Power-efficient Electronic Burst Switching for Large File Transactions

Ilijc Albanese[1], Sudhakar Ganti[2] and Thomas E. Darcie[1]
[1]*Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada*
[2]*Department of Computer Science, University of Victoria, Victoria, BC, Canada*

Keywords: IP Networks, Burst Switching, High Speed DWDM Optical Transmission, Optical Network, Overlay Network, Low Power Networking, Electronic Buffering, Router Power Consumption.

Abstract: Much of the growth in bandwidth demand and power consumption in today's Internet is driven by the transport of large media files. This work presents a power-efficient overlay network specifically designed using electronic burst switching for these large files. The two approaches are presented in which electronic bursts or media frames (MF) containing >1Mb are routed in a manner similar to UDP or concatenated into periodic semi-transparent chains and routed using a two-way reservation protocol. Utilization, blocking, delay and buffer size are compared to UDP/IP by means of simulation. Both approaches dramatically reduce header-related power consumption. Concatenation also reduces significantly the amount of buffer space required. A representative router design is evaluated showing a potential energy saving of roughly 80% relative to standard IP routers.

1 INTRODUCTION

Despite ongoing improvements in bandwidth capacity and power efficiency, power consumption in router networks continues to be a concern. Estimates suggest that Internet-based communication technologies consume between 2-3% of power generated globally and that this number is increasing at a rate of 16-20% per year (Fettweis and Zimmermann, 2008). As a result, extensive effort is ongoing to develop new techniques for minimizing power consumption in future Internet technologies.

Of particular interest is power consumed in electronic routers. Numerous studies have shown that a significant portion of power consumed can be attributed to header processing on each packet. With the increasing popularity of bandwidth intensive applications such as streaming video and the sharing of large files, studies support the general observation that file sizes are growing rapidly. Given the large number of IP packets required for these transactions, header related power consumption is a key contributor to the rapid growth of power consumption in routers. Energy efficiency improves if larger packets are used for large file transfers. In response, the maximum IP packet size was extended from 1500 bytes (B) using Jumbo and Super Jumbo frames pushing packet sizes to 64KB. However, these specifications are not in widespread use due to problems related to backwards compatibility (psc.edu, 2012) and network latency arising from integration of very large packets with smaller packets within the same links. Integration of large (up to 19.5KB) and standard packets within the same network was considered (Divakaran and Altman, 2009) showing that the use of larger packets may reduce power consumption and computational load required from the network hardware. However, coexistence of large and small packets in the same links results in unfairness and higher drop rates for both types of packets, leading to inefficient bandwidth and computing power utilization.

This raises the question as to the potential value of using a separate overlay network for the traffic associated with large file transfers. While clearly the introduction of a new overlay network would have to be predicated on compelling value, it would be unwise to suggest that, given power consumption and scaling considerations of IP, a next generation non-IP switching/routing approach cannot exist.

What might be the form of this new overlay network? On one extreme, numerous optical

Albanese I., Ganti S. and E. Darcie T..
Power-efficient Electronic Burst Switching for Large File Transactions.
DOI: 10.5220/0004357400430052
In *Proceedings of the 2nd International Conference on Smart Grids and Green IT Systems* (SMARTGREENS-2013), pages 43-52
ISBN: 978-989-8565-55-6

networking approaches offer up to an entire wavelength for some time, through which GigaByte (GB) files can be delivered. Optical burst (OBS) (Jue and Vokkrane, 2005) or flow switching (OFS) (Chan, 2010), or user-controlled end-to-end lightpaths (e.g. CAnet4, MONET, CORONET and GRIPhoN (Mahimkar et al., 2011)) have been explored fully. For example, an OBS approach (Yong et al., 2010) uses concatenated data bursts where the data units are organized as non-contiguous and non-periodic series of concatenated timeslots (bursts), which are then handled as a whole in an all-optical network infrastructure. These optical approaches are not embraced by industry, in part because the power efficiency of optical switching is questionable (Tucker et al., 2009), (Tucker, 2006) and optical buffers, widely used in most OBS proposals, have not yet offered a commercially viable alternative to electronic buffers. Also, while capable of supporting large bandwidth, targeted implementations are in the interconnection of specialized nodes rather than broadly distributed Internet users.

Hybrid architectures have been studied in which both electronic and optical switching are combined (Aleksic et al., 2011) to simultaneously handle packets, bursts and TDM circuits. A large reduction in the power consumption is achieved by selecting adaptively which part of the node to activate based on a per-flow evaluation of the data to be routed while the other blocks are put in sleep or low-power mode. While potentially powerful, this approach requires the complex integration of disparate switching and control elements, some of which (like OBS and optical delay lines) have not proven compelling individually.

A more incremental overlay network approach is electronic burst switching (EBS) (Peng et al., 2010). Following the OBS model, bursts are assembled at edge burst switches and switched electronically at core switches. It was concluded that using large bursts (> 1 Mb) may lead to reduction in header-related power consumption in core switches, but the power consumed by burst assembly negates much of the advantage gained in core switching.

