Optical Friendly HiSpeed File Transfer Protocol For Enabling Next Generation Nomadic Virtual PC Services

J. Fullaondo*, V. Lopez*, J. P. Fernandez-Palacios, J. Aracil***

**Universidad Autonoma de Madrid, **Telefonica I+D, Madrid, Spain*

{jaime.fullaondo, victor.lopez, javier.aracil}@uam.es, jpfpg@tid.es

Abstract: This work defines a HiSpeed File Transfer Protocol (FTP) designed for subwavelength metro optical networks. HiSpeed FTP improves TCP throughput, achieves a high rate in 10Gbps links and avoids penalizing transmission of small packets.

OCIS codes: (060.4259) Networks, packet-switched; (060.2330) Fiber optics communications; (060.4256) Networks, optimization;

1. Introduction

The implementation of Network Centric Services (NCS) is interesting for network operators, because it enables creating new business opportunities. A NCS combines IT and network resources. Examples of NCS are: Virtual PC, Video on demand (VoD), SaaS or gaming. These services combined with Internet traffic growth impact the network cost of the metro network architecture. EU FP7 MAINS proposes a novel metro architecture to deal with the new requirements for metro networks [1].

We propose a distributed data center model for some of these NCS due to its benefits to metro network resource utilization and user experience (reduced latency and jitter). This model requires content distribution to the access nodes. In burst switching networks, tunable lasers have a tuning time of 50ns to 100ns, while the transmission time for a 60 byte packet in a 10Gbps environment is 48 ns. Furthermore, in connection-oriented networks featuring fast and dynamic lightpath allocation the connection setup time is clearly larger and the traffic burstiness should be kept low in order to increase channel utilization [2]. We have developed a HiSpeed File Transfer Protocol to optimize content delivery through these optical metro networks.

2. MAINS reference network and service model

EU FP 7 MAINS project is focused on a new multi-service metro network architecture that allows the application/service layer to access subwavelength optical layer resources on-demand. Fig. 1 shows the MAINS reference network architecture. End-users connect to the metro nodes with a GPON network featuring the OLT at the head-end. The OLT is connected to a local multi-purpose server and to the subwavelength OPST node. A GMPLS control plane is in charge of establishing and releasing subwavelength connections. To request a new connection in the metro network a Network-Service Interface (NSI) is defined (Fig. 1). The scope for the GMPLS control is from the NSI head-end (calling party) up to the NSI far-end (called party). The NSI is a direct interface to the optical layer enabling a new "network-centric" approach, which can distribute content to multiple servers in the metro network improving current centralized server architectures [1].

The "thin-middleware" is the entity that distributes the content based on the users requests and network and IT resources. The middleware requests connections based on server location and IT resources through the NSI. The content may be, for instance, an HD video to be streamed, or a virtual machine image to be run for a Virtual PC service. When a user request an application, the middleware sends the content to the optimal location and a server there runs the requested application. Fig. 2 shows the modules for the middleware and application servers. The middleware provides five services to the application layer: (1) job scheduling, (2) content location, (3) optimal job execution server location, (4) connection management and (5) content transfer through the OPST/OBST transport layer [3, 4].

The middleware has two types of entities: central server and local sites. The central server receives user application requests via the web interface, schedules the jobs in its database and finds the content source and optimal location to process the users' application requests. The middleware consists of three modules: broker, HiSpeed file transport protocol and NSI agent (Fig. 2). The broker module is called from the central server when there are new jobs, it then

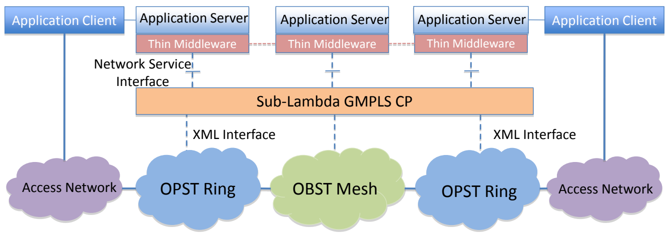

Fig. 1. MAINS Reference Network Architecture

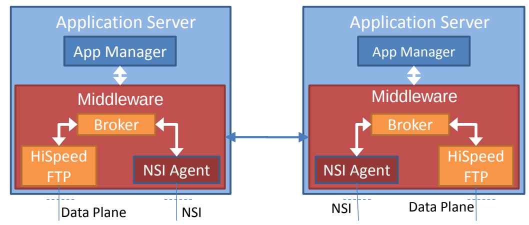

Fig. 2. Application middleware definition

calls the NSI agent to establish a new connection. The NSI agent manages the connections, while the HiSpeed FTP enables efficient data transmission through the subwavelength data plane. Upon successful content delivery, the broker calls the application manager which starts the application for the end-user. For real time applications, the connection with the end-user is done directly through the user's GPON access network, minimizing jitter and delay, and thus reducing bandwidth consumption in the metro network.

3. Virtual PC Use Case

The distributed server approach allows using servers for multiple applications. A Virtual PC service consists of virtual machines (VMs) executed on a remote host, and accessible through "remote desktop" protocols (readily available on Windows, MacOS, etc). This service would allow the user to replace a costly physical PC with a thin-client (ie. tablets, mobile phones). The VM presents a scalable virtual system maintained by the service provider (backups, upgrades, etc). The MAINS network architecture allows a high-bandwidth with a flexible architecture enabling advanced services like a "nomadic Virtual PC" which may roam with the user. When the user logs onto the Virtual PC service from a new location, his data will be transferred to the nearest multi-purpose server: an edge node on his access network. His connection to the VM host server will therefore benefit from reduced jitter and delay, improving the user experience. The steps of this process are: (1) the user requests his VM via HTTP, (2) the central server locates the content source and optimal job execution site, (3) the broker establishes a connection via the NSI, (4) the HiSpeed FTP transfers the content from the original to the final location, (5) the destination middleware triggers the application manager to run the VM.

