Flow Network Model of Bulk Transmission

Elizabeth Varki
University of New Hampshire
varki@cs.unh.edu

Adam Villa
University of New Hampshire
varki@cs.unh.edu

Abstract

Bulk transmission refers to the mailing of big data sets from a sender's network to the receiver's network via the internet. Whereas conventional routing algorithms focus on finding quick routes to the receiver, bulk routing algorithms focus on finding cheap, high bandwidth routes to the receiver. This paper is the first to formulate a comprehensive problem statement of bulk transmission via the internet. For a given internet platform and bulk data set size, the paper shows that the transmission start time is the key parameter that determines the cost/time of transmission between the sender and receiver. For each initiate time, the underlying time-varying flow network of the internet changes, necessitating the solution of several time-varying flow models to find the optimum initiate time and routing path. A key contribution is the result that flow networks fail as a model for generation of bulk transmission routing. This negative result lays the groundwork for the development of a new model for routing in global systems.

1 Introduction

This paper formulates the problem of transmitting bulk data set from a sender's network to a receiver's network via the internet. The bulk data sets are in the tens of gigabyte to the terabyte range. With the proliferation of big data, bulk transmission now encompasses the transmission of petabyte data sets. Bulk transmission first gained prominence when the LHC project started; the project was expected to generate petabytes of data that had to be transmitted to researchers around the globe [4]. The LHC project is now on-line with its data sets being transmitted on private high-speed optical links. Research labs that are on this high-speed network can electronically access LHC's data. A lab that is not hooked up to the high-speed network has to rely on the internet or postal mail for access to LHC's big data sets.

Bulk transmission has received attention in recent years. Cloud players like Amazon, Microsoft, Google, Yahoo!, Akamai, and Facebook have to move bulk data sets between data centers. Corporations that transmit bulk data regularly between their sites could justify building, purchasing, or leasing network links. However, bulk transmission between cloud providers and their clients is a growing necessity, and it is not cost effective to have private high-speed links to every client's network. Moreover, ordinary users - in homes, offices, labs, and schools - may also occasionally want to transmit a larger than normal volume of data. In these cases, one has to rely on the internet or postal mail for transmission.

The internet is the largest and most diverse distributed system. Each internet application competes for bandwidth to transmit its files. For a fixed bandwidth, transmission time increases as the file size increases. For a given file size, transmission time decreases as the transmission bandwidth increases. Consequently, for fast transmission, bulk data sets require a disproportionately large fraction of the internet's bandwidth. The internet is a shared resource, and if an application grabs a large fraction of the bandwidth, then this could negatively impact the performance of other internet applications. The objective of availing large bandwidth without degrading the performance of other internet applications is the essence of the challenge of bulk transmission.

The study of bulk transmission via shared, public networks is fairly new, so the existing literature is limited. The bulk transmission protocols can be divided into two categories based on whether bulk data are transmitted directly from sender to receiver, or whether bulk data are transmitted from sender to intermediate storage nodes and from there to the receiver. That is, bulk transmission protocols are categorized as either end-to-end or store-and-forward. The current trend is in favor of store-and-forward protocols for bulk transmission [8]. However, there is no clear understanding of when and why one type of protocol is better than another. There is no common framework to evaluate the two categories of transmission techniques. Moreover, the papers are primarily about store-and-forward versus end-to-end for bulk transmission, not about how the transmission should be routed along the internet. There is a clear need for a modeling tool to evaluate the two categories of protocols and generate optimum routing paths for bulk data.

This paper is the first to formulate a problem statement of the bulk transmission problem, and develop a mathematical model that is convenient for analysis. Prior papers [5, 7] have developed routing algorithms for transmission of bulk data sets, which is the step that comes after problem formulation and model development. Therefore, this paper is a step back and is the prerequisite to development of routing algorithms. The motivation for this paper is that a representative model of bulk transmission could lead to cleaner algorithms with lower complexity than the current state-of-the-art. The contributions of the paper are:

1. detailed formulation of the bulk transmission routing problem, and development of a comprehensive model;

2. presentation of a common framework to compare existing bulk transmission protocols;

3. proof that with regard to bulk transmissions, end-to-end protocols cannot outperform store-and-forward protocols;

4. proof that for a given internet platform and bulk data set size, transmission start time is the key input parameter that determines performance;

5. generation of quick performance bounds for bulk transmission; and

6. proof that time-varying flow networks fail as a model for bulk routing algorithms of global systems.

This last contribution is an oxymoron - on the one hand, the contribution of the paper is the comprehensive flow model, on the other hand, we prove that the model fails as a routing tool for transfers that span continents. The essence of the bulk transmission problem is finding routing paths that satisfy time and cost constraints, and the model fails at this task for transfers that span the globe. This negative result is a contribution of this paper since it highlights the need for a new modeling tool for flow routing in global networks.

2 Metric to evaluate bulk routing

The internet is a collection of independent networks, where each network is an Autonomous System (AS). Big data transmission, also known as bulk transmission, is the routing of a bulk number of packets from the sender's AS to the receiver's AS via one or more transit ASs. The fundamental difference between standard and bulk internet transmissions is the size of the transmitted data set - say, 100 MB vs. 100 GB. Standard internet transmissions are routed using routing algorithms with objectives to minimize hop count and business cost. If performance of a routing algorithm is measured by metrics such as throughput (or transmission time), then the standard routing algorithms would perform poorly for bulk transmissions. For example, consider 20 Mb/s and 2 Gb/s links: while the 100 MB file is transmitted in less than a minute over either link, the 100 GB file takes 11.36 hours over the 20 Mb/s link and less than 7 minutes over the 2 Gb/s link, a reduction of 99% of transmission time. For bulk transmissions, routes with maximum capacity (throughput) are optimal. Thus, the optimal route for bulk transmissions is not necessarily the route selected by standard internet routing. Consequently, routing of bulk transmissions should be treated differently from routing of standard internet transmissions.

The performance metric of relevance to bulk transmission is the throughput - higher the throughput of the transmission, the smaller is the time to transmit a bulk data set. Therefore, the objective of a bulk routing algorithm is to find the highest capacity links between sender and receiver. Multiple transmission paths can be opened between sender and receiver with concurrent transmission along all paths. For example, suppose there are 2 transmission paths between sender and receiver: sender-transitA-receiver, sender-transitB-receiver. If the sender and receiver have 2 Gb/s links but the transit networks only have 1 Gb/s links, then by opening both paths between sender and receiver, it is possible to avail of the maximum 2 Gb/s capacity. Some parallel transmission protocols are GridFTP [2], BitTorrent [11], and Slurpie [12]. The advantage of parallel transmission protocols is best seen when sender and receiver ASs, along with transit ASs have high bandwidth availability at the same time.

