Modeling Maximum-Flow over the Internet with a Star Graph

Brief Announcement

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

Abstract

Maximum-flow is the logistical problem underlying the transmission of large files over the Internet. The time to transmit a large file depends primarily on the bandwidth capacity of the Internet links. The physical capacity is fixed, but the available capacity of links varies during the day depending on Internet traffic. The transmission protocol divides the file into varying sized segments as determined by the maximum-flow algorithm. The segments are routed independently; some segments may be temporarily stored in transit data centers until bandwidth becomes available. File segments may be further divided into smaller segments at transit data centers from where they are routed independently. Upon arrival of all file segments at the receiver, the file is reassembled. The flow network modeling large file transmission, commonly referred to as bulk transmission, consists of nodes that represent data centers and arcs that represent end-to-end bandwidth capacity between the nodes. The inter-connectivity of the Internet ensures that there is end-to-end flow between any two nodes. Consequently, the flow network of the Internet is a complete graph. The number of nodes in the graph is O(n) and the number of arcs is $O(n^2)$. Since bandwidth capacity along arcs varies by time-of-day, the flow network models flows over time. In a discrete time model, the arc capacity is parameterized for each flow instant t=0, 1, ..., T-1. The time expanded graph of the flow network has $O(nT)$ nodes and $O(n^2T)$ arcs. This paper develops a star graph model of the flow network for bulk transmission. The star graph flow network is equivalent to the complete graph flow network. The time expanded star graph has $O(nT)$ nodes and $O(nT)$ arcs. The complexity of routing algorithms is proportional to the size of the search space. This paper reduces the complexity of bulk routing algorithms by developing a graph model with lower space complexity.

1 Introduction

Maximum-flow is the problem that underlies bulk transmission via the Internet. Bulk transmission refers to the transmission of a large file - whose size may range from tens of Gigabyte to the Petabyte scale - from a sender's network to a receiver's network. The Internet parameter of significance is the bandwidth capacity available to the transmission. For example, a 100 MB file is transmitted in less than a minute over a 20 Mb/s, while a 100 GB file takes 11.36 hours over the 20 Mb/s link and less than 7 minutes over a 2 Gb/s link, a reduction of 99% of transmission time. The physical bandwidth capacity is fixed, but the available bandwidth varies during the day depending on Internet traffic. The network traffic pattern is dependent on time-of-day; bandwidth usage increases during the day reaching peak usage around 8:00 PM, bandwidth usage decreases after 8:00 PM and reaches minimum around 3:00 AM. There is maximum bandwidth availability during the early morning hours of 1:00 AM to about 9:00 AM. This is similar to roadways, where roads are congested during rush-hour traffic.

The dependence of bandwidth availability on time-of-day may result in low end-to-end bandwidth when the sender and receiver are located in different time zones. Flow rate is limited by the smallest pipe, so a sender with ample bandwidth at 3:00 AM (sender's local time) may be forced to transmit at low rates due to insufficient bandwidth availability at the receiver (where the local time is 8:00 PM). This problem can be addressed by transmitting to intermediate data centers that have high bandwidth capacity at the same time as the sender and the receiver. Instead of transmitting directly (end-to-end) from the sender to the receiver, it is faster to transmit by a store-and-forward protocol.

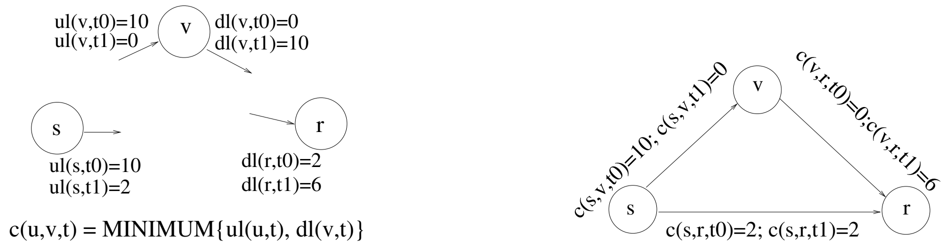
Netstitcher [2] is a store-and-forward bulk transmission protocol where the underlying flow network is modeled as a time expanded graph and the routing is generated by Ford and Fulkerson's maximum-flow algorithm [1]. The routing is determined before the start of transmission; the file is divided into varying sized *segments* as determined by the routing algorithm. Each file segment travels along its predetermined path which may include transit data centers were a segment is stored until bandwidth is available. En route, at a transit data center, a file segment may be divided into further file segments that are routed separately. Once all the file segments arrive at the receiver, the file is reassembled.

