An analysis of international academic research network traffic between Japan and other nations

Toshiya Asaba
I S R,
Recruit Co.,Ltd.
1-13-1 Kachidoki
Chuo-ku,Tokyo 104
Japan
asaba@isr.recruit.co.jp

K. Claffy*
Dept. of CSE,
Univ. of California,
at San Diego
La Jolla, CA 92093
USA
kc@cs.ucsd.edu

Osamu Nakamura
Computer Center,
University of Tokyo
2-11-16 Yayoi
Bunkyo-ku, Tokyo 113
Japan
osamu@wide.sfc.keio.ac.jp

Jun Murai
Department of Environmental
Information, Keio University
5322 Endo,
Fujisawa, Kanagawa 252
Japan
jun@wide.ad.jp

Abstract

International IP network traffic is evaluated for applications utilizing the TCP transport protocol. This evaluation is based on an analysis study performed in January 1992 across an infrastructural connection between Japan and the PACCOM network. This traffic measurement allows an examination of the current status of international information flow. Such analyses are a first, and essential, step toward a better understanding of the type, quantity, and breadth of international traffic flow. They can provide not only confirmation or refutation of intuitions and existing paradigms related to data exchanges, but also a fundamental framework upon which to base future network management, planning and design.

1 Introduction

The dramatic proliferation of computer networking infrastructure, as exemplified by the dramatic growth of the global Internet user base, exposes a previously unexperienced need to support networking clients with robust and predictable networking environments. For example, a distributed operating system may need the capability to define network performance requirements, such as latencies, data loss rates, throughput, or error acceptability (data integrity, lack of data sequencing, etc.).

A first step toward accomplishing such goals is a sophisticated understanding of computer network behavior in general. However, a prerequisite is a clear understanding of traffic trends in restricted domains, which will allow the development of models which can be applied to larger arenas in the future. Such models not only allow one to effectively define network performance characteristics, but also will lead to new insights into the design of the international networking environment.

The effort dedicated to traffic analysis across wide area networks is limited at best. One of the few documented studies is that of Jon Crowcroft and Ian Wakeman of the University College London in England [1]. The paper, a presentation of selected traffic measurements between the UK and the US academic communities, outlines significant first steps toward understanding international traffic behavior.

The services provided by the Internet, such as E-mail, USENET News, remote login, and file transfer, have become an essential basis for information exchange and resource sharing among research communities. Traffic introduced by these services destined to locations outside of Japan from within WIDE[2] are routed through a single link, between PACCOM[5] and WIDE. The Japanese endpoint of this link is a router at the WIDE Network Operations Center at Keio University, connected to the undersea cable between Hawaii and Japan.

Generally, these international links are long-haul, point-to-point circuits with relatively low bandwidth and high cost. Effective utilization of such restricted resources is critical to maintaining and improving the availability of the above-mentioned user services.

The objectives of this research are:

- to evaluate the current utilization of the international link between PACCOM and WIDE
- to evaluate the scope of international information exchange grounded in these application protocols

The methodology used is as follows. We monitored TCP/IP[6] packet headers belonging to the relevant application protocols and analyzed the collected header in-

*also with San Diego Supercomputer Center, PO Box 85608, San Diego, CA 92186-9784 formation. Since most of the application protocols of interest use TCP as the transport protocol, TCP traffic was concentrated on in this analysis.

It must be emphasized that this analysis can yield only initial insight into international traffic flows. Although these observations can provide a framework for optimizing information flow among nations, successful network planning will require applying such investigations to a broader environment, as well as replicating such investigations regularly. An effective methodology of such observations will need to extend to traffic analysis on domestic levels as well.

The first two sections of the paper describe the network environment configuration and the monitoring and analysis tools. The following sections provide details of the analysis. Following the analysis, some conclusions and directions for future research are presented.

2 Internet connectivity from Japan

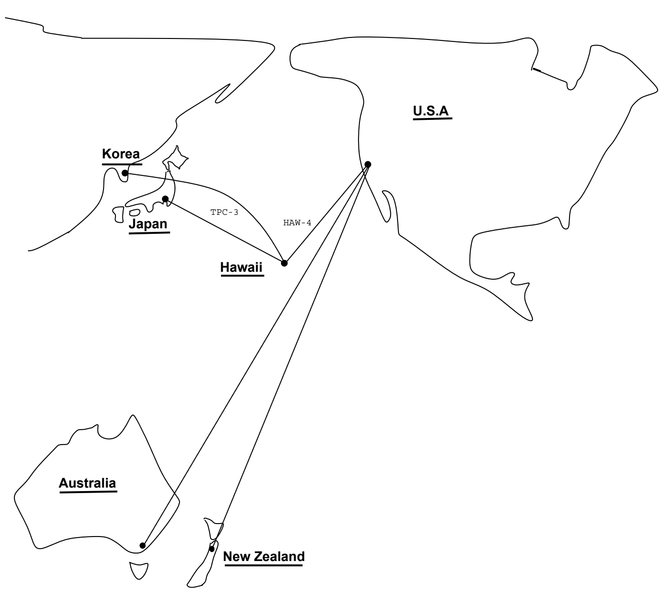

Figure 1: PACCOM Topology

This section describes the current network topology connecting the applicable JP-US networks (WIDE and NSFNET). The interconnection is supported by PACCOM, a larger Pacific regional internetworking effort funded by NASA, NSF and the State of Hawaii. Fig-1 presents a detailed diagram of PACCOM topology.

