A new model to evaluate bulk transmission over the internet

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 links to the receiver, bulk routing algorithms focus on finding cheap, high capacity bandwidth links to the receiver. This paper develops a a comprehensive model for bulk transmission using time-varying flow networks, and identifies parameters that determine the performance of the transmission. Unfortunately, the global span and inter-connectivity of the internet overwhelms the flow network model, leading to impractical routing algorithms with high computational complexity. Consequently, flow networks fail as a model for evaluating bulk transmission and generating routing paths. The paper proposes a new modeling construct, clock nets, for evaluating bulk transmission over the internet. With clock nets, the complexity of bulk routing algorithms is bounded by the number of world time zones. An infeasible NP-complete problem in the domain of conventional flow networks is a simple problem in the domain of clock nets.

1 Introduction

This paper addresses the problem of efficiently transmitting a bulk data set from a sender's network to a receiver's network. 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 [6]. 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 built data centers to store bulk data sets. In addition, research labs, universities, and even ordinary users are generating bulk data sets. Corporations that transmit bulk data regularly between their sites may build, purchase, or lease network links. However, bulk transmission to-and-fro 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 transmissions 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 to a few protocols. 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 ei-

ther end-to-end or store-and-forward. The current trend is in favor of store-and-forward protocols for bulk transmission [10]. 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 develop a comprehensive flow network model of bulk transmission. Using this model, we identify the internet and application parameters that are essential to determining the performance of bulk transmission. Moreover, our model provides a common framework for comparison of end-to-end and store-and-forward bulk transmissions. We prove that the performance of end-to-end can at most equal the performance of store-and-forward. Flow networks are the natural way to model bulk transmissions, but the complexity of flow network algorithms is dependent on the number of nodes and links in the network. The global span and connectivity of the internet translates into a complex flow network with a jumble of nodes and links. Since the corresponding routing algorithms are computationally infeasible, the flow network fails as a modeling tool for bulk transmission routing. Therefore, we propose a new model, the clock net, for bulk transmissions. Regardless of the distance between sender and receiver and the complexity of the intervening networks, the clock net ensures that the complexity of the transmission algorithm never exceeds the number of world time zones. The clock net is proposed as an alternative and an assistant to conventional flow networks as a model for evaluating internet applications.

2 Bulk transmission platform

The internet is a collection of independent networks, where each network is an Autonomous System (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 sender and receiver ASs are stub networks that may be multi-homed. Internet Exchanges (IXs) and Network Access Points (NAPs) allow one network to connect to another network; data between 2 networks flows via these interconnects. In this paper, AS refers to end/transit networks, IXs and NAPs.

Routing in networks occurs at 2 levels, namely, routing within each AS (intra-AS routing) and routing between ASs (inter-AS routing). Routers within an AS use an Interior Gateway Protocol for determining paths within the AS. The intra-AS routing algorithms are driven by metrics such as hop count. Routers between ASs use the Border Gateway Protocol for exchanging routing information between ASs. ASs are independent administrative entities, and traffic exchange between two ASs is controlled by administrative/pricing policies. The inter-AS routing is driven by business policy constraints. Thus, intra-AS routing is metric driven, while inter-AS routing is business driven.

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 that use a combination of minimal hop count and business cost. If performance is measured by metrics such as throughput (or transmission time), then bulk transmissions could perform very poorly with standard internet routing. For example, with 20 Mb/s and 2 Gb/s links, the 100 MB file is transmitted in less than a minute over either link; on the other hand, 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 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.

An issue that should be addressed by the bulk routing algorithms is the shared nature of the internet. 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 time-critical applications by ensuring that bulk transmissions only avail of the remaining "free" bandwidth. Therefore, the objective function of bulk routing algorithms is modified as follows: find the links with the highest free capacity

between sender and receiver.

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 mimics the users' diurnal sleep-wake cycle [8]. For a given network, the valleys in bandwidth usage correspond to peaks in free capacity. For example, a network with 10 Gb/s capacity may limit bulk transmissions to 50 Mb/s during peak bandwidth usage times, but allow 8 Gb/s during the low bandwidth usage times. If bulk transmissions are scheduled during valleys in the bandwidth usage, then the throughput of the transmission increases without decreasing the performance of other transmissions.

The available bandwidth has a diurnal wave pattern depicted in Figure 1. Since bandwidth usage cycle mimics the users' sleep-wake cycle, bandwidth availability at universal time (UTC) is dependent on the location of the AS. 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 reason is that 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 transmission has been proposed for bulk transmissions. Instead of directly transmitting the bulk data set from sender to receiver, data sets are temporarily stored in staging servers 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 staging server 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.

