュリティ情報サイト等から、セキュリティの脆弱性やウィルスに係わる情報を収集し、情報基盤センターのweb ページへの掲載や部局CERTへのメールによって、セキュリティ対応に関し注意喚起を行った。
・学内と学外との通信におけるネットワークを介したインシデントの監視
　IDSや異常トラフィックシステムで学内と学外との通信を監視し、セキュリティインシデントの検知を行った。
・セキュリティインシデントの部局CERTへの連絡
　学内と学外との通信の監視やUT-CERTへの学内外からの連絡等によって判明したセキュリティインシデントについて、該当する部局CERT担当者へメールで対処を依頼した。また、インシデントレポートシステムでの報告を求めた。
・インシデントレポートシステムの運用管理
　部局CERTから、発生したセキュリティインシデントの内容や対処等を定型的な書式で報告できるように、インシデントレポートシステムを開発し、運用管理した。
・インシデントレポートの集計
　部居CERTから報告されたインシデントレポートについて、毎月集計しUT-CERT のホームページ（http://park.itc.u-tokyo.ac.jp/ut-cert/）に掲載した。（表1参照）
・セキュリティインシデントの当該IPアドレスのネットワーク遮断
　部局CERTに連絡したセキュリティインシデントのうち、発見時点から事象が連続している場合は、緊急措置として、当該IPアドレスのホストによる通信をUTnet機器でネットワーク遮断した。
・セキュリティインシデントの部局CERTの対策への連携協力
　部局CERTに連絡したセキュリティインシデントのうち、具体的な対処方法等について、部局CERTからUT-CERTに協力の依頼があった場合、UT-CERT及び部局CERTと連携協力して対応した。

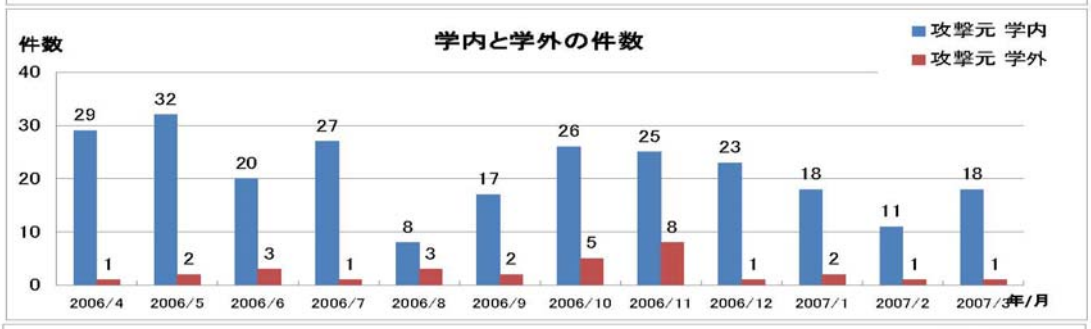
2　講習会・研究会開催報告

（東京大学キャンパスネットワークシステム（UTnet3）の運用管理の 2. を参照）

3　サービス統計

セキュリティインシデント件数





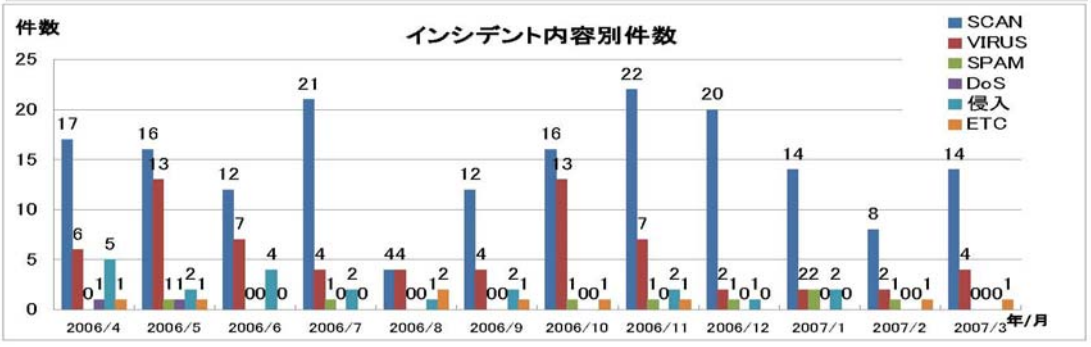


表１　2006 年度の学内のセキュリティインシデント件数

学内ソフトウェアライセンス

早野裕士　　佐野徹　　坂井朱美

1　運用報告

学内での研究教育の円滑な推進を達成することを目的として、利用が多いソフトウェアのうち、可能な範囲でサイトライセンス契約を提供業者との間で結び、安価でかつ容易に利用できるよう図っており、具体的には、以下の各種ソフトウェアの配布サービスとライセンスの管理を行った。

(1)ウィルス対策ソフトウェア
　Windows パソコン、ファイルサーバ、メールサーバ等のコンピュータウィルス対策ソフトウェア。

(2)Pro/ENGINEER
　3次元のCAD のソフトウェア。

(3)JMP
　統計解析ソフトウェア。

(4)SAS9
　統計解析ソフトウェア。

(4)MATHEMATICA
　数値計算や数式処理用のソフトウェア。

(5)Buzzsaw
　設計レビューと建設プロセスを自動化・標準化しながら、プロジェクトに関連するドキュメントや設計情報を一元管理することができるソフトウェア。

(6)Adobe CLP
　Adobe 社の Acrobat、Photoshop 等の製品が安価に購入できるライセンス。

学内共同利用３部門共通　講習会等

学内共同利用3部門共通 講習会等

情報メディア教育部門　　図書館電子化部門　　キャンパスネットワーキング部門

講習会・研究会開催報告

平成 18 年度東京大学技術職員研修（コンピュータ関係）

期　間
　平成 18 年 11 月 28 日（火）～平成 18 年 11 月 30 日（木）
目　的
　コンピュータ及びコンピュータネットワークの安全な利用方法と運用・管理方法を習得することによって、職員の能力・資質等の向上を図るとともに、東京大学のコンピュータ環境の維持向上に資すること。
対象者
　職員のうち、研究室やセンター等においてコンピュータやコンピュータネットワークの初心者レベルの利用者、及びそれらの運用管理業務やセキュリティ対策業務に携わる初心者レベルの管理者で、特にセキュリティ対策に興味のある者。
　なお、コンピュータを初めて使う者向けの講義「コンピュータの使い方」等は本研修に含まれない。
　注：研修名は「技術職員研修」となっているが、事務職員の方の参加も歓迎する。
会　場
　情報基盤センター4階講義室、5階演習室
研修内容・講師
（11 月 28 日（火））
　9:00 ～ 10:00　受付、開講式、オリエンテーション、自己紹介　若原教授
 10:00 ～ 12:00　講義　ネットワーク入門　加藤助教授
 13:00 ～ 14:00　講義　情報倫理とセキュリティポリシー　山口教授
 14:15 ～ 16:00　講義　ネットワークセキュリティ基礎　中村(誠)助手
 16:00 ～ 17:00　実習　ケーブル作成　佐島主任、坂井技術職員
（11 月 29 日（水））
　9:00 ～ 12:00　実習　wiki, blog, CFIVE　田中助教授、尾上講師、関谷(貴)助手
 13:00 ～ 14:00　講義　安全な通信について　田中助教授、関谷(貴)助手
 14:00 ～ 16:00　講義　安全なサーバ運用　尾上講師、関谷(勇)助手
 16:00 ～ 17:00　講義　迷惑メール・フィッシング対策　安東助手
（11 月 30 日（木））
　9:00 ～ 10:30　講義・実習　PKI システム　佐藤(周)助教授
 10:30 ～ 12:00　講義　SSL-VPN による学術情報データベースの学外利用
　　レファレンスサービスの自動化に向けた取り組み　清田助手
 13:00 ～ 16:00　講義　不正アクセスの実状とウィルス体験等　中村(文)助手
 16:00 ～ 16:30　見学　情報基盤センター2階ネットワーク機器室　早野係長
 16:30 ～ 17:00　レポート・アンケート作成、閉講式　米澤センター長、尾上講師

コンピューター・ネットワーク利用セミナー

本セミナーは、本学の教職員、学生等を対象としてコンピューターやネットワークを適切に利用するために必要な技術等に関する講習会で、学内共同利用3部門が協力して実施している。

第 52 回:オープンソース OS 公開講座～OpenSolaris のインストールから活用まで

日時

2006 年 6 月 21 日(水) 9:30 ～ 16:00

場所

本郷会場:情報基盤センター 4F 413 遠隔講義室他

定員

午前の部　約 50 名

午後の部　約 20 名（参加者個々に、インストール用の PC を用意して頂きます。）

概要

東京大学情報基盤センターでは、サンマイクロシステムズの協力を得て、標記セミナーを開催することになりました。

本セミナーでは、Sun のスペシャリストを情報基盤センターにお招きし、セキュリティと安定性に定評がある Solaris OS の説明及び個人 PC への Solaris のインストールを支援して頂きます。

午前のセッションで Solaris OS の新機能の説明と現場での活用方法について紹介して頂き、午後のセッションでは、セミナー参加者が個別に用意した PC を用いて実際に Solaris のインストール作業を行います。

また、インストール後には、Solaris 10 の新機能である Solaris コンテナの作成や Dtrace の使用法についての講習もご用意しております。

OpenSolaris にご興味のある方は、是非、この機会にご参加下さい。

（本セミナーは公開で行いますので、他大学の学生、教職員の方もご参加頂けます。）

内容

9:30- 9:45	受付
9:45-10:00	主催者挨拶等
10:00-10:30	初めての人のための OpenSolaris 入門
10:30-11:00	進化する UNIX --- OpenSolaris
11:10-12:10	先進的なオペレーティング機能のご紹介 ・OS のなかで OS を動かす --- Solaris コンテナ ・OS の挙動を把握/アプリケーションをデバッグする --- Dtrace ・Solaris コンテナ、Dtrace のデモ
Track1	
13:00-16:00	OpenSolaris 実機実習(インストールから活用まで) - 13:00-15:00 OpenSolaris をインストールしよう - 15:00-16:00 質疑応答、簡単な実習 ・ネットワークに接続してみよう ・Solaris コンテナを作成してみよう ・Dtrace を使ってみよう ・開発 Tool をインストールしてみよう

Track2(Track1 でインストール作業が早く終了した方が対象です)
14:00-15:00　OpenSolaris、Java の技術者支援 Tool のご紹介
・開発者コミュニティ(SDC)とは何か?
・Free の開発 Tool のご紹介- 開発ツールで何ができるか?