Within Japan, the domestic academic research network consists of three separate networks: TISN, JAIN, and WIDE. Of these three, TISN carries 14 Internet domains, and has its own 128 Kbps logical undersea link (via the TPC-3 transpacific cable) to Hawaii. The other two networks, JAIN and WIDE, carry 22 and 34 Internet domains, respectively, and share a 192 Kbps logical undersea link to Hawaii via the same physical cable.

From Hawaii to the US mainland, however, JP–US academic research traffic converges on the same (HAW–4) 512 Kbps[1] undersea link. This link connects the University of Hawaii (specifically, the ICS department) to one of the interconnection points for US federal networking activities (FIX-West) at NASA-Ames, located at Moffet Field in Northern California. A few other countries also have links to the PACCOM network: Australia and New Zealand are connected to NASA-Ames, and Korea is connected to the University of Hawaii. All traffic between Japan and these countries utilizes this same link between Japan and the University of Hawaii.

The data analysis in this paper is restricted to the WIDE–to–Hawaii 192 Kbps link. This link carries approximately 60% of the inter JP–US Internet traffic, and thus provides an adequate domain for traffic characterization.

3 Monitoring and Analysis environment

All traffic between Japan and other nations passes through an Ethernet inside the WIDE Network Operations Center at Shounan Fujisawa Campus, Keio University(WNOC–SFC). Data collection was performed from a machine (jp–tap.wide.ad.jp), SparcStation1+, on this local Ethernet. Fig-2 and fig-3 depict the data collection environment. The subsequent analysis of the data is a resource intensive effort, and could not have been done in real-time.

Virtually all of the traffic consisted of packets of the Internet Protocol Suite, in particular application protocols such as: FTP, SMTP, TELNET, DNS[6].

The tool used for data collection was **tcpdump**[8], developed by Van Jacobson of the US Lawrence Berkeley Laboratory, et. al. This tool allowed for filtering of traffic pertinent to this study.

The relatively low speed of the link allowed the use of **tcpdump** on a standard workstation, with little data loss due to processing overhead, about 0.2% on average. The data for analysis was collected from January 19th to 25th, 1992. The data volume exchanged via TCP protocol in this one week period was about 2Gbytes and the header information collected was about 280Mbytes

[1] It has been upgraded to T1

after compression. Analysis was performed using the **perl**[11] programming language to process the logged **tcpdump** data.

To investigate the application traffic, the traffic of actual TCP data will be concentrated on, excluding all non-TCP traffic in the following discussion. The byte counts do not account for IP and TCP headers, and the packet counts do not include zero-sized TCP packets, such as simple acknowledgements.

The information used to trace each TCP connection was[7]:

- Ethernet source/destination address
- Source and destination IP address / TCP port
- TCP flags, i.e. SYN, FIN, etc.
- Sequence number
- Acknowledgement number

They are used to:

1. count the number of bytes and packets sent from and into Japan,
2. trace each TCP connection,
3. determine the direction of connection request,
4. count the number of bytes and packets sent forward or backward to the direction of the initial connection request, and
5. evaluate the amount of information exchanged in each application protocol.

In addition, the inverse mapping mechanism of **DNS** and the **whois** data base were used to map numeric IP addresses to the corresponding domain names and to determine the source and destination nations of information exchanged within each application protocol.

In the following sections, the results of these analyses will be discussed.