3 Related Work

The performance of bulk transmission is dependent on bandwidth availability between sender and receiver, and some papers focus on protocols to access maximum bandwidth. Another angle studied is how to ensure that high bandwidth usage by bulk transmissions does not negatively impact other internet applications. Below, we list prior papers that address the first, second, and both issues.

The direct approach to getting maximum bandwidth is to open multiple transmission links between sender and receiver. Some parallel transmission protocols are GridFTP [4], BitTorrent [13], and Slurpie [14]. 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.

The direct approach to minimizing the negative impact of bandwidth usage is to allow other traffic to go ahead. The Qbone protocol [16] 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 [5, 17]. Another approach is to find the least congested paths from sender to receiver. The most promising of these approaches is OpenFlow [11], where routing is performed by a centralized software router with knowledge of entire network traffic, rather than by routers with local traffic awareness.

In order to avail of maximum bandwidth without being detrimental to other applications, a bulk transmission has to either use private links or use shared links during low traffic times. Research in this direction is targeted to protocols that take advantage of already-paid for bandwidth that is available during the sleep phase of the bandwidth availability distribution. An early paper proposed the advantage of store-and-forward protocols over end-to-end protocols for synchronizing sleep time misalignment between sender and receiver networks [10]. The idea is to transmit from the sender during its sleep period, and to temporarily store data in various transit nodes until bandwidth is available in the receiver. Another paper [15] used simulations to show that peak traffic and cost are reduced with store-and-forward, albeit with a slight increase in latency when compared to end-to-end protocols. Laoutaris et al. [9] and Chhabra et al. [7] address routing of store-and-forward bulk protocols. Both 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 understand-

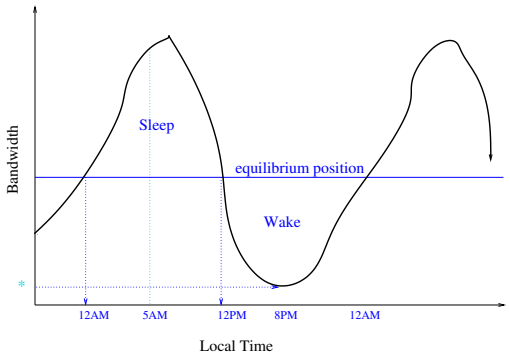


Figure 1: Available bandwidth distribution

ing of how bulk transmissions fundamentally differ from standard transmissions. Subsequently, the research issues pertaining to this area are not clear. 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. There is no mathematical framework to answer questions pertaining to performance of bulk transmissions. In order to develop efficient bulk transmission protocols, it is necessary to 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 attempts to develop such a model.

4 Flow network model

Bulk transmissions ply over links with varying free capacities, where an AS's free capacity follows a diurnal wave pattern. Figure 1 shows bandwidth availability distribution by time of day for a typical stub AS; Table 1 tabulates free bandwidth; the distribution has been deliberately simplified since it is used in examples through the paper. In order to model bulk transmissions plying over a dynamic internet, time-varying flow networks [18] in which the capacity of the edge varies with time-of-day are 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.

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

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;* $t = 0, 1, 2, \cdots, T$*, 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 the clock stops at the end of hour $T = 23$ 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 2:00AM, then t = 0 at 2:00AM, t = 5 at 7:00AM, 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 bandwidth 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)

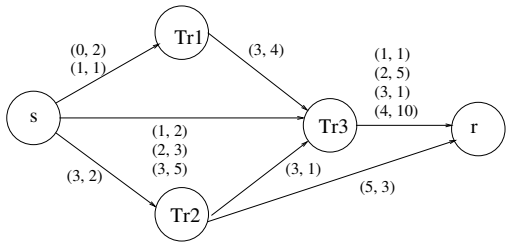


Figure 2: Time varying flow network modeling bulk transmission

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 10:00AM LTC. Then, t=0 at LTC=10:00AM t=1 at LTC=11:00AM, ..., t=23 at LTC=9:00AM. 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=5. If bulk transmission is not permitted along an edge (x, y) during a hour t, $t \leq T$ (*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 $p =< v_0, v_1, v_2, \cdots, v_k >$ 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 $t \leq T$. 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. 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. There are algorithms for computing the maximum flow and universal maximum flow in time-varying flow networks [18]. Equivalently, these algorithms can be used to compute the routing of bulk transmissions over the internet.

