Journal of Southeast University (English Edition) Vol. 23, No. 2, pp. 226 – 231 June 2007 ISSN 1003—7985

Accelerated transmission in peer-to-peer network

Zhuang Yanyan Liu Ye Niu Lin

(Key Laboratory of Computer Network and Information Integration of Ministry of Education, Southeast University, Nanjing 210096, China)

Abstract: The design and evaluation of accelerated transmission (AT) systems in peer-to-peer networks for data transmission are introduced. Based on transfer control protocol (TCP) and peer-to-peer (P2P) substrate networks, AT can select peers of high performance quality, monitor the transfer status of each peer, dynamically adjust the transmission velocity and react to connection degradation with high accuracy and low overhead. The system performance is evaluated by simulations, and the interrelationship between network flow, bandwidth utilities and network throughput is analyzed. Owing to the collaborative operation of neighboring peers, AT accelerates the process of data transmission and the collective network performance is much more satisfactory.

Key words: peer to peer network; transfer control protocol (TCP); available bandwidth

Internet is such an infrastructure that, common users first get access to a regional network connection, such as an Internet service provider (ISP), and from there to the global Internet. It is often the case that the performance levels of network connections are based on their bandwidths, since more bandwidth normally means higher throughput and better quality-of-service to an application. As Internet grows in size and expands in scale, different people from different organizations may want to share electronic documents, video or audio materials. These files may be confidential so as not to be interpreted by a third party. Moreover, files may be so huge in size that the transmission may last for quite a long time. In such a scenario, secure and accelerated transmission has become a crucial concern.

In most cases, user access bandwidth, for example the 10/100 Mbit/s Ethernet, is adequate for most applications. However, when people from different organizations start to communicate with one another, bandwidth bottleneck exists in the Internet "clouds" which build up the whole communication path from the source to the destination. Concerning the application scenario, accelerated transmission (AT) takes into consideration both the "peer selection" and the real-time bandwidth measurement to fit the requirements of large file transmissions. Moreover, AT does not require any routing or addressing capabilities beyond unicast forwarding, and, by collecting feedback in the course of transmission, it provides significant benefits in terms of swiftness, security, and stability.

Received 2006-08-21.

Foundation items: The National Natural Science Foundation of China (No. 60573133), the National Basic Research Program of China (973 Program) (No. 2003CB314801).

Biography: Zhuang Yanyan (1983—), female, graduate, zhuangyanyan@seu.edu.cn.

1 Related Work

Peer-to-peer (P2P) systems have gained tremendous momentum in recent years. Peers communicate directly with one another for the sharing and exchange of data as well as other resources such as storage and CPU capacity. While AT is a novel file transmission system originated by the P2P research group at Southeast University, its idea comes from several earlier works by studying their efficacious mechanisms. PROMISE[1] deals with P2P real-time media streaming that poses more stringent resource requirements for real-time media data transmission. A novel P2P service called CollectCast is developed, which has a pattern of "one receiver collecting data from multiple senders". Similarly, Concast[2] enables a single receiver to treat a large group of senders as a single entity using a group identifier such as multicast. When multiple senders (members of the Concast group) transmit identical datagrams to a single receiver, at most one copy is delivered.

Unlike PROMISE, Concast or other P2P file sharing applications such as BT, peers that once participated in the bygone sessions were not allowed to cache files for possible file retrieval in the future due to security considerations. Besides, AT aims at the alleviation of link bottlenecks along the transmission path. It does not see all senders as a whole but treats them individually to achieve overall maximum bandwidth utilization.

2 Assumptions and Definitions

In this paper, we focus on long-distance, large-size file transmission in a peer-to-peer context. Due to potential large file sizes and thus long transfer time, bandwidths in the transmission path are of primary concern.