The discrete time flow network underlying bulk transmission is modeled as a time expanded graph which contains a copy of the set of nodes and the set of arcs at every flow time step. Let n represent the number of data centers - sender, receiver, and transits - participating in the bulk transmission, and let T represent the number of discrete time steps in the flow duration. Since any two networks (data centers) on the Internet are connected, there are links between every pair of nodes in the graph. Thus, the time expanded graph for bulk transmission has O(nT) nodes and $O(n^2T)$ arcs [2].

This paper reduces the complexity of bulk routing, not by developing a new algorithm, but by reducing the number of nodes and arcs in the underlying graph model. The paper reduces the state space by modeling Internet flow with a star graph rather than a complete graph. The number of nodes in the star graph is $n + 1$ representing the n data centers and the internal node. We show that the star directed graph is equivalent to the complete directed graph modeling Internet flow. The star graph has O(nT) nodes and O(nT) arcs as opposed to the complete time expanded flow network with O(nT) nodes and $O(n^2T)$ arcs.

2 Complete graph model of bulk routing

The complete graph is the standard model for store-and-forward bulk routing. The nodes represent data centers and the arcs represent end-to-end bandwidth capacity between the nodes at each flow instant. The flow network is represented by a graph G=(V,E,T) where V is the set of nodes representing data centers, E is the set of arcs representing end-to-end bandwidth between nodes, and T is the total flow time. A node $v \in V$ has uplink and



(i) Uplink/Downlink bandwidth

(ii) Complete graph

Figure 1: Complete graph modeling bulk transmission

downlink bandwidth capacity, ul(v,t) and dl(v,t), at flow instant t. The sender, s, has zero downlink capacity; the receiver, r, has zero uplink capacity. Since there is end-to-end Internet flow between any two networks, the graph is complete. There are incoming and outgoing arcs between all internal nodes (nodes other than sender and receiver), outgoing arcs from the sender to all nodes, and incoming arcs from all nodes to the receiver. Define the capacity, c(u,v,t), of arc (u,v) at time t as:

c(u,v,t) $\leq$ minimum$\{$ul(u,t), dl(v,t)$\}$.

The capacity c(u,v,t) represents an upper bound on the end-to-end bandwidth available for transmission from u to v at flow instant t=0, 1, ..., T-1. The arc transit time is zero, so flow moves instantaneously along a path at time t. The nodes have storage capacity; it is assumed that storage capacity at the data centers is not a bottleneck.

Figure 1 presents an example to illustrate the complete graph model of bulk transmission. Figure 1(i) shows the uplink and downlink capacities of nodes, s, r, and a transit node v. Figure 1(ii) shows the complete graph for store-and-forward flow corresponding to the nodes. For this simple example, it is easy to compute maximal flow from s to r. At time t0: f(s,r,t0)=2; f(s,v,t0)=6; At time t1: f(s,r,t1)=2; f(v,r,t1)=6. The value of flow $\mid f \mid$ = f(s,r,t0)+f(s,r,t1)+f(s,v,t0)=10 units of flow during flow duration. But the flow value of 10 exceeds the bottleneck capacity 8 (= minimum $\{$ul(s,T)=12, dl(r,T)=8 $\}$). The reason that flow exceeds bottleneck capacity is that the complete graph does not model the bottleneck constraints of the end networks. For example, while the uplink from s at time t0 equals 10 units, the complete graph models an uplink of 12 units at t0 (2 units along (s,r) and 10 units along (s,u)). Thus, the complete graph accurately models the end-to-end capacity between the nodes, but it does not model the end network bottleneck constraint.

3 Star graph model of bulk routing

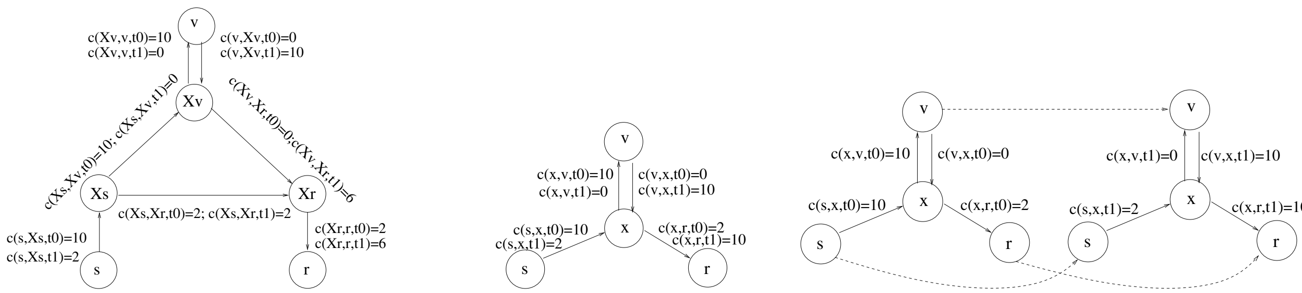
The bottlenecks are modeled in the complete graph as follows: link each data center node, v, to an exchange node, x_v. The capacity of arcs (v, x_v) and (x_v, v) are set to be:

$c(v, x_v, t) = \mathsf{ul}(v,t)$; $\quad c(x_v, v, t) = \mathsf{dl}(v,t)$.