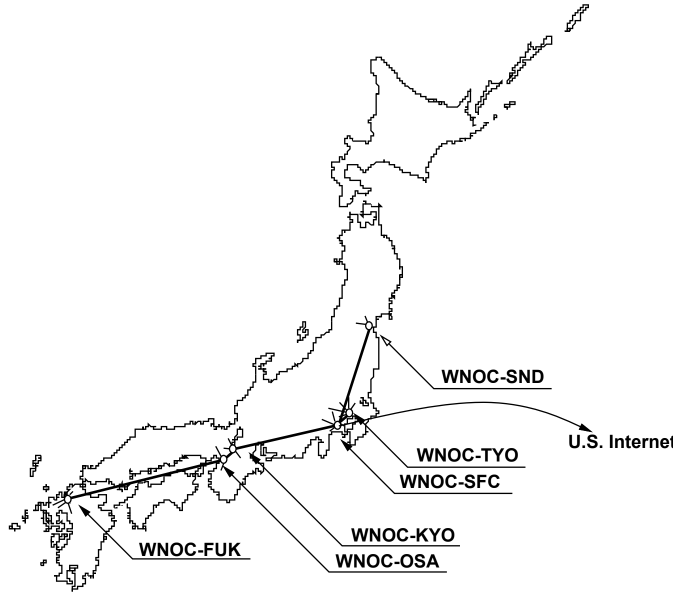

Figure 2: Configuration of the Japanese WIDE Network Environment

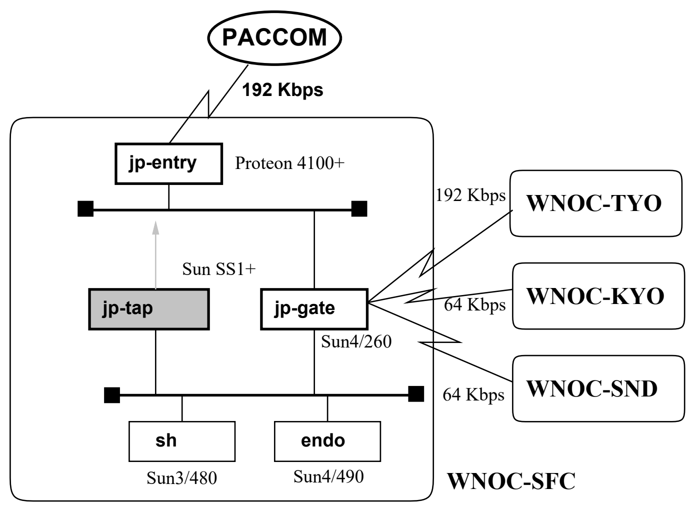

Figure 3: Configuration of Data Collection Environment

4 Inbound and Outbound TCP Traffic

4.1 Utilization of the International Link

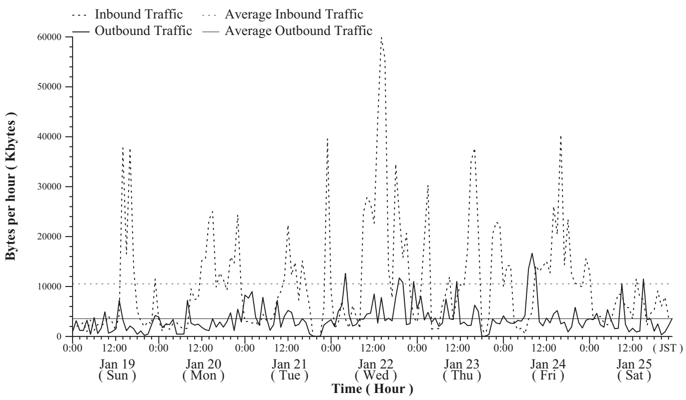

Figure 4: Inbound and Outbound Traffic rate in bytes per hour

Table 1: The Data Volume and the Data Rate

Date	**Bytes Sent (Kbytes)**		
	Inbound	**Outbound**	**Total**
1 / 19	166,711	48,403	215,114
1 / 20	229,702	61,679	291,381
1 / 21	190,187	85,995	276,182
1 / 22	427,900	129,265	557,165
1 / 23	308,425	88,587	397,011
1 / 24	305,437	108,736	414,174
1 / 25	139,933	73,978	213,911
Total	1,768,296	596,643	2,364,938

Date	**Ave. Line Utilization in Kbps**		
	Inbound	**Outbound**	**Total**
1 / 19	17.06(8.9%)	5.21(2.7%)	22.27(5.8%)
1 / 20	23.80(12.4%)	6.85(3.6%)	30.65(8.0%)
1 / 21	23.66(12.3%)	11.34(5.9%)	35.00(9.1%)
1 / 22	43.15(22.5%)	13.97(7.3%)	57.11(14.9%)
1 / 23	34.87(18.2%)	10.82(5.6%)	45.70(11.9%)
1 / 24	31.38(16.3%)	11.78(6.1%)	43.17(11.2%)
1 / 25	14.64(7.6%)	8.08(4.2%)	22.72(5.9%)
Total	26.91(14.0%)	9.67(5.0%)	36.58(9.5%)

Table-1 describes the daily TCP traffic volume and the daily average data rate in both directions. Fig-4 shows hour-by-hour traffic flow into and out of Japan.

During this seven-day period, equivalent to 168 hours of measurement, approximately 2 Gbytes of actual TCP data were exchanged. Unfortunately, there were two brief stoppages of the gateway machine during which international communication was not available: from 18:20 to 22:08 on January 21st and from 18:05 to 20:15 on January 23rd. So the exact observation period was 162 hours and 2 minutes.

Each TCP packet has at least 20 bytes of IP header information and 20 bytes of TCP header information. There were approximately 7.7 million packets which carried actual data; the average packet size was 313 bytes, excluding the headers. Thus, adding a constant 40 bytes for the IP and TCP headers, the average data transfer rate for TCP traffic was about 36.6 Kbps. This rate was obtained using the following formula:

$$\frac{(total\ volume + number\ of\ packets \times header\ size) \times 8\ (bits/byte)}{total\ number\ of\ seconds\ in\ interval}$$

$$= \frac{(2,364,938 + \frac{7,729,973 \times 40}{1024}) \times 8}{583,320} = 36.58 Kbps$$

Since the international link capacity is 192 Kbps in each direction, the TCP traffic occupied an average of approximately 9.5% of the link over the seven days, a relatively light load. However, according to the Fig-4, the peak hour for the seven-day period shows considerably higher traffic. Inbound TCP traffic constituted approximately 133 Kbps during its peak hour; outbound TCP traffic was 37 Kbps. This constituted a significant load at the peak hour, especially considering that the analysis ignored not only non-TCP traffic, but also TCP packets which did not carry actual data. At an even finer time granularity, such as every 15 minutes, even higher peaks in traffic volume must be seen.