2.1 Bandwidth definitions

We define a network path as the sequence of links that forward packets from the sender (source) to the receiver (destination). In this paper, we use the following bandwidth definitions[3]:

- Link bandwidth The data transmission rate of a certain link. We call the lowest link bandwidth in a network path the bottleneck link bandwidth.
- Capacity The maximum IP-layer throughput that the path can provide to a flow, when there is no competing traffic load (cross traffic).
- Available bandwidth The maximum IP-layer throughput that the path can provide to a flow, given the path's current cross traffic load.

The link with the minimum transmission rate determines the capacity of the path, while the link with the minimum unused capacity limits the available bandwidth. Performance of AT largely depends on the available bandwidth that a network path can provide.

2.2 Network model

Link available bandwidth is easily affected by cross traffic. There are two kinds of cross traffic in an actual situation[4]: path persistent cross traffic and one-hop persistent cross traffic. As shown in Fig. 1[4], path persistent cross traffic runs through the entire transmission path while one-hop persistent cross traffic occurs in only one link.

As a matter of fact, cross an traffic in actual network is inevitably a combination of both one-hop persistent cross traffic and path persistent cross traffic, which makes the situation complex. We will see in section 5 that cross traffic conforms to some traffic distribution models so we may go deeper into its behavior monitoring.

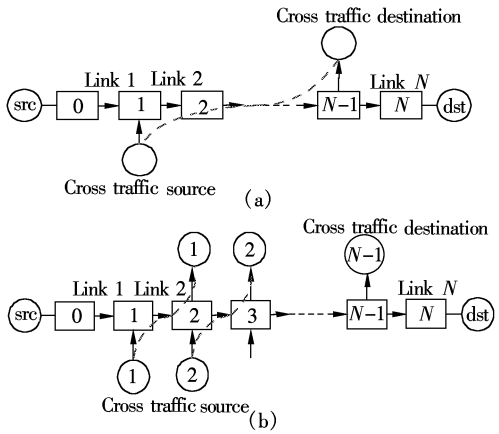

Fig. 1 Cross traffic. (a) One case of path persistent cross traffic; (b) One case of one-hop persistent cross traffic

3 Overview of AT

Traditionally, Internet offers two types of best-effort channels to applications: point-to-point unicast and point-to-multipoint multicast. Symmetry considerations suggest that Internet should support a third type of channel: multipoint-to-point. When information needs to flow in the direction opposite the multicast, the only option is to use multiple unicast. In P2P systems, peers directly communicate with each other via specific routing protocols. Thus, we are able to make possible the multiple point-to-point communication channels by virtue of P2P.

The AT architecture consists of a set of peers interconnected through a P2P substrate (see Fig. 2). The P2P substrate maintains connectivity among peers, and manages peer membership. AT operation is independent of the underlying P2P architecture; therefore, it can be deployed on the top of P2P substrates such as Gnutella[5], CAN[6], Chord[7] or Pastry[8].

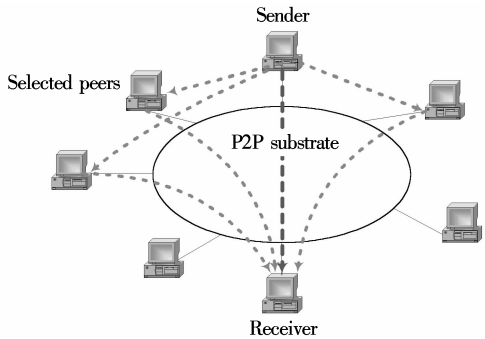

Fig. 2 Structure of AT

Organized by a peer-to-peer substrate, AT reflects the P2P philosophy of dynamically aggregating the limited capacity of peers to perform a task, i. e., the file transmission, which is traditionally performed by two dedicated entities.

4 Design of AT Operation

The feasibility of AT makes it useful in the context of long-distance and large-size file transferring. For practical consideration, the file to be sent is equally fragmented beforehand so that active peers may collectively send the file segment by segment. All segments are assigned to participating peers according to

their bandwidth capacities towards the receiver. A file transmission session in AT is established as follows.