In this paper we continue along the path of EBS. Users share the bandwidth of an overlay network, which we presume to be statically provisioned, using electronic switches or routers specifically designed to handle large file transactions. Unlike (Peng et al., 2010), we eliminate burst assembly at edge switches and consider direct end-to-end delivery of large “media” frames (MF) (roughly 1-10Mb) to users through an overlay to next-generation optical access networks. Free from the constraints of coexisting with highly granular and dynamic IP traffic, this EBS overlay network can be designed specifically for the efficient delivery of the large data transactions that did not exist when the Internet was conceived. Compared to traditional IP routers, switch reconfiguration can be far less dynamic since only very large packets are supported. Unlike proposed optical alternatives, this can be accomplished using available electronic buffers in a form that is entirely compatible with today’s highly efficient cross-point switch arrays.

Our objective is to enable a significant reduction in power consumption of network hardware while optimizing the use of resources. We first explore routing MFs using a standard UDP protocol (MF-UDP). UDP is selected for this study, rather than TCP, as this avoids numerous complexities that add little insight to a comparison with conventional IP and, as discussed later, gives the best case scenario for IP. Based on simple hardware considerations, network performance simulations and comparison with traditional UDP, we arrive at the anticipated conclusion that router power consumption is reduced dramatically, but performance is otherwise unaffected and larger buffers are needed.

We then consider using concatenations of MFs into periodic semi-transparent chains (MFCs) and the scheduled transmission of these MFCs using a two-way reservation mechanism. While such a scheduling mechanism would be inappropriate for traditional IP traffic, the large size of each MFC (e.g. 1 GB) makes scheduling both manageable and worthwhile. Also, the structure of an MFC makes it easy to condense information on its configuration with minimal control plane information, minimizing the amount of information to be processed at each node to schedule the chain and reducing the probability of control plane collisions. Simulation results show increased utilization efficiency and decreasing buffer requirements in comparison to MF-UDP as well as standard UDP. An MFC router is designed based on a commercially available cross-point switch array and power consumption is estimated to be roughly 20% of that of a standard IP router.

The paper is organized as follows: Section 2 provides an overview of the reference network architecture in the context of transactions of large files. Section 3 compares, using OMNeT++ simulation, traditional UDP to MF-UDP in supporting representative large transactions. In Section 4, chains of media frames (MFCs) are introduced, along with exemplary admission control and scheduling algorithms, and compared to UDP and MF-UDP. Router implementation is discussed in Section 5, where implications on power consumption are considered and a representative MFC router is evaluated.

2 NETWORK TOPOLOGY AND LARGE FILE FLOWS

Our discussion is framed by the reference network architectures shown in Figures 1 and 2. Fig. 1 shows a hierarchical network representative of the current state-of-the-art comprising various sizes of routers (access, edge, and core) connected to a transport network through various sizes of add-drop multiplexers.

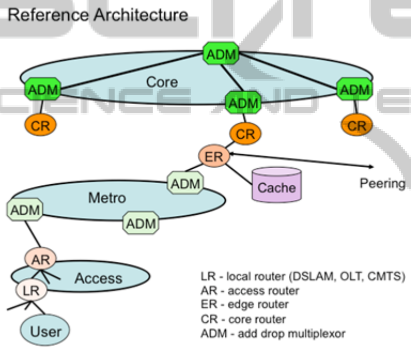

Figure 1: Reference Architecture.

Primary examples of large-file transactions can be super- imposed onto this reference architecture. These examples include: 1) Regionally-cached download: In this case large media files are downloaded from regional cached distribution servers at the end points of Content Delivery Networks (CDN), through Metro and access to end-users. Driven by rapid proliferation of video-related download applications, these downloads represent a significant fraction of traffic growth, and therefore are the focus of this study. Other important transactions include: 2) Source-to-cache distribution: To deliver and update content to CDN servers, large files must be distributed from sources, typically across a core network, to the regional cache; 3) End-to-end file transfers: For peer-to-peer applications, or for files for which widespread distribution is not anticipated, the regional cache is bypassed and files are transacted through Metro and access, possibly across the core network, directly to an end-user.

An overlay network designed specifically for large file transactions might look like Fig. 2. Each of the “media” boxes parallels a present-day IP counterpart and supports the origination or termination (media client interface (MCI)), access bypass (media access bypass (MAB)), admission (media access router (MAR)), and efficient end-to-end routing (media backbone router (MBR)), in accordance with the principles described below for MFs. In addition, it is convenient to consider regional storage of MFs in a media frame cache (MCa).