4.2 Direction of the *service* requests

Fig-5 shows the number of connection requests occurring from and to Japan during each hour. This graph indicates specific traffic characteristics. In particular, many connection requests originated in Japan from around noon to midnight Japanese time, and significantly fewer connection requests occurred during the morning. This pattern reflects the typical working hours of Japanese researchers.

However, from the perspective of the *service* request, this graph is not quite accurate, for the following reason. During the connection set-up phase of a TCP application, the site requesting service sends a connection request, in the form of a SYN packet, to the remote site. **FTP-Data** connections, however, are a special case, as the the actual connection requester is not the *service* requester. The **FTP** server side always sends the **FTP-Data** connection request. Thus, if a host A sends a SYN packet to host B using the **FTP-Data** protocol, it does not mean that host A wants a service from host B, but rather that host B wants to use the file transfer service on host A. Reflecting this idiosyncrasy of the **FTP-Data** connection, The connection request direction for **FTP-Data** was reversed to determine the service request direction, as shown in Fig-6.

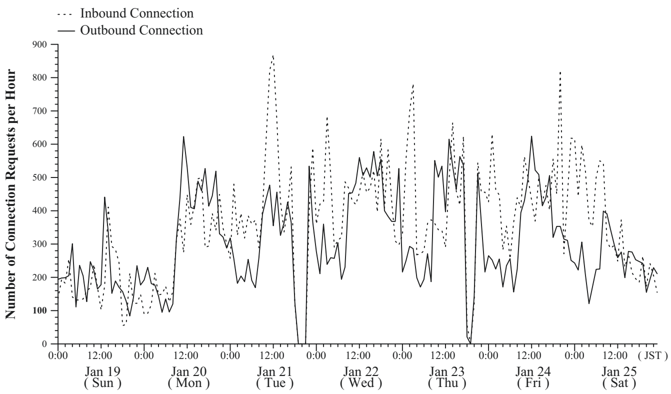

Figure 5: Number of Inbound / Outbound connection request per hours

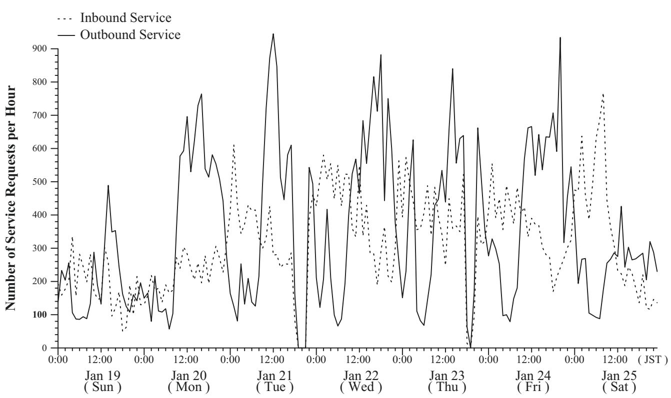

Figure 6: Number of Inbound / Outbound service request per hours

The graph presented in Fig-6 shows an even more extreme pattern in the number of outbound service requests than that observed in the outbound connections curve in Fig-5. In addition, the pattern shown here is complementary to the inbound request pattern, relatively low in the afternoon and higher in the morning. The time difference between the two countries appeared to have a positive side-effect of balancing the load on the link utilization.

The packet size distribution in both directions exhibits comparable bimodal patterns: heavily distributed around very small values and 512 bytes, the latter a commonly used maximum packet size between IP networks.

Table 2: Distribution of Packet Size for Inbound/Outbound Packets (Bytes)

Direction	**Statistics and Percentiles**					
	Average	**5%**	**25%**	**50%**	**75%**	**95%**
Inbound	365.185	1	33	512	512	512
Outbound	220.439	1	3	50	512	512
Total	313.287	1	13	512	512	512

Table-2 shows that the average packet size of the inbound traffic is larger than that of the outbound traffic. The 50th-percentile of the inbound packet size is 512 bytes, while that of the outbound packet size is 50 bytes, indicating that the proportion of inbound packets belonging to the "bulk data transfer" style protocols, e.g. **FTP-Data** or **NNTP** or **SMTP** is greater than the corresponding proportion of outbound packets.

4.3 Per protocol breakdown

Table 3: Per protocol In/Out Data volume in Kbytes and percentage to total

Protocol	**Inbound**		**Outbound**	
FTP-Data	1,345,224	77.16%	402,474	70.84%
SMTP	121,164	6.95%	65,880	11.60%
NNTP	142,038	8.15%	44,252	7.79%
Telnet	42,694	2.45%	30,026	5.28%
Domain	15,063	0.86%	347	0.06%
FTP	5,270	0.30%	4,381	0.77%
Finger	978	0.06%	1,192	0.21%
Others	71,049	4.08%	19,619	3.45%
Total	1,743,480	100.00%	568,171	100.00%

To investigate the connection direction, connections of which the **SYN** packet was missed were excluded from the statistics from here. The data volume included in those excluded connections was 2% of the byte total shown in the Table-1.