4.1 Preparative work

1) Before sending the destined file, it is required that the sender split the file into equal-length data segments, altogether F segments. The name of each segment file comprises two parts: the original file name and the partition sequence number (from 1 to F). The sequence number becomes the postfix of each file segment.

2) The sender selects n peers from its neighbors as its transmission relay. Then it tells these n peers the IP addresses of the receiver.

Up till now, we have made adequate preparations for accelerated transmission. Each file segment is suffixed by partition sequence number so that not a single fellow peer, except sender and receiver, would be able to interpret the original file. Without all the segments, peers cannot reconstruct the entire file. When the session is over, all the collaborating peers cannot cache the file fragments for future use and any session-related information is obliterated permanently, all of which guarantees the confidentiality of the session.

In this paper, however, we do not take into account packet losses or peer failures, which are due to network fluctuations and limited peer reliability. As can be seen in section 6, they are left for future research.

4.2 Data assignment and transmission

First we assume that routes from candidate peers to the receiver do not change during the course of the session. This indicates that the inferred topology is a stably structured graph. Previous studies adopted the same assumption, which has been verified by Internet measurement studies. For example, Ref. [9] indicates that the end-to-end Internet paths often remain stable for a significant period of time.

4.2.1 Bandwidth measurement

There are $n+1$ peers sending fragmented file to the receiver, including sender and its n selected peers. Each peer is assigned a number of fragments to send, in proportion to each sender's actual bandwidth capacity towards the receiver. Therefore, available bandwidth measurement has become the crucial operation.

In AT, both the sending and receiving parties are simply ordinary peers in distributed environments. Neither of them, nor the selected peers is to undertake the task of central server. Therefore, we may well presume that no peers will response ACK to WWW requests. Each active sender continuously sends TCP SYN packet pairs to port 80 of the receiver. The length of these packets P_i, is set to Ethernet MTU—1 500 bytes. As a non-web server, the receiver will reply to each SYN packet with TCP RST, of which the arriving time is T_i. The time interval of two adjacent samplings is $\Delta_i = T_{i+1} - T_i$. The available bandwidth B_i, that determined by the current tightest link, is then calculated as

$$B_i = \frac{P_i}{T_{i+1} - T_i} \tag{1}$$

The bandwidth information is collected by a daemon running on each participating peer. The n peers return the collected results to the sender for dynamic analysis and fragment allocation.

4.2.2 Data assignment

The following algorithm is fulfilled by the sender according to the statistics collected from participating peers. B_i denotes the available bandwidth from the i-th peer to the receiver. If the sending peer is the original sender, then the bandwidth value is denoted by B_A. F is the number of total file fragments to be sent.

Algorithm 1 Data assignment

While (Transmission_Done == FALSE)

$F = \lceil F/2 \rceil$;

For i in 1 to $n+1$

If i is the sender

Then $F_A = \left\lceil \frac{B_A}{B_A + \sum_{i=1}^{n} B_i} F \right\rceil$;

Else $F_i = \left\lceil \frac{B_i}{B_A + \sum_{i=1}^{n} B_i} F \right\rceil$ $(1 \leqslant i \leqslant n)$;

Send these allocated fragments to the receiver;

If ($F == 0$) Transmission_Done == TRUE.

By this mechanism, the sender judiciously chooses the sending peers and orchestrates them in order to yield the best quality for the receiver.

4.2.3 Data collecting

When transmission is done, the receiver reconstructs the fragmented file into an intact whole, by concatenating all the fragments according to their caudal sequence numbers.

5 Evaluation of AT

In this section we evaluate the performance of AT, using the Network Simulator (NS2)[10], the second edition. Our goals in conducting this evaluation study are two-fold. First, place the performance of AT in the context of idealized schemes, serving as a sanity check

for the intuition behind AT. Second, understand the impact of dynamics, such as network fluctuations, on AT.

Though the SPIS[11] system has been implemented by our P2P research group as a P2P substrate, the code is still under refinement due to some practical and security considerations.