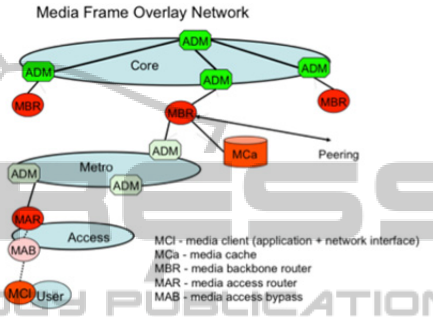
Media Frame Overlay Network
Core
ADM
ADM
ADM
ADM
MBR
MBR
MBR
MCa
Peering
ADM
Metro
ADM
ADM
MAR
MAB
Access
MCI
User
MCI - media client (application + network interface)
MCa - media cache
MBR - media backbone router
MAR - media access router
MAB - media access bypass

Figure 2: Media Frame Overlay Architecture.

Our intention is not to restrict application to specialized nodes (e.g., campuses), but rather to reach broadly distributed consumers, methods must be established to transport MFs through access.

Two possible solutions are represented in Fig. 2. A conservative approach involves an application operable between the MCI and the MAR such that using traditional broadband access alone, MFs and MFCs are assembled or disassembled at the MAR. This functionally is parallel to the burst assembly routers proposed for use in OBS and EBS. A preferred approach involves engaging the evolution of optical access standards, where standards for 10 Gb/s PON have emerged recently, to enable higher dedicated bandwidths perhaps through wavelength overlays. MFs and MFCs would then be assembled at the user end point or client directly.

3 MEDIA FRAMES VERSUS CONVENTIONAL IP

We first explore the issues associated with migrating very large file transfers to a separate overlay network in which the standard unit of bandwidth is a media frame (MF) containing roughly 1-10 Mb of data plus overhead. In (Peng et al., 2010) it was concluded that although using large packets would lead to considerable power savings, increasing the frame size beyond 1Mb would only marginally increase the energy efficiency of an EBS node. However, using larger frames also reduces the required reconfiguration speed of the switch fabric, minimizing requirements on switching speed and inefficiency introduced during transitions.

Overhead may include address, priority, concatenation details (specifying MFCs, as discussed later), coding, guard time and management information. Given the very large capacity within each MF, considerable header information (e.g. 10 KB) can be included with minimal impact on throughput. Structure may be defined to facilitate error correction, segmentation, security, file compression, and easy assembly from a large numbers of smaller IP packets.

An obvious method for networking with MFs is to adopt the same concepts as used with TCP/UDP, allowing each MF to be routed in accordance with predefined routing tables through a connectionless queuing network. While the dynamics are considerably different than with < 1500 B UDP packets, the underlying issues are the same.

To explore this in detail, both UDP and MF-UDP network simulators were built using OMNeT++ (OMNeT++, 2012) and the performances compared in terms of link utilization, buffer space occupied and delay per GB of data transferred. UDP was simulated using 1500B packets and MF-UDP using 1Mb packets. Droptail queues were used for both 1Mb and 1500B UDP packets. For standard UDP packets the maximum buffer size was set to 1000 packets, corresponding to roughly 1.5MB. Buffers of the same size were used for MF-UDP.

A dumbbell topology was considered for the simulations (Figure 3). While more complex topologies could be simulated, this represents the case of our reference network (Figure 2) with a congested link between multiple source servers and end users.

The capacity of each link is set to 10Gib/s (i.e. 10*2^30 Gb/s according to IEC standard) and each source is offering the network an average load of roughly 1.33 Gb/s. Various load conditions were tested by activating more source-destination pairs and the offered load was made to vary from 25% to 150% of the bottleneck link capacity (corresponding to from 2 to 12 source-destination pairs). Each source transmits data to one destination only and all sources compete for the same bottleneck link.

Compared to UDP, the router switch fabric becomes far less dynamic. Reconfiguration occurs far less frequently (by 3 orders of magnitude).

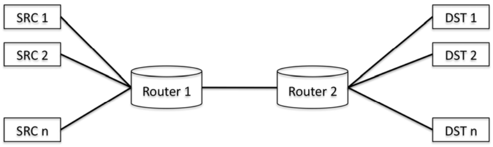

Figure 3: Topology Used for Simulations.

It remains to be seen if an MF-router can be designed to exploit this less dynamic reconfiguration and negligible header processing with a sufficiently large net increase in power efficiency to justify a separate overlay network. This is addressed further in Section 5.

4 CONCATENATED MEDIA FRAMES

While UDP-like MF routing dramatically eases processing (header and switch reconfiguration) over traditional IP, performance is otherwise essentially the same. We now consider the potential impact of concatenation of MFs into larger structured chains (MFCs). Concatenation creates single entities that would contain an entire large transaction, for example a multi-GB movie download. Discussion centers on two key considerations: scheduling and transparency. Scheduling and admission control become worthwhile for such large transactions and these can be used to increase resource utilization and minimize buffer size. However, serving such large transactions continuously in time introduces significant latency for waiting transactions and is incompatible with the lower end-user client and access network throughputs. Making each MFC partially transparent overcomes both problems. We limit our discussion here to a simple functional description of a representative methodology, including MFC structure, transparency, signaling and scheduling algorithm, then compare performance to conventional UDP and the MF-UDP described in Section 3.