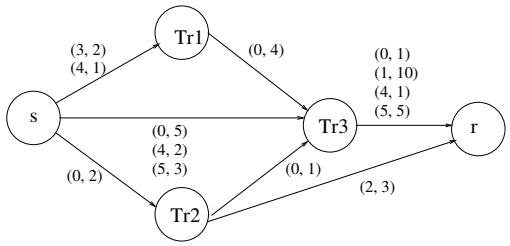


Figure 3: Transformed flow network of Figure 2 when t shifted by 3

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 staging 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. For the flow network in Figure 2, $f(\lambda_\varepsilon, 1) = 1$, $f(\lambda_\varepsilon, 2) = 4$, $f(\lambda_\varepsilon, 3) = f(\lambda_\varepsilon, 4) = f(\lambda_\varepsilon, 5) = 5$; $f(\lambda_S, 5) = 15$, $f(\lambda_S^*, 4) = 14$.

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

In an earlier paper [9], 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

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 3 *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 4 *Suppose edges (s, v) and (x, y) have identical distribution of free capacity given by Table 1. That is,* $c(s, v, t) = c(x, y, t) \ \ 0 \leq t \leq 23$. *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,* $N = \{N^l = (V, E, b) | l = 0, 1, ..., T\}$ *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 $f(\lambda^{max}, T, \theta) \geq f(\lambda^{max}, T, l), \forall 0 \leq l \leq T, l \neq \theta$.*

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

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

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, universal maximal flow (*i.e.,* the minimum time to transmit a data set) requires objective function 4. For store-and-forward, initiate time is relevant to both maximal flow and universal maximal flow.

5.1 Experiments

We experimentally evaluate the impact of initiate time and transit nodes on the performance of bulk transmission. We use OPNET, a commercial simulator capable of simulating a wide variety of network components and workloads [12]. The experimental platform consists of the sender, receiver, and three transit nodes configured as follows: sender to transit-1; transit-1 to transit-2; transit-2 to receiver. Thus, there are a total of 3 edges, where each edge has the same physical capacity. The background traffic emulates network usage based on DE-CIX traffic statistics [2]. Therefore, the edge capacities are all equal, and the available capacity varies according to the sleep-wake diurnal cycle.

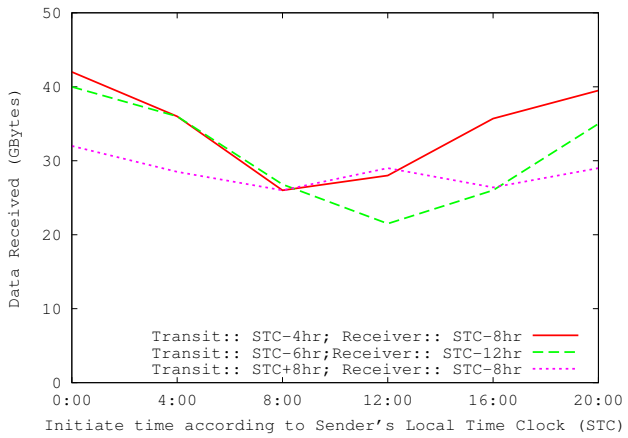
A parameter varied in our experiments is the locations of the sender, receiver and transit nodes; the location displacement is represented by the local time at the receiver and transit nodes with respect to the sender's local time. The first graph in Figure 4 has three lines, each representing a different relative placement of sender, transit, and receiver. The solid line and the dotted line represent the amount of data transmitted as initiate time varies, when the receiver's time is 8 hours behind the sender's time. The transit placement for the solid line experiments is 4 hours behind the sender; and the transit placement for the dotted line experiments is 4 hours ahead of the sender. The dashed line represents the total data transmitted when the receiver is 12 hours behind the sender, and the transit is 6 hours behind the sender. The initiate time is varied in each experiment and the total data transmitted to the receiver over a 24 hour period is plotted. As one can see, the total data transmitted depends on the initiate time and relative displacement between sender, receiver, and transit nodes.

The second graph in Figure 4 shows the bounds generated using only the sender and receiver nodes (no transit nodes). As explained earlier, end-to-end bound is a straight line since the end-to-end transmission during the 24 hour cycle is not dependent on initiate time. The store-and-forward is a tighter bound; the degree of tightness depends on whether the transit bound is a bottleneck. For this experiment, the the store-and-forward bound is statistically identical to the sender-transits-receiver result.

