5 Development of Data Network Systems for Global Environment Measurements

5-1 Development of SALMON system and the environment data transfer experiment

OYAMA Shin-ichiro, MURAYAMA Yasuhiro, ISHII Mamoru, and KUBOTA Minoru

We develop SALMON (System for Alaska Middle Atmosphere Observation Data Network) at Communications Research Laboratory for the network technology and development to smoothly deliver measured environment data, which is very important to study global environment issues. This system is developed using environment data measured at Alaska and the high-speed computer network of APAN and TransPAC. This paper shows the SALMON system structure and shows results from the data transfer experiment using the SALMON system together with development elements for the environment data transfer technology.

Keywords

Global environment, Network, Data transfer technology

1 Introduction

Global environment issues, representing a growing focus of worldwide attention, represent a challenge for the entire international community. To address this problem, researchers have been planning and conducting research activities using environmental measurement data obtained with satellites and ground-based instruments throughout the world. Continuous observations are conducted to monitor long time-scale variations in the environment. Furthermore, data related to global environmental issues tends to be stored at the various individual observation sites. In light of the above, it is clear that today we require: (1) technology for the efficient collection and exchange of distributed data; (2) technology for the distribution of data to numerous institutes and researchers; (3) data-processing technology for various users; (4) technology for the efficient analysis of diverse data; and (5) promotion of joint research and collaboration between different areas, in order to make effective use of these information/communication technologies.

The Communications Research Laboratory (CRL) has been developing the “System for Alaska Middle-atmosphere Observation Data Network (SALMON),” taking advantage of evolving computer and network technology. The goal of the system is to provide the quick transfer of an increasing amount of data that is used for analysis, research, and application. To develop a system suiting the characteristics of environmental measurement data (e.g., large amounts of data and distributing storage), we have used the measurement data obtained in the arctic region, mainly at Poker Flat, Alaska. Moreover, we have established a high-speed network link connected to the Asia-Pacific Advanced Network (APAN) and TransPAC, coupled to a very high performance Backbone Network Service (vBNS), proposed and developed by Indiana University for the National Science Foundation in the

United States. At the same time, we have been developing network technology that can transfer massive amounts of data over a long distance. This paper describes the configuration of the SALMON system and initial results on data-transfer experiments; we then discuss development points for the data-transfer system.

2 Configuration of the SALMON system

The SALMON system consists of three servers for data transfer: (i) a server in Alaska, which receives observation data from various instruments in Alaska on a quasi-real-time basis, (ii) a second Alaskan server, which receives the data from the first server, and (iii) a server in Japan, which registers the data transferred from the second Alaska server via network into a database following sorting and classification (Fig.1).

As shown in Fig.1, nine observation instruments have been installed in Alaska to perform continuous measurement. After the observation data obtained with these instruments has been transferred to the Alaskan server, the received data is compressed and screened according to data type and then sent to the second Alaskan server via a microwave 1.5 Mbps. The second Alaskan server was installed as a transit server to transfer the data to servers in Japan, because APAN, which is used as part of the Alaska-Japan network, allows only packets that are sent by servers having specific subnet addresses.