Table-3 shows that the observation presented according to Table-2 is correct. The percentage of data volume sent by a bulk data transfer protocol such as **FTP-Data** was much higher in the inbound than the outbound direction. Traffic using the **FTP-Data** protocol constituted 77% of the total inbound traffic volume, and 71% of the total outbound traffic volume.

Table-4 shows the number of the connection requested both in the inbound and outbound direction for each protocol as well as the data volume sent forward and backward to the connection request direction. Note that the relation between the direction of connection request and the direction of data flow varies by application.

Table 4: Per protocol In/Out Connections

Protocol		Num.	Data sent(Kbytes)	
			Forward	Backward
FTP-Data	IN	22,593	1,251,408	4,218
	OUT	15,181	398,256	93,816
SMTP	IN	21,260	115,505	8,066
	OUT	16,888	57,814	5,658
NNTP	IN	6,270	46,587	5,184
	OUT	8,969	39,067	95,451
Telnet	IN	1,736	1,000	29,224
	OUT	1,492	802	41,694
Domain	IN	550	15	261
	OUT	2,798	86	15,048
FTP	IN	2,598	896	3,130
	OUT	3,333	1,251	4,374
Finger	IN	4,013	24	1,183
	OUT	1,469	9	954
Others	IN	649	5,717	13,219
	OUT	1,065	6,400	65,331
Total	IN	59,669	1,421,153	64,485
	OUT	51,195	503,686	322,327

In the connection of interactive application such as **Telnet** and **FTP** as well as the inquiring application such as **Domain** and **Finger**, much more data flows in the backward direction than in the forward direction. On the other hand, in **FTP-Data** and **SMTP**, much more data flows in the forward direction than in the backward direction. In **NNTP** protocol, however, the data flow direction does not depend on the connection request direction. It rather depends on the configuration of the news forwarding path.

Table-5 presents the packet size distribution for each protocol. In the connection of interactive application such as **Telnet** and **FTP**, the packet size distributes around very small value. On the other hand, in the bulk data transfer protocol such as **FTP-Data**, it distributes heavily around the fixed length of 512 bytes.

Table 5: Distribution of Packet Size per protocol

Protocol	Average	Percentiles				
		5%	25%	50%	75%	95%
FTP-Data	529.403	512	512	512	512	536
Domain	425.633	23	512	512	512	512
NNTP	324.769	8	43	512	512	512
SMTP	220.427	6	33	60	512	512
Finger	128.772	2	2	9	210	512
Telnet	53.915	1	1	2	9	512
FTP	30.293	6	16	25	30	73
others	121.698	1	2	14	121	512
total	313.287	1	13	512	512	512

5 Information Exchange between Japan and other nations

Table 6: All TCP Traffic Flow

Domain	Bytes		Cons.	
	IN	OUT	IN	OUT
edu	1,081,824,425	269,582,416	28,242	22,144
com	253,380,216	75,163,919	14,146	12,069
org	140,661,183	12,156,135	1,435	277
gov	77,792,355	24,711,322	1,789	2,433
net	77,181,168	14,219,613	3,060	2,980
de	42,092,581	17,499,247	1,313	1,349
fr	21,256,861	21,649,227	364	694
au	16,374,621	21,465,957	933	1,089
ca	17,216,893	19,904,945	3,631	1,226
uk	6,657,069	18,977,853	950	1,633
mil	17,874,860	6,368,826	850	308
se	19,763,793	3,350,355	371	482
dk	1,976,102	14,151,067	257	241
il	44,100	12,827,143	46	125
fi	4,328,088	2,828,460	524	479
ch	2,268,435	3,573,455	202	602
nl	1,624,967	2,690,110	317	429
at	359,268	3,326,400	338	330
it	243,684	3,122,141	123	406
nz	104,019	2,302,672	53	60
no	326,597	1,866,997	117	170
br	161,578	1,629,413	71	90
hk	11,973	1,602,993	14	98
es	529,612	744,111	60	114
sg	72,464	1,125,344	59	42
us	246,963	940,123	40	97
kr	277,716	835,935	110	361
mx	22,405	828,502	14	10
gr	12,981	567,699	32	102
ie	15,200	328,880	5	78
hu	10,204	166,954	0	29
pl	117,404	58,830	27	42
za	17,744	58,782	8	4
is	14,909	35,756	3	9
be	12,549	21,344	11	22
pt	1,371	8,694	0	4
tw	2,878	196	0	5
Total	1,784,879,236	560,691,816	59,515	50,633

In order to investigate the direction of information flow in detail, the IP addresses of each end of all TCP connections were examined and their domain names were obtained using the inverse mapping mechanism of the **DNS**. The rightmost component of these domain names was used as the country name. Unfortunately, **DNS** failed to resolve many of the IP addresses, so the

whois database was also used to determine to which country these IP addresses belong. There still existed "orphan" IP addresses which could not be mapped to a particular nation; connections involving these addresses were excluded from the statistics. Fortunately, these unresolved connections comprised only 0.6% of the total connections and those connections contained only 0.9% of the total data volume in section-4.

