Peer-to-Peer Traffic in Metro Networks: Analysis, Modeling, and Policies

Takeo Hamada*, Kaoru Chujo*, XiangYing Yang**, Takafumi Chujo*

* Fujitsu Laboratories of America
1240 East Arques Avenue, M/S 345, Sunnyvale, CA 94085-5401
{thamada,kaoru,chujo}@fla.fujitsu.com

**University of Texas at Austin

Abstract— **Peer-to-Peer (P2P) services have emerged as a popular way to share large multimedia contents, giving rise to large traffic volume and new traffic patterns different from conventional Internet applications. In this paper, we report our studies on P2P traffic analysis, modeling, and evaluation of policies for the management P2P traffic in metropolitan networks. We conducted crawler-based measurement and analysis of P2P traffic using Gnutella. Based on the empirical parameters obtained through the measurement, known modeling techniques of P2P peers and contents, and GeoPlot geomapping technology, we built a P2P traffic simulator using JavaSim, a Java-based network simulation tool. The simulator is scenario-driven, allowing its user to customize network settings and service environment to examine the behavior of P2P traffic. With the P2P simulator, we were able to recreate spatio-temporal traffic pattern characteristic to P2P services observed in August 2002. We evaluated three different P2P traffic management policies using the simulator, which shows that traffic localization using a peer selection policy at super peers is possible, containing P2P traffic to the local metro network as much as 40%. We also report our preliminary analysis on the impact of FTTH broadband deployment to P2P traffic. With increasingly bulkier and burstier P2P induced traffic, the study suggests that a new class of metro data networking infrastructure may be needed in the near future.**

Keywords — Peer-to-Peer, Traffic Management, Crawling, Gnutella, Metro Network, Content Delivery Policy, Traffic Simulation, Geomapping, JavaSim.

** Work done during internship at Fujitsu Laboratories of America

1. INTRODUCTION

Peer-to-Peer (P2P) file sharing services have emerged as a popular way to share personal audio and video files. P2P services such as Gnutella and Kazaa have gained large service share and mindshare among young population, and new P2P services with different service flavor emerge somewhere in the Internet almost every month. In a recent study, .it has been reported that P2P traffic regularly accounts for 60% of the total Ethernet traffic, and the ratio goes as high as 70% in peak hours. The total volume of P2P traffic is estimated to be 80% of the total downloading traffic in the Internet [1]. In ISP domains, the ratio of P2P traffic to the total Internet traffic has been constantly on the rise. The technology has brought several social and legal implications. For example, recent discussions on intellectual property rights of digital contents have been significantly motivated by the evolution of P2P contents exchange. For the telecommunication industry, however, the rapid growth of P2P traffic has become at the same time a blessing and a curse. On one side, P2P has become the main driving source of the Internet traffic growth, which is roughly estimated to be doubling annually. On the other side, there are two major problems with the P2P traffic. The first one is that it does not contribute to the ISP revenue, and therefore to the telecommunication service provider revenue. The second problem is that the P2P services tend to carry very large data objects, incurring very long downloading time. Congestion caused by P2P downloading traffic has become to be seen increasingly as one of the major sources of service quality degradation of other Internet services.

Despite its urgency of the issue, it might be said that comparatively little has been known of the nature of the P2P traffic in contrast to other telecommunication services such as telephone voice. The root of this issue lies in the fact that the P2P does not fit the conventional telecommunication concept of service, as P2P traffic is mainly generated by self-replicating digital contents and automatically downloading PCs. Nor its usage pattern and its traffic pattern do not connect well with the established traffic engineering techniques. Since P2P services are themselves not so sensitive to network QoS; making the situation possibly worse for other network QoS-sensitive services such as VoIP and video streaming. In particular the situation may become especially grave when all networking services are being consolidated to be Internet-based.

Although efforts are on-going at various fronts to address the P2P traffic issue, two things need to be addressed in order to provide a better service environment for the Internet-based telecom, or telecommunication network fully integrated with the Internet. The first issue is to better understanding of the P2P traffic behavior. Through analysis of its temporal and spatial traffic patterns, we have developed a P2P traffic simulation model, taking into account P2P user behavior and their geometric distribution in the Internet. In particular, we focused on the P2P traffic patterns in the metropolitan networking.

Once the P2P simulation model is established, the second step is to study desirable P2P traffic management techniques. Since management policy needs to be tested and evaluated using the simulation, it is essential that the P2P traffic model is accurate.

2. PREVIOUS WORKS ON P2P TRAFFIC ANALYSIS

Our P2P traffic simulation model study started from previous empirical researches on P2P traffic. Gerber et al. [2] report behavior and its analysis of P2P traffic of two different broadband regions. Sen and Wang [3] report measurement and analysis of three P2P systems, which include FastTrack, Gnutella, and DirectConnect. Both works [2-3] use Cisco Netflow to measure flow-based P2P traffic. Ripeanu et al. [4-5] studied Gnutella by using a crawler, which collected topology information and connectivity among peers. Leibowitz et al., [6] studied another popular P2P service, Kazaa, obtaining measurement data from a large Israel ISP. Saioru et al. [7] studied four representative contend delivery system, namely WWW, Akamai, Gnutella, and Kazaa, by analyzing trace of

Internet traffic obtained at University of Washington.

3. MODELING OF PEER-TO-PEER SERVICES AND SPATIO-TEMPORAL TRAFFIC PREDICTION

Built upon the previous works on P2P traffic behavior [1-7], we set upon building a set of behavior models in order to simulate P2P traffic, and to derive and to predict spatio-temporal P2P traffic pattern in the metro networking. Although existing traffic analysis works already revealed general pattern of P2P traffic either at aggregated flow level, as it can be measured by router-resident monitoring programs such as Netflow [2-3, 6-7], or at service overlay network level, which is collected by crawler and similar programs [4-5], one piece of information was yet to be discovered to use these information for our simulation model. It is geometric distribution of peers, which determines spatial distribution of P2P traffic. It is of our particular interests, since provisioning and management in the metro networking is constrained by the physical capacity of the network, and it is bound by the real topology of the network and geometric distribution of the users.