High bandwidth is required for big data transmissions to complete in a reasonable amount of time. However, greedy protocols that grab the maximum possible bandwidth would have a negative impact on the performance of other internet applications. This could result in network administrators shutting down the bulk transmission. Therefore, bulk routing algorithms must address the shared nature of the internet. The direct approach to minimizing the negative impact of bandwidth usage is to

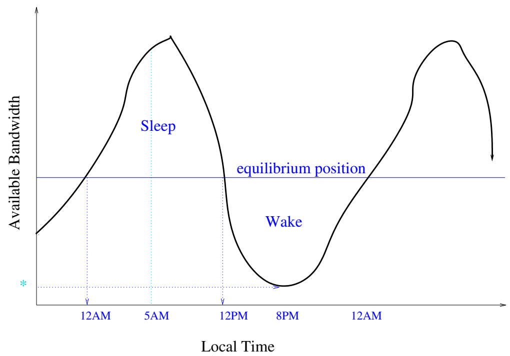


Figure 1: Available bandwidth distribution

allow other traffic to go ahead. The Qbone protocol [14] achieves this objective by lowering the priority of bulk packets. Lowering priority of standard TCP packets has also been presented as a viable option for bulk transmissions [3, 15]. Another approach is to find the least congested paths from sender to receiver. The most promising of these approaches is OpenFlow [9], where routing is performed by a centralized software router with knowledge of entire network traffic, rather than by routers with local traffic awareness.

Bulk transmissions share the internet with other applications; many of these applications are real-time and have QoS requirements of low latency and low jitter. During certain hours there is heavy bandwidth usage from these time-critical applications. Bulk transmissions are delay tolerant, so priority should be given to these time-critical applications by ensuring that bulk transmissions only avail of the remaining “free” bandwidth. Therefore, the objective function of bulk routing algorithms is to find links with the highest free capacity between sender and receiver.

3 Available bandwidth distribution

The objective of bulk transmission routing is to find high capacity links that are not congested. High available capacity is required in order to guarantee high throughput for bulk transmissions without disrupting other internet applications. The physical capacity of a link is fixed, but free capacity is variable. Free capacity is a function of bandwidth usage - higher the bandwidth usage, lower the free capacity. The bandwidth usage in an AS has a diurnal wave pattern that mimics users’ sleep-wake cycle [6]. There is more bandwidth during the early AM hours and less bandwidth during the PM hours. Thus, available bandwidth has a diurnal wave pattern where the valleys in bandwidth usage correspond to peaks in free capacity. Figure 1 shows bandwidth availability distribution by time of day for a typical AS. Table 1 tabulates free bandwidth; the distribution has been deliberately simplified since it is used in examples through the paper.

Table 1: Simplified example highlighting the diurnal wave distribution of bandwidth availability. LTC = Local Time Clock; BW/hr = Bandwidth per hour; * = available base bandwidth; C = Constant value

LTC	BW/hr	LTC	BW/hr
12:00 AM	* + 10C	12:00 PM	* + 8C
01:00 AM	* + 14C	01:00 PM	* + 7C
02:00 AM	* + 16C	02:00 PM	* + 6C
03:00 AM	* + 17C	03:00 PM	* + 5C
04:00 AM	* + 18C	04:00 PM	* + 4C
05:00 AM	* + 18C	05:00 PM	* + 3C
06:00 AM	* + 17C	06:00 PM	* + 2C
07:00 AM	* + 16C	07:00 PM	* + 1C
08:00 AM	* + 14C	08:00 PM	*
09:00 AM	* + 12C	09:00 PM	* + 1C
10:00 AM	* + 10C	10:00 PM	* + 4C
11:00 AM	* + 9C	11:00 PM	* + 8C

For a given network, the valleys in bandwidth usage correspond to peaks in free capacity. For the example network in Table 1, the maximum bandwidth is available during hours 4:00-6:00 AM, and minimum bandwidth is available during peak usage time of 8:00 PM. For optimal performance in terms of cost and time, bulk transmissions should be scheduled during valleys in the bandwidth usage. However, the valleys in bandwidth usage at the various ASs along the path from sender to receiver may not be synchronized. The bandwidth usage cycle mimics the users’ sleep-wake cycle which is dependent on the location of the ASs. If the sender and receiver are situated in different time zones, then the high free capacity times of the sender, receiver and transit ASs do not coincide. Consequently, even if the sender, receiver, and transit networks all have equal physical capacity links and similar bandwidth usage pattern, it is still possible that the throughput of the bulk transmission is far below the peak free capacity of the ASs.

The throughput of an end-to-end routing algorithm depends on the smallest free capacity link on the route. For example, suppose the sender and receiver networks have 2 Gb/s free capacity during the hours 1:00 AM - 8:00 AM; during the rest of the day, bulk transmission is limited to 20 Mb/s (set by the network administrators to ensure QoS of time-critical internet applications). Assume that the transit ASs have ample bandwidth. If there is a 8 hour time zone difference between the sender’s AS and the receiver’s AS, then the high capacity times at the sender and receiver are out of sync, and transmission bandwidth is 20 Mb/s, not 2 Gb/s (for a total transmission time of 11.36 hours).

To overcome temporal non-synchronization of free capacity at sender/receiver ASs, store-and-forward routing algorithms have been proposed for bulk transmissions [8]. Instead of directly transmitting the bulk data set from sender to receiver, data sets are temporarily stored in storage hubs along the path until bandwidth opens up in the forwarding AS. Reconsider the previous example: with store-and-forward, a 100 GB file would be transmitted from the sender AS to a storage device in a transit AS; the transmission takes 6.67 minutes at 2 Gb/s. The data set is later transmitted to the receiver AS during its high free capacity time of 2 Gb/s. Thus, the actual transmission time at any network is 6.67 minutes for a total transmission time of 13.34 minutes, as opposed to 11.36 hours with end-to-end.

A few papers have evaluated store-and-forward approach for bulk transmission. An early paper [13] used simulations to show that peak traffic and cost are reduced with store-and-forward, albeit with an increase in latency when compared to end-to-end protocols. Laoutaris et al. [7] and Chhabra et al. [5] have studied routing of store-and-forward bulk protocols. These papers use flow networks to generate routing paths that minimize response time of the transmission. Simulations are used to compare their protocols against random store-and-forward protocols (selecting routing paths without the objective of minimizing response time) and end-to-end protocols like BitTorrent.