The analysis presented in this section focuses on the quantity and type of information exchanged between researchers in Japan and researchers in other nations. It is assumed that there is a considerable degree of asymmetry in the data flow between Japan and other nations, in that the flow of data is largely from other nations to Japan rather than in the opposite direction. However, this research indicates that there is also substantial data flow from Japan to other nations.

Table-6 presents all TCP data exchanged between Japan and other nations. This raw data is ordered by the number of bytes exchanged; the first column lists the domain. Although most of the data was exchanged between the USA and Japan, at least 13% of the total data was exchanged with 30 other nations during the observation period.

Clearly the communication range of the Japanese academic society is not restricted to the USA, but extends to many nations in the Pacific Rim and Europe.

5.1 Commonly used service protocols

SMTP and **NNTP** are the most commonly used protocols for interpersonal communication. **FTP**, in contrast, is used to exchange data or distribute public domain software. In the following sections, each of these commonly used application protocols will be examined individually.

5.1.1 SMTP Traffic

Tables-7 shows the **SMTP** traffic flows in each direction. It is notable that the domains in Japan have more connections with other nations via the **SMTP** protocol than with any other protocol. About 34% of TCP connections belonged to this protocol. Such data confirms the intuitions that electronic mail is the most popular form of international communication. With some nations, the volume of **SMTP** data sent from Japan was comparable to that sent to Japan.

Table 7: SMTP Traffic Flow

Domain	Bytes		Cons.	
	IN	OUT	IN	OUT
edu	58,073,056	26,496,811	10,287	6,878
com	32,841,366	21,019,158	5,167	4,862
gov	10,654,102	1,990,317	649	351
net	3,772,570	3,292,139	531	895
uk	1,958,345	2,590,771	327	637
ca	2,819,436	1,622,761	746	432
au	2,797,095	1,224,208	327	336
org	2,237,973	468,004	748	96
nl	1,196,560	1,456,917	197	166
de	1,478,962	1,155,673	273	275
se	1,323,454	896,382	166	78
mil	1,464,593	629,316	578	129
fi	790,418	687,994	205	185
fr	358,308	408,936	135	174
at	316,232	426,281	283	105
dk	378,208	253,572	118	40
kr	206,621	397,324	71	126
ch	296,570	293,712	84	428
us	245,476	189,559	34	72
no	207,112	166,551	54	53
it	124,671	248,329	50	95
ie	14,200	256,546	2	63
br	95,564	103,155	43	38
il	39,660	142,141	19	41
hu	10,204	166,954	0	29
es	44,485	101,518	22	37
nz	68,149	72,279	38	33
sg	68,558	62,133	32	26
za	17,744	58,782	8	4
is	9,896	35,658	3	8
hk	7,053	21,936	3	16
mx	20,476	7,820	9	3
be	11,613	14,998	7	7
pl	12,212	7,577	9	2
pt	1,371	8,694	0	4
tw	250	49	0	1
Total	123,962,563	66,974,955	21,225	16,725

Table-7 shows, however, that with some nations, the volume of inbound data flow was twice that of outbound data flow. It seems strange for a person-to-person communication protocol such as **SMTP** to generate such an unbalanced exchange of traffic volume. One reason is that many mailing lists maintained at sites outside of Japan send mail to Japanese participants, and that much more mail originates from colleagues of the list outside of Japan. Such **SMTP** communication is not exactly on a person-to-person basis. Another reason for the asymmetric volume is the existence of "mail-ordering" information systems such as the RFC request system. In this case, ordering a piece of mail requires transmission of only a few lines of data but the response tends to be significantly larger.

5.1.2 NNTP Traffic

Table-8 describes the traffic flow of another interpersonal communication mechanism, the Bulletin Board System. The **NNTP** protocol is often used to distribute public domain software.

NNTP connections are not arbitrary, but rather depend on the administrative configurations of news feed servers in each nation. Indeed, the number of nations with which Japan may communicate by the Bulletin Board system is not restricted. Once news articles are forwarded outside of Japan, intermediate **NNTP** servers can propagate them across the whole Internet. Additionally, the Japan domestic news group "fj" now supports subscribers at many sites outside of Japan. Traffic for "fj" groups, written in Japanese, can thus be forwarded to as well as sourced outside of Japan.

Table 8: NNTP Traffic Flow

Domain	Bytes		Cons.	
	IN	OUT	IN	OUT
edu	95,281,960	20,872,134	12	3,252
com	48,792,354	11,863,547	6,221	4,418
net	252,884	7,269,830	0	1,125
uk	68,499	3,031,062	14	160
au	883,833	7,809	0	9
mil	144,844	709	0	5
Total	145,424,374	43,045,091	6,247	8,969

5.1.3 FTP traffic flow

User **FTP** sessions are subdivided into **FTP-Control**[2] and **FTP-Data** connections. The **FTP-Control** connection handles the transactions related to interactive user commands, such as the commands which initiate data transfers. All actual data, including responses to control commands such as "dir" or "get", travel across **FTP-Data** connections. Table-9 shows the traffic flow of the **FTP-Control** connections and Table-10 shows that of the **FTP-Data** connections.

