OASIS: High-Speed Network Possibilities

V.V. KLUEV and AKIYASU HAYASHI
The Core and Information Technology Center
The University of Aizu
Tsuruga Ikki-machi Aizu-Wakamatsu City
Fukushima 965-8580
JAPAN

Abstract: - This paper offers an improved model of the OASIS system, which is a distributed domain specific search system in the Internet. The main thrust of our approach is to connect OASIS servers with a high-speed network based on ATM technology. This high-speed network allows communication between servers. Here we present methods of connecting OASIS servers with the ATM network. Obtained results indicate low response time of the OASIS system for the end user. Our simple approach requires only a minimum amount of hardware to be implemented.

Key-Words: - Search Engine, Distributed System, High-Speed Network, ATM Technology

1 Introduction
Effective retrieval of relevant information from the rapidly growing Internet is a difficult task: The end user must formulate queries very carefully to avoid a large number of useless documents. General purpose search engines help only a little. There are a number proposed approaches to improve this situation [6], [7], [8], [9]. These offer different methods of searching for information. The questions concerning architecture of search systems are explained in citations [4], [5]. Most of the presently working systems have a centralized architecture. The idea of distributed searches [1], [10] is getting more popular, because it is impossible to create an index of the whole net on one computer due to the limitations of hardware and network recourses of any given server. The OASIS system offers a promising solution of a distributed search.

OASIS (Open Architecture Server for Information Search and Delivery) was developed by an international consortium in the framework of the INCO Copernicus program of the Commission for the European Communities1 in 1997 - 1999. The system is open source software and can be obtained from http://www.oasis-europe.org. The OASIS service presents a distributed system of Internet search engines. The system provides search services for plain text PS, PDF and HTML documents stored on publicly accessible HTTP and FTP servers on the Internet.

Every OASIS server keeps a local index of a relatively small portion of documents available on the Internet. As most users’ queries are subject oriented, topic specific indexes are required for scalable distributed query processing. OASIS servers are not required to be mutually exclusive in terms of the topic areas they cover. It is possible to have more than one server that covers or overlaps the same topic area. A user can connect to any OASIS server with a search query. Such queries are then processed locally by the same OASIS server or automatically propagated to a subset of OASIS servers, which might contain the requested information. To select a subset of OASIS servers, the global OASIS directory of available collections is used. This directory assists the OASIS server in the selection of a set of collections for query propagation. If the query is propagated from one server to another, then the first one acts as a client of the second server. Before returning a result data set to the user, the client server eliminates duplicate records and sorts the results by their relevance score. The OASIS system is optimized for processing poorly specified search criteria by paying special attention to user ranking of results and to using this relevance feedback for the improvement of search result accuracy. More details about the OASIS system can be found in [1], [3]. The main bottleneck of the system is its relatively slow response to the end user, which is a consequence of queries that are usually propagated to other servers using an ordinary Internet connection.

---

1 OASIS, Project PL 1116-96, INCO Copernicus, Framework IV.
Our proposal is to connect OASIS servers with a high-speed network to increase performance of the system. In this case, servers can communicate with each other and with the directory through this network. The OASIS service will be provided as usual though the Internet.

The paper is organized as follows. Information concerning the high-speed network used in our experiments is provided in the next section. Following this, we describe how to connect the OASIS server with this network. After that we clarify the details of our experiments. Final remarks conclude the paper.

2 High-Speed Network Used in Experiments

The Special ATM Line [11], [12] connecting the University of Aizu with High Tech Plaza in Koriyama (Japan) has been used in our experiments. We utilized it to connect OASIS servers. The following information relates to the ATM configuration.