※ Track2 は、受講される方の予備知識やご希望に従って変更する場合がありますので、あらかじめご了承ください。

第 53 回:情報基盤センターサービスの紹介

日時

2006 年 10 月 25 日(水) 10:30 ～ 16:30

場所

本郷会場:総合図書館 3F 大会議室
駒場会場:情報教育棟 4F E44 室(テレビ会議システムによる中継、定員 10 名程度)

概要

このセミナーは、情報基盤センターが展開しているサービスについて分かりやすく紹介し、教職員、学生等の皆様に利用できるサービスを見つけて頂くためのものです。

特に今回は、教育用計算機、ネットワーク、図書館電子化などの話はもちろんのこと、全国共同利用のスーパーコンピューターの利用に関してや、研究プロジェクトとして進めている PKI による認証サービスの試行についての説明も予定しています。

内容

13:10-13:30　受付
13:30-13:35　開催の挨拶
13:35-14:05　教育用計算機と e ラーニング関係のサービスについて
　情報メディア教育部門　山口教授
14:05-14:35　スーパーコンピューティング部門の紹介
　スーパーコンピューティング部門　有賀係長
14:45-15:15　UTnet3 の概要／セキュリティ対策支援の概要／
ソフトウェアライセンスについて
　キャンパスネットワーキング部門　早野係長、中村係長
15:15-15:45　東京大学の認証サービスについて
　PKI プロジェクト　佐藤(周)助教授
15:45-16:15　東京大学学術機関リポジトリについて
　図書館電子化部門　鳴係長
16:15-16:30　質疑応答

第 54 回:平成 18 年度東京大学技術職員研修(コンピュータ関係)公開講義

日時・講師

(11 月 28 日(火))
10:00-12:00　講義　ネットワーク入門　加藤助教授
13:00-14:00　講義　情報倫理とセキュリティポリシー　山口教授
14:15-16:00　講義　ネットワークセキュリティ基礎　中村(誠)助手
(11 月 29 日(水))
13:00-14:00　講義　安全な通信について　田中助教授、関谷(貴)助手
14:00-16:00　講義　安全なサーバ運用　尾上講師、関谷(勇)助手
16:00-17:00　講義　迷惑メール・フィッシング対策　安東助手

（11 月 30 日（木））
9:00-10:30　講義・実習　PKI システム　佐藤(周)助教授
10:30-12:00　講義　SSL-VPN による学術情報データベースの学外利用
レファレンスサービスの自動化に向けた取り組み　清田助手
13:00-16:00　講義　不正アクセスの実状とウィルス体験等　中村(文)助手
場所
情報基盤センター4階講義室

第 55 回：IIJ の迷惑メール対策　～ここまでやる現場での対策～
日時
2007 年 2 月 6 日(火) 14:00 ～ 15:00
場所
情報基盤センター 4F 413 遠隔講義室
定員
約 40 名
講師
株式会社インターネットイニシアティブ　山本 功司　氏
協力
財団法人インターネット協会

スーパーコンピューティング部門

スーパーコンピューティング業務部門概要

本セクションでは，2006 年度のスーパーコンピューティング部門における業務概要について述べる。本部門の業務は，他の3部門と同様に，職員と教員の密接な連携によって円滑に実施されている。

2006 年度末に 2005 年 3 月導入のスーパーコンピューターSR11000 の中間レベルアップにより性能を約 3 倍に増強する作業が行われた。また別システムの 2001 年 3 月導入のスーパーコンピューターSR8000/MPP は 6 年間のレンタル契約が 2007 年 3 月で終了し，運用支援システムと共にその役目を終えた。なお，後継機については 2008 年 3 月導入に向け 2006 年度から仕様策定等の調達プロセスを開始した。仕様策定の前段階として，本センター教員が中心となって筑波大学及び京都大学と共同研究を行い，その成果について記者発表(2006.9.5)を行うと共に，公開ワークショップを開催(2006.9.6)した。この共同研究及び公開ワークショップの内容については，本稿スーパーコンピューティング研究部門で詳しく述べている。

本文では，2006 年度に行った SR11000 及び SR8000/MPP のサービス状況，次期スーパーコンピューターの調達関連の状況等について報告する。また，技術職員による研究開発活動についても併せて報告する。

スーパーコンピューティング業務

１．2006年度のシステム整備状況

2006 年度に行ったシステム整備を以下に述べる。利用者の利便性を高めるため，新たな利用環境等の整備及び次年度に向けてのシステム整備，並びに従来から継続してきた大規模（超並列）ジョブ実行サービスや教育利用への支援等を行った。また長年サービスしてきたオープンフロアサービスは，ネットワーク利用の拡大に伴い近年その利用が激減したため，施設有効利用の観点からスーパーコンピューティング専門委員会の議を経て 2006 年度末でサービスを終了した。

1.1 SR11000 システム増強

ベクトル並列型スーパーコンピューター（SR11000/J1）は中間レベルアップにより， 2007 年 3 月 25 日から 4 月 2 日にかけて 44 ノードモデルから 128 ノードモデル（いずれも 16CPU 構成）へ増強を行い，4 月 2 日からサービスを開始する。（増強後（128 ノード）のシステムを SR11000/J2 と表記する。） 性能を表1にまとめる。

表1. システム性能比較

	SR8000/MPP	SR11000/J1	SR11000/J2
ノード数	144 (8CPU)	44 (16CPU ※)	128 (16CPU)
1ノードあたりの 理論演算性能	14.4 GFLOPS	121.6 GFLOPS (16CPU ※)	147.2 GFLOPS (16CPU)
総理論演算性能	2,073.6 GFLOPS	5,350.4 GFLOPS	18,841.6 GFLOPS
1ノードあたりの 主記憶容量	16 GB	128 GB	128 GB
総主記憶容量	2,304 GB	5,632 GB	16,384 GB
ノード間ネットワーク 性能	(片方向) 1.6 GB/秒 (双方向) 3.2 GB/秒	(片方向) 12 GB/秒 (双方向) 24 GB/秒	(片方向) 12 GB/秒 (双方向) 24 GB/秒
磁気ディスク装置	4.7 TB	32.3 TB	94.2 TB
サービス終了・開始時期	2007 年 3 月 7 日 8:30 サービス終了	2007 年 3 月 25 日 8:30 サービス終了	2007 年 4 月 2 日 9:30 サービス開始

※SR11000/J1 では物理ノードを論理分割し，8CPU で構成する SMP を1ノードとして運用している。
これにより，1ノード当たりの理論演算性能，主記憶容量は 73.6GFLOPS，64GB となる。

1.2 SR8000/MPP，運用支援システム，オープンフロアのサービス終了

超並列型スーパーコンピューター（SR8000/MPP）は 2007 年 3 月 7 日，運用支援システム（OS:VOS3/FS-JSS4， HI-OSF/1-MJ）および本センター1階オープンフロアサービスは 3 月 6 日をもってサービスを終了した。

SR8000/MPP と m-unix の 2 月末時点での/home 上のファイルは，センターが SR11000/J2 に移行を行った。

1.3 SR11000/J1 ノード単体性能向上

2006 年 11 月末に SR11000/J1 のノード単体性能を 60.8GFLOPS（8CPU）から 73.6GFLOPS（同）へ性能向上を SR11000/J2 導入に先行して行った。また，OS およびコンパイラーのバージョンアップを行った。

1.4 SR11000/J1 personal キュー新設

2006 年 4 月 3 日より, SR11000/J1 においてデバッグ, テスト実行用キューとして personal キューを新設し, サービスを開始した。本サービスはパーソナルコースに登録している利用者のみを対象とした。

personal キューの各制限値は以下のとおり。

キュー名　　:personal(パイプキュー名:personal)
制限時間　　:30 分
仮想メモリー　　:標準 56GB(最大 56GB)
使用可能ノード数:2 ノードまで

1.5 SR11000/J1 P008 キュー及び open キュー試行サービス開始

それまで実施していた P008 キューの試行サービスを拡大し, 2006 年 6 月より SR11000/J1 にて 8 ノード並列ジョブクラス(P008 キュー)の常時サービスを開始した。また, 期間限定で全ての利用者が実行可能な 8 ノード並列ジョブクラス(open キュー)の試行サービスを開始した。サービスの詳細は以下のとおり。

表２．P008, open キューサービス

サービス名	P008 キューサービス	open キューサービス
バッチキュー名 （パイプキュー名）	P008 (parallel)	open (open)
制限時間	8 時間	1 時間
仮想メモリー	標準 56 GB (最大 56 GB)	
使用可能ノード数	8 ノードまで	
実施期間	常時 (ただし右記期間を除く)	毎月第 4 金曜日 9:00 ※1 ～ 翌月曜日 9:00 ※2
対象利用者	パーソナルコース 3 (SR11000 オプション 4 ノード) 及びグループコース利用者	全利用者

※1 P008 キューで実行中のジョブが終了次第, open キューサービスに移行。
※2 open キューで実行中のジョブが終了次第, P008 キューサービスに移行。

1.6 サービス状況

上記 1.1～1.3 に記載した 2007 年 3 月におけるシステム変更以外は変更を行っていない。2006 年度は 2005 年度のシステム提供形態を継続した。

SR11000/J1 のノード稼動率は, 年平均 82.2%, 最大月(11 月)で 87.3%, SR8000/MPP は, 年平均 73.7%, 最大月(12 月)で 78.4%に達し, ほぼ飽和状態であった。

1.7 超並列ジョブ実行サービス

SR8000/MPP の 128 ノードによる超並列ジョブ実行サービスも前年度に引き続き, 原則として毎月第一週の金曜日夕方から翌週月曜日早朝まで行った。処理件数は合計で 1,008 件であった。

1.8 教育利用

前年度に引き続き，教育利用のためのシステム提供(試行)を行い SR11000/J1 による 1 件の利用があった。

・東京大学理学系研究科地球惑星科学専攻にて大学院生向けの授業で，SR11000 を利用した並列プログラミング教育

1.9 参考

システム構成図を図1，図2に，ソフトウェア一覧を表3に示す。

図1．システム構成図(2006 年 4 月)

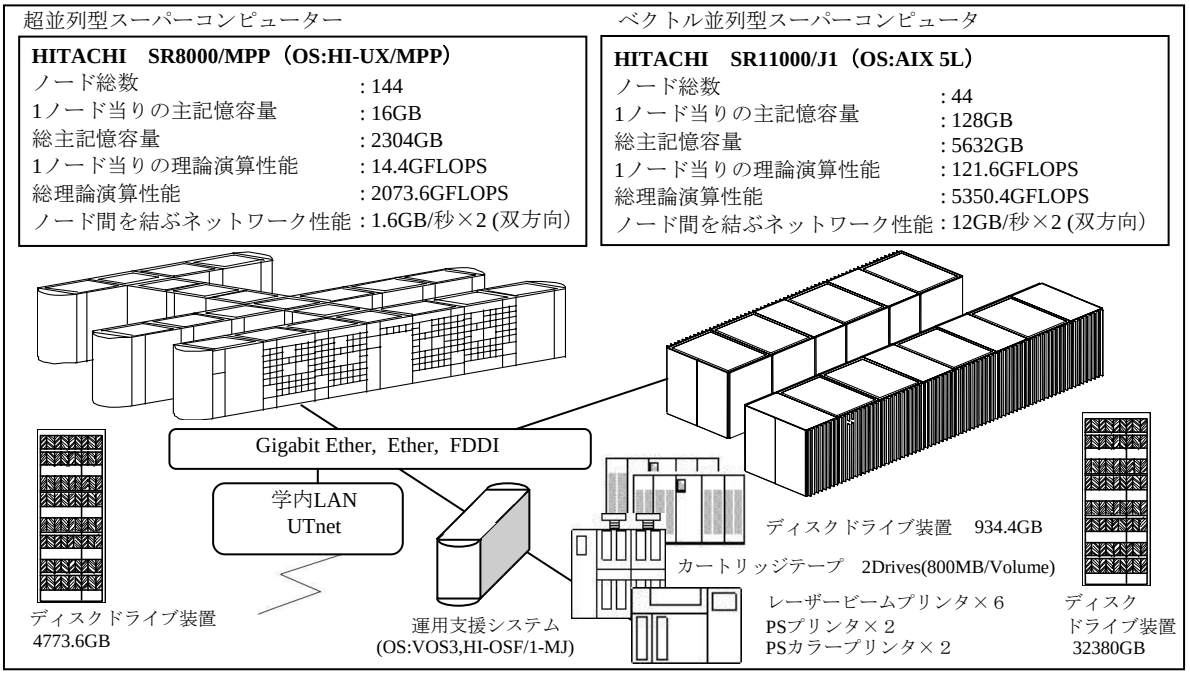


図2. システム構成図(2007 年 3 月末)