These tables shows that a remarkable amount of data, about 22% of the total **FTP-Data** traffic flow, was sent from Japan to other nations. Within some domains, the **FTP-Data** traffic flow leaving Japan was actually higher than in the opposite direction.

A comparison of Tables-9 and Table-10 shows that the proportion of **FTP-Data** connection requests sourced in Japan, relative to those destined for Japan, is the inverse of the same ratio for **FTP-Control** connections. These values also reflect the amount of data sent in each direction.

[2] Usually, **FTP** means **FTP-Control**

Table 9: FTP Traffic Flow

Domain	Bytes		Cons.	
	IN	OUT	IN	OUT
edu	3,279,348	2,294,206	1,092	2,181
com	417,750	415,581	382	267
net	434,258	161,387	17	271
de	217,727	225,562	152	99
gov	207,119	129,486	43	89
uk	68,891	183,130	159	14
ca	89,611	153,080	91	56
au	88,852	131,604	98	43
org	141,291	49,597	11	84
fr	53,884	104,615	76	28
fi	99,351	34,483	12	51
se	51,415	67,861	39	36
mil	59,376	37,276	25	38
it	19,327	63,552	47	6
nl	27,394	53,847	42	20
ch	19,165	46,162	41	4
dk	20,904	43,905	43	10
at	15,430	43,867	23	1
no	12,849	41,938	37	4
kr	9,711	28,894	20	1
gr	8,664	25,188	20	2
il	4,433	25,229	25	0
hk	4,798	17,147	9	0
br	7,706	9,358	6	10
us	1,347	6,882	4	0
nz	2,993	5,086	8	1
es	2,225	5,104	8	1
sg	1,117	3,637	4	0
be	936	3,287	4	0
mx	1,006	2,541	2	1
pl	2,192	1,296	2	7
ie	905	2,487	1	0
tw	236	68	0	2
Total	5,372,211	4,417,343	2,543	3,327

An interesting usage of many FTP implementations is the support for "anonymous" FTPs, in which the **FTP-Control** connection request end typically retrieves large amounts of data using subsequent **FTP-Data** connections. Researchers often establish an anonymous FTP server at their research sites to allow interested parties to retrieve relevant electronic materials.

From the perspective of information and software distribution, how widely the server is known is important. Table-11 shows the number of bytes transferred by the server, the number of connections received by the server, and the number of different client hosts which access the server. Each row presents the top five servers in terms of the number of service requests they accepted. The top five servers inside JP domain accepted 1,290 **FTP** connection requests from 893 hosts in other nations, and transferred data to those hosts. The 893 data requesting hosts spanned 31 different top-level domains.

Table 10: FTP-DATA Traffic Flow

Domain	Bytes IN	Bytes OUT	Cons. IN	Cons. OUT
edu	844,769,610	193,374,712	14,677	7,588
com	158,639,413	33,370,384	1,651	1,534
org	136,345,557	5,354,684	555	37
gov	53,595,550	21,280,459	984	450
net	71,355,544	3,409,280	2,182	108
de	33,830,493	14,974,922	848	803
fr	20,701,781	20,899,762	128	474
au	11,416,664	19,197,477	326	481
ca	13,197,590	16,803,173	252	643
mil	16,049,449	5,693,688	241	88
uk	2,325,839	11,549,226	106	659
dk	1,011,020	12,549,906	31	180
il	0	12,658,232	0	84
se	9,227,302	1,969,278	154	234
fi	2,516,179	2,055,505	296	62
ch	1,906,889	1,754,727	40	165
at	0	2,825,456	1	213
it	8,810	2,799,327	4	281
nz	448	2,212,790	4	19
no	98,142	1,658,061	25	110
hk	0	1,563,644	0	82
br	38,198	1,516,540	22	37
nl	288,011	1,113,722	35	220
sg	0	861,311	0	16
mx	316	817,853	2	3
us	0	743,206	0	25
gr	4,224	541,581	9	100
es	129,643	296,647	4	14
kr	2,615	394,364	7	134
ie	0	69,677	0	15
pl	7,844	4,414	8	3
be	0	3,059	0	15
Total	1,377,467,131	394,317,067	22,592	14,877

Table 11: Top 5 FTP servers inside / outside JP

FTP Server	Service Requests	Client Hosts	Bytes Transferred
Inside JP	1,290	893	237,548,327
Outside JP	573	208	377,447,509

This table lists the top five servers both inside and outside of Japan. Actually, there were 151 servers in Japan which served requests from other nations, and 627 servers in other nations which served requests from Japan. According to Table-10, the top five servers within Japan transferred 60% of the total data sent from Japan to other nations, and the top five servers outside of Japan transferred 27% of the total data sent into Japan. If copies of the data and files provided by the top five servers outside of Japan were made available from within Japan, it would reduce the volume of inbound traffic to 81% of the existing volume of inbound traffic, by the following reasoning. The inbound **FTP-Data** traffic was 71% of the total inbound traffic. Multiplying 71% by 27% gives 19%. Thus 19% of the total inbound traffic could be reduced. Using the same reasoning for the outbound traffic yields a 46% potential reduction.