To understand the spatio-temporal pattern of the P2P traffic, one has to know the geographic distribution of traffic sources and their destination, and their temporal behavior patterns. We took a new approach, which combines knowledge on P2P traffic sources obtained from previous works, empirical modeling parameters obtained by monitoring Gnutella control traffic, various published social/economic statistics such as population density, and the geomapping technique. We used Gnutella as a representative of P2P services, as it has remained one of the most well known, widely accepted, and popular P2P services. Analysis of Gnutella control traffic also enabled us to collect information on P2P data objects and their distribution among peers.

3.1 P2P Peer Behavior Model

In studying behavior of P2P peers, we used an open source crawler software Gnutellavision [8], which was developed at UC Berkeley. We modified the original Gnutellavision so that we are able to passively monitor and log all users requests (queries) on the Gnutella control channel and sharing content information matching any specific query. We ran the Gnutellavision-based crawler during April 2002 through July 2002 and observed total 6,861,645 peers information. Based on the activity log of peers on Gnutella control channel and results from previous analyses [1-2], we extracted an daily fluctuation pattern of Gnutella peers as illustrated in Figure 1.

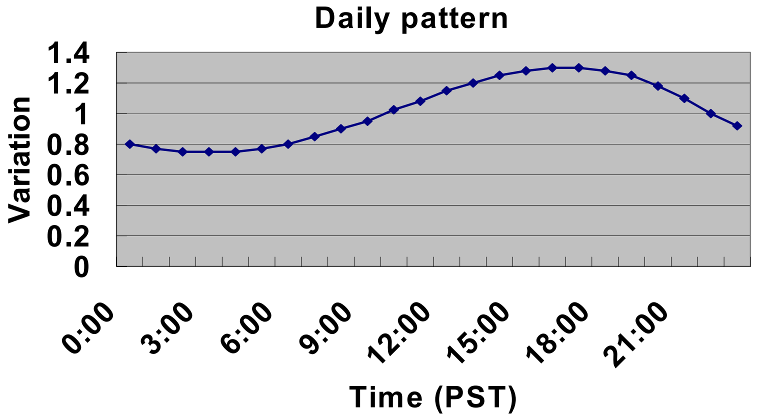

FIGURE 1: DAILY FLUCTUATION PATTERN OF GNUTELLA PEERS

Note that a daily pattern of P2P traffic is clearly visible in the figure, but much less so compared to those known for Web traffic or telephone traffic. We suspect that due to very large data objects of P2P services and automatic downloading, peer activities are averaged over 24 hours. We observed similar daily fluctuation pattern for other P2P services such as Kazaa. With this peer activity pattern and geographic information on peer population distribution obtained by geomapping, it becomes possible to estimate time-varying spatial traffic induced by P2P services. The total number of P2P service users fluctuates, but in general was on the increase during the period we conducted our measurement. Statistics on major P2P service users (peers) is available at Slyck home page [9]. Over 3,000,000 users were constantly observed for the top five P2P services during the period we conducted the measurement in August 2002.

3.2 Content Size Distribution

Using the crawler, we first collected 171,498 queries and observed 60% of queries have file extension somewhere in the query. We categorized these queries into 5 content types, audio, video, image, document, and software. We observed percentages of user requests for the above five 50%, 38%, 2%, 1%, and 9%, respectively. Then we collected total 175,647 contents (audio: 82,801, video: 78,366, software: 14,480) using file extensions as their keys. We found average file size of audio, video, and software contents are about 4.5MB for audio, 52.5MB for video, and 34.5MB for software. Their respective distributions are illustrated in Figures 2-4. Although audio files occupies the largest body of request from peers, its impact on the traffic is much smaller compared to video objects, as the average object size is significantly smaller. Based on these distributions, we derived the P2P content size distributions as a combination of log-normal distribution and uniform distributions.

FIGURE 2: GNUTELLA CONTENT SIZE DISTRIBUTION OF VIDEO OBJECTS (AUGUST 2002)

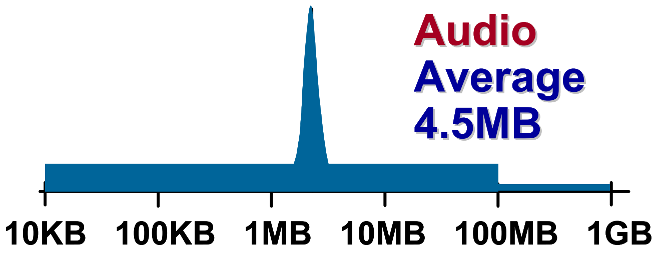

FIGURE 3: GNUTELLA CONTENT SIZE DISTRIBUTION OF AUDIO OBJECTS (AUGUST 2002)

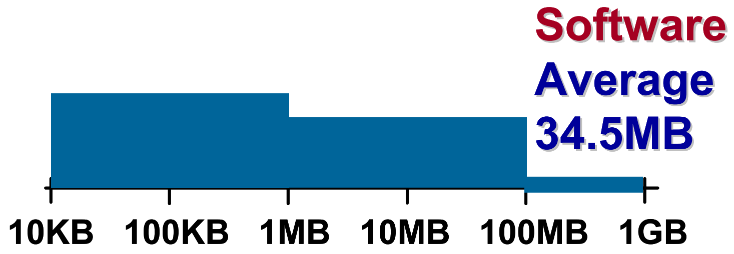

FIGURE 4: GNUTELLA CONTENT SIZE DISTRIBUTION OF SOFTWARE OBJECTS (AUGUST 2002)

In Gnutella, a peer sends a query to the control channel. Other peers respond to the query by returning a list of contents matching to the query string. The requesting peer selects an object from the list, starting a content transfer (downloading). Therefore, volume of the P2P traffic and its spatial pattern is determined by the following three parameters; (1) geometric position of the requesting peer, (2) geometric position of the selected peer, (3) size of the content to be transferred. The information available on the crawler do not provide any of the above information directly, as it only provide queries issued by peers and responses to the queries by other peers; they do not reveal how or whether the two communicating peers actually start content transfer. To overcome the situation, we adopted the following hypotheses, in our simulation model.