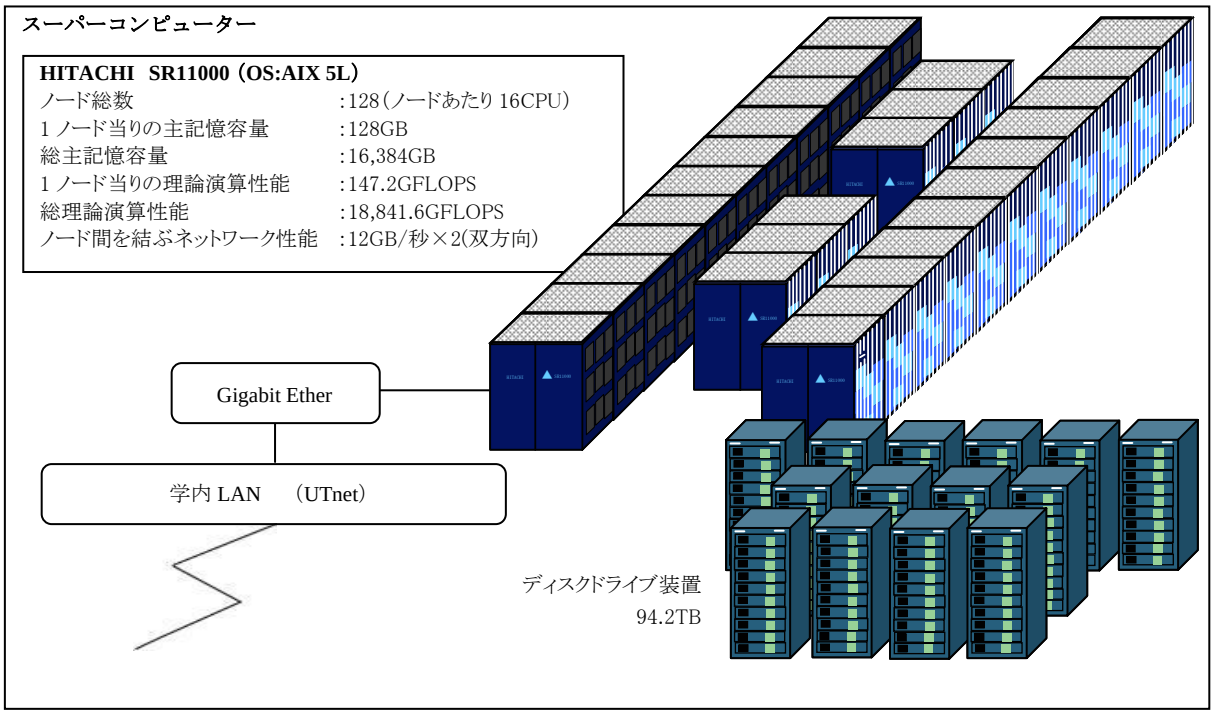


表3. ソフトウェア一覧

超並列型スーパーコンピューター	: HITACHI SR8000/MPP
オペレーティングシステム	: HI-UX/MPP システム
シェル (コマンドインタプリター)	: sh, csh
言語プロセッサー	: 最適化 FORTRAN77, 最適化 FORTRAN90, 最適化 C, 最適化 C++, 他
数値計算	: MATRIX/MPP, MATRIX/MPP/SSS, MSL2, BLAS, LAPACK, ScaLAPACK
分子計算	: GAUSSIAN98
並列化支援	: MPI(MPI-2), PVM, Parallel FORTRAN(HPF), PARALLELWARE(Express), リモート DMA 転送, 他
ベクトル並列型スーパーコンピューター	: HITACHI SR11000
オペレーティングシステム	: AIX 5L システム
シェル (コマンドインタプリター)	: sh, csh
言語プロセッサー	: 最適化 FORTRAN77, 最適化 FORTRAN90, 最適化 C, 最適化標準 C++※, 他 ※ 国際標準規格 ISO/IEC 14882:1998 準拠 及び 旧言語仕様 (ARM 仕様) 対応
数値計算	: MATRIX/MPP, MATRIX/MPP/SSS, MSL2, BLAS, LAPACK, ScaLAPACK
分子計算	: GAUSSIAN03
並列化支援	: MPI(MPI-2)

2. 2007年度のシステム整備計画

2007 年度のサービスに向けた計画を以下に述べる。SR11000 の中間レベルアップに伴うフェーズ 2（SR11000/J2）への移行計画，並びに SR11000/J2 のサービス開始及び SR8000/MPP 契約終了によるサービス終了に伴う計算サービス内容の変更，新たな利用者層に対するサービスメニューの追加等を，スーパーコンピューティング専門委員会の議を経て計画した。

2.1 SR11000 システム増強

ベクトル並列型スーパーコンピューター（SR11000/J1）について，2007 年 3 月 25 日から 4 月 2 日にかけて 44 ノードモデルから 128 ノードモデル（いずれも 16CPU 構成）へ増強を行い，4 月 2 日からサービスを開始する。

2.2 2007 年度のサービス変更

2.2.1 主な変更点

2007 年 4 月以降の SR11000 の利用に関する主な変更点は以下のとおり。詳細については次項以降を参照のこと。

- 提供システムリソースの拡大
- 「計算リソース追加オプション」の新設
- 「研究室コース」の新設
- 「利用負担金」額の改正

2.2.2 サービス変更点

SR11000 システムのサービスの変更点について以下に記載する。なお，システムのホスト名，IP アドレス，ファイルシステム名（/batch，/para-io）は変更ない。

① 16CPU/node による運用

これまでのシステムは 8CPU/node で運用してきたが，2007 年 4 月以降はシステムの大部分を 16CPU/node で運用する。これにより 16CPU/node で利用するノードにおいては，112GB/node のメモリーが利用できる。複数ノード利用のキューについては，この 16CPU/node を充てる。なお，1 ノード利用のキューとしては P001(16CPU/node)，H1(8CPU/node) 両方用意する。

② P016 キューの新設

これまでは並列用ジョブキューとして P008(8CPU/node) が最大だったが，P016(16CPU/node) キューを新設する。これにより演算量・メモリー量とも 4 倍のリソースが利用できることになる。なお，これまであった P002，open，personal キューは廃止する。

③ debug キューの新設

全ての利用者（研究室コース（後述）・教育利用（試行運用，無料）を除く）が利用可能な debug キューを新設する。4 ノード(16CPU/node)まで利用可能。

④ single キューの再編

single キューはこれまで 6 段階（A～F）だったが，これを4段階(A～D)に変更する。また，ES キューは廃止する。

⑤ 計算リソース追加オプション

計算リソース追加オプションとして，1･4･8 ノード専用キューを新設する。詳しくは「2.2.4. 計算リソース追加オプション」を参照のこと。なお，SR8000/MPP のグループコースで利用可能だったグループ用専用キューについては用意しない。

⑥ 大規模ジョブサービス

大規模ジョブサービスとして 64 ノード利用可能な P064 キューを月1回（原則として最終金曜日午後～翌月曜日朝）行う。このサービスは 5 月から開始する予定。

⑦ ファイルシステム

- /batch の上限値は現行どおり 10GB とする。グループコースは 1 グループ（利用者番号 10 個まで）で 200GB とする。追加オプションによりディスク容量を増やすことができる。
- /short/batch は廃止する。
- /para-io はインタラクティブから参照可能とする。（書き込みは NQS からのみとする。）
  保存期間は 5 日間のままとする。ファイル上限値設定は行わない。

ジョブクラス制限値は表 4 のとおり。

2.2.3 利用負担金

利用負担金について，基本的にはこれまでの定額制を継続する。主な変更点は以下のとおり。

- 現行のパーソナル･グループコースに加え，研究室コースを新設する。
- パーソナルコースでは 4 コースを，グループコースでは 2 コースを用意する。
- 追加オプションによる計算リソース（1･4･8 ノード専用キュー）を新設する。

利用負担金表は表 5 のとおり。

2.2.4 計算リソース追加オプション

追加オプションにより Q001(48 時間)，Q004(24 時間)，Q008(24 時間) キューが利用できるサービスを 2007 年 5 月から行う。制限時間が一般のキューP001(24 時間)，P004(12 時間)，P008(8 時間)より長く設定されている。また，人数制限されているためジョブが実行しやすくなっている。

① 利用見込み額

- ･利用申し込みのためには，利用見込み額の設定が必要。
- ･新規および継続利用申込書には，利用見込み額の記入欄がある。
- ･利用見込み額の追加は，Web ページまたは専用申込書（書面）により行える。
- ･利用見込み額として記入された金額が課金されるのではなく，②に示す Web による利用申し込みをした時点で課金対象となる。
- ･利用見込み額を超えた利用申し込みはできないが，利用見込み額を追加することで利用申し込みが可能になる。

② 利用申し込み

- ･利用申し込みは，本センタースーパーコンピューティング部門の Web ページ（http://www.cc.u-tokyo.ac.jp/）から行う。
- ・ 利用申し込みは，月単位とする。3ヶ月先まで予約可能とする。

・それぞれのキューについて各月の上限を10名とする。(複数のキューを利用したい場合はそれぞれ申し込みが必要。Q008を申し込めばQ001,Q004も利用できるわけではない。)

③ ジョブの実行

・NQSでのキューの指定は“#@$-q Q001”のようにキュー名を指定する。
・1利用者番号あたり投入可能なジョブは1とする。
・ジョブの実行はFIFO(First In First Out)とする。つまり,待ちキューに並んでいる順番に実行される。
・ジョブスケジューリングシステムにおけるCPU使用量の積算は行わない。

2.2.5 研究室コース

グループコースに登録するためには,10 人程度の利用者を集める必要があり,かつ 100 万円以上の利用負担金が発生していた。このため数人からなる小規模研究室では利用しにくい状況にあった。これらを解消するため,「研究室コース」を新設した。利用負担金表(表 5)の注意事項にある要件を満たすことにより3名以上で安価な利用負担金で利用できる。利用可能なキューについてはジョブクラス制限値(表 4)を参照。