5.1 Simulation setup

The topology used in the simulation is depicted in Fig. 3, which is generated by NS2. In this scenario, peer 2 is the sender and peer 1 is the receiver. Peers 5, 8, 11 are selected as relay nodes by peer 2. Peers 1, 2, 5, 8 and 11 belong to the same per-to-peer substrate (Gnutella). Other peers, namely peers 3, 6, 9 and 12, are responsible for cross traffic in each transmission path. In Fig. 3, flows that are in dashed lines belong to cross traffic, whereas those in dash-dotted lines are the actual data packets being sent. We assume that there only exists path persistent cross traffic. As for one-hop persistent cross traffic, which only runs through one data link and makes the simulation model more complicated, its impact on our system will be left for future work due to some practical and technical problems.

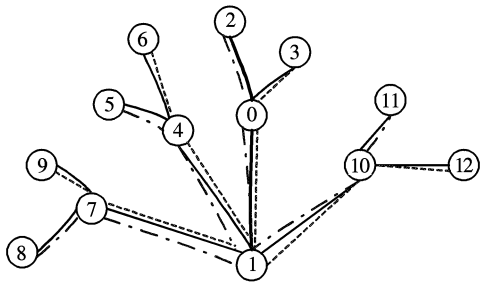

Fig. 3 Simulation scenario

Network flow of ordinary transmission, e. g., file sending and receiving between the sender, receiver and relay nodes, follows the constant bit rate (CBR) model, which sends data packets at a constant bit rate. This is reasonable in most common situations. Cross traffics are modeled by four kinds of random variables, namely exponential, Pareto, Poisson, and CBR, just for universality and comparison between different traffic models. We study the impact of these models on the AT transmission system, assessing the swiftness and stability of AT in different scenarios. The parameters of each link are shown in Tab. 1. As can be seen, bottlenecks inevitably dwell in the transmission path. Due to these tight links, transmission efficiency and bandwidth utilities are degraded. In the latter part of this section, we are to be convinced of the performance and efficiency advantages of AT over traditional file transferring.

Tab. 1 Link parameters

From_Node	To_Node	Link_Bandwidth	Actual_Bandwidth	Traffic_Model
2	0	200	70	CBR
3	0	150	30	CBR (cross traffic)
0	1	100(bottleneck)	100	Compound
5	4	100	50	CBR
6	4	150	30	Exponential (cross traffic)
4	1	80(bottleneck)	80	Compound
8	7	160	30	CBR
9	7	120	10	Pareto (cross traffic)
7	1	50(bottleneck)	40	Compound
11	10	80	20	CBR
12	10	10	≤10	Poisson (cross traffic)
10	1	30(bottleneck)	≤30	Compound

5.2 Performance analyses

5.2.1 Network flow and bandwidth utility

In Fig. 4, there is only one transmission path, from peer 2, peer 0 to peer 1, and not a single neighboring peer is selected. The network flow would be quite smooth. The bandwidth, however, is never to be fully utilized. Some of the link capacity is available but wasted and the overall throughput is determined solely by transmission bottleneck. In AT, however, the sender judiciously orchestrates selected peers for sending the whole file, for achieving the best service quality.

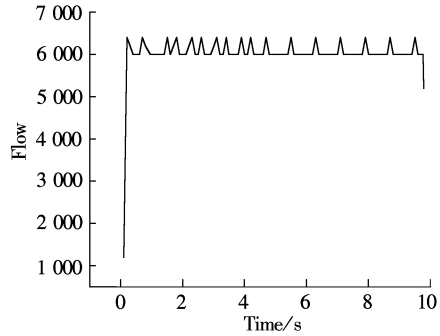

Fig. 4 Network flow of ordinary transmission (CBR)

Fig. 5 plots some examples of coordinate collaboration of peers. From Fig. 5, the fluctuant extent of the actual sending flow is proportional to the cross traffic. That is, the more drastically the cross traffic fluctuates, the more vertiginous the actual sending flow will be. In Figs. 5 (a) and (b), neither the changes in Exp and Pareto cross traffic models are very acute, thus, the corresponding actual flow fluctuates regularly. In some cases such as with the Poisson distribution, where the cross traffic becomes stochastic and changes randomly, the resulting actual data flow fluctuates drastically (see Fig. 5 (c)).