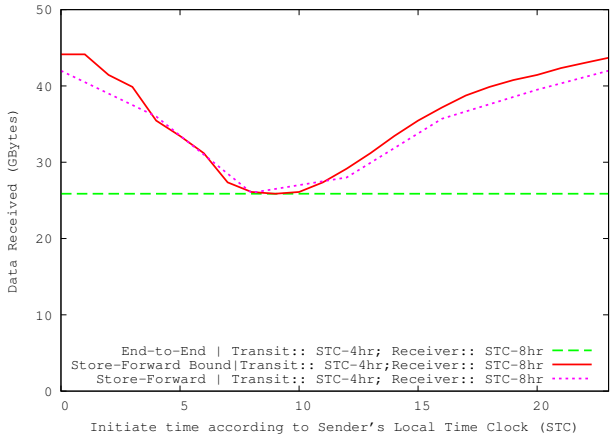
6 Issues with the flow model

By showing that the problem of bulk routing is equivalent to maximal flow in time-varying flow networks, it would seem that understanding bulk transmission and developing efficient transmission protocols would be easy. However, there are challenges to using flow networks for bulk transmissions. Some of the most critical issues are:

1. getting input data to construct the flow network underlying bulk transmission;
   The ASs are managed by independent organizations who are in the business of global telecommunications. These businesses are unlikely to hand over



Impact of start time and transit nodes on performance



End-to-end, Store-and-forward bounds

Figure 4: OPNET experiments

data regarding free capacity. The transit links are upgraded quite regularly, with new interconnections and faster links. Consequently, the map of global internet bandwidth usage is dynamic and needs updations on a somewhat regular basis.

2. the complexity of the internet;
   The internet is a densely interconnected system which spans the globe. Typically, there are several paths between any two nodes. The number of nodes and links in the internet keep increasing with time. Even if one is able to construct an accurate map of the internet, the ensuing flow network is a hodgepodge of nodes and links. The flow network reflects the intricacy of the internet, and it fails as a model since it is very difficult to extract meaning from complexity. When a routing algorithm outputs an optimal path, it is difficult to comprehend why this path is better than others.
3. 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 [18]. 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 results in having to search for a solution from potentially exponentially many solutions.

Finding an optimal bulk routing solution between any two nodes using the tried and tested method of time-varying flow networks is an inherently intractable problem. In the next section, we propose a new modeling construct that is suitable to "big transmissions" on the internet.

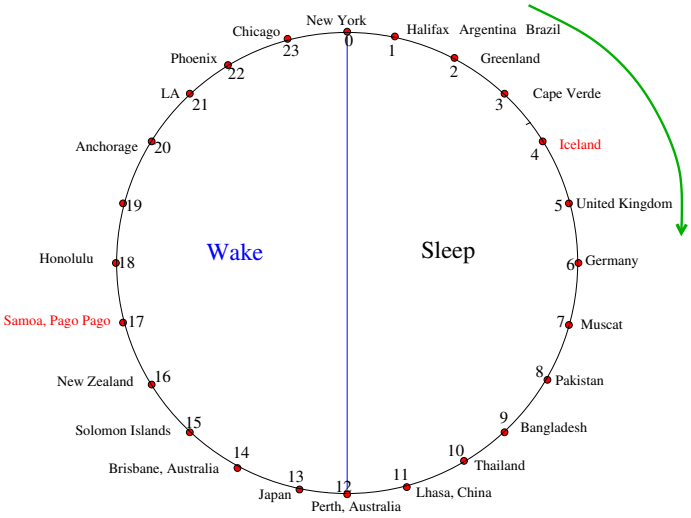


Figure 5: Time-zone clock: Each zone is referred to by the name of a country within that zone. The number represents local time in the node. The clock (numbers) are fixed, the nodes rotate in the clockwise direction

7 Time zoned flow network - clock net

The flow network model is an established mathematical framework to evaluate the bulk transmission problem. Using this model, one can frame performance issues, estimate the significance of parameters like initiate time, and establish the relationship between end-to-

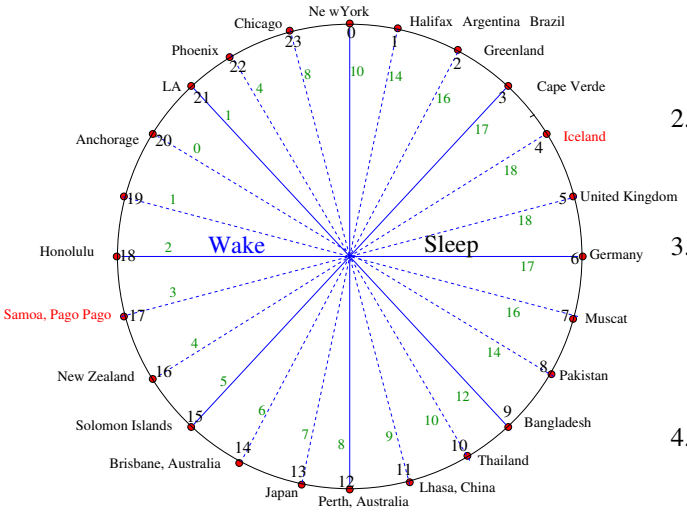


Figure 6: Clock net: the outer number represents the local time in the countries within the time zone; the inner number represents the free capacity in the time zone.