4.1 Media Chain Transparency

A MFC with transparency degree 3 (defined as the number of interstices between two consecutive MFs in a chain plus 1) is illustrated in Figure 4.

Using periodic semi-transparent chained data structures provides five primary functions. First, given the large size of each MF, concatenations of multiple MFs without transparency would introduce substantial latency by occupying network resources

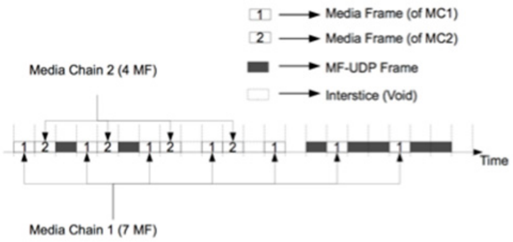

Figure 4: An Example of How Two MFCs, MF-UDP Frames and Voids Fit in a Channel.

for substantial durations. The use of transparency allows servicing multiple MFCs without introducing long delays, albeit at a slower data rate. Secondly, a fixed transparency simplifies scheduling of large amount of data with minimal computational load. Third, the semi-transparent structure of each MFC allows the use of buffering as a means to affect the relative timing of an MFC with respect another in order to interleave MFCs competing for the same link, but without buffering entire chains. Our hope is that this use of the buffer will enable significant savings in buffer space occupied relative to the UDP and MF-UDP of Section 3. Fourth, we anticipate bit rates of 10Gb/s in access networks at a time when bandwidths in core networks will be 100Gb/s. Transmission from access must then be up-shifted in rate, resulting in, for this case, 10% time occupancy. This would be accommodated naturally by a core transparency degree 10 times higher than that used in the access. Finally, for file transfer applications one desires to send as much information as fast as possible. For this transparency is a disadvantage. However, streaming applications have evolved to deliver segments of large files spread out over long time periods. A large transparency degree can be used to emulate streaming delivery while retaining the other worthwhile attributes described above.

4.2 Signalling and Control

Signaling is needed to establish and update the network state, schedule MFC transmission, acknowledge receipt, and many other functions. Information about an entire MFC, including length, scheduling information, priority, etc. can be easily contained within each MFC, MF in an MFC, or a small separate control packet. Signaling could be ‘in-band’ using periodic time slots within the MF transport structure, or ‘out-of-band’. Our preference is to exploit the ubiquitous availability of traditional IP networks for out-of-band signaling. In what follows, we assume that each of our media access and backbone routers (MAR, MBR in Fig. 2) are able to signal through a suitable IP network.

Global Control: Since each signaling event corresponds to of order 1 GB of data, the number of signaling events is small. It is then reasonable to use a centralized ‘state server’ to provide each router with global path, timing, and occupancy information. Each router updates status to the state server regularly, and the state server is able to calculate paths and approve initiation of a request for MFC scheduling, as discussed below. The state server must know the topology and is then able to make globally informed decisions to queries from routers. It is also useful to know the propagation time between nodes for efficient scheduling. Numerous methods can be implemented, like the ranging protocol used in PON, to estimate these times and report them to the scheduling server.

Distributed Control: Each router communicates directly with its neighbors and the state server. Each router continually updates the state server of status and load, and MARs request path and approval for MFCs from the state server. Approval does not guarantee success, but suggests high probability. Communication between routers along the path determines ultimate success, as described below. This minimizes latency in each MFC request-grant negotiation.

4.3 Scheduling

The objectives of scheduling are to organize transmission of concatenated chains in such a way as to minimize hardware complexity and power consumption, to maximize link utilization and to minimize buffer requirements. All of these objectives can be addressed through the use of an Expected Arrival Time (EAT_i) of the MFC to the next node in its path. This is computed based on the physical distance between the nodes, which is assumed to be known by each scheduling node, and included in a control packet CP (< 1 Kb). The CP is used to reserve resources for its associated MFC along its path. Many variations of the scheduling algorithm can be considered. For purposes of simulation, we defined an example that comprises the following basic steps:

- MFCs assembled at end user machine.
- A request packet containing at least the length of MFC, transparency degree, source and destination address is sent to media access router (MAR).
- MAR queries state server for path, propagation time associated with each hop in path.
- MAR estimates a Time-to-Transmit (TT) parameter. All routers along path use TT to search

for available time slots. TT is determined based on transmission and propagation delay from user to the MAR, hop distances along path and processing time for control packets at each node.

- MAR generates control packet (CP) and sends it to next node along path. CP contains sender/destination address, length of MFC, transparency degree, expected arrival time (EAT[Ni]), and ID that associates each CP to an MFC. EAT[Ni] indicates to the node receiving the CP the amount of time, after the reception of the control packet, before the arrival of first bit of the MFC .
- EAT field in CP is updated before forwarding CP to next hop node, in order to account for the additional transmission, buffering and processing delays at each node. This continues until destination (egress MAR) is reached or until the reservation process fails.
- If reservation succeeds, confirmation packet is sent over IP network directly to source node which starts transmission of MFC. If reservation fails, a "NACK" packet is sent over reverse path to free resources and source will retry after random back-off time.