According to [3], a significant percentage of FTP file transfers are duplicates, i.e., they are FTP-ed more than once. If this is the case with traffic across our international link, replicating the more frequently retrieved files on ftp servers within Japan leads to the economization on bandwidth, obviating the need for international FTP transfers across a relatively slow link.

Actually, this software distribution methodology performed well during the X11R5 release. Copies of X11R5 distribution kits were made before the distribution began and it was announced that X11R5 would be available from several domestic FTP servers.

By monitoring the servers which distributed X11R5 inside of Japan and observing international traffic patterns for one week from the release, it was found that only one site had retrieved the X11R5 distribution kit from an FTP server in the U.S., while 36 domestic sites retrieved X11R5 from one of the ftp servers in Japan. Instead of monitoring the names of files being retrieved during FTP sessions, the quantity of data retrieved in each FTP-Data connection was calculated to determine whether the FTP session was related to X11R5 distribution, and it was assumed that fixed length data retrievals of 524,288 bytes were specific to the X11R5 distribution.

Without the preconfiguration and announcement of the X11R5 distribution server within Japan, those 36 sites would have been forced to retrieve it from outside Japan, using the undersea link which is the focus of this paper.

At the same time, it must be considered that software distribution can occur via several different protocols. For example, software is often distributed both from an anonymous FTP server via the **FTP** protocol and from USENET News as an article,[3] via the **NNTP** protocol. Identical data is then transferred on the same international link via two different protocols. Ideally, traffic related to a software distribution should traverse

[3] For example, in the *comp.sources.unix* news group.

the international link only once, and subsequently be duplicated in several locations within Japan as necessary.

6 Conclusions and Future Work

This paper investigates the characteristics of TCP service provision between Japan and other nations. Assuming that TCP connections represent "services", how these services have become essential to the work of many Internet users is examined. As the demand for such network services grows, network utilization increases and communication scope expands. People in Japan now communicate with others in various nations via E-mail and network news. One objective of this research was to evaluate the breadth of the reachability of traffic from these services. As the demand for these services increases, so will the load on the relatively slow, expensive, and difficult-to-upgrade link between Japan and the outside world.

The work presented in this paper represents only the initial steps in an investigation of traffic patterns across wide area networks in the international arena. The study presented here has lent both credence and suspicion to various intuitions regarding the traffic behavior in the investigated domain. However, this analysis constitutes only a small step in the effort required for workload characterization of international networks. Analysis using more sophisticated statistical analysis techniques at multiple levels of granularity of the data will allow greater insights into detailed traffic characteristics, including correlations, trends, and dependencies.

A robust global Internet will also require comparison of data from Japan to similar studies from other nations. Building a public library of usable scripts for analysis in different nations would greatly assist this enterprise. Establishing a set of measurement servers at strategic locations, which monitor the network for both one shot and longitudinal studies, would further contribute to rational analysis of the network by multiple parties.

We firmly advocate undertaking a collaborative effort within the community toward establishing such standard mechanisms. Only concerted investigation in the above areas, according to common methodologies agreed upon in an international forum, will allow effective progress into the higher speed, more ubiquitous, and increasingly powerful networks which we are facing everyday.

7 Acknowledgements

We would like to thank the members of the WIDE project for their support and helpful comments. We would also like to thanks Hans-Werner Braun of San Diego Supercomputer Center for his networking expertise and advice with the data analysis and presentation process.

References

[1] J. Crowcroft, I. Wakeman, *"Traffic Analysis of some UK-US Academic Network Data"*, Proceedings of INET'91, Jun. 1991.

[2] J. Murai, A. Kato, H. Kusumoto, S. Yamaguchi, T. Sato, *"Construction of the Widely Integrated Distributed Environment"*, Proceedings of TENCON'89, Nov. 1989.

[3] D. Ewing, R. Hall, M. Schwartz, *"A Measurement Study of Internet File Transfer Traffic"*, Technical Report of University of Colorado, CU-CS-571-92, Jan. 1992.

[4] V. Cerf, *"Guidelines for Internet Measurement Activities"*, RFC1262, SRI-NIC, Oct. 1991.

[5] J. Quarterman, *"The Matrix"*, Digital Press, 1990.

[6] D. Comer, *"Internetworking with TCP/IP volume I"* Prentice Hall International, 1991.

[7] D. Comer, D. L. Stevens *"Internetworking with TCP/IP volume II"* Prentice Hall International, 1991.

[8] V. Jacobson, et al, *"TCPDUMP (1)"*, Unix Manual Page.

[9] J.B. Postel, J.K. Reynolds, *"File Transfer Protocol "*, RFC959, SRI-NIC, Oct. 1985.

[10] J. Plank, *"jgraph"*, Unix Manual Page.

[11] L. Wall, R. L. Schwartz, *"programming Perl"*, O'Reilly & Associates, Inc, 1990.