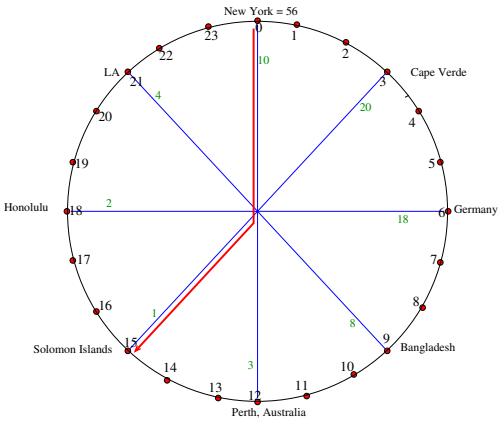


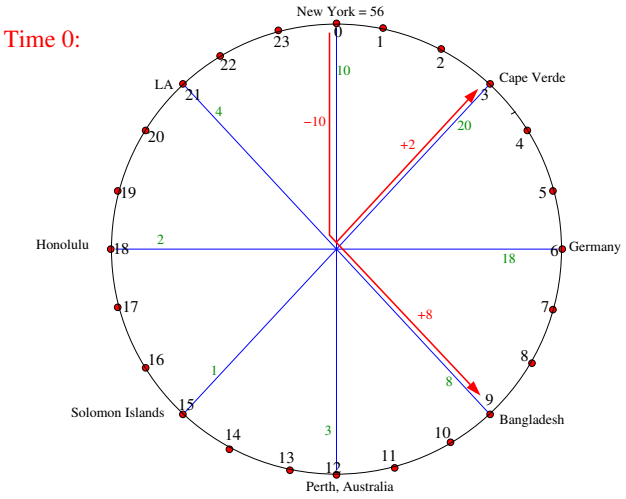
Figure 7: Objective: (56)New York−− >Solomon Islands

end and store-and-forward. However, the routing algorithm requires constructing a flow network of the internet with all the ASs and their available bandwidth capacities, an unfeasible proposition given the scope, complexity, and changeability of the internet. Once the internet map is constructed, solving the resulting flow network is computationally expensive. The flow network is a good model for small systems with few nodes and links, but it fails for large systems due to the exponential complexity of the search algorithms. Therefore, we propose a different model that incorporates the features specific to bulk transmission over the internet. These features are:

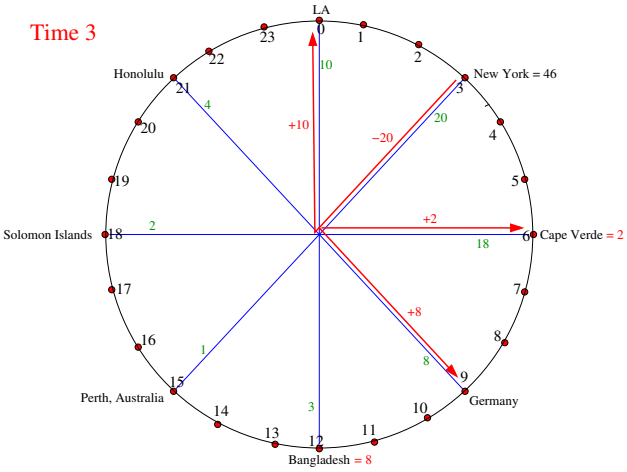
1. the sender and receiver nodes are the only invariant nodes in the network. The transmission can never exceed the capacity of the minimum of these 2 nodes. They limit the performance of the transmission. (The bounds are not shown here due to page limits.)
2. the transit nodes are typically interconnected to several nodes, and there are several paths between any two nodes. The connectivity of the internet allows flexibility in the selection of transit nodes.
3. the global span of the internet, which results in a large number of nodes and links. The model should have the power to reduce the number of nodes and links while retaining the essential characteristics of the network's bandwidth capacity and its connectivity.
4. the bandwidth availability typically follows the sleep-wake cycle. This has been shown in end, transit, and exchange nodes [1, 2, 3].
5. the relative distances between nodes determines the degree of synchronization between bandwidth availability cycles, so the model should incorporate this feature.
6. the initiate time is critical to the performance of the transmission, and a robust model should provide insight into a suitable initiate time.
7. the similarity of the internet to other stable, globally interconnected distributed systems such as roadways, airlines, and railways. As these systems mature and reach steady state, the capacity in equivalent zones equalizes to allow free flow.