Bulk transmissions over the internet is a fairly new area of research. Currently, there is a lack of understanding of how bulk transmissions fundamentally differ from standard transmissions. Subsequently, the research issues pertaining to this area are ill-defined. While it seems that store-and-forward should be the underlying transmission approach, the reasons for the advantages over end-to-end are murky since simulations do not provide explanations. A couple of prior papers [5, 7] have developed routing algorithms from flow networks, but the algorithms have high complexity and are not scalable to the global level. In order to develop efficient bulk transmission protocols, it is necessary to first develop a robust model that can be used to understand and evaluate the characteristics of big data transmission over a global, shared internet. This paper develops a complete flow model of bulk transmissions.

4 Flow network model

The goal of this section is to develop a mathematical framework for bulk transmission routing. The internet can be represented by the sender and receiver ASs, and the transit ASs that connect them. The key feature of an AS that is relevant to bulk transmissions is the total physical bandwidth link to the internet (*i.e.,* , to other ASs). For example, if a sender AS has a single 10 Gb/s link to a transit AS, then bulk transmission cannot exceed this capacity; or, if a transit AS on the path from sender to receiver has maximum of 5 Gb/s incoming link and 50 Gb/s outgoing link, then the throughput routed via this transit AS cannot exceed 5 Gb/s.

Bulk transmissions ply over links with varying free capacities, where an AS's free capacity follows a diurnal wave pattern. In order to model bulk transmissions plying over a dynamic internet, time-varying flow networks in which the capacity of the edge varies with time-of-day is used. The source and sink vertexes of the flow network model the sender and receiver ASs, while the intermediate vertexes model the transit ASs. Each vertex, except the sender, has one or more incoming edges that represent the incoming internet flow into the AS from other ASs along a path from sender to receiver. Each vertex, except the receiver, has one or more outgoing edges that represents the outflow from the AS to other ASs in the internet routing path. Since ASs experience diurnal bandwidth usage, the modeling covers a period of 24 hours starting from the time that the transmission is initiated.

We now present translation of the bulk transmission routing problem to the mathematical framework of flow networks.

Definition 1 *Bulk transmission routing is modeled by a time-varying flow network, N = (V, E, b), where V is the set of vertexes representing ASs, E is the set of edges representing internet links between the ASs, b(x, t) ≥ 0 is the storage capacity of vertex $x \in V$ at time t, and b(x, y, t) is the free bandwidth capacity of edge $(x, y) \in E$ at time t. The time t is relative to transmission initiation time, so t is equal to the transpired time since initiation at t = 0; $t = 0, 1, 2, \cdots, T-1$, where T is the cycle time and is the maximum allowable flow time from the sender s vertex to the receiver r vertex.*

Thus, b(x, y, t) ≥ 0 is the maximum amount of bulk data flow from x to y, when the flow departs from x at time t. The clock starts when bulk transmission is initiated at t=0. The unit for time can be 1 second, 5 minutes, 1 hour, or any appropriate time division. The value of T is set according to the chosen unit for t. For example, if hour is chosen as time unit, then T=24 since the distribution of free capacity has a diurnal wave pattern. Throughout this paper, for continuity and readability, we use hour as the time unit. For example, b(x, y, 5) is the (x, y) edge capacity at hour 5, where transmission is initiated at start of hour 0. Referring to Table 1, suppose transmission is initiated at 2:00 AM, then t = 0 at 2:00 AM, t = 5 at 7:00 AM, and $b(x, y, 5) = * + 16C$, where * is the base bandwidth on the link.

The vertex capacity b(x, t) is relevant only to flows that wait at the vertex when there isn't sufficient band-

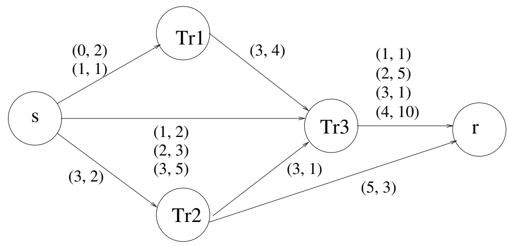


Figure 2: Time varying flow network modeling bulk transmission

width to move forward. This happens when the outflow capacity is less than the sum of inflow and stored capacity at vertex x during time t. The edge capacity b(x, y, t) represents the capacity along edge (x, y) from the start of hour t until the end of hour t. The total free capacity is the integral of the wave function representing free bandwidth during t. From the perspective of the internet, this represents the amount of bulk data that can be transmitted from AS x to AS y during 60 minutes of hour t. For example, if administrators of AS x only permit fixed 20 Mb/s during hour 8 (relative to when the transmission is initiated) to AS y, then b(x, y, 8) = 9 GB.

Example 1 *Consider the bandwidth distribution shown in Table 1. For simplicity, let base bandwidth * be 0 and let C be 1. Suppose bulk transmission is initiated at 10:00 AM LTC. Then, t=0 at LTC=10:00 AM, t=1 at LTC=11:00 AM, ..., t=23 at LTC=9:00 AM. b(x,y,0)=10, b(x,y,1)=9, b(x,y,2)=8, b(x,y,3)=7, b(x,y,4)=6, b(x,y,5)=5, b(x,y,6)=4, b(x,y,7)=3, b(x,y,8)=2, b(x,y,9) = 1, b(x,y,10)=0, b(x,y,11)=1, b(x,y,12)=4, b(x,y,13)=8, b(x,y,14)=10, b(x,y,15)=14, b(x,y,16)=16, b(x,y,17)=17, b(x,y,18)=18, b(x,y,19)=18, b(x,y,20)=17, b(x,y,21)=16, b(x,y,22)=14, b(x,y,23)=12.*

The next example presents a flow network with cycle time T set to 5.

Example 2 *Consider the time-varying network N shown in Figure 2. Here, s is the sender vertex, r is the receiver vertex, Tr1, Tr2, Tr3 are the transit vertexes. The two numbers inside each pair of brackets associated with an edge (x, y) are t, and b(x, y, t) respectively. For example, (0, 2) near edge (s, Tr1) means that during hour 0, the time at which the transmission is initiated, at most 2 units of bulk data can be transmitted from s to Tr1. The maximum flow time T=6. If bulk transmission is not permitted along an edge (x, y) during a hour t, t ¡ T* (i.e., *b(x, y, t) = 0), then the bracket is not shown in the figure. For example, Edge (s, Tr1) has no capacity set aside for bulk transmissions during hours 2, 3, 4, 5, so these are not shown.*

A path, p, of length k in N is a sequence $p = \langle v_0, v_1, v_2, \cdots, v_k \rangle$ of vertexes such that $v_o = s$, $v_k = r$, and $(v_{i-1}, v_i) \in E$ for i = 1, 2, ..., k. Let f(x, y, τ) be the value of the flow departing x at time τ to traverse the edge (x, y); and let f(x, τ) be the value of the flow stored in vertex x at the end of time τ. Let λ specify a set of paths from sender to receiver, and f(λ, t) be the total flow with solution λ within the time limit $t < T$. Then,

$$f(\lambda, t) = \sum_{(x,r) \in E, \tau \leq t} f(x, r, \tau) \quad (1)$$

and

$$f(\lambda, t) = \sum_{(s,x) \in E, \tau \leq t} f(s, x, \tau) - \sum_{x \in V - \{s,r\}} f(x, t) \quad (2)$$

where

$$f(x, t) = \sum_{(v,x) \in E, \tau \leq t} f(v, x, \tau) - \sum_{(x,v) \in E, \tau \leq t} f(x, v, \tau) \quad (3)$$

It follows that $f(\lambda, T)$ is the value of flows sent from s to r within the time limit T. From the view point of bulk transmission routing, $f(s, 0^-)$ is the size of the bulk data set at the sender node just before transmission at time t=0. $f(\lambda, T)$ is the size of the data set that can be transmitted from sender vertex s to sender vertex r using routing paths λ within time T.