2.2.6 参考

表4. ジョブクラス制限値

キュー名 ※3	CPU数	最大ノード数	制限時間	メモリー(GB) ※1	パーソナル				グループ		研究室コース
					コース1	コース2	コース3	コース4	コース1	コース2	
(single)											
A	1	−	1 時間	2(56)	○	○	○	○	○	○	○
B	〃	−	10 時間	〃	○	○	○	○	○	○	○
C	〃	−	50 時間	〃	○	○	○	○	○	○	○
D	〃	−	225 時間	〃	○	○	○	○	○	○	○
S1	8	1	2 時間	56(56)	○	○	○	○	○	○	○
S4	〃	4	2 時間	〃	×	×	×	×	×	×	○
H1	〃	1	24 時間	〃	×	○	○	○	○	○	×
(parallel)											
P001	16	1	24 時間	112(112)	×	○	○	○	○	○	×
P004	〃	4	12 時間	〃	×	×	○	○	○	○	×
P008	〃	8	8 時間	〃	×	×	×	○	×	○	×
P016	〃	16	6 時間	〃	×	×	×	○	×	○	×
P064 ※2	〃	64	2 時間	〃	×	×	×	○	×	○	×
debug	〃	4	5 分	〃	○	○	○	○	○	○	×
Q001	〃	1	48 時間	〃	×	△	△	△	△	△	×
Q004	〃	4	24 時間	〃	×	×	△	△	△	△	×
Q008	〃	8	24 時間	〃	×	×	×	△	×	△	×
インタラクティブ	1	−	18 時間	512MB (1GB)	○	○	○	○	○	○	○

△　追加オプションにより申込可

※1　1ノードあたりの標準値(括弧内は最大値)

※2 P064 キューは月1回のみサービス(原則として最終金曜日午後～翌月曜日朝)
※3 NQS でのキューの指定方法について
・single(A～D), parallel(P001～P064) 利用については, これまでどおり“#@$-q single”, “#@$-q parallel”と指定する。
・それ以外のキュー(S1, S4, H1, debug, Q001, Q004, Q008) については, キュー名をそのまま指定してする。(例:“#@$-q S1”,“#@$-q debug”)
・キュー名は, 大文字・小文字を区別するので, 注意が必要。

表5. 利用負担金表

コース			負担金額(税込)	/batch	備考
パーソナルコース		コース1	20,000 円／年	10GB	計算リソース追加オプションは不可
		コース2	80,000 円／年		P001 キュー
		コース3	200,000 円／年		P004 キュー
		コース4	400,000 円／年		P064 キュー
グループコース		コース1	1,500,000 円／年	200GB	P004 キュー 利用者番号 10 個まで
		コース2	3,000,000 円／年		P064 キュー 利用者番号 10 個まで
		追加オプション(1利用者番号につき)	150,000 円／年	+20GB	コース1用利用者番号追加
			300,000 円／年		コース2用利用者番号追加
研究室コース(1利用者番号につき)			50,000 円／年	10GB	S4 キュー 計算リソース追加オプションは不可
追加オプション	計算リソース	1ノード専用キュー	20,000 円／月		パーソナル2・3・4, グループ1・2 利用者が申込可
		4ノード専用キュー	50,000 円／月		パーソナル3・4, グループ1・2 利用者が申込可
		8ノード専用キュー	80,000 円／月		パーソナル4, グループ2 利用者が申込可
	ディスク		600 円/(GB*年)		GB 単位で申込可

- P001,P004,P064 キューは各コースにおける利用可能なキューで, 最大利用可能ノード数(1,4,64)を示す。その他及び詳細は, 表 4 ジョブクラス制限値を参照。
- 利用期間は 4 月から当該年度末サービス終了日までの 1 年間を基本とし, 年度途中で利用登録を取り消した場合であっても利用負担金額の変更は行わない。
- 利用負担金は, 原則として以下のとおりそれぞれ初旬に一括して請求する。
  - パーソナルコースは, 利用開始月が 4 月から 5 月までは 7 月, 6 月から 8 月までは 10 月, 9 月から 11 月までは1月, 12 月から 3 月までは 3 月。

  - グループコース・研究室コースは，利用開始月が 4 月から 6 月までは 7 月，7 月から 9 月までは 10 月，10 月から 12 月までは1月，1 月から 3 月までは 3 月。
  - 計算リソース追加オプションについては，申し込み月直後の7月，10 月，1月，3月。
  - 上記以外の月に請求を希望する場合は，別途相談に応じる。
- パーソナルコース（ただし，本センターのスーパーコンピューターシステムに初めて登録された利用者）においてのみ，利用開始月の翌月末日までに利用を中止することができる。利用負担金はパーソナルコースの利用開始月 3 月の金額を適用し，請求する。
- 利用負担金が減額となる変更はできない。
- コース間の変更については，利用負担金が増額になる場合のみ別途相談に応じる。（ただし，利用者番号変更の場合がある。）
- グループコースのディスク量は，グループ全体の上限値である。
- 研究室コースの申し込みは，以下の要件を満たす必要がある。
  - 同一研究室内にてグループを構成する。
  - グループは 3 名以上（教員・学生各 1 名以上）で構成される。（教員がパーソナルコース 2 以上に登録している場合は，学生 2 名以上のみの登録を可能とする。）
  - 登録人数の上限はない。
  - 教員がパーソナルコース 2 以上に登録している場合は，当該教員に対して S4 キューの実行を可能とする。
- 計算リソース追加オプションの申し込みは，月単位とする。それぞれのキューについて，各月の上限を 10 名とする。申し込み者の人数により，お断りする場合がある。利用申し込みは，スーパーコンピューティング部門の Web ページから行う。

3. 超並列型スーパーコンピューターシステムの調達

超並列型スーパーコンピューターシステム SR8000/MPP 契約終了に伴い後継機の選定を行っているところであるが，文部科学省を中心に進められている「次世代スーパーコンピューター」の動向を踏まえた仕様検討，システム規模の大幅な増強及びセンター設備の老朽化に伴い施設・設備改修が必要となることから，スーパーコンピューティング専門委員会の議を経て，次期超並列型スーパーコンピューターシステムの調達を当初 2007 年 3 月導入の予定から 1 年延期することとした。以下にこれまでの経緯等を述べる。

3.1 概要・日程

2006 年度中の日程は以下のとおりである。

(1) 第 5 回スーパーコンピューティング専門委員会（2006.6.30）
次期システム導入の延期，仕様検討にあたり 3 大学（筑波大学・東京大学・京都大学）で次世代スーパーコンピューターのアーキテクチュアーを検討する共同研究プロジェクトの承認。
(2) 記者発表（2006.9.5），ワークショップ開催（2006.9.6）
共同研究プロジェクトに基づき，9 月 5 日記者発表「オープンスパコン仕様策定へ」，9 月 6 日ワークショップ「大学の知を支援するセンターマシン」を開催。
(3) 第 6 回スーパーコンピューティング専門委員会（2006.10.3）
次期スーパーコンピュータの要求要件（概要）の承認，仕様策定委員及び技術審査委員の選任。
(4) 資料招請に関する公表（官報掲載，2006.10.30）
(5) 導入説明会の開催（2006.11.9）
資料及びコメントの受付期限（2006.12.11）
(6) 全システム構成を提案した業者に対するヒアリングの実施（2006.12.25）
(7) 第 1 回超並列型スーパーコンピュータシステム仕様策定委員会の開催（2007.2.8）
(8) 第 2 回超並列型スーパーコンピュータシステム仕様策定委員会の開催（2007.2.15）
(9) 第 7 回スーパーコンピューティング専門委員会（2007.3.5）
仕様策定委員会の仕様策定状況と仕様書原案の報告。
(10) 仕様書（案）説明会の開催案内状の送付（2007.3.12）
(11) 仕様書（案）説明会の開催（2007.3.23）
照会及び提案の申出・修正期限（2007.5.14）

3.2 資料招請

2006 年 10 月 30 日に官報公示，11 月 9 日に導入説明会を行った。導入説明会参加は 18 社 23 名であった。12 月 11 日までに資料の提出があったのは 13 社であった。このうちシステムの全体について資料提供のあった 5 社に対して，12 月 25 日にヒアリングを行った。

3.3 仕様策定

2007 年 1 月 5 日に仕様策定委員会を設置した。その後，2 月 8 日（第 1 回），2 月 15 日（第 2 回）に仕様策定委員会を開催し仕様書原案を策定した。この間にライブラリー利用状況を把握するため，本センターのスーパーコンピューター利用者に対しアンケート調査を行った。3 月 5 日の第 7 回スーパーコンピューティング専門委員会に報告の後，3 月 23 日に仕様書（案）説明会を行い，参加は 19 社 23 名であった。仕様策定委員会委員は以下のとおり。

氏　　名	所　　属　・　職　　名
◎　平　尾　公　彦	大学院工学系研究科　教授
木　本　昌　秀	気候システム研究センター　教授
加　藤　千　幸	生産技術研究所　教授
石　川　　裕	情報基盤センター　教授
佐　藤　文　俊	情報基盤センター　助教授
黒　田　久　泰	情報基盤センター　助手
菊　池　　力	情報基盤センター　事務長
大日方　一　男	情報基盤センター　専門職員

◎　委員長

4. 専門委員会

本センター運営委員会のもとにスーパーコンピューティング専門委員会を設置し，全国共同利用スーパーコンピューターシステム及びその運用に関する事項について，企画，立案及び審議を行っている。2006 年度に開催した同専門委員会の主な議事内容は以下のとおりである。

スーパーコンピューティング専門委員会

第 5 回 スーパーコンピューティング専門委員会

日　時: 2006 年 6 月 30 日(金)16:00～
議　題:・次期超並列型スーパーコンピューターの更新について
・共同研究プロジェクトについて
・平成 17 年度スーパーコンピューティング部門決算報告について
・各種報告(稼動統計，プログラム相談状況，教育利用)

第 6 回 スーパーコンピューティング専門委員会

日　時: 2006 年 10 月 3 日(火)10:00～
議　題:・次期超並列型スーパーコンピューターの更新について
・SR11000 の中間レベルアップ計画について
・SR11000 の来年度運用方針について
・プログラム指導員体制の見直しについて
・各種報告(全国共同利用情報基盤センター長会議，稼動統計，プログラム相談状況，教育利用)

第 7 回 スーパーコンピューティング専門委員会

日　時: 2007 年 3 月 5 日(月)10:00～
議　題:・SR11000 の来年度の運用につい
・スーパーコンピューターシステム利用規程の改定
・超並列型スーパーコンピューターシステムの導入(仕様書原案など)
・各種報告(稼動統計，プログラム相談状況，教育利用)

5. システムの利用状況

本部門は，学術研究及び教育等に供することを目的とし全国共同利用サービスを行っている。提供するシステムは，SR8000/MPP 及び SR11000/J1 のスーパーコンピューター，運用支援システムの汎用計算機である。2006 年度におけるジョブ統計は以下のとおりである。スーパーコンピューターの利用状況については実利利用者が 360 名を超え，特に性能の高い SR11000/J1 の実利用者数，計算件数及び演算時間共に増加している。

5.1 SR8000/MPP (HI-UX/MPP) のジョブ処理状況

年月	登録者数	実利用者数	TSS件数	NQS件数	接続時間(時間)	CPU時間		ﾉｰﾄﾞ稼動率(%)	/home ファイル使用量(MB)	/short ファイル使用量(MB)
						mpp-s(時間)	NQS(時間)			
0604	765	165	3,845	3,699	4,568.01	169.64	493,331.36	70.2	365,768	31,300
0605	775	188	3,981	2,976	4,517.38	204.81	437,768.84	75.0	360,053	39,086
0606	781	185	5,659	3,527	5,738.74	247.26	550,723.99	72.8	356,924	42,270
0607	794	168	4,996	3,432	5,321.14	235.60	556,661.79	73.0	371,109	50,283
0608	800	167	3,805	2,848	4,390.85	175.47	419,280.27	76.1	738,607	213,448
0609	806	167	5,024	3,900	5,497.60	265.28	581,945.92	75.7	733,423	250,692
0610	818	165	5,117	3,706	5,700.67	265.78	474,401.24	70.0	729,392	267,993
0611	836	172	5,131	3,785	5,809.60	207.73	553,091.94	73.5	737,559	263,786
0612	852	181	5,705	3,769	5,851.97	216.25	580,394.82	78.4	753,279	511,826
0701	855	164	5,004	3,437	6,115.17	229.37	602,681.24	76.9	712,216	369,607
0702	862	160	3,808	2,765	4,276.31	143.67	561,801.58	75.5	690,890	808,381
0703	862	112	642	598	868.70	42.52	72,298.04	44.2	673,329	766,372
合計	862	353	52,717	38,442	58,656.14	2,403.38	5,884,381.03			