Bandwidth availability at any (UTC) time is a function of the location, and a natural way of depicting this graphically is using a time-zoned clock. Figure 5 is a 24 hour clock depiction of the world time zones an hour apart [0,23]. The node at a zone, referred to by the name of a country within that zone, shifts in the clockwise direction as the earth rotates. The numbers inside the clock represent the local times in the zones in which the nodes are located. The clock is fixed and the nodes rotate in the clockwise direction at the rate of earth's rotation around its axis. The 24 time zones can be divided into 2 regions based on sleep-wake cycle of bandwidth availability. The sleep region encompasses zones [0,12[, the wake region encompasses zones [12,0[. As shown in Figure 1 (Section 4), a node has higher free bandwidth in the sleep region than when the node is in the wake region. The time-zoned clock encapsulates both time and relative location.

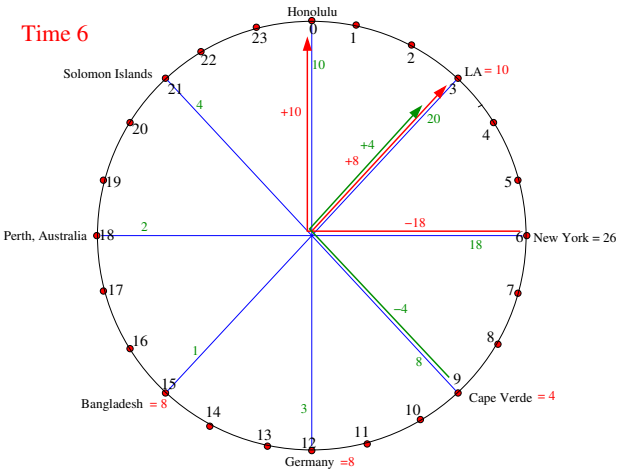
Each node represents the internet in its time zone. For example, the node for New York, located in time zone UTC-5, represents the internet for the entire UTC-5 zone. The zone could encompass a vast region with a large number of ASs or it could encompass a small region with fewer ASs close to the storage node. Each node must capture the essence of the zone's internet that is of significance to the given bulk transmission. This parameter



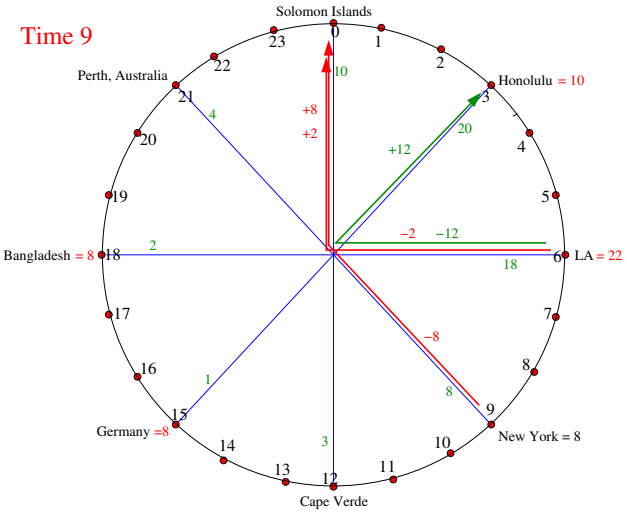
Time 0: (56)NY−− >Cape Verde(2);
NY−− >Bangladesh(8)



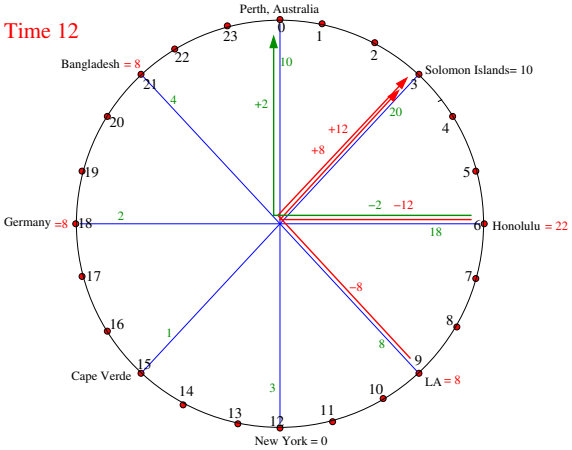
Time 3: (46)NY−− >LA(10);
NY−− >Cape Verde(2); NY−− >Germany(8)



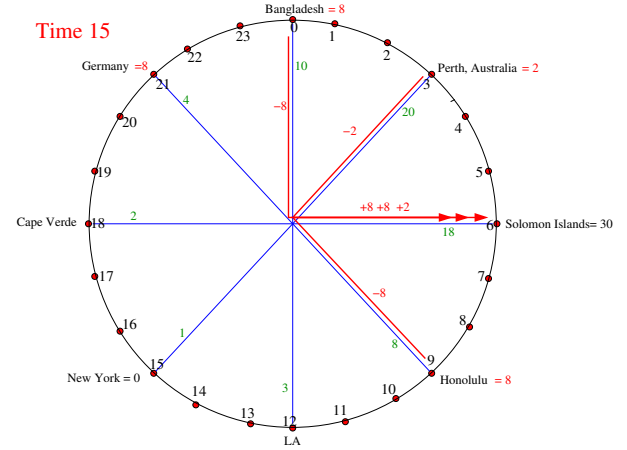
Time 6: (26)NY−− >Honolulu(10);
NY−− >LA(8); (4)Cape Verde−− >LA(4)