4. Hi-Speed File Transfer Protocol

In the context of content delivery, TCP's acknowledgements for file transfers are of minimum packet size because file transfers are inherently one-way communication, and therefore acknowledgements cannot piggyback data frames. The huge amount of small packets transmitted and the traffic burstiness due to congestion control, which is unnecessary in MAINS' optical metro network, hurts network utilization.

To optimize bulky file transfer over optical networks, we have developed a transfer protocol specifically for this purpose. We scrap the idea of packet acknowledgement altogether, while retaining reliability by coalescing retransmission requests for missing packets (file chunks in practice) into large MTU-sized packets. The protocol has been implemented both over UDP (in case routing is required), and directly over Ethernet to reduce unnecessary bloat in P2P transfers. Our implementation eliminates the need to acknowledge received segments. We also intend to amortize laser tuning times and data transmission time slots by sending packets exclusively at size MTU (excluding the protocols handshakes).

The protocol as such is stateful. Consisting of two phases: in the first phase the file is sent sequentially in chunks of size MTU. If frames are missed, in the second phase, the missing file chunks are requested in batches using frames of MTU size. The coalescing of retransmission requests into large frames allows us to request hundreds of missing packets in a single packet. Handshakes also play an important role. The initial handshake allows the sending and receiving side to agree on a common MTU, and a unique connection ID (ie. ephemeral "port") is issued for the file transfer. Additionally, the filename and filesize are transmitted, allowing the receiver to create a bitmap with which to track the received file chunks. Obviously, both the initial and termination handshakes also signal the readiness and successful completion of the transfer.

5. Results

Experiments were conducted between two i7 920@2.67Ghz machines with 6GB RAM running CentOS 5.5 (kernel 2.6.18-194). 10GE dual-port Intel 82598 NICs were used and linux network parameters tweaked to allocate additional memory for network sockets: `net.core.{rmem_max, wmem_max, rmem_default, wmem_default}` = 16777216. File transfers were performed RAM-to-RAM due to the HD bottleneck, and the file was 2GB in size. An MTU of 9000 bytes was used. Figures reflect the average results for 10 executions.

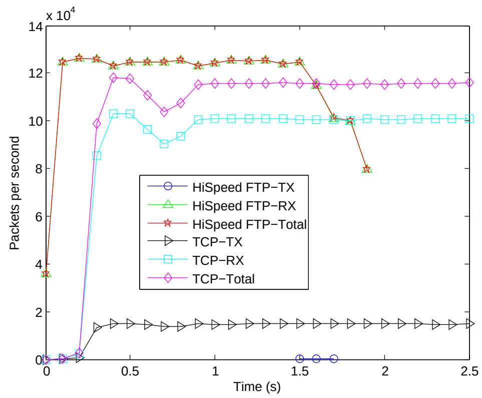

Fig. 3. HiSpeed FTP vs. TCP: Pkts/sec

Protocol	packet count
HSFTP-TX	24
HSFTP-RX	230223
TCP-TX	33814
TCP-RX	228568

Fig. 4. HiSpeed FTP vs. TCP avg. pkt count

We have managed to surpass the performance of TCP (SACK enabled) for file transfers, achieving an average effective throughput of 8.8 Gbps as opposed to TCPs 7.1 Gbps, and an improved packet/sec rate (Fig. 3). Additionally, as can be seen on Fig. 4, and excluding handshakes, we reduce the number of small packets in the network to zero, and the number of packets in the“anti-transfer” direction to the order of tens of packets for the aforementioned 2GB file transfer employed in our tests. TCP on the other hand injects tens of thousands of tiny 60 byte acknowledgement packets for the same file (30000 packets in our tests).

6. Conclusions

Full-Mesh OPST/OBST architectures such as the one described for MAINS will be deployed in metro networks. Simultaneously, “the cloud” appears to be the new computing paradigm, and services such as a Virtual PC and SaaS, where computing becomes a new ubiquitous utility seem to be blooming. It is therefore important to utilize these optical network resources to the fullest. TCP has been, and will continue to be, a cornerstone for connection oriented communications. Its characteristics however do not suit certain applications such as file transfers (involved in VM sync/restoration, backup services, etc) over an OPST metro network. The HiSpeed File Transfer Protocol, has been developed with these networks in mind and, at least for services involved in distributed data centers, should improve on both throughput and network utilization as opposed to TCP.

Acknowledgments

This work has been supported by the EC’s Seventh Framework Programme through IST STREP project MAINS (INFSO-ICT-247706).

References

1. J. Fernández-Palacios et al., “Metro Architectures enabliNg Subwavelengths: Rationale and Technical challenges,” in “Proc. Future Network and Mobile Summit,” (2010).
2. P.M. Santiago del Río et al., “Application of internet traffic characterization to all-optical networks,” in “Proc. of the International Conference on Transparent Optical Networks,” (2010).
3. J. Fernández-Palacios et al., “IP Offloading over Multi-granular Photonic Switching Technologies,” in “Proc. of the European Conference and Exhibition on Optical Communications (ECOC),” (2010).
4. J. Triay et al., “GMPLS/PCE/OBST Architectures for Guaranteed Sub-Wavelength Mesh Metro Network Services,” in “Proc. of the Optical Fiber Communication (OFC) Conference,” (2011).