The capacity of arcs between two exchange nodes x_u and x_v are set to be:

$c(x_u, x_v, t)$ = minimum $\{\mathsf{ul}(u,t), \mathsf{dl}(v,t)\}$.

That is, the capacity of arc (u, x_u) is set to the uplink from u, the capacity of arc (x_u, u) is set to the downlink from u, and the capacity of arc (x_u, x_v) is set to the end-to-end capacity from u to v. By construction, the value of flow $\mid f \mid \leq$ minimum $\{$ ul(s,T), dl(r,T) $\}$. Figure 2(i) incorporates bottlenecks into the complete graph of Figure 1(ii). The maximum flow from s to r during flow time t0 and t1 is: At time t0: f(s,r,t0)=2, f(s,v,t0)=4; At time t1: f(s,r,t1)=2, f(v,r,t1)=4. Thus, f(s,r)=8 units of flow during t0 and t1, which equals the bottleneck capacity at r.



(i) Bottleneck modeling in complete graph (ii) Equivalent star graph (iii) Time expanded star graph

Figure 2: Flow network modeling bulk transmission

In the bottleneck modeling, the arc (x_u, x_v) is redundant since capacity of augmenting path, p_{uv}, from u to v, $c(p_{uv}) =$ minimum $\{c(u, x_u), c(x_u, x_v), c(x_v, v)\}$ = minimum $\{\mathsf{ul}(u), \mathsf{dl}(v)\}$. The complete graph consisting of exchange nodes and the arcs between the exchange nodes is redundant, and can be replaced by a single exchange node, x, that is linked to all the data center nodes. (Proof omitted due to page constraints.) This leads to a star graph flow network where the n data center nodes are leaves and the exchange node x is the internal hub. Figure 2(ii) presents the star model corresponding to Figure 2(i). The nodes s, r, v are leaves while x is an interior node. The capacity of arc (v,x) is equal to the uplink capacity of v, and the capacity of (x,v) is equal to the downlink capacity if v.

The assumption underlying the model is that the bandwidth bottleneck (*i.e.,* minimum cut) for end-to-end transmission from u to v is the minimum of u's uplink bandwidth and v's downlink bandwidth. The uplink (downlink) bandwidth of a node is at most equal to the minimum of the node's Internet link and the node's backbone bandwidth. This is a reasonable assumption given that backbone transit providers must provide the bandwidth connectivity paid for by end networks, and in a network as dense and interconnected as the Internet, it is possible to find multiple paths between two end networks.

The star graph modeling bulk routing is defined by G=($V \cup x$,E,T) where V has n nodes representing s, r, and the (n-2) transit nodes; x is the central exchange node. All nodes $v \in V$ have storage capacity; x has no storage capacity. The arc set E = $\{(s, x), (x, r)\} \cup \{\ (v, x), (x, v)\ \mid\ \forall v \in$ V-s-r $\}$. The capacity at flow instant t=0,1, ..., T-1, is given by

$c(x, v, t) = \mathsf{dl}(v, t) \quad \forall (x, v) \in E$

$c(v, x, t) = \mathsf{ul}(v, t) \quad \forall (v, x) \in E$

Figure 2(iii) shows the time expanded graph; there is a copy of the node set and the arc set for each flow instant. Let v(t) represent the node at flow instant t. There are holdover arcs (v(t),v(t+1)), t=0, 1, .., T-2, $\forall v \in V$; the capacity of holdover arcs is infinite since it carries excess flow to the next time instant. When the assumption of bottleneck capacity holds, it can be proved that the star graph flow network is equivalent to the bottleneck complete graph flow network.

References

[1] L.R. Ford Jr. and D.R. Fulkerson. Constructing maximal dynamic flows from static flows. *Operations Research*, 6:419–433, 1958.

[2] Nikolaos Laoutaris, Michael Sirivianos, Xiaoyuan Yang, and Pablo Rodriguez. Inter-datacenter bulk transfers with netstitcher. In *Proceedings of the ACM SIGCOMM 2011*.