Bulk transmission routing algorithms address the following optimization problems:

1. maximize the size of the bulk data set that can be transmitted from sender to receiver within cycle period; and

2. minimize the time to transmit a bulk data set from sender to receiver.

In the domain of flow networks, the problems translate to the following objective functions.

Objective function 1 *Generate* λ^{max} *such that* $f(\lambda^{max}, T) \geq f(\lambda, T), \forall \lambda$ *in N.*

Objective function 2 *For a given* $f(s, 0^-)$, *generate* λ^*, *where* $f(s, 0^-) = f(\lambda^*, \tau) > f(\lambda, t)\ \forall \lambda$ *in N when* $t < \tau$.

Objective function 1 is equivalent to finding the maximum flow in a time-varying network. Objective function 2 is equivalent to finding the universal maximum flow in a time-varying network. There are algorithms for computing the maximum flow and universal maximum flow in time-varying flow networks [16]. Equivalently, these algorithms can be used to compute the routing of bulk transmissions over the internet. Note that there are other objective functions, such as latest send time so that file arrives within T; if objective functions 1 and 2 are solved, then so can the others.

4.1 End-to-end vs. Store-and-forward

Bulk data could be directly transmitted from sender to receiver using an end-to-end protocol, or bulk data could be transmitted to intermediate storage servers in transit ASs using a store-and-forward protocol. Both approaches are modeled using the flow network N.

Definition 2 *For end-to-end bulk transmission,* $b(x,t) = 0, \forall x \in V$. *For store-and-forward bulk transmission,* $b(x,t) \geq 0, \forall x \in V$.

For store-and-forward bulk transmissions, it is usually assumed that each AS has plentiful (infinite) storage. To distinguish the two flows for network N, let λ_ε and λ_S represent end-to-end and store-and-forward solutions, respectively.

Example 3 *Solve the first objective function, namely, maximum flow with end-to-end transmission for the flow network in Figure 2. The solution* λ_ε*:*

```
t=1:1 unit of flow on path p1=(s,Tr3,r).
t=2:3 units of flow on p2=(s,Tr3,r).
t=3:1 unit of flow on p3=(s,Tr2,Tr3,r).
```

$f(\lambda_\varepsilon, 1) = 1$, $f(\lambda_\varepsilon, 2) = 4$, $f(\lambda_\varepsilon, 3) = f(\lambda_\varepsilon, 4) = f(\lambda_\varepsilon, 5) = 5$

Example 4 *Solve objective function 1, namely, maximum flow with store-and-forward transmission for the flow network in Figure 2. The solution* λ_S*:*

```
t=0: f(s,Tr1,0)=2; f(Tr1,0)=2.

t=1: f(s,Tr1,1)=1;
     f(s,Tr3,1)=2; f(Tr3,r,1)=1;
     f(Tr1,1)=3; f(Tr3,1)=1, f(r,1)=1.

t=2: f(s,Tr3,2)=3; f(Tr3,r,2)=4;
     f(Tr1,2)=3; f(Tr3,2)=0, f(r,2)=5.

t=3: f(Tr1,Tr3,3)=3; f(Tr3,r,3)=1;
     f(s,Tr3,3)=5;
     f(s,Tr2,3)=2;
     f(Tr2,3)=2; f(Tr3,3)=7; f(r,3)=6.

t=4: f(Tr3,r,4)=7;
     f(Tr2,4)=2; f(r,4)=13.

t=5: f(Tr2,r,5)=2;
     f(r,5)=15.
```

$f(\lambda_S, 5) = 15$.

Example 5 *Solve objective function 2, namely, minimum time to transmit data set of size 14 with store-and-forward transmission for the flow network in Figure 2. The solution* λ_S^**:*

```
t=0: f(s,Tr1,0)=2; f(Tr1,0)=2.

t=1: f(s,Tr1,1)=1;
     f(s,Tr3,1)=2; f(Tr3,r,1)=1;
     f(Tr1,1)=3; f(Tr3,1)=1, f(r,1)=1.

t=2: f(s,Tr3,2)=3; f(Tr3,r,2)=4;
     f(Tr1,2)=3; f(Tr3,2)=0, f(r,2)=5.

t=3: f(Tr1,Tr3,3)=3; f(Tr3,r,3)=1;
     f(s,Tr3,3)=5;
     f(s,Tr2,3)=2; f(Tr2,Tr3,3)=1;
     f(Tr2,3)=1; f(Tr3,3)=8; f(r,3)=6.

t=4: f(Tr3,r,4)=8;
     f(Tr2,4)=1; f(r,4)=14.
```

$f(\lambda_S^*, 4)$ *= 14, and t = 4 is the minimum time. Note that the above solution is the maximum flow in time t=4.*

The solutions in Example 4 and Example 5 are different. The solution in Example 4 does not give the maximal flow of 14 at time 4. If the solution in Example 5 is extended to time 5 by adding the flow b(Tr2, r, 5) = 1, then this solution gives the the maximal flow at time 5. One can check that this solution is maximal $\forall t \leq 5$. This second solution is the universal maximal flow.

Our model has provided a common framework to compare the two major types of bulk transmissions. Using this framework, end-to-end is a special case of store-and-forward, and $\lambda_\varepsilon \subseteq \lambda_S$.

Result 1 $f(\lambda_\varepsilon, t) \leq f(\lambda_S, t)\ \forall t < T$

In earlier papers [7], simulations showed that NetStitcher, a store-and-forward bulk transmission protocol, outperformed BitTorrent, an end-to-end routing protocol. Result 1 is the theoretical basis for the superior performance of NetStitcher.