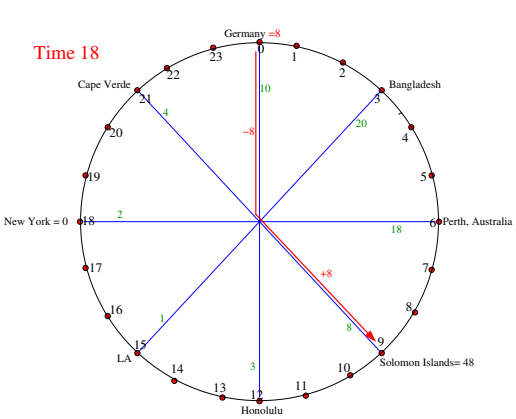
Time 9: (8)NY−− >Solomon Islands(8);
(22)LA−− >Solomon(2);



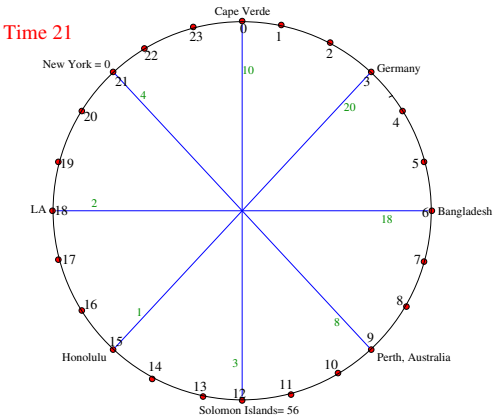
Time 12: (8)LA−− >Solomon(8);
(22)Honolulu−− >Solomon(12); Honolulu−− >Perth(2)



Time 15: (8)Honolulu−− >Solomon(8);
(8)Bangladesh−− >Solomon(8); (2)Perth−− >Solomon(2)



Time 18: (8)Germany−− >Solomon(8)



Time 21: (56)NY−− >Solomon(56) completed.

Figure 8: (Continued from previous page) Depicting the generation of the routing path using clock net

is the inflow and outflow capacity given to the bulk transmission within the zone during each hour.

Since the internet is fully connected, all the nodes of the zone clock are connected as shown in Figure 6. The lines connecting the nodes represent network links, and the numbers along the lines represent the free capacity available for bulk transmission during the corresponding local time. In the figure, the free capacity varies according to the distribution given in Table 1 when C is set to 1. The diagram represents a time zoned flow network - a clock net - where the links represent edges that connect the nodes, and the numbers represent available inflow and outflow capacity along the edges. The edges are fixed (along with the clock times), while the nodes rotate clockwise moving to the next edge in an hour. From edge to edge, the capacity varies according to the sleep-wake cycle. (In the diagram, all zone ASs have the same distribution, but this is not a requirement.) For example, Figure 6 shows Iceland at time 4 with edge capacity of 18; after 4 hours, Iceland will be at time 8 with edge capacity of 14.

We present a motivating example, to explain how the clock net can be used to select transit nodes and in the process generate a bulk transmission route.

Example 5 *The objective function is to transmit 56 units from NY to Solomon Islands in less than 24 hours. The problem is depicted in Figure 7. Figure 8 give the solution where 56 units are transmitted in 21 hours.*

Some observations on the solution:

1. For clarity of exposition, the clock net of Figure 7 is demarcated into 8 zones, each representing 3 hours. From a realistic perspective, this could model the scenario when data centers (storage nodes) are only available in the 8 zones shown in the figure.
2. The bandwidth for each zone is the number written next to the link. Therefore, the bandwidth capacity for ASs spanning local times 0-3 is 10. That is, 10 units of flow is permitted during times 0-3.
3. From the clock net, one can see that relatively little capacity is available during the wake cycle. Consequently, for this example, bulk transmission is permitted only during sleep times [0-12[. The restriction of transmission during sleep hours is not a property of the clock net; rather it is a constraint imposed by us while computing the solution.
4. There is only 1 node and 1 edge per modeled time zone, but this does not necessarily imply that the networks in 2 zones interconnect directly. It is possible that there are several ASs and several paths between the two zones, and clock net implicitly models this scenario. For example, to transmit from NY to Bangladesh, the intermediate nodes in the clockwise direction are Cape Verde, Germany; the intermediate nodes in the anti-clockwise direction are LA, Honolulu, Solomon Islands, Perth. An end-to-end transmission between NY and Bangladesh may have to pass several of these intermediate nodes. Suppose NY's local time is 0, while Bangladesh's local time is 9 (Figure 7). A transmission from NY to Bangladesh in the clockwise direction at time 0, would transmit 8 units of data which is the minimum bandwidth along this path. On the other hand, an end-to-end transmission from NY to Bangladesh in the anti-clockwise direction would result in only 1 unit of data being transmitted (assuming that Solomon is a transit node on this path).
5. End-to-end is used when transmitting between zones that are both in the sleep region. The transit ASs of this transmission are always in zones located in better parts of the sleep-wave availability