(注) ・/short の値は /short/mpp-s と /short/mpp-bt の合計である。
・ノード稼動率とは 1 ノードが 100%稼動したと仮定した場合のサービスノードに対する稼動率である。
・3 月分は 7 日までの数値である。

5.2 SR11000/J1 (AIX 5L) のジョブ処理状況

年月	登録者数	実利用者数	TSS件数	NQS件数	接続時間(時間)	CPU時間		ﾉｰﾄﾞ稼動率(%)	/batch ファイル使用量(MB)	/short ファイル使用量(MB)
						sr11000-s(時間)	NQS(時間)			
0604	765	165	3,654	5,398	6,688.99	121.74	345,210.40	80.2	2,722,836	352,400
0605	775	159	3,442	4,465	5,793.06	122.86	261,436.74	71.3	3,672,226	336,652
0606	781	179	4,568	5,579	7,647.34	140.25	389,894.67	84.1	3,668,309	599,133
0607	794	184	4,328	4,528	6,285.12	113.45	398,923.26	83.5	3,778,564	797,203
0608	800	181	3,398	3,472	4,237.71	75.36	272,656.60	82.0	3,649,809	608,394
0609	806	192	4,053	4,249	5,032.21	138.46	390,329.74	83.6	3,598,319	775,789
0610	818	196	4,294	3,987	4,953.41	112.90	349,110.06	85.2	3,682,046	1,098,665
0611	836	210	4,873	5,512	6,324.99	220.91	415,610.29	87.3	3,580,149	1,511,616
0612	852	213	5,034	6,960	6,326.60	111.27	337,023.21	75.6	3,780,331	1,281,943
0701	855	213	5,081	5,220	7,056.83	116.19	394,933.53	82.2	3,949,321	1,251,088
0702	862	219	4,962	3,243	5,340.28	64.79	372,654.35	83.3	4,002,928	1,825,541
0703	864	212	4,847	3,148	4,737.55	62.13	326,047.20	81.2	4,087,691	1,177,992
合計	864	363	52,534	55,761	70,424.09	1,400.31	4,253,830.05			

(注) ・ノード稼動率とは 1 ノードが 100%稼動したと仮定した場合のサービスノードに対する稼動率である。
・3 月分は 25 日までの数値である。

5.3 運用支援システム (VOS3) の利用者数及びジョブ処理件数

年月	登録者数	実利用者数	総処理件数	一日平均処理件数	総CPU時間（時間）	一人あたりのジョブ件数	TSS 最大同時接続数
0604	394	30	868	31	17	29	4
0605	401	26	604	26	3	23	4
0606	403	26	705	24	2	27	4
0607	410	32	854	28	6	27	5
0608	413	33	2,075	99	17	63	10
0609	419	33	1,521	52	3	46	4
0610	430	28	745	30	8	27	3
0611	438	31	1,232	42	3	40	5
0612	442	34	1,013	34	4	30	5
0701	442	46	2,528	84	37	55	8
0702	447	66	2,848	105	13	43	8
0703	447	59	1,145	229	2	19	10
合計	447	119	16,138		115		

(注) ・総処理件数ならびに総 CPU 時間には TSS 及びバッチの処理件数, CPU 時間が含まれる。
・3 月分は 6 日までの数値である。

5.4 運用支援システム (VOS3) のTSS処理件数

年月	入出力形態別処理件数			時間		TSSの比率	
	インターネット	TSS	合計（件）	接続時間（時間）	CPU時間（時間）	件数比（%）	CPU 時間比（%）
0604	344	146	490	62	1	56.5	5.8
0605	179	119	298	35	1	49.3	33.3
0606	194	154	348	66	1	49.4	50
0607	241	245	486	138	5	56.9	83.3
0608	1,308	237	1,545	196	15	74.5	88.2
0609	989	181	1,170	116	2	76.9	66.7
0610	283	98	381	86	7	51.1	87.5
0611	482	116	598	131	1	48.5	33.3
0612	422	104	526	122	1	51.9	25
0701	1,254	151	1,405	327	7	55.6	18.9
0702	1,771	370	2,141	373	7	75.2	53.8
0703	792	120	912	166	1	79.7	50
合計	8,259	2,041	10,300	1,818	49		

（注） ・TSS の比率とは総処理件数及び総 CPU 時間に対する TSS の比率である。
・3 月分は 6 日までの数値である。

5.5 運用支援システム (VOS3) のバッチ処理件数及び入出力機器利用状況

年月	バッチ投入件数	CPU時間（時間）	バッチの比率		プリンター出力件数	磁気テープ利用件数
			件数比 (%)	CPU 時間比 (%)	センター内	
0604	378	16	43.5	94.2	107	2
0605	306	2	50.7	66.7	107	2
0606	357	1	50.6	50	140	7
0607	368	1	43.1	16.7	291	13
0608	530	2	25.5	11.8	182	0
0609	351	1	23.1	33.3	220	4,962
0610	364	1	48.9	12.5	261	0
0611	634	2	51.5	66.7	142	0
0612	487	3	48.1	75	136	2
0701	1,123	30	44.4	81.1	337	4
0702	707	6	24.8	46.2	393	74
0703	233	1	20.3	50	124	28
合計	5,838	66			2,440	5,094

（注） ・TSS の比率とは総処理件数及び総 CPU 時間に対する TSS の比率である。
・3 月分は 6 日までの数値である。

5.6 運用支援システム (HI-OSF/1-MJ) のジョブ処理状況

年月	登録者数	実利用者数	TSS件数	NQS件数	接続時間（時間）	CPU 時間		/home ファイル使用量 (MB)	/short ファイル使用量 (MB)	プリンター出力枚数
						TSS（時間）	NQS（時間）			
0604	187	19	175	0	266	0	0	5,909	0	134
0605	187	18	125	2	1,064	0	0	5,928	0	349
0606	187	18	119	0	589	0	0	5,988	0	225
0607	189	19	123	0	170	0	0	5,966	7	140
0608	189	17	97	2	94	0	0	5,968	8	43
0609	189	20	143	0	1,237	0	0	6,014	0	20
0610	189	19	113	0	748	0	0	6,051	0	262
0611	189	16	76	0	74	0	0	6,051	0	334
0612	189	16	109	0	444	0	0	6,038	0	511
0701	189	17	130	0	132	0	0	5,674	0	546
0702	190	23	170	0	380	0	0	5,631	0	364
0703	190	11	41	3	49	0	0	4,697	0	185
合計	190	42	1,421	7	5,247	0	0			3,113

（注） ・プリンター出力枚数は unix システム(m-unix, SR8000/MPP, SR11000/J1)の合計である。
・3 月分は 6 日までの数値である。

6. プログラム指導員と刊行物

スーパーコンピューターサービスの一環で，利用者にプログラム相談を行っている。プログラム相談は，センターオープンフロアに設置したプログラム指導員室において面談によるプログラム相談，電子メールによるプログラム相談，利用者の多い連絡所にもプログラム指導員を設置している。なお，来年度（2007 年度）については，オープンフロアサービスの終了等に伴い，スーパーコンピューティング専門委員会の議を経てプログラム指導体制の見直しを行う。

6.1 2006年度プログラム指導員

6.1.1 センター付きプログラム指導員

氏　　名	所　　属	職名等
尾崎　敬二	国際基督教大学・教養学部	教授
黒田　久泰	東京大学・情報基盤センター	助手

6.1.2 E-MAIL プログラム指導員

氏　　名	所　　属	職名等
岩井　浩一	茨城県立医療大学	教授
林　　隆史	会津大学・コンピュータ理工学部	教授
中村　文隆	東京大学・情報基盤センター	助手

6.1.3 連絡所付きプログラム指導員

氏　　名	所　　属	職名等
植田　毅	千葉大学・総合メディア基盤センター	助教授

6.2 プログラム指導員体制の見直し

2007 年度（2007 年 4 月）からプログラム指導員体制を以下のように変更する。

	現　行	変更後	備　考
指導員室 （センター１階）	指導員室での面談	事前予約による面談，センター教職員が対応	2007 年度からオープンサービスを廃止することに伴う
メール指導員	メールによる質問の回答	soudan@cc.u-tokyo.ac.jp は継続し，センター教職員が回答	
連絡所付指導員	各連絡所での面談	（継続）	

6.3 スーパーコンピューティング部門発行「利用の手引き」及び他の刊行物

2006 年度に刊行したものは次のとおりである。

スーパーコンピューティングニュース　Vol.8　No.3　～　Vol.9　No.2　(2006 年度)
スーパーコンピューティングニュース　Vol.8　No. Special Issue 1

以下はオンライン・マニュアル（VOS3: MANUAL コマンドでプリンターに出力）である。
オープン磁気テープシステム利用の手引
VOS3 ASPEN HELP（日本語）
VOS3 ASPEN MESSAGE（英語）
図形出力ライブラリ GPSL 利用の手引

7. 職員による研究開発活動

技術系職員は，日々利用者サービスの向上やシステム運用の効率化・高度化を図りながら業務に携わっており，その成果を発表することで更なる技術向上及びインセンティブ向上等の動機付けになるものと期待される。このような成果の発表の場として，全国共同利用情報基盤センター長会議のもとに開催している研究開発連合発表会講演会への参加，その他に各大学等の機関で行っている総合技術研究会にも積極的に参加している。

7.1 第 28 回全国共同利用情報基盤センター 研究開発連合発表講演会

大阪大学サイバーメディアセンターの当番で，第28回全国共同利用情報基盤センター研究開発連合発表講演会が 2006 年 11 月 28 日に開催され，本センターからも職員が参加し発表を行った。

本講演会は全国共同利用情報基盤センター長会議（全国共同利用の7大学センター及び情報・システム研究機構国立情報学研究所で構成）のもとに毎年開催しているもので，日々急激に進展するスーパーコンピューターを中心とした計算機，ソフトウェア及びネットワークに関する最新の技術について知識・サービスについて発表・討議を行い，情報基盤センター群相互の情報交換を図ることを目的としている。

＜本センターの研究開発論文＞
（東京大学情報基盤センター）
- 「スーパーコンピューターにおけるバッチジョブのスケジューリングについて」
  －田川善教(発表者)，石崎勉－
- 「東京大学認証局(UT-CA)構築に向けて」
  －大島大輔(発表者)，西村健，佐藤安一郎，佐藤周行－
- 「IC カードを使用したパスワード通知書発行システム」
  －宮嵜洋，小川大典－