5 Optimal initiate time

The last section assumes that the initiate time is a fixed input parameter. Here, we relax this assumption to determine whether the initiate time makes a difference to the

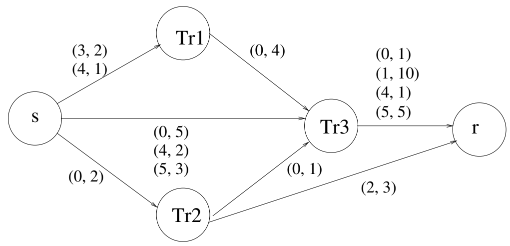


Figure 3: Transformed flow network of Figure 2 when t shifted by 3

maximal flow and the universal maximal flow. If this is the case, then initiate time should be chosen so that the objective functions are maximized.

How does the choice of the initiate time impact the underlying flow network and consequently the bulk transmission routing? The free capacity of a link is a function of the local time at the corresponding AS, as shown in Figure 1 and Table 1. We introduce new notation to show the link between free capacity and local time at the corresponding AS. Let $c(x, y, t)$ represent the free capacity along (x, y) at x's local time t.

Example 6 *Consider the bandwidth distribution shown in Table 1. As in Example 1, for simplicity, let base bandwidth * be 0 and let C be 1. Then, c(x,y,0)=10, c(x,y,1)=14, c(x,y,2)=16, c(x,y,3)=17, ..., c(x,y,20)=0, c(x,y,21)=1, c(x,y,22)=4, c(x,y,23)=8.*

The edge capacity for (x,y) in the flow network is represented by b(x, y, t) where t is the time relative to transmission initiate time. Thus, b(x,y,t) is the edge capacity with relation to initiate time, while c(x,y,t) is the edge capacity with relation to x's local time. In order to relate the two distribution, we introduce another notation. Let l(x) represent the local time at $x \in V$ when transmission is initiated at sender s; l(s) is the local time at sender s when transmission is initiated, but when it is clear from the context, l, not l(s), is used. It follows that

$$b(x, y, t) = c(x, y, (t + l(x)) \text{mod } 24), \quad 0 \leq t \leq 23$$

Example 7 *Suppose edges (s, v) and (x, y) have identical distribution of free capacity given by Table 1. That is, $c(s, v, t) = c(x, y, t) \quad 0 \leq t \leq 23$. Suppose x is 4 hours ahead of the sender. When the sender initiates transmission at 10:00 AM, it is 2:00 PM at x. For the sender s, l(s)=10: b(s,v,0)=10, b(s,v,1)=9, ..., b(s,v,21)=16, b(s,v,22)=14, b(s,v,23)=12. For vertex x, l(x)=14: b(x,v,0)=6, b(x,v,1)=5, b(x,v,2)=4, b(x,v,3)=3, ..., b(x,v,20)=10, b(x,v,21)=9, b(x,v,22)=8, b(x,v,23)=7.*

Changing the initiate time transforms the flow network since the distribution of edge capacity changes. Figure 3 shows the flow network of Figure 2 with the initiate time shifted by 3. If the flow network is mapped to a coordinate system with time and capacity on the axes, then shifting the initiate time is equivalent to a translation of the flow network. For the transformed flow network in Figure 3, $f(\lambda_\varepsilon, 3) = 1$, $f(\lambda_\varepsilon, 4) = 2$, $f(\lambda_\varepsilon, 5) = 5$; and $f(\lambda_S, 5) = 12$. Thus, this transformation changes the maximal flow and the universal maximal flow.

The initiate time can be set to any hour of the day, and the flow network models the state of the ASs linking the sender and receiver ASs for a 24 hour period from this initiate time. For each value of initiate time, t=0, 1, 2, .., 22, 23, a transformed flow network exists and has to be solved for maximal/universal flow. The updated definition of the flow network modeling bulk transmission is:

Definition 3 *Bulk transmission routing is modeled by the set of time-varying flow networks, N = $\{N^l = (V, E, b) | l = 0, 1, ..., T-1\}$ where V, E, b are as defined earlier, and l is the local time at the sender s when transmission is initiated.*

Let $f(\lambda, T, \theta)$ represent the flow in time T when transmission is initiated at s's local time θ. With varying initiate time, the objective functions for bulk transmissions are:

Objective function 3 *Find λ^{max} with initiate time θ such that $f(\lambda^{max}, T, \theta) \geq f(\lambda^{max}, T, l), \forall 0 \leq l < T, l \neq \theta$.*

Objective function 4 *For a given $f(s, 0^-)$, find λ^* with initiate time θ such that $f(s, 0^-) = f(\lambda^*, \tau, \theta) > f(\lambda, t, l)$, $\forall 0 \leq l < T$, $l \neq \theta$ when $t < \tau$.*

The initiate time is included in the notation only when it is relevant to the computation. For example, the next result shows that for end-to-end transmission, the initiate time has no impact on the maximal flow.

Result 2 *$f(\lambda_\varepsilon^{max}, T, \theta) = f(\lambda_\varepsilon^{max}, T)$ where θ is the local time at s when transmission is initiated.*

Proof: For end-to-end,
$b(x, t) = 0 \implies f(x, t) = 0 \;\; \forall x \in V - \{s, r\}$.
It follows that

$$\sum_{(s,x) \in E} f(s, x, \tau, \theta) = \sum_{(x,r) \in E} f(x, r, \tau, \theta) \quad 0 \leq \tau < T$$

where τ is the time relative to the initiate time θ. Since f(x,t) = 0, the flow arriving at the receiver at any time t is dependent only on the edge capacities at time t, not on the flow at time prior to t. Thus,

$$
\begin{aligned}
f(\lambda_{\varepsilon}^{max}, T, \theta) &= f(\lambda^{max}, 0, \theta) + \\
&\quad f(\lambda^{max}, 0, (\theta+1) mod\, 24) + \\
&\quad \cdots + f(\lambda^{max}, 0, (\theta+23) mod\, 24) \\
&= f(\lambda_{\varepsilon}^{max}, T)
\end{aligned}
$$

□

In order to solve objective functions 3 and 4, solutions must be found for all 24 flow networks in set N. For the original objective functions 1 and 2, only a single flow network relating to the fixed initiate time is to be solved. While maximal flow for end-to-end is not dependent on the initiate time (Result 2), universal maximal flow (*i.e.,* the minimum time to transmit a data set) is dependent on the initiate time. Finding the optimum initiate time requires the solution of T networks (objective function 4). For store-and-forward, initiate time is relevant to both maximal flow and universal maximal flow.

6 Invariant bounds

The down side of the flow network model is that the interconnectivity and and span of the internet is reflected in the model. The complexity of the routing algorithm is directly proportional to the number of links and nodes in the flow model, making the model too complex to be practical for routing. Here, we evaluate the model to see if it can be used to generate estimates on the performance of routing algorithms.

