

## コンテンツ容量から見た情報モビリティに関する検討

中川晋一(通信総合研究所/北陸先端大学院大学) 杉浦一徳、井上潤(通信総合研究所)  
木村美恵子(タケダライフサイエンスリサーチセンター) 土池政司(ネクサス)  
E-mail: [snakagaw@crl.go.jp](mailto:snakagaw@crl.go.jp)

### 要約

近年大容量ハードディスク等の低価格化大容量化が加速している。発表者らはわが国におけるWebコンテンツの容量推定を行った。超大容量通信回線を維持するコストと一時的な情報到達性からモデルを比較し、ユビキタス情報モビリティのための検討を行った。

キーワード: ネットワークリソースマネジメント、多階層ネットワーク構造、ユビキタスネットワーク、次世代インターネット

### Information Mobility and Transport Model based volume of Internet Contents

Shin-ichi Nakagawa (CRL/JAIST), Kazunori Sugiura, Jun Inoue(CRL),  
Mieko Kimura (Takeda Life-Science Research Center), Seiji Tsuchiike (Nexus Corp.)

### Abstract

Recently, portability and cost of the hard disks for the mobile computer is getting more cost-effective. The authors tried to estimate the surface Internet Web's volume of contents at 2002<sup>nd</sup>. From that result, it is suggested that the more cost effective model would be required among the data transport cost and information connectivity as the ubiquitous information network resource management.

Key words: Network Resource Management, Multi Layered Network, Ubiquitous Networking, Next Generation Internet

### はじめに

インターネットホームページ(Web ページ)は、1995 年頃から一般化し現在インターネットにおける最も一般的な情報検索収集手段の一つとして認識されている。中島らは2002年2月わが国のインターネットコンテンツ統計を行い[1]、Web コンテンツの総量はサーバ数 197000、ページ数 65,550,000、ファイル数 173,880,000、データ量 5,002 ギガバイトであると推定した。また、同報告で近年 1998 年調査開始時サーバ数 36,000、データ量 305 ギガバイトであり、ページ数の増加が 1998 年から 99 年にかけての増加率が約 3 倍、サーバ数が 2 倍に増加したと報告した。Lyman は 2001 年、全世界人口一人当たり約 250 メガバイト、全世界合計で 1 から 2 エキサバイトのユニークな情報がインターネットの情報量として

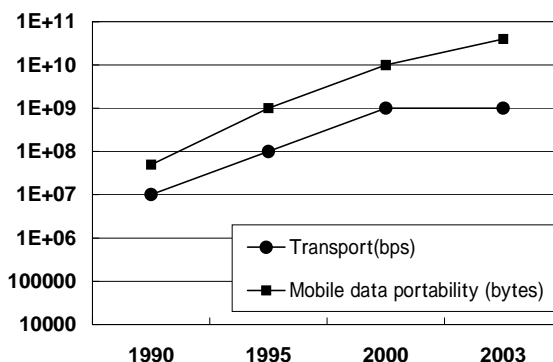


Fig.1: Trends of Internet Data transport rate and Mobile data portability

生産されていると推定した。前者はオートパイロットを用いた実測値からの推定であり国民一人当たり約 0.05 メガバイト、後者は社会学的基礎データからの推定値であり 250 メガバイトである。パーソナルコンピュータ用ハードディスクの容量は、1 年で

約 2 倍の伸びを示し、ラップトップコンピュータのハードディスク容量は 1996 年当時 2 G バイトであったが、現在 80 ギガバイトまで向上しており、今後、情報モビリティの観点からインターネットのコネクティビティがある時点で必要とする情報を事前にダウンロードして持ち歩く事も可能になる可能性がある。1990 年から 2003 年におけるネットワーク伝送性能とパーソナルコンピュータ用ハードディスク容量の最大値を Fig 1 に示した。今回発表者は、教育機関と医療機関を対象として情報モビリティの可能性を検討する為、2 都道府県（合計人口 750 万人）地域に存在する機関の発信するインターネット Web ページ全量の計測を行い、情報モビリティの検討を行い、情報モビリティとユビキタスネットワークの実装に関する基礎的検討を行った結果を報告する。