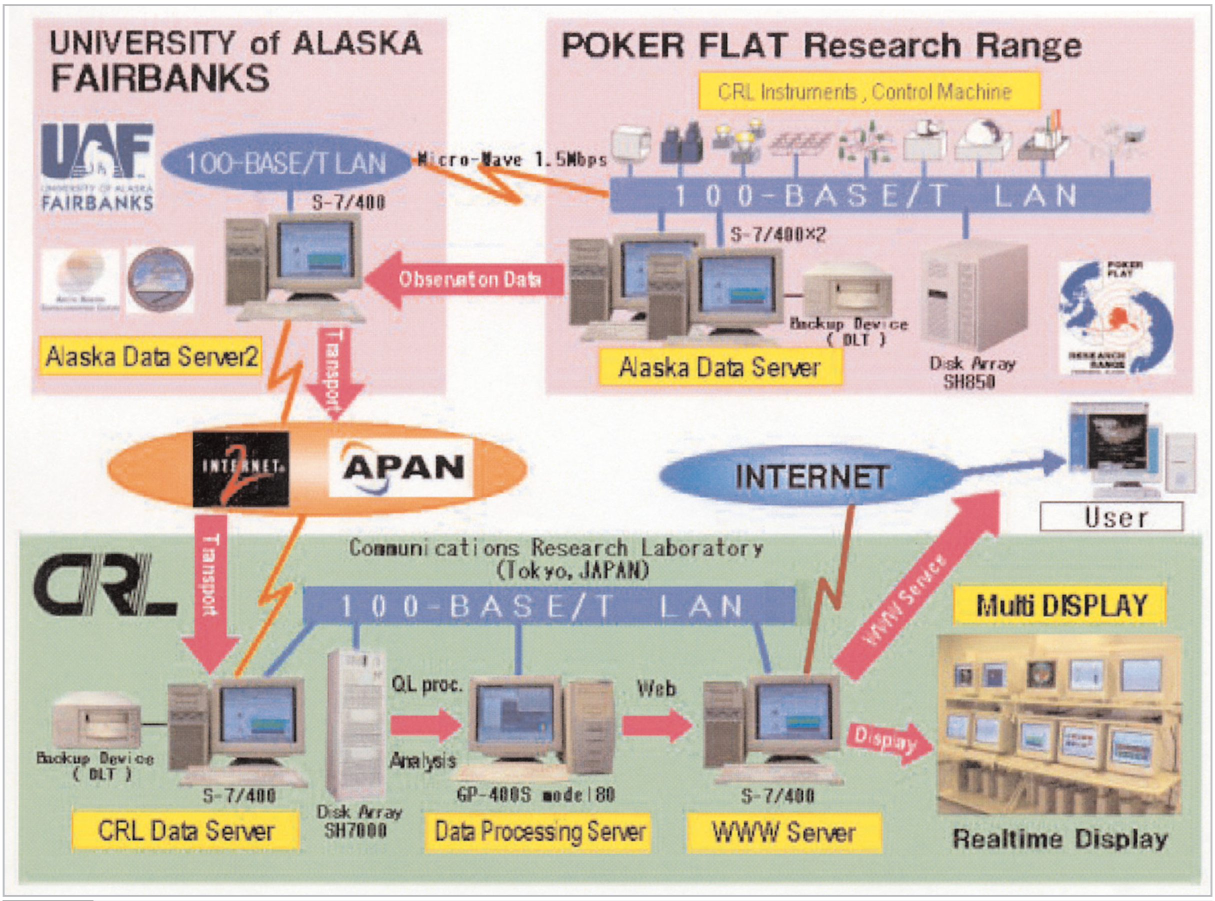


Fig.1 *Structure of SALMON system*

This schematic drawing shows data flow: The environmental measurement data obtained using the nine instruments (Fourier transform infrared spectrometer, Fabry-Perot interferometer, all-sky imager, imaging riometer, MF radar, millimeter radio meter, Rayleigh lidar, multi-wavelength lidar, and Super-DARN HF radar) installed by the CRL mainly at the Poker Flat Research Range is automatically transferred to the server in Japan via APAN and TransPAC. At the Japanese sever, the data sets are processed into a useful format for users in analysis and research.

The data server in Japan registers the observation data in a database, after sorting and separation. The data is automatically transferred to the application server, then converted into a useful format for studying the atmospheric environment, after which it is transferred to the web server to be posted on the homepage. On the homepage (http://salmon.crl.go.jp), users can view and download data as desired.

3 Dependence on TCP parameters for FTP data transfer

It is very important to regulate the network parameters as well as to adjust the expression system to arrange amount of data set and process them into a useful format for the average user. Currently, a high-speed link has been established using APAN and the TransPAC international network. Although the data-transfer rate via this network has not reached the expected value, the current SALMON system has succeeded in transferring and processing the data obtained in Alaska. The system, however, requires further improvement to ensure the smooth exchange, handling, and distribution of measurement data for every network environment. Particularly to develop network data-transfer techniques, it is important to understand the network characteristics and to estimate the bandwidth for the end-host link[1]. We measured the data-transfer rate in our system, varying the system parameters for TCP. We employed the Solaris 2.6 standard FTP daemon and booted up the client program in Japan and the server daemon program in Alaska. The data-transfer rate was determined as the speed at which the FTP client sent data for printout to a standard device.