The two invariant ASs in bulk transmission are the sender and receiver ASs. The routing algorithm selects transit ASs between the sender and receiver, but the end ASs are fixed inputs and cannot be substituted. The bulk transmission application is run on behalf of customers of these end-user ASs, and the administrators have incentive, financial or otherwise, to reveal information on bandwidth availability for bulk transmission. A flow network constructed with bandwidth availability data from the sender and receiver ASs will bound the performance of the complete flow network. While it is obvious that the bounded flow network cannot be used for routing algorithms, this bounded network can be used to cheaply compute bounds for the objective functions. Moreover, data from the end ASs can be used to compute the optimum initiate time, regardless of the transit ASs.

Definition 4 *For flow network, N = (V, E, b), its bounded flow network is given by*
$N_{bn} = (\{s, Tr, r\}, \{(s, Tr), (Tr, r)\}, b)$*, where*
b(Tr, t) ≥ 0 *is the storage capacity of vertex Tr at t,*
$b(s, Tr, t) = \sum_{\forall (s,x) \in E} b(s, x, t),$

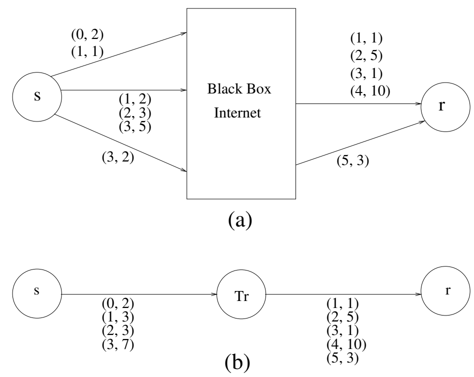


Figure 4: Time varying bounded flow network for Figure 2

$b(Tr, r, t) = \sum_{\forall (x,r) \in E} b(x, r, t),$

$t = 0, 1, 2, \cdots, T-1$*, where T is the maximum allowable flow time from the sender s vertex to the receiver r vertex via the transit Tr vertex.*

Figure 4 is the bounded network for Figure 2. The bounded network is formed from two cutsets, $\{b(s, x, t) \in E | t = 0, 1, .., T-1\}$ and $\{b(x, r, t) \in E | t = 0, 1, .., T-1\}$. The maximum capacity of the network cannot exceed the capacity of the minimum of these cutsets. The bound cutsets are not necessarily the minimum cutset, since the minimum cutset can only be generated by evaluating the complete flow network. However, the sender and receiver cuts are invariant cutsets for bulk transmission. It is very simple to compute objective functions for the cutsets. For end-to-end, the flow is:

$$f_{bn}(\lambda_{\varepsilon}, T) = \sum_{t=0}^{T} \text{minimum } \{b(s, Tr, t), b(Tr, r, t)\}$$

For store-and-forward, the flow is:

$$f_{bn}(\lambda_S, T) = \sum_{t=0}^{T} b(s, Tr, t) - f(Tr, T)$$

where $f(Tr, T)$ is computed as:

$$f(Tr, t) = \begin{cases} \max\{0, f(Tr, t-1) + b(s, Tr, t) - \\ \quad b(Tr, r, t)\} \quad \forall 0 < t < T \\ \max\{0, b(s, Tr, t) - b(Tr, r, t)\} \quad t = 0 \end{cases}$$

Result 3 *For both end-to-end and store-and-forward,*
$f(\lambda^{max}, T) \leq f_{bn}(\lambda, T)$
If $f(s, 0^-) = f_{bn}(\lambda, \tau) = f(\lambda^*, t)$ *then* $\tau \leq t$*.*

Result 3 states that the maximal flow and universal maximal flow of a network is at most equal to the corresponding flow of its bounded network. If the sender or the receiver is the bottleneck, then the maximal flow and minimal time of the bounded network is equal to that computed by the objective functions for the complete flow network. If the transit ASs are the bottleneck, then the bounded network generates an optimistic bound.

Corollary 1 *The optimal initiate time for the bounded network is the optimal initiate time for the corresponding flow network if the transit ASs are not the bottleneck.*

The usefulness of the bounds stems from the property that bulk transmission is essentially a first-mile, last-mile problem. Therefore, ensure that data gets out of the sender when there is bandwidth at the sender, and data gets transmitted to the receiver when bandwidth is available. Once the time constraints at the sender and receiver are known, transit nodes that synchronize their inflow with the sender and their outflow with the receiver should be selected. In the next section, we experimentally showcase the paper's theoretical results.

7 Experiments

This section experimentally evaluates the impact of initiate time and route selection on the performance of bulk transmission. We use OPNET, a commercial simulator capable of simulating a wide variety of network components and workloads [10]. The background traffic emulates network usage based on DE-CIX traffic statistics [1] which follows the sleep-wake cycle. A parameter varied in our experiments is the locations of the sender, receiver and transit nodes relative to each other. The location displacement is represented by the time difference of the receiver and transit nodes with respect to the sender's local time. For example, if the sender is in LA and the receiver is in Germany, then the receiver is 9 hours ahead of the sender (or, equivalently, 15 hours behind the sender). Routes from sender to receiver are selected by varying the locations of transit nodes.

The first graph in Figure 5 varies the receiver and the transit nodes, while keeping the sender fixed. Therefore, each plot represents the maximum data transmitted over a period of 24 hours, from sender to different locations, as the initiate time varies along the X axis. The second graph in Figure 5 plots a store-and-forward transmission along a random path where the receiver is 8 hours behind the sender. This routing path is compared against the best route represented by the store-and-forward bound. The end-to-end bound is also plotted (dashed line).

The next 4 graphs fix the sender and receiver, and plot data transmission along various paths. The best path is again represented by the store-and-forward bound (solid line); the end-to-end bound is also plotted (dashed line). Figure 6 sends data from LA to Germany via routes with transit nodes in various time zones. The first graph plots data transmitted over a period of 24 hours, the second graph plots the total time to transmit a 2 TB file from sender to receiver while the transmission start time from sender is varied. Figure 7 shows similar graphs with receiver 6 hours behind sender.