## 方法

### 1. 教育機関ホームページ全数調査

文部科学省学校基本調査[4]に基づいて県下/府下の公立学校数を母数とした。また、個々の学校名称のリストは、県/府教育委員会や各市町村役所や教育委員会のホームページから母数との照合を行いつつ作成した。学校名称を元に、研究目的で K12 サイトのポータルサイトを運営している URL[5]、民間検索エンジン（Yahoo, Infoseek, Google 等）に作成した学校名称リストを直接打ち込んで検索、サーチ結果

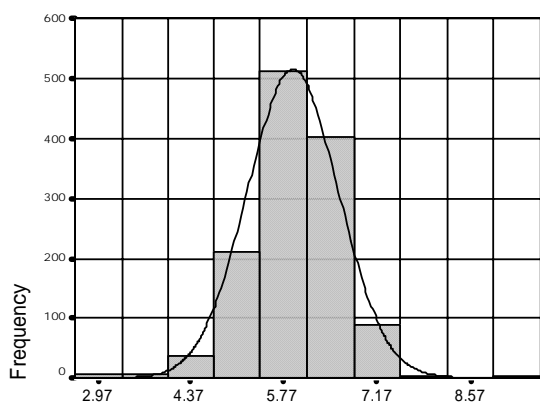


Fig.2: Frequency of Web contents Volume (log scale) of 1264 schools in H and K pref. at 2002.

から一般名称として学校のホームページ以外で用いられている場合を手作業で除外、該当するホームページを探索し URL を同定した。両府県とも大学での開設率は 100%であった。2002 年、国内 2 つの府県（555 万人の H 県と 220 万人 K 府）の教育機関全数（H 県、K 府あわせて 2308 機関）の調査を行い、ホームページの存在した 1264 個の URL ツリーを解析した。

### 2. 医療機関ホームページ全数調査

2002 年、厚生労働省の作成した全国医療機関名簿から H 県と K 府の有床診療所以上の医療機関（病院）のリストを作成し、既存の民間検索エンジン（Google）を用いてそれぞれの病院名の検索結果リストを該当する機関のページの検出が行なえるまで一つ一つクリックし、その機関のホームページが存在した場合、trace-route し運用形態を検索する一方、該当するサーバから全ファイルをダウンロードして容量とファイルの構成を検索した。

## 結果と考察

教育機関容量：全体（1264 URL）コンテンツデータ容量の頻度分布を Fig. 3 に示した。府県別、4 種（小中高大）の教育機関種別でも同様に指数分布であり、URL ツリーの情報量がこの対象では指数関数分布であること、Tab 1 に示すように小学校から高校まで教育機関の Web コンテンツの平均値は 1

Pref.	Category	N of C	Mean(byte)	S.D.(bytes)	Sum(bytes)
H	E	301	1.8E+06	3.0E+06	5.5E+08
	JH	179	1.7E+06	3.3E+06	3.0E+08
	H	181	2.1E+06	3.1E+06	3.8E+08
	U	39	1.5E+08	6.5E+08	5.9E+09
	Total	700	1.0E+07	1.5E+08	7.2E+09
K	E	290	2.4E+06	4.5E+06	6.8E+08
	JH	147	1.3E+06	1.6E+06	1.8E+08
	H	99	2.2E+06	3.5E+06	2.1E+08
	U	28	1.0E+07	2.8E+07	2.9E+08
	Total	564	2.4E+06	7.3E+06	1.4E+09
H+K	E	591	2.1E+06	3.8E+06	1.2E+09
	JH	326	1.5E+06	2.7E+06	4.9E+08
	H	280	2.1E+06	3.2E+06	6.0E+08
	U	67	9.3E+07	5.0E+08	6.2E+09
	Total	1264	6.8E+06	1.2E+08	8.5E+09

Tab.1: Statistics of Web contents volume of each category of education in two pref..