Despite the flow fluctuation, spare bandwidth resources belonging to others can be adequately made use of. In this way, AT dynamically aggregates the lim-

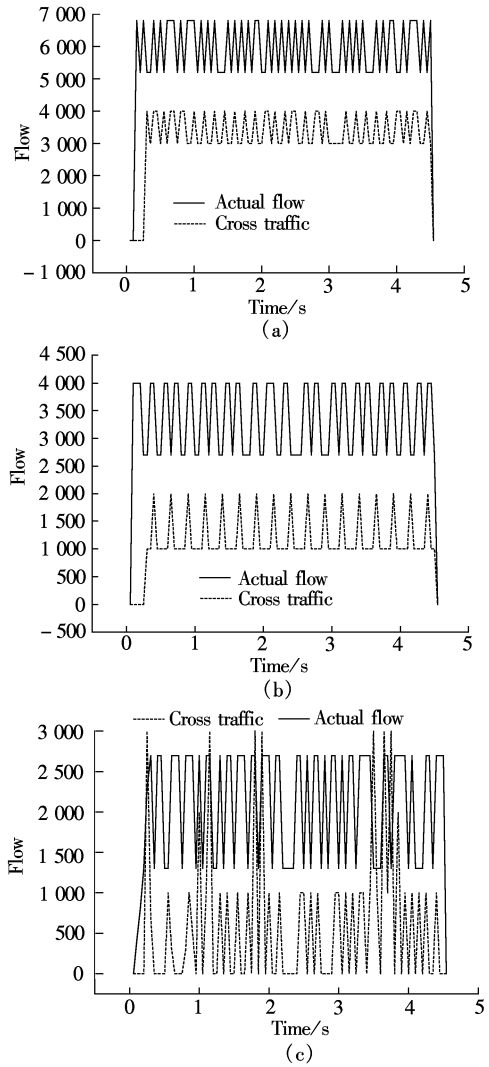

Fig. 5 Peer collaborations. (a) Exp cross traffic; (b) Pareto cross traffic; (c) Poisson cross traffic

ited capacity of peers to perform the mutual transferring task.

5.2.2 Throughput

By definition, throughput is the mount of data transferred from one place to another or processed in a specified amount of time. In Fig. 6, we compare the overall throughput in two scenarios: accelerated transmission and ordinary transmission (OT).

In Fig. 6, the OT throughput is also rather smooth, whereas that of AT is unstable, fluctuating as time elapses. However, the average performance level in the latter scenario is almost twice or three times as much as that of the former. If the entire transmission duration of OT had lasted 10 s, as can be seen, the same task would have been accomplished by AT by the time of 4.5 s already.

It is worth noting that, when the selected neighbors belong to different ISPs other than that of the

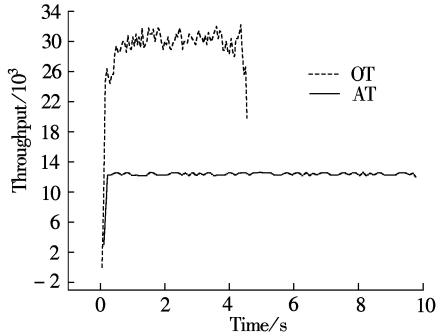

Fig. 6 Throughput in two scenarios

sender, system performance is sure to be boosted owning to different bandwidth capacities and qualities of service that different ISPs can provide. Under this circumstance, AT is especially useful for practical applications and widespread deployment. In short, AT is in possession of the attractive features of swiftness, security and stability, making the system highly practicable and effective.