(1) Queries and content downloading have high correlation. In other words, peers almost always start content transfer as soon as they receive query results from other peers.

(2) Statistical properties of queries and those of contents resemble each other. In other words, a popular query string indicates popularity of the intended content, and vice versa.

(3) Query results are returned in random and the requesting peer selects a peer from the responding peers in random.

The last hypothesis is actually a subject of our research. As we discuss in Section 5, peer selection process can be controlled by management policies of the domain, which may influence the peer selection process at peers and possibly content caching at super peers. For the time being, we assume that hypothesis (3) holds for the current P2P traffic. Kunwadee [10] observed that distribution of query strings generally follows a power law, which is also known as generalized Zipf's law. Applying this observation to the second hypothesis, we assumed that the popularity distribution of downloading contents also follows the generalized Zipf's law, with the same power law coefficients as the query strings themselves. The power law appears in other areas of P2P networking. For example, Jovanovic et al [11] studied topology of Gnutella in 2000, discovering that the power law also applies to its node degree distribution.

3.2 Geomapping of Peers

As we discussed in the previous section, we assume that a peer that issued a query starts content transfer as soon as they receive responses to the query from other peers. What it implies is that once geographical position of requesting peers and the responding peers are identified, we are able to infer the actual content transfer traffic pattern. The technique that allows this mapping, which establishes correspondence between given an IP address and its geographical location, is called geomapping, also known as geotargetting. The technique has been applied to various applications, from customer service to SLA management. We used GeoPoint, a commercial geomapping product by Quova, Inc. [12] GeoPoint allowed us to identify geographical locations of Gnutella peers at the resolution of ZIP address level. Using geomapping in our crawler-based Gnutella monitoring process, we observed that more than 70% of peers consistently reside in the US. We also observed that peers are distributed over 2200 ZIP areas, whereas more than 45% of peers are concentrated in the top 50 ZIP areas, implying that the majority of Gnutella users live in big cities. The 20 metropolitan areas considered in our study are; Atlanta, Boston, Buffalo, Chicago, Dallas, Denver, Houston, Kansas City, Los Angels, Miami, Milwaukee, Minneapolis, New York, Philadelphia, Phoenix, St. Louis, San Diego, San Francisco, Seattle, Washington DC, in alphabetical order.

We speculate that it corresponds to availability of broadband connection in the respective regions; with such large average content sizes, obviously P2P services do not make much sense without broadband. We also found that only 15% of peers are willing to share contents with others, while the ratio is fairly constant across all the metropolitan areas. Other additional important information obtained from GeoPoint is that it was able to tell types of access lines, such as dial-up modem, cable, or DSL. This information is useful when correlating empirical observation on P2P services with broadband deployment in the region, in particular when evaluating impact of different broadband scenarios to P2P traffic growth and impact of different policies to P2P traffic patterns. Another interesting finding is that only 6-11% of links in Gnutella control channel stays within the same metro networking area. In other words, P2P services are very likely to induce a large amount of inter-metro content transfer traffic, contributing increase in long distance, Internet backbone traffic.

4. P2P SIMULATOR USING NETWORK ABSTRACTION

We built our P2P traffic simulation model using JavaSim [13], a component based, object-oriented simulation environment written in Java originally developed at Ohio State University. The idea of the simulation is to simulate and to recreate traffic demand matrix from grounds up, from synthesized behavior of P2P peers, based on empirical observations. Each control channel requests/messages and content transfers among peers are to be simulated as they would have occurred in the real P2P services. Since our purpose was to derive the traffic demand matrix with reasonable accuracy, our simulation did not simulate at the detailed protocol level, such as TCP. Instead, each P2P service is represented by two abstract phases, a request phase using the control channel and a content transfer phase. Each of the two phases incurs certain traffic among peers. In particular, the content transfer phase starts transferring a large volume of data between selected peers, associated with a delay of the content size divided by an estimated bandwidth between the peers. The estimated bandwidth is given by the smaller of the two access bandwidths of the selected peers, one near-end and the other far-end, as access links are almost always the bottlenecks in the Internet end-to-end transfer.