機関あたり約2メガバイト、大学で93メガバイトであり、大学を加え1264機関では全量で8.5ギガバイト、平均6.8Mバイトであった。H県の1大学が合計1ギガバイトを超えるデータ容量を示したが、今回は1機関あたりのWebによる情報発信量の平均値を求めることを目的としたため、解析からは外さず採用した。また、以前の報告では人口当たりの情報発信量の平均値を求める事を目的としたため、ホームページを持たないものを0バイトとして解析に採用したが、今回は1機関あたりの情報発信量を推定する目的からホームページを持たない機関をMissing Valueとして解析から除外した。H,Kそれぞれの教育機関種別の1URLツリーの含むWebコンテンツのデータ量の平均標準偏差ならびにそれぞれの種別の合計容量を表2に示す。また、各教育機関種別のコンテンツ容量を比較するため、H,Kそれぞれにおいて小学校(ES),中学校(JH),高等学校(HS),大学(Univ.)を独立因子とする1元配置分散分析を行った。その結果、ES,JH,HSの三者では等分散性がなかったのに対してUnivが他の教育機関種に比べ、有意にコンテンツ量が多いことがTukey, Scheffe両方で $p<0.05$ で示唆された。

医療機関容量：H,K両府県に存在する医療機関合計538に対する調査の結果をTab.2にまとめた。全医療機関の約50%がドメイン名を保持していた機関293のうち、実際にインターネットから到達可能であったものは、290であった。Webコンテンツの容量は $1.3 \pm 2.3$ メガバイトの指数分布を示した。全Web容量の約85%はGIF, JPEGのファイルであり、ファイル数における画像ファイルの割合75%に比べ、高値であった。両府県における医療機関のWebページ保持率はカイ2乗検定で有意にHが高く、Webによる情報発信の地域性存在の可能性も示唆された。

以上のことから、教育機関ならびに病院におけるWebデータコンテンツ容量はそれぞれ多く見積もって小中高で5メガバイト、大学で100メガバイト、病院が5メガバイトであるとする、学校数6万(文部科学省学校基本調査による)、大学1290、病院1万機関として合計480ギガバイトであることが推定された。今後、書き換え頻度、それぞれのWebページのリンクの深さとコンテンツ内容の質の検討は必要ではあるが、約500ギガバイトのディスク容量でこのカテゴリーのWebページが可搬できることが示唆された。

	H pref.	K pref.	total
Own domain and home page	132	99	231
Provided home page	48	15	63
Not exist	172	72	244
total	352	186	538

	No of case	Mean	St. Dev.
Numbers of Hits	290	32.1	59.1
Total Volume(bytes)	290	1377632.0	2300315.5
Total HTML file volume(bytes)	290	220853.4	350982.5
Total contents volume except for HTML (bytes)	290	1156778.7	2032658.0
Total volumes of GIF files (bytes)	290	448047.4	723440.3
Total volumes of JPEG files (bytes)	290	613617.9	1260710.4
Total numbers of files	290	129.6	164.4
Numbers of HTML files	290	31.7	43.0
Numbers of files except for HTML	290	97.9	126.3
Numbers of GIF files	290	65.1	92.8
Numbers of JPEG files	290	31.0	46.8

**Tab. 2: Summary of Web contents volume survey of Hospitals in H and K prefecture at 2002.**

	10MB	100MB	1GB	100GB	1TB
54Kbps	1.5E+03	1.5E+04	1.5E+05	1.5E+07	1.5E+08
64Kbps	1.3E+03	1.3E+04	1.3E+05	1.3E+07	1.3E+08
128Kbps	6.3E+02	6.3E+03	6.3E+04	6.3E+06	6.3E+07
1.5Mbps	5.3E+01	5.3E+02	5.3E+03	5.3E+05	5.3E+06
8Mbps	1.0E+01	1.0E+02	1.0E+03	1.0E+05	1.0E+06
22Mbps	3.6E+00	3.6E+01	3.6E+02	3.6E+04	3.6E+05
100Mbps	8.0E-01	8.0E+00	8.0E+01	8.0E+03	8.0E+04
1Gbps	8.0E-02	8.0E-01	8.0E+00	8.0E+02	8.0E+03

Tab. 3: Calculated Elapsed Time (second)of Transport for each contents volume by each Bandwidth