6 Conclusion and Future Work

Traditionally, file transmission is performed by two dedicated entities. AT is a novel peer-to-peer file transmission system that constitutes several sending parties, monitors transfer behaviors of peers and reacts to connection degradation dynamically.

The simulation results show that in a particularly complex network environment, AT is competent for large file transmissions, which brings us an analytically tractable model for analyzing network flow, bandwidth and overall throughput. Extensions of this work to further discussion on the impact of one-hop persistent cross traffic, packet losses and peer failures are under investigation now and the refinement of the SPIS system in the ongoing procedure is left for future research.

References

[1] Mohamed Hefeeda, Ahsan Habib, Boyan Botev, et al. PROMISE: peer-to-peer media streaming using CollectCast [C]//*Proceedings of the* 11*th ACM International Conference on Multimedia*. Berkeley, CA, USA, 2003: 45 – 54.

[2] Calvert L, Griffioen J, Mullins C, et al. Concast: design and implementation of an active network service [J]. *IEEE Journal on Selected Area in Communications*, 2001, **19**(3): 426 – 437.

[3] Melander Bob, Bjorkman Mats, Gunningberg Per. A new end-to-end probing and analysis method for estimating bandwidth bottlenecks [C]//*IEEE Global Telecommunications Conference*. San Francisco, CA, USA, 2000: 415 – 420.

[4] Liu M, Li Z C, Guo X B, et al. An end-to-end available bandwidth estimation methodology [J]. *Journal of Software*, 2006, **17**(1): 108 − 116. (in Chinese)

[5] The Gnutella protocol specification v4.0 [EB/OL]. (1999-04-17) [2005-06-10]. http://www9.limewire.com/developer/gnutella_protocol_0.4.pdf.

[6] Ratnasamy S, Francis P, Handley M, et al. A scalable content-addressable network [J]. *Computer Communication Review*, 2001, **31**(4): 161 − 172.

[7] Stoica I, Morris R, Liben-Nowell D. Chord: a scalable peer-to-peer lookup service for Internet applications [C]//*Proceedings of the ACM SIGCOMM'*01 *Conference*. San Diego, California, USA, 2002: 149 − 160.

[8] Rowstron A, Druschel P. Pastry: scalable, distributed object location and routing for large-scale peer-to-peer systems [C]//*Proc of* 18*th IFIP/ACM International Conference on Distributed Systems Platforms* (*Middleware* 2001). Heidelberg, Germany, 2001: 329 − 350.

[9] Zhang Yin, Duffield Nick, Paxson Vern, et al. On the constancy of Internet path properties [C]//*Proc of ACM SIGCOMM Internet Measurement Workshop*. San Francisco, CA, USA, 2001: 197 − 211.

[10] Sandeep Bajaj, Lee Breslau, Deborah Estrln, et al. Improving simulation for network research, 99-702b[R]. University of Southern California, 1999.

[11] Research Group of P2P Technology of Key Laboratory of Computer Network and Information Integration of Ministry of Education. Detailed designing report of the SPIS system [R]. Nanjing: Southeast University, 2006. (in Chinese)

P2P 网络中的加速传输服务

庄艳艳　刘　业　钮　麟

(东南大学计算机网络和信息集成教育部重点实验室,南京 210096)

摘要:介绍了加速传输服务系统的设计. 借助 TCP 协议和 P2P 网络的底层路由机制,系统能够选择高性能的对等节点,协调各个节点的传输状态,在网络和节点性能发生变化时进行动态调整传输速率,从而以较低的开销动态适应网络性能的变化. 通过仿真对系统性能进行了评价,并对网络流和带宽利用率、网络吞吐量之间的关系进行了分析. 结果表明:在加速传输系统中,接收方之间的相互谐调和多个邻居节点间的相互协作,使得数据传输速度得以加快,系统总体性能得以提高.

关键词:P2P 网络;TCP 协议;可用带宽

中图分类号:TP393