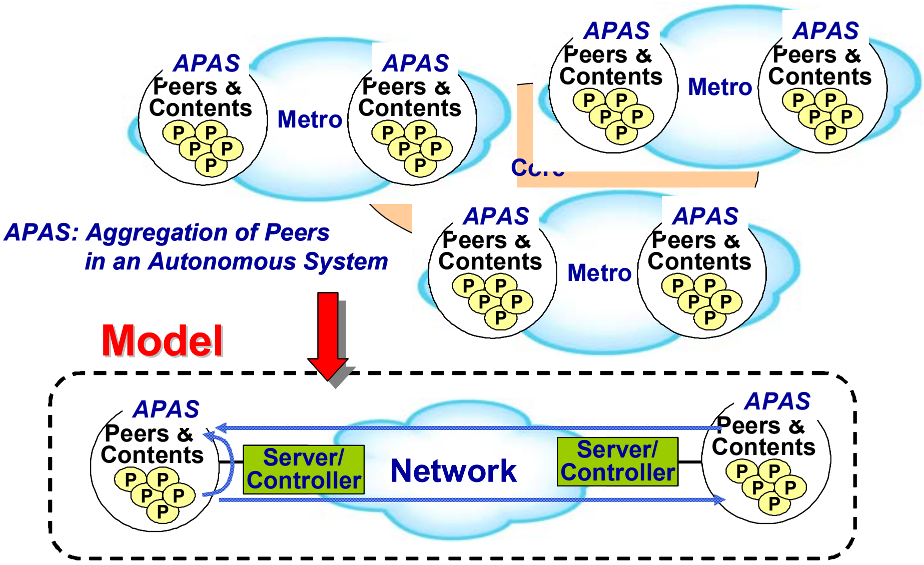

FIGURE 5: P2P SIMULATION ARCHITECTURE USING NETWORK ABSTRACTION

Figure 5 illustrates architecture of the P2P simulation. Although it is ideal to simulate individual behavior of P2P peers, due to the sheer number of 3,000,000 P2P peers, certain practical considerations were necessary. We introduced an aggregation of peers at Autonomous System (AS) level, which we called APAS (Aggregation of Peers in an Autonomous System). With APAS, individual responses of P2P peers are abstracted out to be collective responses at APAS level, and the probability distributions at the peer level are appropriately transformed so that the collective response and their induced traffic demand are still correct at the APAS level analysis. Since JavaSim is Java-based, object-oriented nature of the system makes it fairly routine to realize hierarchical composition of simulation components, contributing to the reduction of storage space for the simulation and indirectly to the reduction of simulation running time.

Figure 6 shows the component architecture of the simulation. Using the network abstraction by APAS, the simulator contains six different types of components.

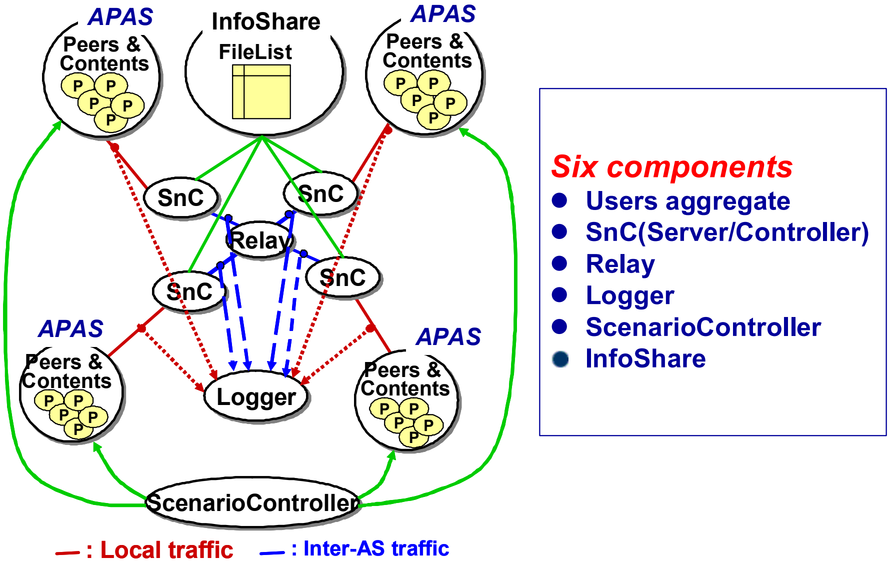

FIGURE 6: P2P SIMULATION COMPONENT ARCHITECTURE

- *UserAggregate:* an internal component within APAS, which represents collective user (peer) behavior within the AS such as content request and downloading. As we noted earlier, individual peer behavior is replaced by an abstract level interactions between UserAggregate components in respective APASes. At this abstraction level, all P2P services are considered to behave invariably the same, which repeats request-downloading cycles with presupposed probabilistic distributions.
- *SnC:* a component representing P2P servers and policy controllers within P2P services. Although P2P services are implied to be server-less, certain point of entries is necessary to provide a list (directory) of on-line peers. In Gnutella, it is called servant. In Kazaa, a super peer provides a similar role as Gnutella servant, providing directory entries of popular peers/contents. As we will discuss in Section 5, these super nodes or super peers can serve as policy control points for P2P traffic management.
- *Relay*: a component representing transmission characteristics over the Internet such as transmission errors and delay.
- *Logger*: a component logging all the messages between APASes.

- *ScenarioController:* this component controls behavior of simulation components through various behavioral parameters such as requesting pattern of peers, daily change pattern of the number of online peers, and daily update frequencies and popularity changes of online contents. Through this component, management policy changes are also easily realized in the simulation, so that impact of different policies on the traffic demand matrix can be experimented, as we will explain in Section 5.
- *InfoShare:* this is a component to provide information on content files dispersed over the network. For example, number of replicas of a particular content and their file size are stored in this component.

Our first goal in this simulation study was to recreate the P2P related traffic in the Internet as accurately as possible, so as to check our simulation assumptions on P2P systems and their behavior. We assumed that 3 million peers are online daily on average, which is based on the P2P service statistics obtained at the time (August 2002) [9]. Note that the number is not only for Gnutella, but all top five P2P services combined that existed at that time. Total 162 APASes; represented world wide distribution of P2P peers in this simulation, which include top 7 IP domains, top 20 US metropolitan areas listed in section, and other significant ASes in the Internet. Based on the simulation, our estimation on P2P content transfer traffic within the US turned out to be 60PB per month. Following an RHK report on the Internet traffic [14][1] studying the same time period, it was estimated that P2P accounts for 46% of total Internet traffic. Of the P2P induced content transfer traffic, 80% was estimated to be video files. Although there is no accurate statistics on the Internet P2P content transfer traffic, its ratio to the total Internet traffic fairly matched to the numbers reported, based on the measurements on selected access/local networks [2, 7].