Since the estimated EAT for the MFC is computed using the physical hop delay (known globally) and the expected arrival time is carried in the CP, there is no need for network-wide synchronization. A local timer at each node keeps track of the time differences between the reception of the CP and the expected arrival time of its relative MFC. The guard time in each MF compensates for the time uncertainties in this estimation process.

4.4 Simulation

The scheduling algorithm was implemented (details to be published) using OMNeT++ and the same topology used in Section 3. Each MF is assumed to contain 1Mb of data and each MFC is defined as the concatenation of 8000MF (i.e. 1GB of data). A fixed transparency degree of 8 was chosen for all MFC simulations. The performances of the MFC-based transport are then compared to those of UDP sources using MF (1Mb) and standard UDP packets (1500B).

Given the functioning of the algorithm presented in Section 4.3, the payload bits arrive at the node only if the reservation process was successful. In order to compare this to a connectionless protocol as in Section 3 (UDP) it was assumed that each source-destination pair would *attempt* to transmit on average the same amount of data. Hence for the UDP cases (both standard and MF-UDP) the "offered load" is the load physically reaching the bottleneck node. For the MFC case, the offered load is computed based on the number of reservation attempts per second. The maximum buffer size allowed for the bottleneck router (router 1 of Figure 3) was set to be 1.5MB for all cases.

4.5 Simulation Results

Link Utilization: As can be seen in Figure 5 the link utilization is very close for all 3 approaches tested for load conditions up to ~74%. Beyond this point the higher cost of dropping larger frames comes into play and the bandwidth efficiency of MF-UDP is reduced. For load greater than ~90% load packet dropping also affects standard UDP and its utilization drops below that of MFC. MFC utilization is ~9% higher for higher loads.

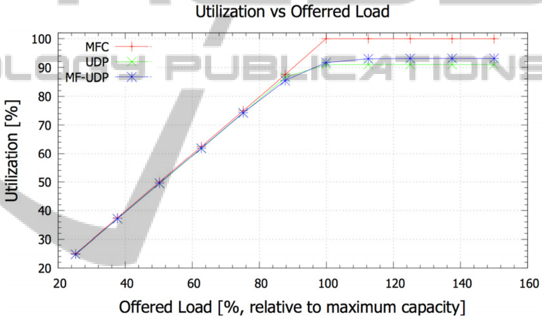

Figure 5: Link Utilization Vs Offered Load.

The highly structured MFC and the scheduling algorithm allow interleaving large amounts of data with link utilization similar to that of time division multiplexed systems and performance consistently better than both other cases under high load conditions.

Packet Dropping and Blocking: Packet dropping (two UDP cases) and blocking (MFC) are presented Figure 6. The random nature of the arrival for UDP packets allows for the possibility of filling the queue at the bottleneck node even if maximum load has not yet been reached. In the reservation-oriented MFC system call blocking only occur after the maximum bottleneck link capacity has been effectively reached.

It is important to note that when a packet is dropped, payload bits are discarded and these may have already used resources along their path (buffer space, switching power, bandwidth, etc.). When an MFC request is blocked, we are simply rejecting an *attempt* to transfer the data and not the data itself.

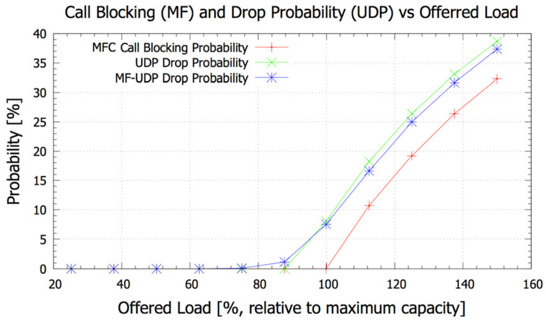

Figure 6: Packet Dropping and Call Blocking Vs Offered Load.

This advantage of reservation-based mechanisms in terms of efficiency of resource usage is well known.

Delay: To compare the delay performance a 1GB transaction was taken as a reference quantity. For MFCs, upon failing a resource reservation attempt, the source simply backs off for a random amount of time before re-attempting the reservation procedure. The back-off time is exponentially distributed with an average equal to the duration of an entire chain. Re-transmission attempts were also taken into account in the delay performance measurement, as shown in Figure 7. Up until 75% load, the delay experienced by the MFC system is virtually identical to that of both UDP cases. Beyond this point using MFCs reduces delay. Besides offering equal delay performances to UDP for the majority of the range of operation of the network, MFC also provides reliable data transfer, which cannot be achieved by UDP due to the statistical nature of arriving packets.