＜他機関センター等の研究開発論文＞
（北海道大学）
- ビデオ会議を利用した遠隔授業およびキャリア支援への活用
- メールヘッダを利用したスパムフィルタとその結果について

（東北大学）
- Eduroam の導入と日本版アレンジ
- 新並列コンピュータシステムについて
- インシデント情報管理システムの運用性の向上に関する報告

（名古屋大学）
- 名古屋大学 ID 管理・運用ツールの開発
- Windows 操作の自動化について

（京都大学）
- コンテンツ作成支援サービスの試行について
- SSO の導入とその課題
- HPC の効率的運転とスケジューラ開発

（大阪大学）
- 大阪大学 Grid Operation Centre の取り組み
- 学術文献データベースの提供方法について

（九州大学）
・RSS を利用した九大ポータルサイトの構築
・XML データ処理による Web 用外国語学習教材作成・管理システム Web Drill の実装
（国立情報学研究所）
・NAREGI が拓くサイエンスグリッドの世界
・次世代学術情報ネットワーク(SINET3)について

第28回全国共同利用情報基盤センター研究開発連合発表講演会

スーパーコンピューターにおけるバッチジョブのスケジューリングについて

田川 善教，石﨑 勉
東京大学情報基盤センター

1. はじめに

東京大学情報基盤センター（以下，本センター）では2004年度からスーパーコンピューターの利用負担金を定額制としているが，これに伴いNQS（Network Queueing System）によるバッチジョブのスケジューリング方法をユーザー毎のシステム利用実績を加味したものに変更し運用している。ここでは，そのスケジューリング方法について説明する。

2. 背景

本センターでは表1に示すスーパーコンピューターシステムにより，全国共同利用サービスを行っている。利用形態としては，バッチジョブ利用とインタラクティブ利用の大きく2つに分かれる。このうちバッチジョブ利用については，１つまたは複数のノードを専有しての計算が主となる。利用にあたっては各スーパーコンピューターのインタラクティブシステムから表2に示すバッチキューにジョブを投入して計算を行う。このときバッチキューに並んだ各ユーザーのジョブをどのような順序で実行させるかが今回のテーマである。特にSR11000/J1では投入されるジョブが多く，実行を待たされることも多いため，本センターが公平なサービスを行っていく上で重要である。

表1 スーパーコンピューターシステムの性能

スーパーコンピューターシステム	SR8000/MPP	SR11000/J1
ノード数（ノードを構成するCPU数）	144（8CPU）	44（16CPU）
１ノードあたりの理論演算性能	14.4GFLOPS	121.6GFLOPS※
総理論演算性能	2,073.6GFLOPS	5,350.4GFLOPS
１ノードあたりの主記憶容量	16GB	128GB※
総主記憶容量	2,304GB	5,632GB

※SR11000/J1は物理的な1ノード(16CPU)を論理分割し1ノード(8CPU)×2として運用しているため，1ノードあたりの理論演算性能，主記憶容量は60.4GFLOPS，64GBとなる。

表2 ジョブクラス制限値（ノードを専有する主なバッチキュー）

バッチキュー名	ノード数	SR8000/MPP		SR11000/J1	
		制限時間	メモリー	制限時間	メモリー
P001	1	64時間	13.5GB	64時間	56GB
P002	2	32時間		32時間	
P004	4	16時間		16時間	
P008	8	8時間		8時間	
P016	16	4時間			

3. ジョブスケジューリング方式

基本となる考え方は，ユーザーに有効かつ公平な計算機資源の提供を行うことである。各ユーザーの CPU の使用量（時間）を平準化すること，およびユーザーが支払う利用負担金額に応じた CPU 資源を提供することが目的である。

本ジョブスケジューリング方式では，実行済みのバッチジョブで使用した CPU の使用量（以下，CPU 使用量とする）と，投入したジョブに設定した実行予定時間より算出される CPU の予定使用量（以下 CPU 予定使用量とする）から，各ジョブの優先度を決定する。

・バッチキュー毎に，実行待ちのジョブにそれぞれ優先順位（63 が最優先で 0 まで）を付け，順位の高いジョブから順に実行する。ユーザーはジョブの優先順位を qstat コマンドで確認することができる（以下は実行例）。

```
% qstat
NQS schedule stop time : 2006/09/29 (Fri)  8:30:00 (Remain: 497h 35m 27s)
   REQUEST      NAME      OWNER      QUEUE     PRI NICE   CPU    MEM    STATE
81947.n81       job_b     xxxxxx     P004      62  0   unlimit    56GB  QUEUED
```

→ PRI の値が 62 なので優先順位は第 2 位（63 が最高）。

・順位付けの要素は，年度当初（4 月）から現在までに実行されたバッチジョブの CPU 使用量をユーザー毎に積算した値と，投入したジョブで要求している CPU 予定使用量を合計したものであり，その値が少ないジョブを優先して順位付けを行う。CPU 使用量の初期値は 0 である。

・CPU 使用量および CPU 予定使用量の算出は以下のように行う。CPU 使用量についてはシステム（SR8000/MPP，SR11000/J1）別に，ユーザー毎に積算する。

CPU 使用量　　：ジョブの実行時間 × 8（CPU）× ノード数 × コース別係数※1
CPU 予定使用量：ジョブの実行予定時間※2 × 8（CPU）× ノード数

※1 コース別係数を表 3 に示す。スーパーコンピューターを利用するためのコースは，パーソナルコースとグループコースがあり，それぞれ係数を設定している。多くのノードを同時に使用可能な上位のコースは，下位のコースに比べて利用負担金額を高く設定しているが，その分 CPU 使用量の積算を軽減し，ジョブの実行に有利となるようにした。

※2 ジョブスクリプトに記述した値（#@$-lT XX:XX:XX）を使用する。省略した場合はバッチキューの制限時間が実行予定時間となる。

表 3 コース別係数

パーソナルコース		係数 (SR8000/MPP)
コース 1　(1 ノード)		1/1
コース 2	2 ノード	1/2
	4 ノード	1/4
	8 ノード	1/8
コース 3　(16 ノード)※		1/16

パーソナルコース／SR11000 オプション		係数 (SR11000/J1)
コース 2	1 ノード	1/1
コース 3	2 ノード	1/2
	4 ノード※	1/4

※グループコースは「パーソナルコース 3，SR11000 オプション 4 ノード」と同等

CPU 使用量の算出例

SR11000/J1 の P004 キューで，グループコースユーザーのジョブが 1 時間実行された場合

3,600（秒）× 8（CPU）× 4（ノード）× 1/4（コース別係数）＝ 28,800

・それぞれのシステムで，実行しているジョブが終了する毎に，各ユーザーの CPU 使用量を積算し，優先順位の付け替えを行う。

・大規模計算を行うユーザーが不利とならないよう，一定時間（10 分）毎に CPU 使用量に一定の逓減率（0.9993）を掛ける。これにより時間の経過と共に優先度が回復する。なお，この逓減により CPU 使用量は 1 週間で約半分になる（$0.9993^{6\text{回}\times 24\text{時間}\times 7\text{日}}=0.493690204$）。積算も含めた CPU 使用量の推移は図 1 のようになる。なお，バッチキューに並んでいる際の待ち時間は，順位付けの要素に含まない。

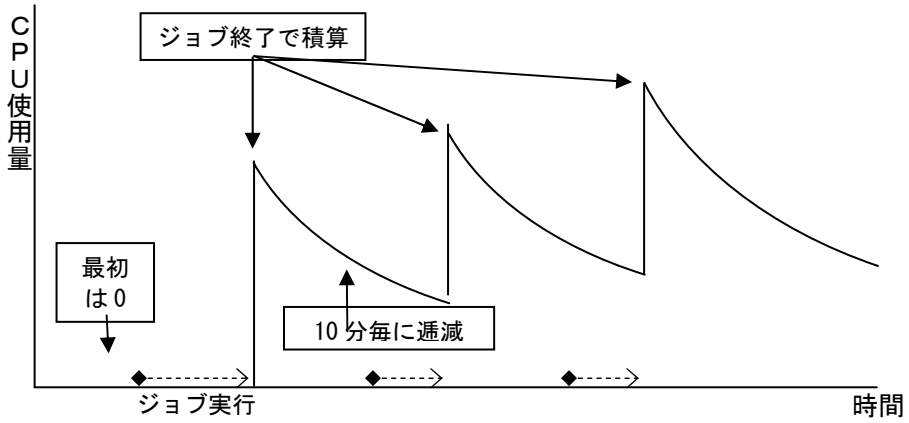


図 1 CPU 使用量の推移イメージ

4. 実現方式

NQS がもつ従来の機能に加え，以下の 3 つの機能を作成することで，本スケジューリング方式を実現した。

・ジョブ情報収集　　各バッチキューで待ち状態のジョブについて情報を収集する。

・CPU 使用量管理　　ユーザー毎の CPU 使用量について計算を行う。NQS の Terminator から起動の場合は CPU 使用量の積算を，cron から起動の場合は CPU 使用量の逓減を行う。

・優先度変更　　各ジョブの CPU 使用量＋CPU 予定使用量を算出し，優先度の値を更新する。NQS の Terminator から起動する。

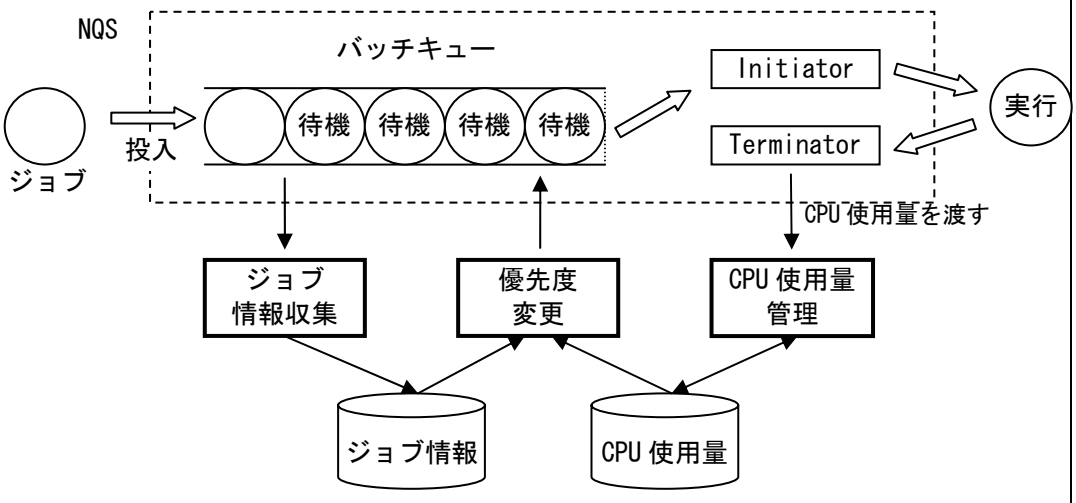


図 2　ジョブスケジューリング機能の構成

5. 評価

ここでは，3 人のユーザーがそれぞれ複数のジョブを投入した場合に，スケジューリング方式によってどのような実行順序になるか比較を行ってみる。

以下の各ユーザーは同一のコースとし，同一のキュー（P001）にジョブを投入する。全てのジョブが投入されたところでジョブの実行を開始する。ジョブは 1 本ずつ順に実行され，CPU 使用量の逓減については，説明を簡単にするため考えないものとする。ジョブの実行時間は，実行予定時間と同じとする。

