Chapter 1

Introduction

In the last decade the Internet introduced millions of users to wide area networks. Despite the growth of the power of Internet routers and links the everyday use of the Internet is very often frustrating because of the poor quality of service (QoS) of its data streams. For example, video-on-demand remains more of a rather promising idea than a reality. Video data stream suffers from poor quality of network links, voice is often incomprehensible, the download of large files often takes too long, etc.

Much research has been conducted to resolve the problems of Internet performance, but the problems are far from being solved. Research on open network architectures (ONA) and its sub-branch, active networks (AN) is one potential answer to the poor Internet performance. The main idea behind these two concepts is the design and real-time deployment of adapters that can resolve the problems of low bandwidth or insecure links. In the last seven years, this area has made great progress, but the problems it faces made the early optimism fade.

One difficult problem for ONA and AN is *planning*. The questions of which adapters to use to solve which problems, where to locate the adapters, and in what order to apply them must be answered very quickly after data transmission is requested. This problem has an exponentially distributed space of possible answers, many of which are inefficient or infeasible. The thrust of this dissertation is to show that it is possible to design and build a planner that can find good solutions to such problems quickly and cheaply.

To validate this thesis we developed such a planner for Panda, an AN middleware system, and tested the planning procedure using Panda to support multimedia connections. This dissertation describes the design, implementation, and evaluation of the Panda planning component.

1.1 Motivation

In this section we consider some Internet-related problems that have motivated ONA research.

1.1.1 Deployment of new protocols

Network technology and applications are changing rapidly, and existing protocols may not operate well in new circumstances. There are a number of examples (transit from IPv4 to IPv6, the rise of real-time and multicast communications using UDP, the rise of applications that frequently open and close TCP connections) that suffer from the inability of traditional networks to deploy and proliferate new protocols rapidly. ONA technology would allow the fast and relatively painless addition of new protocols.

1.1.2 Heterogeneous networks

Modern networks are characterized by the high level of heterogeneity [Yarvis99A]. Conventional wired networks coexist with wireless networks, satellite networks, etc. In the future, the level of network heterogeneity might even increase. A connection that traverses heterogeneous networks implies the coexistence of a number of communication protocols within the same connection. For example, TCP protocol is different in wired and wireless networks. Some wireless networks require extra measures of reliability, such as forward error correction, or security, such as authentication and encryption. ONA allows highly customized session protocols that will serve the current connection best.

1.1.3 User application requirements

Some user applications have extra requirements for data streams. For example, palmtops can only receive and process images in certain formats [Fox98]. These applications may require special treatment of their data streams on intermediate nodes, i.e., conversion of the data from the original format to one that is appropriate for the palmtop. ONA allows the automatic deployment of special format-converting protocols serving the particular needs of that palmtop.

1.1.4 Service for legacy applications

Automatic deployment of data-modifying protocols is the only way to serve the needs of legacy applications, i.e., applications that were designed without awareness of ONA services. Legacy applications, which are still important for the existing market, cannot get the benefits of ONA without special support. They are unable to recognize local communication conditions and therefore cannot adjust their services to the appropriate level. Nor can they instruct the ONA on how to adjust to conditions. ONA technology can automatically adjust the circumstances around these applications to keep the performance above a certain threshold or degrade it gracefully. To support legacy applications, ONA should recognize their needs, decide how to satisfy them, and deploy the corresponding adapters. Therefore, ONA should have access to information about network conditions, information that legacy applications usually do not have. The result of planning by ONA is a sequence of adapters that should be applied to a data stream. The sequence of adapters must preserve the integrity and security of data. That is achieved by the proper selection of adapters and their ordering.

1.2 Open Network Architectures

Traditional data networks passively transport bits from one end system to another. The network is insensitive to the bits it carries, and they are transferred between end systems without modification. The role of computation within such a network is extremely limited, just header processing in packet-switched networks and signaling in connection-oriented networks. However, more extensive use of computations, provided by adapters located in the network, can bring extra advantages to communications. The adapters must use computational resources efficiently and treat transferred data properly. The choice of the computations should take into account both user application and network requirements to compose a properly ordered set of adapters for rerouting or modification of a user data stream.