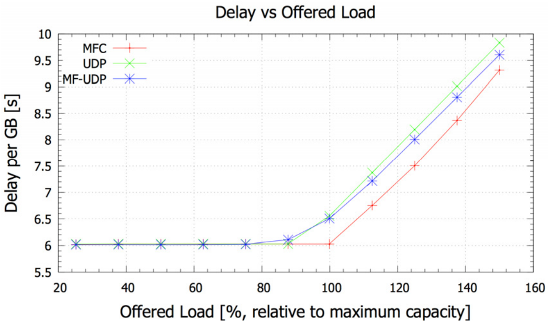

Figure 7: Delay (per GB of data transferred) Vs Offered Load.

Buffer Size: As shown in Figure 8, the buffer size required at low load for MFC and MF-UDP is higher than that occupied by standard UDP (with MF-UDP occupying the largest buffer space). As the load increases (>88%) MFC requires considerably less buffer space than both MF-UDP and standard UDP, which beyond a certain load quickly start to fill buffers up to the maximum capacity.

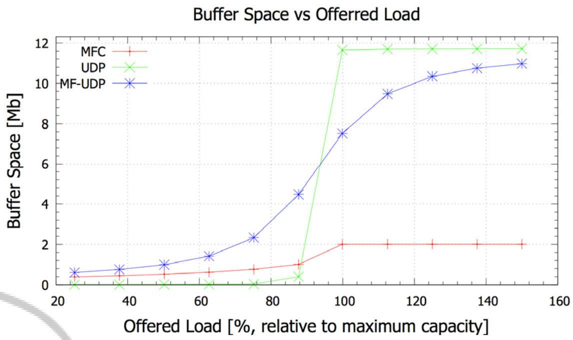

Figure 8: Buffer Occupancy Vs Offered Load.

The periodic data structure of the MFC and the scheduling algorithm bound the required buffer space to about 2-5 times less than that of both UDP approaches.

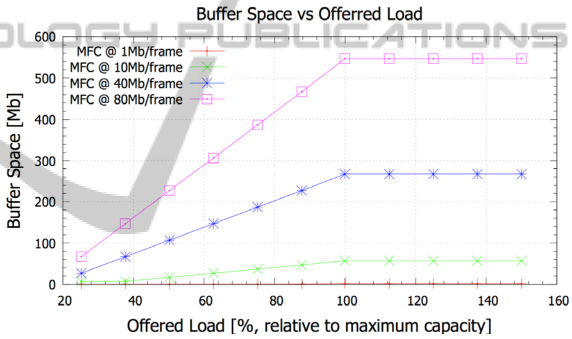

Figure 9: Buffer Occupancy Vs Offered Load for Various Frame Sizes for MFC.

For both UDP and MF-UDP the buffers are required to avoid dropped packets, while for MFC the buffers are used to align MFCs in time with scheduled slots. This results in a buffer size that depends much more on the size of the MF within the MFC than on the offered load. This dependency is shown in Figure 9 where various media frame sizes were tested for MFC. Varying frame size results in virtually identical performances in terms of blocking probability and utilization but a significant increase of the required buffer space. Similarly, reducing the MF size for each MFC can reduce the buffer space. Increasing the frame size for the MF-UDP case would result in buffer sizes that would simply become impractical. In the MFC case using frames larger than 1Mb may enable a relaxation of the reconfiguration speed for the switch fabric as well as a reduction of the CPU utilization (see section 5.1) while keeping the buffer size limited.

5 DISCUSSION AND ANTICIPATED BENEFITS

Results for MFC indicate a considerable advantage in terms of buffer size as well as a large reduction in processing power with respect to UDP. A comparison with TCP would have been useful in that, unlike UDP and more like MFC, TCP can guarantee the correct delivery of the file transferred. TCP acknowledgements and retransmissions would have made the delay per GB much higher and bandwidth utilization much lower than that achieved with UDP. In other words, for our simulation study UDP is the best-case scenario for IP networking.

5.1 Impact on Router Power Consumption

In (Aleksic, 2009) the issue of power consumption in large scale networks is addressed concluding that power consumption of header related functions is far larger than that consumed by the switching fabric. Apparently, most power is consumed in data processing functions (i.e. header parsing, address lookup, etc.), which must be carried out for each packet traversing a router.

MF-UDP and MFC offer significant reductions in the number of packets to be processed. Consider a state-of-the-art router working at 1 Tb/s. Depending on the manufacturer such a router can consume about 4 KW (CRS-3, 2012), (T1600, 2012), or an overall energy per bit of roughly 4 nJ/b.

Estimates of the power consumption of the various functional blocks of a similar IP router can be found in (Tucker et al. 2009). From this study it is clear that, other than power supply and cooling blocks, which are largely dependent on the energy needed by the other blocks, the forwarding engine consumes the most power, using about 32% of the energy supplied to the router.