ユーザー	ジョブ（実行予定時間）		
A	A-1（10 時間），	A-2（10 時間），	A-3（10 時間）
B	B-1（5 時間），	B-2（10 時間），	B-3（7 時間）
C	C-1（1 時間），	C-2（5 時間），	C-3（10 時間）

ジョブの投入（ → の順）

A-1，A-2，A-3，B-1，B-2，B-3，C-1，C-2，C-3

各スケジューリング方式での実行順

・FIFO（先入れ先出し）

A-1，A-2，A-3，B-1，B-2，B-3，C-1，C-2，C-3

投入順と同じ。

・e-time 平等化（以前の方式）

A-1，B-1，C-1，A-2，B-3，C-2，A-3，B-2，C-3

投入ユーザー順（A-B-C の順）に,実行予定時間の短いジョブから順に，平等に実行機会を与える。

・現方式

C-1，B-1，C-2，A-1，B-3，C-3，A-2，B-2，A-3

最初は CPU 使用量が全員 0 で実行予定時間の短い C-1 が最初に実行される。その後は「CPU 使用量＋CPU 予定使用量」の最も少ない順に実行される。

FIFO では投入順にジョブが実行されるため，ひとりのユーザーが一度に多くのジョブを投入した場合など利用が偏る問題がある。仮に投入ジョブ数を 1 とすることで，実行ジョブ数は各ユーザーで平均化されるが，利便性が損なわれるため，全国共同利用には向いていない。

e-time 平等化では実行されるジョブ数は各ユーザーで平均化されるが，CPU 使用量については考慮していないため，やはり利用に偏りが生じる。

現方式ではこれまでの CPU 使用量が考慮されるため，各ユーザーで CPU 使用量が平均化される。これは本センターが目指すところの有効かつ公平な計算機資源の提供に沿うものと考えられる。今回の例では各ユーザーのコースを同一としているが，コースが異なる場合はコース別係数(表 3 参照)が異なり，上位のコースほど CPU 使用量の積算が軽減されるため，より実行され易くなる。利用にかかる金額にも見合うスケジューリング方式となっている。

表 4 はそれぞれのスケジューリング方式において 6 番目のジョブが終了した時点の各ユーザーの実行時間の累計である。FIFO や e-time 平等化に比べて，現方式の値が最も平均している。実行時間累計の最大値と最小値の差は，全てユーザーA と C の差で，FIFO の場合は 30 時間，e-time 平等化では 14 時間，現方式では 6 時間である。

実行されるジョブ数についてユーザーC を見てみると，FIFO の場合は 0 個，e-time 平等化の場合は 2 個であるが，現方式の場合は 3 個という結果になっている。これはユーザーC のジョブの実行時間が他のユーザーのジョブに比べて短いものが多く，優先的に実行された結果である。

表 4　6 番目のジョブが終了した時点の各ユーザーの実行時間累計（時間）

ユーザー	FIFO	e-time 平等化	現方式
A	30	20	10
B	22	12	12
C	0	6	16
（平均値）	17.3	12.7	12.7

※ユーザーのコースが同一であり，コース別係数が同じ値となるため，上記の値は CPU 使用量の比と一致する。

図 3 は実際のユーザー毎の CPU 使用量累計と P004 キューにおける待ち時間累計の分布を示したものである。図中の実線（1 次回帰直線）が示すとおり，CPU 使用量が増えるに従い待ち時間も増えていることがわかる。なお，ここでは CPU 使用量の逓減については考慮せず，CPU 使用量の累計を用いている。

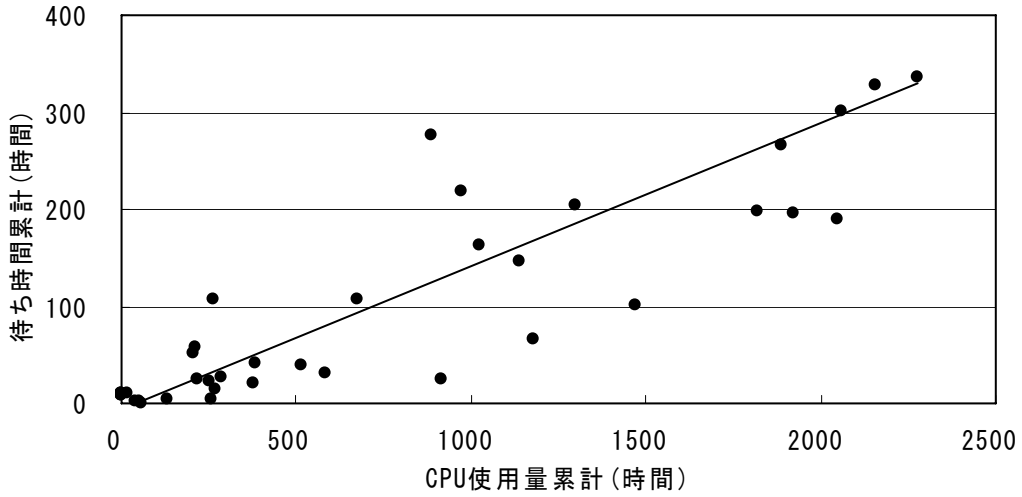


図 3 ユーザー毎の CPU 使用量累計と P004 キューにおける待ち時間累計の分布（2006 年 4 月，SR11000/J1）

現方式においては，同じような結果を求める計算をするのであれば，実行時間（実行予定時間）を短くすることで，次回以降のジョブや当該ジョブの実行に有利となる。そのため，チューニングされた効率の良いプログラムが必須となる。これは品質のよいプログラムを作ることへの動機付けになると思われる。

6. 今後の課題

以下の検討項目がある。

・逓減率，コース別係数の検証
逓減率は各ユーザーの優先度の回復に，コース別係数はコースの内容および金額とのバランスに係わってくる。そのため，これらの値が適切かどうかを検証する必要がある。

・グループコースへの制限
現在グループコースにおいても CPU 使用量の積算はユーザー単位でしか行っていない。当該ユーザーへの積算以外に，同じグループコースの他のメンバーにも若干の積算行うなど，グループ利用を意識した処理が必要かどうか。

・CPU 使用量のリセット時期
各ユーザーの CPU 使用量は年度当初（4 月）のみ 0 にリセットしているが，これ以外にも定期的なリセットや，閑散期や繁忙期など利用状況によるリセットを行うかどうか。

・年度途中で申し込みをしたユーザーの扱い
年度の途中で利用を申し込んだ場合，CPU 使用量の初期値は 0 である。これは，特に年度当初から継続利用しているユーザーに対しては非常に有利な状況となる。これを解消するためには，年度途中で申し込みをしたユーザーに対して CPU 使用量を初めから増やしておくことなどが考えられるが，その必要があるかどうか。

この他，今回のスケジューリング方式に限らないが，多数のジョブが実行されている環境では一度に大量のノードを確保することが難しくなるため，適度な間隔で大規模ジョブを実行するようなスケジューリング方法の検討が必要である（バッチキューをまたがるスケジューリングなど）。

7. おわりに

バッチジョブのスケジューリングは，そのときのシステムの運用やサービス，予想される利用状況などを考慮する必要がある。

特に本センターのスーパーコンピューターシステムの場合，全国共同利用であり，また定額制でもあることから，多数のユーザーがジョブを投入し利用している。実際，システムの CPU 利用率は概ね 80% 前後であり，CPU の稼動状態はほぼ満杯（実質的な頭打ち状態）である。ユーザーはジョブ投入から実行開始まで長時間待たされることもある。そのため，ユーザーには今回のスケジューリング方式について，ウェブページやスーパーコンピューティングニュース等で説明を行っている。

現状ではジョブスケジューリング方式に対する苦情は特に寄せられていないが，これに満足することなくユーザーが納得できるサービスを行えるよう，また今後の課題となっている機能の提供等を含めより公平で充実したサービスを行えるよう，努めていきたい。

謝辞

今回の開発にあたり，多大なるご協力をいただいた株式会社日立製作所の長谷部浩樹氏に深謝いたします。

PKI プロジェクト

PKIプロジェクト

プロジェクト概要

PKI プロジェクトは、CA（認証局）を安定的に運用するための運用技術を、発足以来重要項目として検討してきたが、いくつかの試行錯誤を経て、UT-CA としてキャンパス PKI のための CA を実際に構築し、運用技術を蓄積するために、インソースでの運用でのプロトタイプを構築した。現在では、UT-CA の実験協力部局は情報学環、情報基盤センター・本部合同になり、新年度は新たに UTnet を仮想的な部局として実験を継続できる運びとなった。

PKI プロジェクトのもうひとつの目標は PKI の需要調査である。特に有力なアプリケーションとして、SSL-VPN を取り上げ、そのテスト運用を行った。特に図書館電子化部門・分散教育メディア研究部門と協力して図書館アクセス認証のゲートウェイとして SSL-VPN を採用し、テスト的に運用を行っている。

1　運用報告

1.1　PKI プロジェクトの実施

情報基盤センターは、2005 年 1 月、PKI に関連する基盤的な技術を蓄積し、あわせて研究を進めるためにセンター内に PKI プロジェクトを立ち上げた。この背景には、東京大学が 2004 年秋、2005 年度からの職員証、学生証の IC カード化に伴い、キャンパス PKI を実現する予定であることを公式にアナウンスしたことがある。また、情報サービスの多様化にともない、特に認証に関してセキュリティを強化する必要に迫られているものが出てきたことも背景にある。

PKI プロジェクトは、CA（認証局）を安定的に運用するための運用技術を、発足以来重要項目として検討してきた。現在、キャンパス PKI の運用については、インソースでの運用、アウトソースでの運用を含めてさまざまなことが国内外の大学で試行されているが、その経験を直接接木できるものでもない。東大のような規模を持ち、また地理的な展開状況にあり、さらに運用管理について「部局の自治」に関して特色ある文化をもつ大学において、最適解は何かということは決して自明ではない。

PKI プロジェクトは、いくつかの試行錯誤を経て、同年 10 月、キャンパス PKI のための CA を実際に構築し、運用技術を蓄積するために、インソースでの運用でのプロトタイプを構築することを決定した。これが UT-CA である。CA の運用においてキーとなるのは、RA（登録局）の運用である。UT-CA は、分散 RA アーキテクチャをとり、各部局で行われている事務フローへ RA 業務を溶かし込むことができるのではないかと考えている。そのためのシステムを設計した。納入を経て 2006 年 7 月にはデモンストレーションを行い、さらに 2006 年 11 月には、情報学環の協力を得て実証実験を開始することができた。多くのシステムでそうであるように、PKI においても、実際に運用を行う際に新たな問題を発見することがある。PKI プロジェクトが発足当時、同プロジェクトでは、PKI の理論はすでに構築され、残りは運用の経験を積むだけであると考えて

いた。しかし、実証実験を企画するにあたって、全学レベルで展開するには運用面でも解決を要する問題が存在することを実感した。

現在では、UT-CA の実験協力部局は情報学環、情報基盤センター・本部合同になり、新年度は新たに UTnet を仮想的な部局として実験を継続できる運びとなった。分散 RA 方式の評価はこれからだが、東大が今後認証局を全学展開するときには有力なアーキテクチャになると考えている。