Our second goal was to recreate daily traffic pattern of P2P services using the simulation model

---

[1] This report is not publicly available.

discussed in Section 3. Figure 7 illustrates a simulation result of first 24 hours for a hypothetical metro network with peer population of 100,000, with eight access routers serving as PoPs for eight different ASes. Each AS is represented as a separate APAS in this simulation, with respective population sizes ranging 30% (heaviest) to 5% (smallest). Figure 7 shows daily traffic pattern between AS_1 and other ASes in the network.

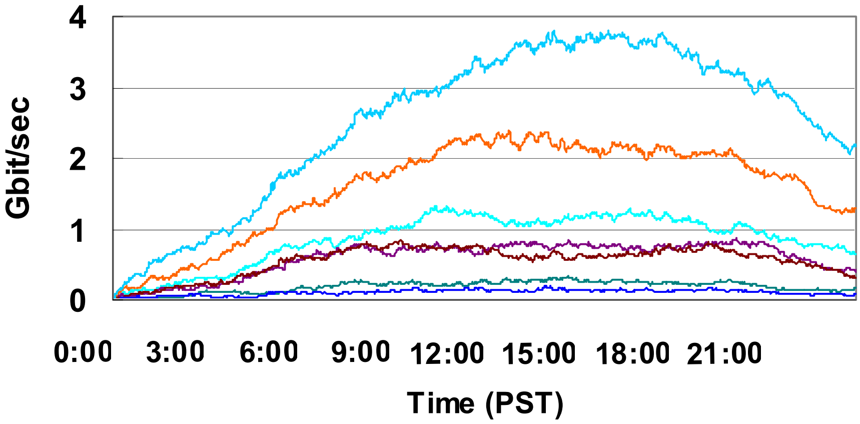

FIGURE 7: P2P TRAFFIC PATTERN BETWEEN AS_1 AND AS_2-AS_7

For a reference, Figure 8 shows an actual measurement of daily traffic pattern including P2P services by Saroiu et al. [7]. We concluded that our simulation model have been able to recreate P2P traffic behavior with reasonable accuracy so that we are able to proceed to study P2P traffic management policies in ISP domains.

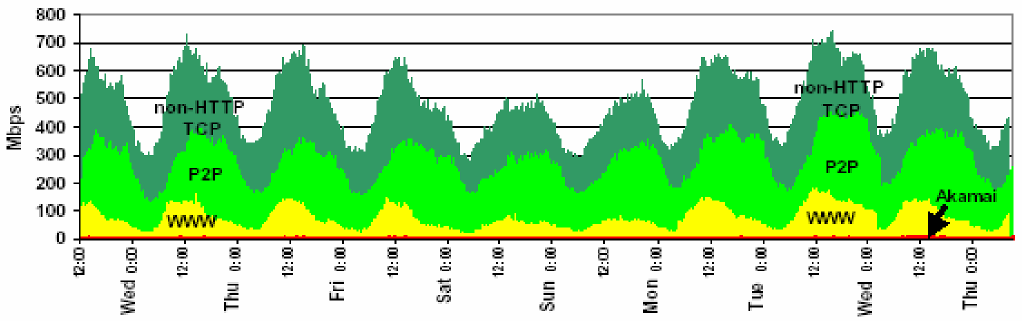

Figure 1. **TCP bandwidth:** total TCP bandwidth consumed by HTTP transfers for different content delivery systems. Each band is cumulative; this means that at noon on the first Wednesday, Akamai consumed approximately 10 Mbps, WWW consumed approximately 100 Mbps, P2P consumed approximately 200 Mbps, and non-HTTP TCP consumed approximately 300 Mbps, for a total of 610 Mbps.

FIGURE 8: MEASURED DAILY PATTERN OF P2P TRAFFIC (BY SAROIU ET AL. [7])

5. P2P CONTENT MANAGEMENT POLICIES AND TRAFFIC PATTERNS

Our primary objective in this study is to understand and to measure the impact of P2P traffic management policies in the metro network service environment. The primary interests in P2P networking is file sharing [15]. An interesting consequence from of this from the management point of view is that popular contents are replicated faster than less popular contents within P2P service, thus consuming bandwidth and collective storage spaces, which hold the replicated copies of the content on individual peers. The process can be seen as if the content is replicating and spreading itself using the network resources, as if it were a virus. Indeed, the mathematical formulation governing the process follows the same type of diffusion equation as virus spreads among population [16]. This phenomena, which creates congestion induced by fast self-multiplying copies of a single content, is also known as Flash Crowd, named after a science fiction published in 1973 [17].

Although negative side effects of Flash Crowd such as bandwidth crunching caused by sudden surge of P2P traffic have been often emphasized, we should also note that it is actually a self-management principle at work, where storage usage and bandwidth consumption scale up and down following the demand. Therefore, our approach toward P2P traffic is not to curb down the traffic but to put mild control over its behavior, through restrictive use of policies using policy control entry points such as super peers [18]. The policies evaluated in this paper deals with locality of P2P traffic in the Internet, as illustrated in Figure 9.