distribution than the minimum of the end ASs.

6. Store-and-forward is used when there is insufficient bandwidth at either of the end nodes or at any of the transit nodes that lie in the path. This usually occurs when transmitting between 2 zones that are so far apart that one lies in the sleep region and another in the wake region.
7. The solution completely bypasses the wake region thereby ensuring that bulk transmission gets large capacity at low cost without disrupting service of other internet applications.
8. The sender and receiver nodes are in the NY and Solomon Islands zone, respectively. The sender node is the source, the receiver node is the sink, and the clock net is the transit nodes of the flow network. In order to find the optimum initiate time, the sender/receiver bounds can be used in combination with clock net. Typically, the sender should start transmitting at time 0:00 to transmit maximum amount of data. However, there are other objective functions of interest, namely, for a given bulk data set, what is the latest initiate time to ensure the earliest arrival time, shortest transmission time, fewer storage hops, cheapest bandwidth, etc. For example, if NY wants to transmit 8 units, then end-to-end transmission when NY's local time is 9:00 and Solomon's local time is 0:00 is optimum with regard to storage hops.
9. Note that several of the simplifying assumptions made in Example 5 are introduced for clarity of explanation, and are not constraints of the clock net. For example, node edges could have varying physical bandwidth. In this case, during each clockwise shift of nodes, the edge capacities have to be updated. In fact, the wave distribution of available capacity is not a necessary condition for the clock net.
10. It is not necessary that the clock should be divided into 24 zones. In the example, the clock net is divided into 8 zones. It is possible to divide the clock into any number of zones, based on the placement of storage nodes, or accuracy required. An interesting issue to be investigated is whether the routing path with 8 zones would be similar to a routing path with 24 nodes, albeit with additional transmissions every hour.

With the clock net modeling construct, the bulk transmission flow network is pruned to a maximum of 24 nodes and 24 edges excluding the sender and receiver nodes.

Clock net is a modeling tool tailored to a global, high-level view of the internet. The clock net can be used in conjunction with conventional flow networks for low level routing within a zone and for end-to-end routing between two zones in the sleep region. A mathematical framework for the clock net will be developed as future work. This is essential for tapping the full potential of the clock net. Since clock net is a special case of time-varying flow networks, all the algorithms for time-varying are valid. By mapping nodes and edges to time zones, thereby limiting the maximum number of nodes and edges, the complexity of these algorithms has become bounded. However, the clock net allows the generation of simpler algorithms than that of standard time-varying flow networks, as hinted to by Example 5.

8 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 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. This model can also be used to develop quick performance bounds using the sender and receiver nodes only.

A key contribution of our time-varying flow network model is proof of self failure - the model proves itself to be a poor tool for generation of routing paths, the essence of the bulk transmission problem. The complexity of the internet is mirrored by its flow network model, and therein lies the reason for this failure. The flow network 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 and nodes, thereby rendering the flow network useless for large, intricate systems.

A second major contribution of this paper is the presentation of a new model, the clock net, for the internet. Whereas the flow network is a good modeling tool for the "small network," the clock net is a good modeling tool for the "large network." Technology makes the world appear flat, but it is still round, and the clock net encapsulates this round world impact on systems that span the globe.

This paper explains the clock net informally. We plan to develop a mathematical definition for the clock net and present its properties. A formal model definition is required in order to understand the full potential of clock net and its limitations. The generation of routing algorithms and the prediction of performance is difficult without this formalization. It is possible that clock nets could be used to model applications besides bulk transmission. Clock nets and flow networks could be used collaboratively to understand big data applications over distributed global platforms.

References

[1] Ams ix amsterdam internet exchange; statistics, https://www.ams-ix.net, 2013.

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

[3] Linx london internet exchange, https://www.ams-ix.net, 2013.

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

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

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

[7] 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.

[8] 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.

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

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

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

[12] 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.

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

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

[15] 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.

[16] 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).

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

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