Tab. 3 に計算上のコンテンツデータ伝送容量（横軸）をそれぞれの帯域（縦軸）で伝送した場合の所要時間の計算値（秒）を示した。これによれば、500G バイトのデータ伝送を行うのに帯域 100 Mbps を要しても 10 時間以上必要であり、現在のインターネットでは望むべくもないが、実際にこの容量のデータを例えば、ハードディスクのコピーを行なうなどのデータ伝送系（ハードディスクのハードコピーを実際の運送手段を用いて配布する）を従来のインターネットと別に想定し、Web キャッシュとして運用することによって、全てのデータコンテンツへのダウンロードコストが軽減できる可能性も示唆される。

発表者らは、1998 年から APEC - APII テストベッドと通信放送機構研究開発用ギガビットネットワーク（JGN）を用い、ストリーム型データ、特に高品質動画像（MPEG2、デジタルビデオ、D1）を対象としたリアルタイムデータ伝送を IPv4 コ

ニキャスト、IPv4 マルチキャスト、IPv6 ユニキャスト、IPv6 マルチキャストにおいて行い、リアルタイムコミュニケーションに関する検討を進めてきた[3-9]。これら実験は原則的にアドホックに実験ネットワークを構築し、全てのセットアップを現地で行うというものであり、コストが大きく、ネットワークのユビキタス環境向上の必要性を実感させるものであった。これらの実験を行った APII テストベッドと研究開発用ギガビットネットワークは、世界各国の NGI（次世代インターネット）テストベッドを対象として発表者らが行った G8-GIBN Survey of Acceptable Use Policy により、Multi-Layer2、原則として各 Layer2 ネットワークパスに対しては帯域保証を行わず、Fig. 3 に示すように Layer3 のポリシーをユーザサイドに委ねるという運営が行われてきた。[10,11]。

JGN 開設当初はそれぞれのユーザグループの PVC, PVP の設定要求を帯域保証コミットメントの判断でパスの設定を行なうことも検討されたが、すでに合衆国 vBNS 等で行なわれていた運営方式でありオペレーションの省力化のため、UBR(Unassigned Bit Rate)でのパスの設定を行い、Layer3 ユーザが開示されているその時点での Layer2 ネットワークの ATM セルロスデータから Layer3 ネットワークの伝送品質を推定して実験を行なう運用形態をとった。Fig.4 のような放送型ストリームデータは接続タイプ以外のインターネットサービスの通信形態とは異なり、Constant bit rate

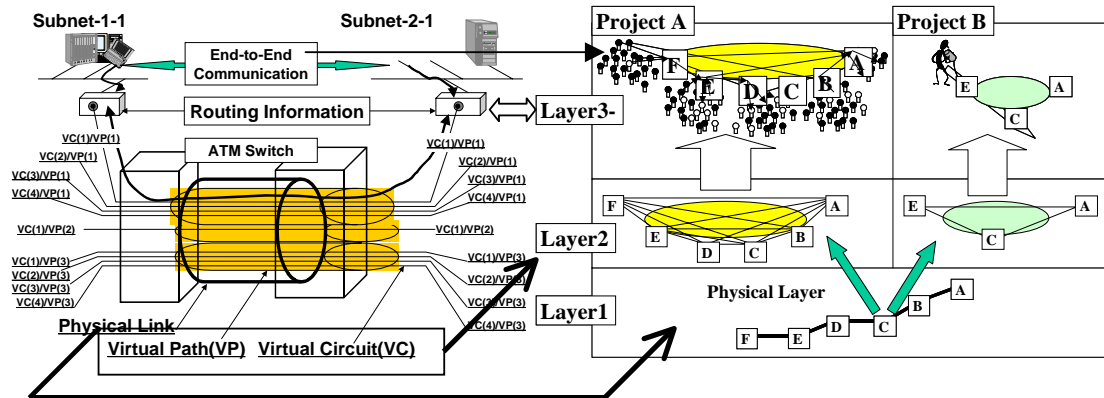


Fig. 3 Open Test-bed Network Concept based on the Layer 2 AUP.