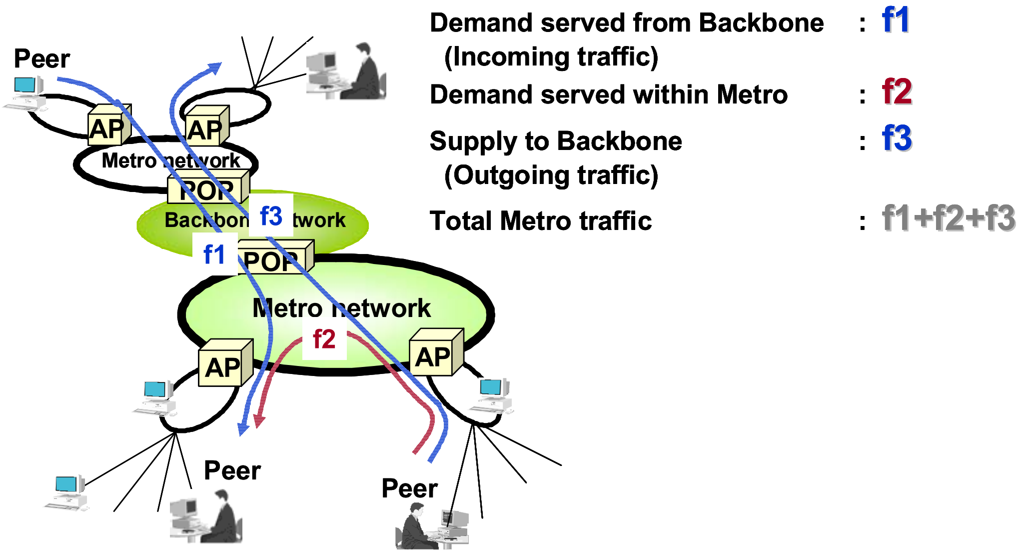

FIGURE 9: THREE TYPES OF TRAFFIC LOCALITY

We distinguished three different traffic types with regards to metro networking; (1) incoming from the Internet backbone (denoted as f1 in Figure 9), (2) served within the metro (denoted as f2), and (3) outgoing to the Internet backbone (denoted as f3). The total metro traffic is the sum of all the three, f1 + f2 + f3. Given that the total metro traffic is the same, a question arises what would be the optimal policy to distribute the traffic among the three classes of the traffic?

The above question bears practical implication when the ISP operating on the metro network has a peering relationship with a Tier 1 ISP operating on the backbone network. When the ISP has a payment condition with the Tier 1 ISP, whose payment is proportional to the traffic volume going to the backbone, the above question immediately translates into an optimal traffic management policy, which balances between short-term payment toward the Tier 1 ISP and added congestion in the metro network, which may eventually lead to additional investment for the ISP to maintain QoS in the metro network. Were all these economic questions set aside, it is still of technical interests to evaluate how effectively can certain traffic control policy at super peers actually route P2P traffic from one locality to another, given an ISP service environment.

For this experiment, we studied the following three policies. A super peer, upon receipt of content request from P2P peers, responds to the query as follows.

- *Policy 0:* (no peer selection policy) under this policy, list of content serving peers are in random order. As a result, the requesting peer selects another peer randomly out of the group. This policy is more likely to select a peer outside of the local peers within the metro network, since the peer population in the Internet far exceeds that within the metro network. In the current Gnutella and many P2P services, there is no specific peer selection policy. Typically, the ratio between metro traffic to the total traffic, f2/(f1 + f2 + f3), is less than 5%.
- *Policy 1:* (preferred ISP policy) under this policy, P2P traffic is preferred to be contained within a given ISP. When a super peer receives a content request from a P2P peer, the super peer always returns a list of content serving peers in the preferred ISP domain if such peers exist.
- *Policy 2:* (preferred metro area policy) under this policy, a particular metropolitan area is preferred when a super peer returns a list of content serving peer to the requesting P2P peer. This policy is similar to Policy 1, except that Layer 3 concept of ISP domain do not necessarily coincide with Layer 2 concept of metropolitan network; the super peer in Policy 2 must be provided with some level of geomapping capability.

Figures 10-12 show results of experiments. The three policies have been applied to the US P2P traffic simulator, which we discussed in Section 4. Assuming online 3,000,000 P2P peers on average and the same broadband deployment status known at August 2002, we evaluated P2P traffic pattern for the 20 US metropolitan areas. The top five US metro areas in terms of the traffic volume are Washington DC, New York, Atlanta, Boston, and San Francisco, in the order. Note that the metro traffic (f2) consistently on the rise across the three policies, from Policy to Policy 2. Note also that the total traffic, f1 + f2 + f3, slightly decreases across the three policies since unit increase in the metro traffic (f1) accounts for decrease of double the amount in the backbone (f2 + f3) traffic.

An interesting question on this result is why these five cities are the busiest in terms of P2P traffic, as they are not necessarily the most populous ones in the US. There seem to be several reasons for this outcome, which include factors such as broadband deployment, locations of major educational institutions, locations of large ISP PoPs, general awareness on Internet and P2P file sharing, etc.

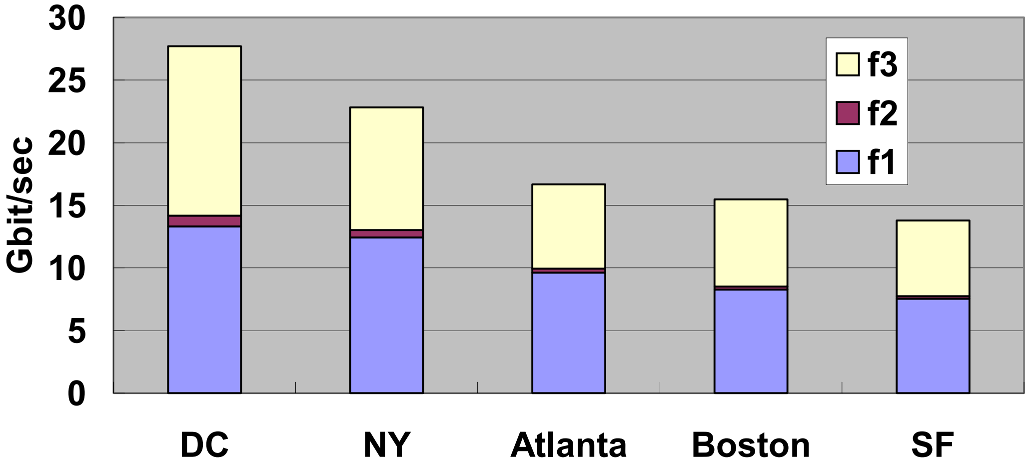

FIGURE 10: TOP 5 US METRO TRAFFIC DISTRIBUTION UNDER POLICY 0

Figure 10 shows the metro traffic distribution under Policy 0. The ratio of metro traffic to the total traffic is typically less than 5%. Figure 11 and 12 show the metro traffic distribution under Policy 1 and Policy 2, respectively. The ratio of metro traffic to the total traffic doubles from 20% under Policy 1 to 40% under Policy 2.