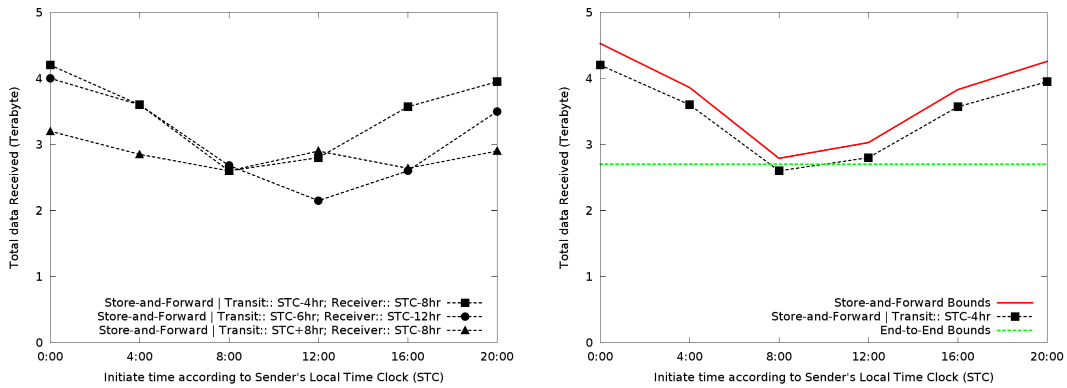
The experiments present a graphical representation of the paper's theoretical results. As one can see, the total data transmitted depends on the initiate time and relative displacement between sender, receiver, and transit nodes. The choice of transit nodes (routes) has a big impact on performance. In fact, the graphs show that rather than choosing random routes, it is often better to transmit directly from sender to receiver. Similarly, the transmission start time determines the duration of transmission. The experiments highlight the best-case performance of store-and-forward bounds, the constant performance of end-to-end transmission with regard to initiate time and total data transmitted during a cycle, the impact of route selection, impact of location of storage nodes, and the impact of initiate time.

8 Reevaluating problem

We reevaluate the objectives of the routing algorithms after taking into consideration the experimental and theoretical results. Our evaluation shows that the location of storage nodes drives performance. A storage hub temporarily stores data until bandwidth becomes available in the forwarding AS. Since bandwidth availability is a function of local time which is a function of geographic time zone positioning, a storage hub should be located in every time zone. It is not necessary to put storage in every node of the flow network, just one storage node in each time zone is required. The flow network representing bulk transmission then consists of nodes that represent storage hubs and edges that represent bandwidth availability in the zones. The objective functions compute maximal flow, universal maximal flow for the flow network that connects the storage hubs between the time zones.

Bulk transmission between data centers is a related problem. This problem is of interest to cloud enterprises who have data centers around the globe. In this case, the underlying flow network has storage nodes only in the data center locations, not in all time zones. The problem is to compute maximal flow and universal maximal flow when the storage hubs are available only in some of the time zones.

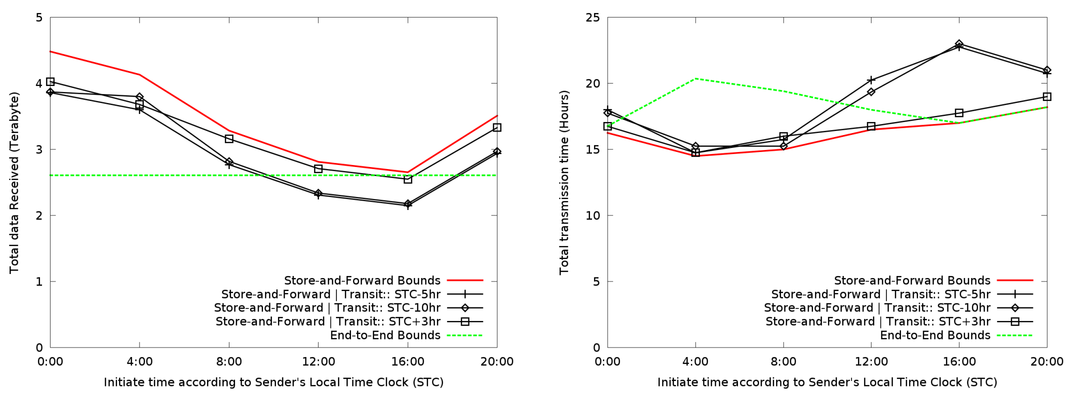
Other objective functions of relevance to the bulk transmission problem are: given a sender and receiver, where should storage nodes be located; how much bandwidth should be purchased in order to ensure that the bulk data arrives within a given time range; what is the latest time to start transmission while ensuring that the transmission completes in time. Since all these objectives are equivalent to maximal flow and universal maximal flow in time-varying flow networks, it is natural



Varying transit and receiver nodes

Receiver is 8 hours behind sender

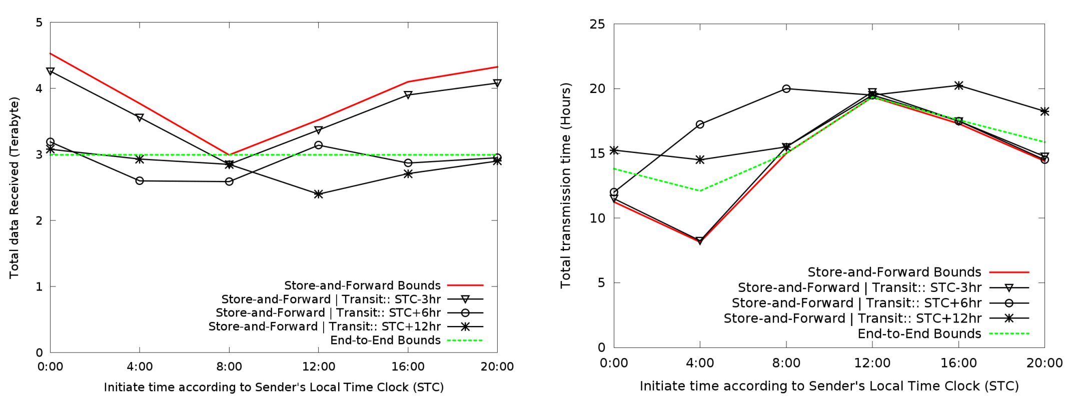
Figure 5: Varying the transmission routes from sender.



Bulk data transmitted during 24 hours

Time to transmit 2TB file

Figure 6: Transmission from LA to Germany along various routes.



Bulk data transmitted during 24 hours

Time to transmit 2TB file

Figure 7: The receiver is 6 hours behind the sender in all experiments; the transit nodes are changed.

to assume that the problem of routing bulk data sets is solved. However, there are challenges to using flow networks for bulk transmission routing. Two of the most critical issues are:

1. the construction of the flow network;

   The internet is a densely interconnected global network. Typically, there are several paths between any two nodes. The number of nodes and links in the internet keep increasing with time. The ensuing flow network is a hodgepodge of nodes and links. The flow network reflects the dense confusion of the internet without providing any order or clarity, and thereby fails as a model.

2. the complexity of time varying flow algorithms.

   While there are several algorithms to solve objective functions 1 and 2, these algorithms have high computational complexity. The first optimization problem, namely, maximal flow, can be solved in polynomial time. However, the second optimization problem, namely, universal maximal flow, is NP-complete [16]. Solutions for objective functions 3 and 4 require solutions to T flow networks. Even if one is able to construct an accurate flow network model, the scale and inter-connectivity of the internet and the diurnal bandwidth availability cycle results in having to search for a solution from potentially exponentially many solutions.