PKI プロジェクトのもうひとつの目標は PKI の需要調査である。PKI が構想され、市場的にも大きな位置を占めるようにはなってきたが、具体的なアプリケーションのご利益は広く実感できていないというのが現状である。この状況に対して、では潜在的な需要はどこにあるのか、現在利用可能なアプリケーションで、PKI が実現したらすぐにでも利用可能になっているものにはなにがあるのか、の調査を行い一部実地で評価してきた。現在、PKI プロジェクトで把握している需要には次のものがある。

1. S/MIME　——　情報学環の協力を得て実験中。このために UT-CA プロファイルの変更を行っている。
2. Web 認証　——　学内で、強い認証を要求するサーバが増えているため。
3. SSL-VPN 認証　——　簡易的なポータルとして利用するのに適している。さらに、学内専用のサーバに外からアクセスするためのポータルとしても利用価値が高い。実際図書館で運用している SSL-VPN サーバは（証明書認証ではないものの）人文社会系の利用が大幅に増えている。
4. ディジタル署名　——　事務のワークフローの電子化を考える人たちからアイデアが出ている。
5. アプリケーションごとの PKI の統制　——　無線 LAN の認証など、アプリケーションごとの認証をおこなう簡易的な PKI を統制するのはテーマの一つであるが、これについては話が始まったばかりである。
6. その他。　——　データの暗号化など。

S/MIME は、情報学環内部で秘匿性の高い情報のやりとりに使われている。SSL-VPN は、サーバ機器を購入し、複数部局に配備して、モニタをしてもらっている。残りについては、需要は確実に見込めるものの、ソフトウェアに対するそれなりの投資が必要になると考えている。われわれは、デモンストレーションの場で解決を必要とするものとしてレガシーアプリケーションを PKI の傘下におくためのコスト問題である「サイボウズ問題」と、秘書等、権限が委譲されていくことが本質的である組織において認証と認可をどう組み合わせるかの問題である「秘書問題」をあげた。これらに解を与えるための研究も進めている。

もうひとつ述べる必要があるのは、コンサルティングを含めた外部との連携である。PKI はプロジェクト開始時にはすでにビジネスとして確立し、そこで行われている慣習の理解は必須であった。さらに国立情報学研究所が 2005 年度から実施している UPKI は、大学間でキャンパス PKI を連携させることを構想している。PKI プロジェクトは、UT-CA の構築当初からコンサルティングを導入し、また UPKI へ参加することにより、これら外部的な動きを正しく理解することで、UT-CA でのさまざまな成果が業界のスタンダードの中で位置づけられるように努力してきた。

1.2 UT-CA の運用時系列

UT-CA の構築に関するイベントを時系列であげる。

2005.11 UT-CA 設計　発注
2006.03 までコンサルティングを受けた
2006.07 UT-CA お披露目デモ
2006.10 S/MIME 対応用にプロファイル変更。主要 MUA 対応確認
2006.11 情報学環 RA 開設
2006.12 情報センター・本部合同 RA 開設
2006.12 運用管理規定 (CP/CPS) 制定
2007.01 学内広報に IA 拇印公開
2007.03 運用の一年延長決定
2007.04 運用の検査検討と予備検査

2007 年 3 月現在、登録利用者数 117、証明書発行数 61 になっている。

1.3 PKI アプリケーションの立ち上げ

PKI の有力なアプリケーションとして、SSL-VPN を取り上げ、そのテスト運用を行った。SSL-VPN を PKI 対応にした上で、センター内のみでサービスしているサーバに接続する運用モデルを構築した。さらに UT-CA の実証実験の一部としてセンターのほかに本部事務と情報学環にサーバを設置し、サービスを提供した。

SSL-VPN 自体は、アクセスポイントとして PKI とは独立に運用されるべきものである。図書館システムでは、学内 IP アドレスでフィルタリングをかけざるを得ないデータベースシステムが存在する。これらのシステムのうち、ライセンス上の問題が発生しないものについて、SSL-VPN をアクセスポイントとして利用することを計画した。2006 年度の運用データの一部を掲載する。ここでは、特にいわゆる「文系」の利用が非常に多いのが観察できる。

部局名	検索回数の合計
教養学部 (前期課程)	2153
教養学部 (専門課程)・総合文化研究科	7373
文学部・人文社会系研究科	10751
法学部・法学政治学研究科	6020
教育学部・教育学研究科	2526
工学部・工学系研究科	1035
公共政策学教育部	2000
その他	3766

1.4 CSI UPKI プロジェクトへの係わり

NII が主導している CSI UPKI プロジェクトが開始した「オープンドメイン証明書発行実験」に参加し、パブリックサーバ証明書発行のための学内の体制を構築した。2007 年度から本格的にパブリックサーバ証明書の配布を開始する。

2 講習会・研究会開催報告

2006 年 7 月 28 日に UT-CA のデモンストレーションを兼ねて「UT-CA デモ報告会」を開催した。

2.1 広報その他

PKI プロジェクトの Web ページは `http://www.pki.itc.u-tokyo.ac.jp/`である。

2.2 業務報告

研究開発をのぞく業務報告として、以下の発表を行った。

1. 大島大輔, 西村健, 佐藤安一郎, 佐藤周行: 東京大学認証局 (UT-CA) 構築に向けて, 第 28 回全国共同利用情報基盤センター研究開発連合発表講演会, 2006 年 11 月.
2. 大島大輔: 東京大学認証局 (UT-CA) 構築と普及向けて, 平成 18 年度名古屋大学総合技術研究会, 2007 年 3 月.

PART 4

その他

平成１８年度委員会委員等

教員名	委員会委員等名	任　期
山口和紀	東京大学保健センター 運営委員会委員	17.4.1～19.3.31
	東京大学 情報セキュリティ委員会委員	17.11.1～19.3.31
	東京大学 国際交流委員会専門委員	18.4.1～19.3.31
田中哲朗	社団法人 情報処理学会 会誌編集委員会委員	18.4.1～20.3.31
	社団法人 情報処理学会 論文誌編集委員会編集委員	18.6.1～20.5.31
尾上能之	東京大学本郷地区交通対策専門部会 浅野キャンパス代表者	18.7.13～20.3.31
安東孝二	財団法人 インターネット協会 迷惑メール対策委員会委員	18.4.1～19.3.31
	CTC アカデミックユーザアソシエーション 運営委員会委員	18.4.1～19.3.31
	トヨタ IT 開発センター 技術アドバイザー	18.4.1～19.3.31
	農林水産省関東農政局 事業総合支援システム研究委員会チームリーダー	18.8.8～19.3.23
関谷貴之	社団法人 情報処理学会 会誌編集委員会委員	18.4.1～20.3.31
中川裕志	日本学術会議 情報学研究連絡委員会委員	15.10.21～18.10.20
	日本認知科学会 運営委員	17.1.1～18.12.31
	富士通総研 研究会委員	17.12.22～18.7.31
	東京大学 キャンパス計画委員会本郷地区部会委員	18.4.1～20.3.31
	東京大学駒場図書館 運営委員会委員	18.4.1～20.3.31
	東京大学柏図書館 運営委員会委員	18.4.1～20.3.31
	社団法人 情報処理学会 自然言語処理研究運営委員会主査	18.4.1～20.3.31
	アジア太平洋機械翻訳協会 特許翻訳研究会委員	18.5.24～19.3.31
	財団法人 国際情報化協力センター 調査委員会委員長	18.10.23～19.3.10
二宮　崇	東京大学本郷地区交通対策専門部会 浅野キャンパス代表者	18.4.1～18.7.12
清田陽司	言語処理学会 年次大会プログラム委員会委員	18.9.26～19.4.30
若原　恭	社団法人 電子情報通信学会 執行委員会ソサイエティ副会長	17.5.28～19.5.27
	社団法人 電子情報通信学会 国際委員会委員	17.5.28～19.5.26
	社団法人 電子情報通信学会 研専運営会議委員	17.5.28～18.5.27
	社団法人 電子情報通信学会 特集号編集委員長	17.7.25～18.4.1

	東京大学 ポータルサイト構築検討委員会委員	17.12.1～19.3.31
	東京大学 キャンパス計画委員会柏地区部会委員	18.4.1～20.3.31
	東京大学 国際交流委員会専門委員	18.4.1～19.3.31
	名古屋大学 情報連携基盤センター運営委員会委員	18.4.1～20.3.31
	国立情報学研究所 学術情報ネットワーク運営・連携本部委員	18.4.1～19.3.31
	国立情報学研究所 学術情報ネットワーク運営・連携本部 ネットワーク作業部会委員	18.4.1～19.3.31
	社団法人 電子情報通信学会 小特集編集委員会委員長	18.4.3～19.4.1
	社団法人 電子情報通信学会 功績賞・業績賞委員会委員	18.10.12～19.2.19
中山雅哉	総務省関東総合通信局 関東JGNⅡ懇話会委員	16.11.16～20.3.31
	社団法人 情報処理学会 研究運営委員会委員	17.5.12～19.3.31
	社団法人 電子情報通信学会 編集委員会編集副委員長	17.5.28～18.5.27
	東京大学 情報セキュリティ委員会委員	17.11.1～19.3.31
	独立行政法人 宇宙航空研究開発機構 宇宙利用推進本部通信・測位利用推進センター主幹開発員	18.4.1～19.3.31
	社団法人 電子情報通信学会 研究専門委員会専門委員	18.5.27～19.5.25
	国立情報学研究所 学術情報ネットワーク運営・連携本部 情報セキュリティポリシー策定作業部会委員	18.8.1～19.3.31
	総務省情報通信政策局 評価検討会構成員	19.2.19～19.3.31
加藤　朗	独立行政法人 情報通信研究機構 研究開発プロジェクト特別研究員	18.2.13～20.3.31
	独立行政法人 情報通信研究機構 次世代高度ネットワーク推進会議メンバー	18.3.6～20.3.31
	慶應義塾大学SFC研究所 上席所員（訪問）	18.4.1～19.3.31
中村文隆	社団法人 電子情報通信学会 ネットワークシステム研究専門委員会委員	16.5.29～18.5.28
	法政大学 非常勤講師	18.4.1～18.9.30
関谷勇司	慶應義塾大学SFC研究所 上席所員（訪問）	18.4.1～19.3.31
金田康正	東京大学気候システム研究センター 研究協議会委員	17.4.1～20.3.31

	東京大学物性研究所 スーパーコンピュータ共同利用委員会委員	18.4.1～20.3.31
	東京大学物性研究所 スーパーコンピュータ共同利用課題審査委員会委員	18.4.1～20.3.31
	京都大学学術情報メディアセンター 全国共同利用運営委員会委員	18.4.1～20.3.31
	大阪大学サイバーメディアセンター 運営委員会委員	18.4.1～20.3.31
	鶴岡工業高等専門学校 非常勤講師	18.4.12～19.3.31
	国立環境研究所 スーパーコンピュータ関連研究ステアリンググループに係る委員	18.9.6～19.3.31
佐藤周行	北海道大学情報基盤センター 全国共同利用委員会委員	18.4.1～19.3.31
	国立情報学研究所 学術情報ネットワーク運営・連携本部 認証作業部会委員	18.4.19～19.3.31

東京大学情報基盤センター年報

第８号

２００６年度

（２００７年６月発行）

編集　東京大学情報基盤センター年報編集委員会

発行　東京大学情報基盤センター

〒113-8658　東京都文京区弥生 2-11-16

電話　03（5841）2710

印刷　よしみ工産株式会社