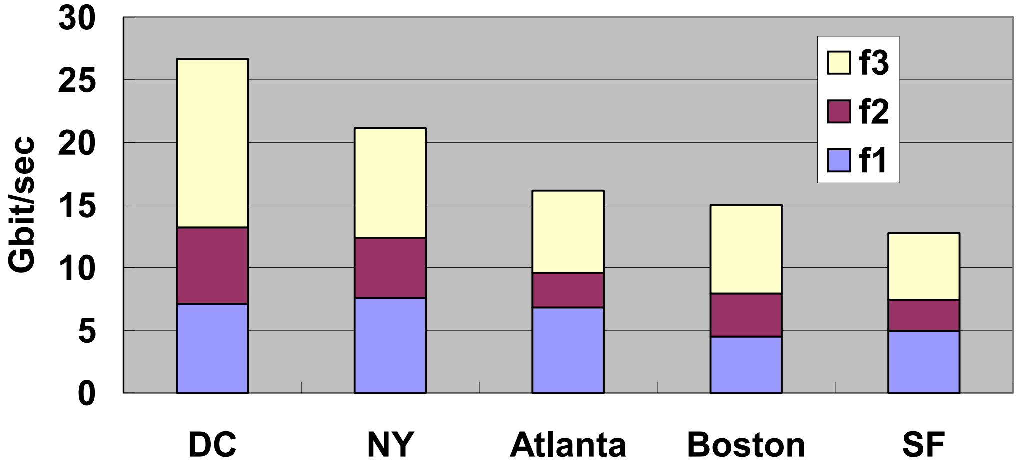

FIGURE 11: TOP 5 US METRO TRAFFIC DISTRIBUTION UNDER POLICY 1

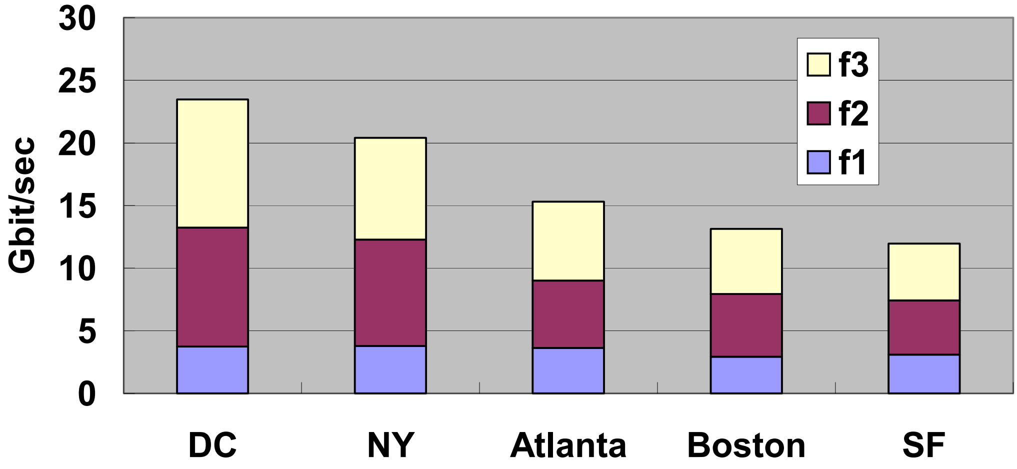

FIGURE 12: TOP 5 US METRO TRAFFIC DISTRIBUTION UNDER POLICY 2

6. P2P CONTENT MANAGEMENT POLICIES AND TRAFFIC PATTERNS

Through our simulations, we noted that level of broadband deployment has high impact on behavior pattern of P2P traffic. Simply put, broadband deployment encourages P2P activities, and P2P services further accelerate demand for broadband deployment. One interesting possibility is that this positive feedback cycle between the P2P service and the broadband deployment may provide further dramatic impact to the characteristics of traffic in the future metro networks. We conducted a preliminary experiment using the same hypothetical network we used in Section 4. Figure 13 shows traffic on the same links between AS_1 and AS_2-AS_7 as in Figure 7, except that access links in Figure 13 scenario is 100Mbps FFTH whereas it was ADSL in Figure 7 (uplink 384kbps, downlink 3Mbps).