The data transfer experiment was performed between servers connected to the same segment (a single switching hub) within a single room. Using three buffer sizes for TCP transmission/reception (Table 1), we sent files of two different sizes (approximately 1 MB and 1 KB) by FTP for each buffer size within a period of two hours. Figs.2 and 3 indicate that in every case, the transfer rate decreases significantly as the buffer size increases. This fact suggests that the settings provided by the vendor upon shipment will be nearly ideal when the inter-server connection is so stable

Table 1 *Parameters for data-transfer experiment in a single segment*

The default buffer size for data transfer and reception is 8 KB, while 16 KB is recommended and 64 MB is the maximum. Cases 1-3 took place separately on February 16, 19, and 20 in 2001. The Alaska-Japan data transfer experiment was carried out under the same conditions, except for the system environment. Cases 1-3 then took place on February 26, 27, and 23 in 2001.

	Host on transmission side		Host on reception side	
Experimental environment	OS: Solaris2.7 ftp: native (/usr/sbin/in.ftpd)		OS: Solaris2.6 ftp: native (/bin/ftp)	
Parameter	"tcp_xmit_hiwat" Send buffer	"tcp_recv_hiwat" Receive buffer	"tcp_xmit_hiwat" Send buffer	"tcp_recv_hiwat" Receive buffer
Case 1	8 KB	8 KB	8 KB	8 KB
Case 2	8 KB	8 KB	8 KB	64 KB
Case 3	16 KB	16 KB	16 KB	64 KB

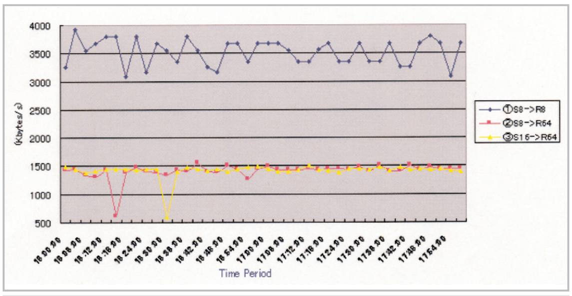


Fig.2 *Experimental results for the case where approximately 1 MB of data was transferred in a single segment using the transmission/reception buffer sizes listed on Table 1*

Although there are some fluctuations in the transfer rate in Case 1 (blue line), this rate is always higher than those of Case 2 (pink line) and Case 3 (yellow line).

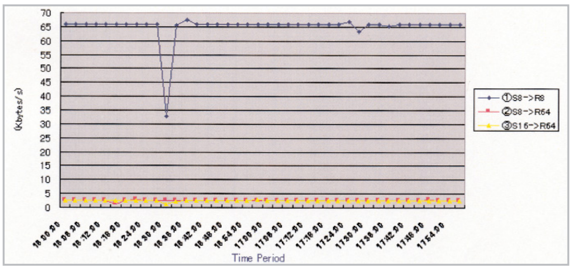


Fig.3 *Experimental results for the case where approximately 1 KB of data was transferred in a single segment*

The format of this figure is the same as that of Fig.2. Although the transfer rate sometimes becomes significantly low in Case 1, the overall rate in Case 1 is higher than the other cases.

that network delay becomes negligible.

The same transmission experiment was performed between the second Alaskan server and the Japanese server. When large files were transferred (Fig.4), the throughput in Case 2 and Case 3 was approximately 1.3 times greater than that in Case 1. On the other hand, there were almost no changes when files were small (Fig.5).

These results indicate that the TCP parameters can be adjusted to maximize the transfer rate in the case of FTP data transfer over a long distance (such as between Alaska and Japan) via IP network althogh the default settings will be nearly ideal when the inter-server connection is good. The throughput is expected to be raised as much as a factor of 1.3 by adjusting the parameters, particularly in the case of abrupt and cotinuous large-sized data transfer.

The data-transfer experiment between Alaska and Japan indicated that the transfer rate did not reach the expected level. The reason for this problem is now under investigation, but it may be because the long-range network is using a TCP/IP connection. Since a TCP session is of the connection-oriented type, it first establishes a physical connection prior to actual data exchange. This connection mode is reliable and suitable for the transmission of measurement data that does not allow data loss. In a long-range network, however, it takes time to get the acknowledgement of connection, and the network may be regarded as congested. As a result, the transfer rate may be actually restricted.