Prior papers have skirted the first issue, namely, determination of what nodes and links to include in the construction of the flow network. However, prior papers talk about the complexity of the routing algorithms. The conclusion is that the complexity of the routing algorithm cannot be avoided given the the complexity of the flow network which reflects the global span, interconnectivity, and time-varying dynamic nature of the internet.

9 Beyond flow networks

A flow network is a graph theoretic model where nodes (vertices) and all edges between nodes are required to describe the overlaying system. In the current state-of-the-art, a time-varying flow network is the only model for generating bulk routing paths, but flow networks perform poorly for the following reasons:

1. the relative positioning between nodes of a flow network must be explicitly entered by showing the links between nodes, and for a global network, this is a non-trivial task; and

2. a flow network does not model passage of time nor the relationship between local time and geographic positioning.

There are three different times of relevance to the bulk transmission problem, namely, the local time at each node, the time difference between the nodes, and the transpired time since the start of transmission. The complexity of the flow network arises from the inability of the flow network to capture the relative positioning between the nodes and the inability of the flow network to capture the relationship between local time, universal time, and transpired time at the nodes.

The main problem is that the flow network is neither a map nor a clock, and modeling global systems requires encapsulation of both geographic positioning and time. An example demonstrating the need for a mapping mechanism in the underlying routing model: bulk data between New York and Germany should be routed either eastward or westward. A flow network routing algorithm checks out all paths, including paths going to Chicago, Texas, and then zigging back to Florida, London, Germany. An example demonstrating the need for a timing mechanism in the underlying routing model: bulk data from New York to Germany where start time is 5:00 AM, should be transmitted westward with possible storage hops in western US, Asia, and Europe rather than eastward since time zones in UK would be entering the peak bandwidth usage times at 5:00 AM EST. A flow network routing algorithm would have to evaluate all routing paths since the sleep-wake bandwidth availability distribution is not an inherent characteristic of flow networks.

A model for routing in global systems should incorporate a map, a clock, and a mapping between geographic positioning, time, and capacity/cost. By its very nature, such a routing model would automatically weed out inefficient paths, thereby pruning the search space for the routing algorithm. With a reduction in search space, the complexity of the routing algorithm would be lower.

10 Conclusion

This paper presents a thorough evaluation of the bulk transmission problem. We are the first to build a comprehensive time-varying flow network model of bulk transmissions over the internet. This model identifies the relevance of initiate time to performance. The paper shows how the initiate time maps to the underlying flow network routing model. The model provides a common mathematical framework for end-to-end and store-and-forward protocols, making it possible to compare and contrast the two types of protocols. Furthermore, the paper develops quick performance bounds using the sender and receiver nodes only.

A key contribution is proof of self failure of flow networks as a model for generation of transmission routes in network systems that span the globe. The complexity of the internet is mirrored by its underlying time-varying flow network, and therein lies the reason for this failure. The flow model provides little clarity into how to search for paths within the jumble of nodes and links. The routing algorithms scale exponentially with the number of links, nodes, and range of initiate times, thereby rendering the flow network useless for large, intricate systems.

Our analysis of the reasons flow networks fail provides the framework for the design of a routing model for global systems. We have developed a routing model that incorporates a global map, a clock, and a mapping between positioning and time. In a future paper, we plan to present this new model and a bulk routing algorithm generated from the model.

References

[1] De-cix where networks meet; statistics, http://www.de-cix.net/about/statistics/, 2013.

[2] BRESNAHAN, J., LINK, M., KHANNA, G., IMANI, Z., KETTIMUTHU, R., AND FOSTER, I. Globus gridftp: What's new in 2007. In *GridNets* (October 2007).

[3] BROSH, E., BASET, S. A., RUBENSTEIN, D., AND SCHULZRINNE, H. The delay-friendliness of tcp. In *2008 ACM SIGMETRICS* (2008), ACM, pp. 49–60.

[4] CERN. Lhc physics data taking gets underway at new record collision energy of 8tev. *http://press.web.cern.ch* (2012).

[5] CHHABRA, P., ERRAMILLI, V., LAOUTARIS, N., SUNDARAM, R., AND RODRIGUEZ, P. Algorithms for constrained bulk-transfer of delay-tolerant data. In *ICC'10* (2010), pp. 1–5.

[6] LAKHINA, A., PAPAGIANNAKI, K., CROVELLA, M., DIOT, C., KOLACZYK, E. D., AND TAFT, N. Structural analysis of network traffic flows. In *SIGMETRICS'04* (2004), pp. 61–72.

[7] LAOUTARIS, N., SIRIVIANOS, M., YANG, X., AND RODRIGUEZ, P. Inter-datacenter bulk transfers with netstitcher. In *Proceedings of the ACM SIGCOMM 2011*.

[8] LAOUTARIS, N., SMARAGDAKIS, G., RODRIGUEZ, P., AND SUNDARAM, R. Delay tolerant bulk data transfers on the internet. In *SIGMETRICS* (2009), pp. 229–238.

[9] LIMONCELLI, T. A. Openflow: A radical new idea in networking. *Queue 10*, 6 (June 2012), 40:40–40:46.

[10] LUCIO, G. F., PAREDES-FARRERA, M., JAMMEH, E., FLEURY, M., AND REED, M. J. Opnet modeler and ns-2. In *ICOSMO* (2003), pp. 700–707.

[11] QI, J., ZHANG, H., JI, Z., AND YUN, L. Analyzing bittorrent traffic across large network. In *Cyberworlds* (2008), pp. 759 –764.

[12] SHERWOOD, R., BRAUD, R., AND BHATTACHARJEE, B. Slurpie: a cooperative bulk data transfer protocol. In *INFOCOM* (2004), pp. 941 – 951 vol.2.

[13] SHI, C., AMMAR, M. H., AND ZEGURA, E. W. idtt: Delay tolerant data transfer for p2p file sharing systems. In *GLOBECOM'11* (2011), pp. 1–5.

[14] TEITELBAUM, B., HARES, S., DUNN, L., SYSTEMS, C., NARAYAN, V., AND NEILSON, R. Internet2 qbone - building a testbed for differentiated services. In *IEEE Network Magazine, Special Issue on Integrated and Differentiated Services for the Internet* (September 1999).

[15] VENKATARAMANI, A., KOKKU, R., AND DAHLIN, M. Tcp nice: A mechanism for background transfers. In *OSDI'02* (2002), pp. –1–1.

[16] XIAOQIANG CAI, DAN SHA, C. K. W. *Time-Varying Network Optimization (International Series in Operations Research & Management Science)*. Springer, 2007.