Assuming an average packet size of 10 Kb (caida.org, 2012) means that a 1 Tb/s router will have to perform header related operations approximately 10^8 times per second (CRS-3, 2012). Organizing data in MFCs carrying roughly 1 GB of payload can reduce this by many orders of magnitude.

In addition, the required processing speed is reduced so that a much slower processing unit can handle the same data throughput. From the study presented in (Wang et al.,2006) it is also reasonable to say that the use of large frames, together with organizing the data within the sources in large blocks (i.e. MFs) will allow a significant reduction in CPU utilization in terms of number of memory accesses and IRQs that the CPU has to handle. This may lead to additional power savings in data storage servers.

Given the large size of the MFs, requirements on the reconfiguration speed are significantly relaxed, as well as the amount of control information needed to drive the switch fabric, with further impact on the power consumption. Further reduction in the amount of control information is achieved when using MFCs, as a result of the predictability of the periodic payload.

5.2 Power Consumption of Example MFC Router

A schematic of an MFC-based router based on a commercially available cross-point switch array is shown in Figure 10.

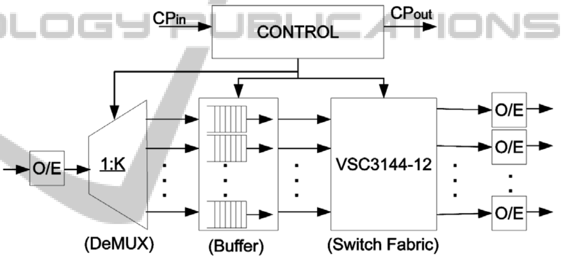

Figure 10: Concatenated Media Frame Router.

Input data is converted into the electrical domain and data streams from each channel are de-multiplexed into their constituent MFCs. Each MFC is then delayed by the amount indicated in its associated CP. The number of buffer queues needed depends on the number of chains the device is able to handle and is given by *# of dedicated buffer queues = # of input channels * # of MFCs per channel*. The total buffer size also depends on the number of simultaneous flows the device must handle and the transparency degrees of the chains.

At the output of the buffer stage, chains competing for the same output channels are synchronized in order to allow interleaving. MFCs are then passed to the switch fabric, which simply routes each MFC to the appropriate output with no further buffering or processing.

In order to estimate the power consumed by our proposed router shown in Figure 10, the VSC3144-11 has a switching capacity of 1.2 Tb/s with a power consumption of about 20 W (Vitesse, 2012).

Regarding the buffer stage, values for the power consumption of electronic memories are largely dependent on implementation and size. If we assume our MFC router would use a memory with similar size and structure to that used in (CRS-3, 2012), using the data on power consumption for a router from (Tucker et al., 2009) gives roughly 200 W for the buffer stage power consumption. Similarly we can estimate the power consumption of O/E/O blocks (including Tx/Rx equipment) to be about 280W.

The forwarding and routing engine, using about 32% and 11% of the total energy consumed by a standard IP router, respectively Tucker et al., 2011) (i.e. a total of ~720W (CRS-3, 2012)), are grouped in the "control" block of Figure 10 and, as a result of the drastic reduction in the number packet processed will be most likely reduced to negligible levels.

The input de-multiplexing stage, which de-multiplexes MFCs from each channel, can be modeled by (Tucker, 2011):

$$E_{DeMUX} = E_0 Log_2 k \tag{1}$$

Where E_0 is the energy per bit for a 1:2 de-multiplexer and k is the number of output ports. Assuming 2010 technology, E_0 = 10 pJ (Vitesse, 2012). We can set k = 144 in Equation (1) obtaining a total energy for the input stage of about 71W or 71 pJ/b.

Aside from power supply and cooling equipment, the power consumed by the MFR would be roughly 571W. Power supply and cooling would consume about 33% of the power consumed by the rest of the device (Peng et al., 2010), leading to total power consumption for the MFR of roughly 759W. This is roughly 19% that of a state-of-the-art IP router working in the same throughput range and load conditions (CRS-3, 1012). This is dominated by the power consumed by the buffer stage. Recognizing that these buffers function more like slowly reconfigurable delay lines than the buffers used in typical IP routers, further study may reveal even more significant reductions in power consumed.

6 CONCLUSIONS

In this paper we study the potential advantages of using an overlay network in which only large media frames (MFs) (1-10Mb) or concatenated frames (MFCs) are used to efficiently transfer large files. Numerical simulation is used to compare the use of MFs using a traditional UDP routing protocol (MF-UDP) to traditional UDP. Little difference is observed in delay, utilization and throughput, while the large packets dramatically reduce header-related processing load.

In an effort to reduce buffer size and improve resource utilization, a reservation-based networking approach is developed using MFCs and compared to MF-UDP and UDP through simulation. A reservation system is defined for scheduling MFC transmission, eliminating wasted network resources since data leaves the source only if service is guaranteed. Results show that buffer size can be reduced by at least a factor of 2 under high load conditions and the scheduling of large transactions can increase efficiency to close to 100%. Further advantages of using periodic, semi-transparent MFCs is the ability to schedule large amount of data with minimal header processing and to reduce the reconfiguration speed requirements of the switch fabric, ultimately reducing power consumption.