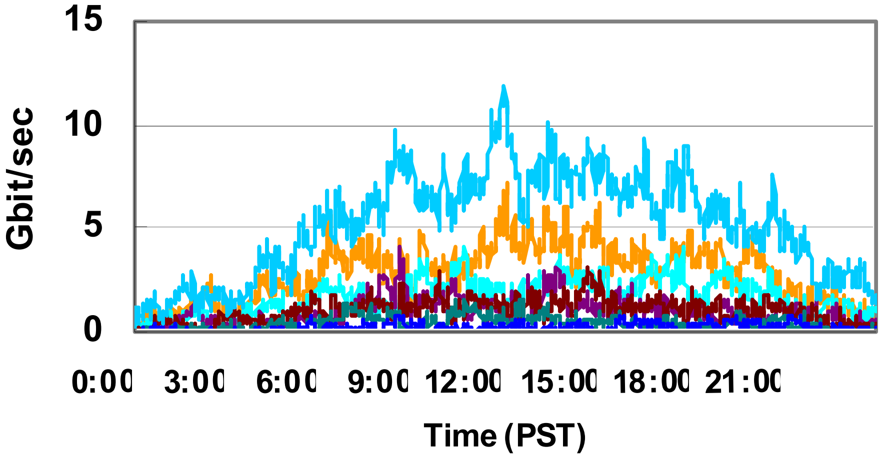

FIGURE 13: P2P TRAFFIC PATTERN BETWEEN AS_1 AND AS_2-AS_7 (BROADBAND)

Comparing Figure 7 and Figure 13, we can immediately notice that not only the peak traffic volume has multiplied by a factor larger than 3, but also the traffic pattern in Figure 13 is much jaggier, with multiple peaks incurred by succession of short bursts. We interpreted it that the P2P services with FTTH broadband deployment may require a new class of metro networking infrastructures than those exist today.

7. CONCLUSION

In this paper, we studied behavior of P2P traffic based on Gnutella. We also studied its traffic models, and have compiled them into a P2P traffic simulator using JavaSim, a Java-based network simulation environment. Based on sampled measurement on P2P traffic and available statistics on the Internet traffic, we judged that the simulator was capable of recreating P2P traffic with reasonable accuracy. We then proceeded to use the simulator to evaluate and to solve P2P content transfer management policy issues in the metro networking. Three different policies have been evaluated. We found that containment of P2P traffic within the metro networking is actually possible, and it can be done fairly effectively, containing as much as 40% of the P2P traffic within the local metro network. For the future works, we intend to use this simulator to solve other relevant traffic management issues of P2P services. We note that broadband deployment and its penetration into the metropolitan areas is the key factor for the rise of P2P services. We also conducted a simulation to evaluate impact of broadband deployment for the future metro networking. Our preliminary analysis using the P2P simulator suggests that the traffic will become much burstier. As a result, large and instantaneous fluctuations in traffic demand would require a new metro networking architecture provided with more dynamic traffic management and bandwidth provisioning schemes than those available today.

REFERENCES

[1] P-cube white paper, “Approaches to Controlling Peer-to-Peer Traffic: A Technical Analysis”, http://www.p-cube.com.

[2] Alexandre Gerber, Joseph Houle, Han Nguyen, Matthew Roughan, and Subhabrata Sen, "P2P The Gorilla in the Cable", National Cable & Telecommunications Association (NCTA) 2003 National Show, Chicago, IL, June 8-11, 2003.

[3] Subhabrata Sen and Jia Wang, "Analyzing peer-to-peer traffic across large networks", in Proc. of ACM SIGCOMM Internet Measurement Workshop, Marseilles, France, Nov. 2002.

[4] Matei Ripeanu and Ian Foster, "Mapping Gnutella Network", 1st International Workshop on Peer-to-Peer Systems (IPTPS’02), Cambridge, Massachusetts, March 2002.

[5] Matei Ripeanu, Ian Foster, and Adriana Iamnitchi, “Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design”, IEEE Internet Computing Journal, vol. 6(1) 2002.

[6] Nathaniel Leibowitz, Matei Ripeanu, and Adam Wierzbicki, "Deconstructing the Kazaa Network", 3rd IEEE Workshop on Internet Applications (WIAPP'03), June 23-24, 2003, San Jose, CA.

[7] S. Saroiu, K. P. Gummadi, R. J. Dunn, S. D. Gribble, and H. M. Levy, "An Analysis of Internet Content Delivery System", in Proc. of the 5th Symposium on Operating Systems Design and Implementation (OSDI), Boston, MA, December 2002.

[8] Dhamija, R., Fisher, D., and Yee, K.-P., "Gnutellavision: Real-Time Visualization of a Peer-to-Peer Network", 2000, http://www.sims.berkeley.edu/~rachna/gtv/.

[9] Slyck home page, statistics on P2P networks, http://www.slyck.com.

[10] Kunwadee Sripanidkulchai, "The Popularity of Gnutella Queries and Its Implications on Scalability," Feb. 2001, http://www-2.cs.cmu.edu/~kunwadee/research/p2p/gnutella.html.

[11] M. Jovanovic, F.S. Annexstein, and K.A. Berman, "Scalability Issues in Large Peer-to-Peer Networks –A Case Study of Gnutella", Technical Report, University of Cincinnati, 2001.

[12] Quova Inc. home page, http://www.quovq.com.

[13] Javasim home page, http://www.javasim.org.

[14] Shing Yin and Muayyad Al-Chalabi, "North America Internet Traffic: Past, Present, Future," RHK report, Aug. 2002.

[15] Kavitha Ranganathan , Matei Ripeanu, Ankur Sarin, and Ian Foster, "To Share or Not to Share: An Analysis of Incentives to Contribute in Collaborative File Sharing Environments", Workshop on Economics of P2P Systems, June 5-6, 2003, Berkeley, CA.

[16] Takeo Hamada, Kaoru Chujo, and Xiangying Yang, "Spatial Diversion Tree-based Restoration and Its Application to P2P Traffic Control", Fujitsu Labs. of America Technical Memo, FLA-INR-TM-03-17, Mar. 2003, unpublished.

[17] Wikipedia entry on Flash Crowd, http://www.wikipedia.org/wiki/Flash_Crowd.

[18] Beverly Yang, Hector Garcia-Molina,"Designing a Super-peer Network", in Proc. Of the 19th Int. Conf. On Data Eng. (ICDE), March 2003, http://www-db.stanford.edu/~byang/pubs/superpeer.pdf.