1.2.1 Adaptation of data streams

Each link in a network may present a different level of bandwidth, latency, jitter, reliability, and security. This level depends on the current communication conditions of the network. Data transmission would be improved if it were possible to adapt the data stream of a particular application to the network conditions. For example, data can be compressed to meet the requirements of reduced throughput of a poor-quality link, or data can be encrypted if security of a link is compromised. ONA should compose data-modifying protocols that adjust the data stream to network conditions, keeping its integrity and security at the appropriate levels.

1.2.2 Active Networks

AN technology allows networks to dynamically deploy adapters. AN is one type of ONA, a broader field that allows improvement of the service through local adjustment of adapters. In AN, adapters are mobile, dynamically deployable components that can be stored on hosts distant from the places where their help is needed. Users can program AN to choose adapters to apply to their data streams. Upon the request of an AN node or a user application, the adapter must be delivered and deployed in real time. It immediately raises the question concerning the selection of adapters that should improve the corresponding connections. The adapters must be properly ordered and located throughout the connection to provide efficiency for the adaptation. The consistency, integrity, and security of user data must be protected at all cost. Therefore, the adaptation of the data stream in AN must be properly planned.

1.3 Planning for Active Networks

Planning is an important tool for AN that guarantees the consistency of adapters and augments their efficiency. Efficiency becomes an issue because the execution of adapters requires computing resources of intermediate nodes, which can be limited, and increases the latency of data delivery. Unnecessary repetition of a particular adapter, as well as improper location of adapters, is highly undesirable. For example, assume that we have a connection that consists of more than two links. Assume that two links of the connection are poor and require compression of the data. It would be a poor idea to compress and decompress the data twice. The data should be compressed once, before the first poor link and decompressed once, after the second poor link. This will save execution resources of intermediate nodes and save time on the compression. Automated planning is intended to solve this and similar problems.

The automated planning consists of three natural consecutive steps. First, planning data must be collected and delivered to the site where the planner is located. Second, the planner calculates a plan using a *planning algorithm*. The plan is a list of adapters that must be executed in a particular order on particular connection nodes. Third, the plan must be deployed, i.e., the adapters must be delivered to the corresponding connection nodes and instantiated. These three steps are called the *planing protocol*.

1.4 Implementation of the Planner

In this dissertation, we will explore various aspects of the planning problem for ONA. We present our implementation of the planner for unicast active network communications. The emphasis is on performance, resource management, and the algorithm for plan calculation. However, issues of security are also discussed. The planning procedure requires specific planning data collection, plan calculation, plan deployment, and adapter design. The adapter design presumes the cooperation of planner and adapter designers. If an adapter designer fails to provide proper planning data for an adapter, the planner can misuse the adapter, possibly disastrously. This and other aspects of the security problem will be discussed in the dissertation.

We will show the results of the planner performance measurements and the planner-equipped connections for multimedia real-time applications.

1.5 Road Map to the Dissertation

In the remainder of the dissertation we will describe planning for ONA that addresses the problems we have described above. We implemented a framework that provides planning for an AN system serving UDP packet streams.

In Chapter 2, we describe the background of active networks. In particular, we describe the environment of AN, typical AN node architecture, Panda, AN-based middleware, and the problems of AN, particularly the problem of planning.

Chapter 3 describes planning alternatives. Here we present various approaches to solving the planning problem. Chapter 4 describes the planning protocol that was implemented for AN connections.

The planning algorithm is described in Chapter 5. We present our approach to the planning problem where the planner uses network conditions and user preferences for the plan calculation.

Chapter 6 presents measurements of the overhead of the planning protocol and the planning algorithm. We provide data on the quality of the plans produced, both in terms of an optimizable heuristic and the actual performance of AN connections that use the planning, including the measurements of QoS for video streams. We demonstrate improvements in the latency of planning over exhaustive search among all possible plans.

Related work is presented in Chapter 7. It describes various ONA systems, AI planning, and ONA planning projects. Chapter 8 presents future work, and we provide our conclusions in Chapter 9.