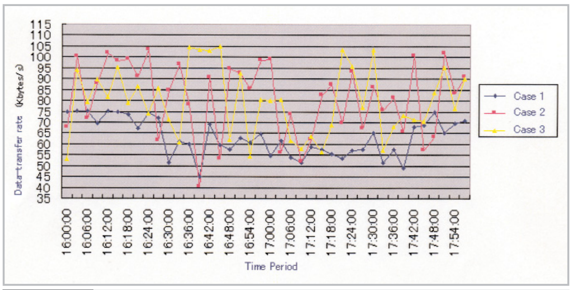


Fig.4 *Experimental results for the case in which approximately 1 MB of data was sent between the second Alaskan server and the Japanese server*

The format of this figure is the same as that of Fig.2. The average data-transfer rates in Case 2 and Case 3 becomes higher than that in Case 1 by changing the transmission/reception buffer size from the default one.

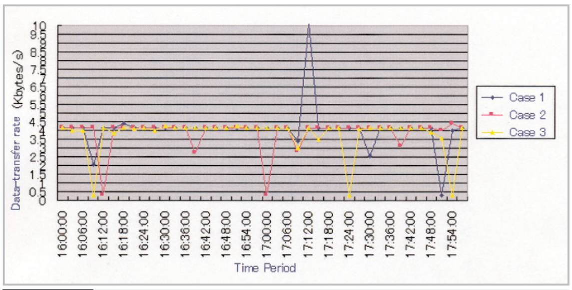


Fig.5 *Experimental results for the case in which approximately 1 KB of data was sent between the second Alaskan server and the Japanese server*

The format of this figure is the same as that of Fig.2. There was no change in transfer rates when the data size was relatively small, which is different from the results shown in Fig.4.

4 Summary

CRL is developing the SALMON system to improve a data-transfer and data-access system for prompt sorting and processing of environmental measurement data by applying a rapidly-evolving network technique. Our research has focused on developing network technology to transfer massive amounts of data via a long-distance network using the data obtained with observation instruments installed in Alaska as sample data, in order to develop a network system that responds to the storage and distribution of global environment measurement data among separate multiple observation locations. Furthermore, utilizing APAN and TransPAC to establish a high-speed link, we have concurrently investigated the technology for developing transmission of a large amount of data over such a broadband network. The present SALMON system has succeeded in smoothly transferring the continuous observation data obtained in Alaska to Japan and in automatically processing data that is useful to users.

As a result of the FTP data-transfer experiment using the SALMON system, the vendor's default parameter settings will be nearly ideal when the inter-server connection is stable. On the other hand, when large file data (of approximately 1 MB) is transferred over a long-distance network (e.g., between Alaska and Japan), throughput can be raised by adjusting TCP parameter settings.

Since this throughput was investigated in a short period of time, the resulting data-transfer rate may strongly influence the temporal network conditions. In order to develop a system that can run in a stable state over the long term, we must have a reliable method of throughput estimation. Further, in order to extend the applicability of SALMON, we must develop data-transfer technology that also deals with non-broadband networks.

In recent years, several technologies - such as those relating to Active Network - have been proposed, technologies which will revolutionize the conventional concept of the information network[2]. Combining such new technologies with SALMON will enable smoother exchange, processing, and distribution of measured environmental data. At the same time, we anticipate that the system will also provide complementary feedback to the field of information technology.

References

1. Matoba, K., S. Ata, and M. Murata, "Measurement based classification of network bottlenecks in the internet", The Transactions of the institute of electronics, information and communication engineers, 65 - 70, Nov. 2000.
2. H. Ikeda, M. Yamamoto, "A Study on active network technology and its applications",Tokyo Joho University Proceeding, 4, 181-193, 2001.

***OYAMA Shin-ichiro**, Ph. D.*

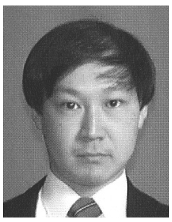
Researcher Fellow, International Arctic Environment Research Project Group, Applied Research and Standards Division

Atmospheric Dynamics, Information Engineering

***MURAYAMA Yasuhiro**, Ph. D.*

Leadar, International Arctic Environment Research Project Group, Applied Research and Standards Division

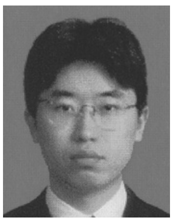
Observational Study of Middle Atmosphere



***ISHII Mamoru**, Ph. D.*

Senior Researcher, International Arctic Environment Research Project Group, Applied Research and Standards Division

Atmospheric Dynamics



***KUBOTA Minoru**, Ph. D.*

Researcher, International Arctic Environment Research Project Group, Applied Research and Standards Division

Atmospheric Dynamics, Optical System, Image Processing