を要求するものであり、一般のインターネットサービスとの混在をお互いに望まないトラフィックであることは周知であるが、今後このようなストリームデータ、Webのような配布型のデータ、DNSのようなインターネット通信を行う上においてリアルタイムでの通信が不可欠である場合など、それぞれのサービス内容によって情報配布の

バリエーション（例えば全量をダウンロードし、必要時に認証するまでそのコンテンツはオペレーションシステムからアクセスできない等の機構）も考えられる。今後、これらの情報配布、データ伝送コスト、利便性、可搬性の適切さを検討する必要があると思われる。

<http://www.sims.berkeley.edu/research/projects/how-much-info>

## まとめ

2002年、合計人口700万人のH、K両府県においてWebによる情報発信を行なっている教育機関（小中高大：1264機関）ならびに医療機関（290機関）を対象としてWebコンテンツの全数調査を行った。その結果、多くともそれぞれの機関のWebコンテンツ量は小中高で5メガバイト、大学で100メガバイト、病院が5メガバイトであり、この領域のWebコンテンツの総量は全国で約500ギガバイトであることが示唆され、インターネットWeb情報を全量可搬することが不可能ではない可能性が示唆された。

## 謝辞

本研究を行うにあたり御助力いただいた通信総合研究所諸氏、武田病院諸氏ならびに通信放送機構諸氏に深謝する。なお本研究は厚生科学研究がん研究助成金「超広域ギガビットネットワークを用いたがん情報収集配布に関する研究」（がんセンター若尾班）独立行政法人通信総合研究所情報通信部門研究費ならびに通信放送機構公募研究「IPv6技術を用いた健康栄養情報のメタデータ流通に関する研究」により行った。

## 参考文献

[1] 中島睦晴, 島田博也. インターネットコンテンツ統計に関する調査研究,  
<http://www.iptp.go.jp/research/monthly/2002/168-h14.09/168-asearch2.PDF>, 2002

[2] Peter Lyman, Hal R. Varian, James Dunn, Aleksey Strygin, Kirsten Swearingen, "How much Information",

[3] S.Nakagawa, Y. Kitamura, T.Fujii, H.Jiang, T.Shioimi, J.Ong, R.Seng and S.Isobe, QoS Evaluation Method for Stream Data Transport with ICMP - An Experiment Networking and Telemedicine Demonstration at APEC TELMIN3 -, Proceedings of IEEE ICOIN-13,11C-1.1-1.6 1999

[4] 植月修志, 小峯隆宏, 町澤朗彦, 杉浦一徳, 勝本道哲, 中川晋一, 久保田文人, 「DV over IP伝送映像の品質評価」, 電子情報通信学会コミュニケーションクオリティ研究会, 信学技法, pp.77-82, 2000.5

[5] 原田 雅博, 古瀬 弘康, 勝本 道哲, 中川 晋一, 「D1 over IPによる高品位動画像転送・蓄積システムの設計と実装」, 情報処理学会研究会報告, vol.2000, No.30, pp.161-166,2000.5

[6] A. Machizawa, K. Sugiura, T. Komine, S. Uetsuki, H. Okazawa and J. Chun and S. Nakagawa, "On the delay and quality of DV transmission Systems using ATM networks", The 15th IEEE-International Conference on Information Networks (ICOIN-15), pp709-713, 2001

[7] Machizawa A., Kitaguchi Y., Sakurada T., Sugiura K., Komine T., Nakagawa S., Deployment of Native IPv6 Wide Area Network over GbE, ICOIN16, 2001

[8] S. Shinomiya, Y. Kidawara, T. Sakurada, S. Tsuchiike, S. Nakagawa, NADIA: Network Accessible Device on the Internet Architecture, Proceedings of ICOIN16, pp 8D4.1-8D4.8, 2002

[9] 櫻田武嗣, 杉浦一徳, 町澤朗彦, 北口善明, 篠宮 俊輔, 八尾 武憲, 田中健二, 中川晋一, ATM NICによるマルチキャストルーティングの限界, 電子情報通信学会情報ネットワーク研

研究会, 2002,(101-639),pp9-15

[10] S. Nakagawa, F.Kubota and K.Kuroiwa, Harmonization of Acceptable Users Policies of the Networks IPSJ Symposium Series, Vol 99-7, pp 357-361, 1999

[11] **S. Nakagawa, Proposal of a Policy Model for the Next-Generation Internet Testbed, Journal of CRL, Vol(48)-2, pp23-34, 2001**

[3] 中川晋一,インターネットコンテンツ容量計測法と次世代インターネット通信モデルの提案, 2002,情報処理学会電子化知的財産・社会基盤研究会EIP18-4,p17-p22