A representative MFC router is designed and power consumption is estimated, under conservative assumptions, to be roughly 20% that of a traditional IP router.

Other ongoing studies include methods to allow coexistence of scheduled MFCs and directly routed MF-UDP within the same routers and links, and extension of the simulations to other network topologies. Our objective is to more definitively articulate the cost-benefit trade-off, where the cost is the rather large barrier associated with the creation of a new overlay network.

In summary, the use of very large packets (MFs) and concatenations of these (MFCs) offers an interesting path to more power-efficient networking for the dominant and rapidly growing portion of Internet traffic that comprises very large (i.e. 1 GB) transactions.

REFERENCES

G. Fettweis, E. Zimmermann. (2008) "ICT Energy Consumption –Trends and Challenges", *11th International Symposium on Wireless Personal Multimedia Communications* (WPMC 2008)

Psc.edu (2012). Retrieved October 25, 2012 from http://staff.psc.edu/mathis/MTU/AlteonExtendedFrames_W0601.pdf

Divakaran, D. M.; Altman, E. (2009) Post, G.; Noirie, L.; Primet, "*From Packets to XLFrames: Sand and Rocks for Transfer of Mice and Elephants", in IEEE INFOCOM Workshops 2009.*

J. P. *Jue, V. M. Vokkarane, (2005)* "*Optical Burst*

Switched Networks", Ed. Springer 2005.

Chan, V. W. S. (2010) "*Optical flow switching*", Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 *Conference on (OFC/NFOEC)*, vol., no., pp.1-3, 21-25 March 2010.

Ajay Mahimkar, Angela Chiu, Robert Doverspike, Mark D. Feuer, Peter Magill, Emmanuil Mavrogiorgis, Jorge Pastor, Sheryl L. Woodward, and Jennifer Yates. (2011). Bandwidth on demand for inter-data center communication. In*Proceedings of the 10th ACM Workshop on Hot Topics in Networks* (HotNets-X). ACM, New York, NY, USA, , Article 24 , 6 pages

Yong Liu, Kee Chaing Chua, Gurusamy Mohan. (2010) "*Achieving High Performance Burst Transmission for Bursty Traffic using Optical Burst Chain Switching in WDM Networks*", IEEE Transactions on Communications, Vol.58, Issue 7 pp. 2127-2136, 2010.

Tucker, R. S., Parthiban, R., Baliga, J., Hinton, K., Ayre, R. W. A., Sorin, W. V. (2009) *Evolution of WDM Optical IP Networks: A Cost and Energy Perspective*, IEEE J. Lightwave Technology, Vol. 27, Iss. 3, Feb. 2009, pp. 243-252.

R. S. Tucker. (2006) "*The role of optics and electronics in high-capacity routers*", J. Lightwave Technol., vol. 24, pp. 4655 - 4673, 2006.

Slavisa Aleksic, Matteo Fiorani, and Maurizio Casoni (2011) "*Energy Efficiency of Hybrid Optical Switching", ICTON 2011*

S. Peng, K. Hinton, J. Baliga, R. S. Tucker et.al. (2010)"Burst Switching for Energy Efficiency in Optical Networks", *OSA/OFC/NFOEC 2010*

OMNeT++ (2012) Discrete Event Simulation Tool, http://www.omnetpp.org.

Aleksic, S. (2009) "*Analysis of Power Consumption in Future High-Capacity Network Nodes*," Optical Communications and Networking, IEEE/OSA Journal of, vol.1, no.3, pp.245-258, August 2009.

CRS-3, single shelf system cisco data sheet (2012) (http://www.cisco.com/en/US/prod/collateral/routers/ps5763/CRS-3_4-Slot_DS.html)

T1600 Juniper networks data sheet (2012) (http://www.juniper.net/techpubs/en_US/release-independent/junos/topics/reference/specifications/t1600-specifications-power-requirements.html)

caida.org (2012). Data retrieved October 25, 2012, from www.caida.org

Wilson Yong Hong Wang, Heng Ngi Yeo, Yao Long Zhu, Tow Chong Chong, Teck Yoong Chai, Luying Zhou, Jit Bitwas. (2006) "*Design and development of Ethernet-based storage area network protocol*", Computer Communications, Volume 29, Issue 9, 31 May 2006, Pages 1271-1283, ISSN 0140-3664, 10.1016/j.comcom.2005.10.004

Vitesse VSC3144-11, Data Sheet. (2012). Retrieved online November 2, 2012 from http://www.vitesse.com

Tucker, R. S. (2011), "*Scalability and Energy Consumption of Optical and Electronic Packet Switching*," *Lightwave Technology, Journal of*, vol.29, no.16, pp.2410-2421, Aug.15, 2011.