- This is a high-speed subnet of the University of Aizu network.
- The line includes two ATM switches (NEC ATOMIS/2000), two ATM routers with a Fast Ethernet interface (NEC MegaAccess MX/4E).
### Table 1 Specifications of the OASIS servers

<table>
<thead>
<tr>
<th>Processor</th>
<th>Memory</th>
<th>Hard Disk</th>
<th>Network Card</th>
<th>Motherboard</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pentium III 667 MHz</td>
<td>196 Mb</td>
<td>20 Gb</td>
<td>Intel 82557 PCI Ethernet Adapter (10/100), Intel PRO/100+</td>
<td>Intel 820</td>
<td>Aizu (Japan)</td>
</tr>
<tr>
<td>Pentium 233 MHz</td>
<td>160 Mb</td>
<td>2.1 Gb</td>
<td>AMD PCNET Family PCI Ethernet Adapter, 3 COM Etherlink 10/100 Mbps PCI NIC with 3XP Processor</td>
<td>Intel 430TX</td>
<td>Koriyama (Japan)</td>
</tr>
</tbody>
</table>

- These routers support IP technology. For this purpose the ATM Adaptation Layer type 5 specification is used.
- A 155 Mbit SC Fiber cable connects the ATM switches with routers. Its length is about 1 meter.
- The line uses point-to-point links (PVCs 155M UBR connection) to support communication between switches.
- The Special Line is a part of the NTT (Nippon Telephone and Telegraph) ATM network. It provides connection at a speed of 135 Mbit/s.
- IP addresses for the hosts and Fast Ethernet interfaces of ATM routers were assigned by the University of Aizu network. This is a network of C class.
- The distance between Aizu and Koriyama is about 60 km.

Figure 2 illustrates the aforementioned configuration.

### 3 Connection of the OASIS Server with the High-Speed Network

Figure 1 presents an improved model of the OASIS system. Servers are linked to the ATM high-speed network. This feature is the main difference from the previous model [1]. There are three possibilities to connect the OASIS server with the aforementioned high-speed network using:

1) One 100BaseT Network Interface Card. (Note, this kind of card can be setup as an ordinary Ethernet card.) This variant is used if the net has a connection with the Internet.
2) One 100BaseT NIC and one Ethernet card. This case should be utilized when a high-speed network is isolated from the Internet, and this net is dedicated only for communication between OASIS servers and OASIS servers with the directory.
3) The same hardware equipment as in item 2. This case should be applied to the local high-speed network in the event that it would be good to have access from this net to the Internet. In this case, ip-forwarding needs to be setup. The simplest way to do this in the OS Linux is as follows:

```bash
ipchains –A forward –j MASQ –s 192.168.X.0/24 –d 0.0.0.0/0 echo 1 > /proc/sys/net/ipv4/ip_forward
```

Here the network identifier should replace «X». This configuration provides access to the Internet from the high-speed network.

### 4 Experiments

Two OASIS servers were used in our experiments. Their specifications are presented in Table 1. To setup a 3 COM Etherlink 10/100 Mbps PCI NIC with 3XP Processor, an experimental driver 3e990 was installed. 3 COM Corporation provided this driver. Linux 7.0 was running on those computers, and they were dedicated for these experiments. The OASIS directory was installed on the server, which was located at the university of Aizu. All three of the aforementioned kinds of connections were tested.

Table 2 describes collections of documents installed on servers. Collections of Programming Languages on both servers were intersected. More details concerning these collections can be found in [2], [3].

The numbers of collection architectures were created for the OASIS system. One of the proposed variants includes the Isearch system, which is a software package for indexing and searching text documents. It supports full text searching and uses a vector space model for internal document representation and an inverted file to index documents. This software is in the public domain also, and it can be downloaded from http://www.etymon.com/Isearch/index.html. The aforementioned Isearch system was used as a local search engine on every server.

A client computer was connected to the server in Koriyama. A number of queries were submitted
from the client computer. Some of those queries were simple: they consisted of only one word, while others included several words and phrases. Response time of the system was significantly improved compared to the configuration of the system, which uses the ordinary Internet connection. Our test also showed that it is better to connect to the high-speed network using the much more powerful server than server on the base Pentium MMX 233 MHz processor.

4 Conclusions
A new model of connection between OASIS servers using the ATM high-speed network was proposed in this paper. Communications between servers and servers with the OASIS directory are managed through this network. Three methods of connections were tested. The results of our preliminary experiments show that this model has promising performance characteristics. Response time for the end user was low.

The OASIS server located at the university of Aizu is publicly accessible: http://oasisntc.u-aizu.ac.jp/oasis.

Our aim in future work is to conduct some experiments with Japan Gigabit Network [13], [14], to test performance of the OASIS system using this network for communication between servers.

References: