BGP-based Locality Promotion for P2P Applications

Peter Racz*, Simon Oechsner†, Frank Lehrieder†
*Department of Informatics (IFI), University of Zurich, Switzerland
†Institute of Computer Science, University of Würzburg, Germany
Email: racz@ifi.uzh.ch, [oechsner|lehrieder]@informatik.uni-wuerzburg.de

***Abstract*—P2P applications attract a lot of users and generate the dominant portion of the overall traffic in the Internet today. On the one hand, this large amount of traffic results in high operational costs for ISPs, mainly because of expensive inter-domain connections. On the other hand, the performance of P2P applications is constricted by suboptimal peer selection or by bandwidth limitations of ISPs. To overcome these problems, the collaboration of P2P applications and ISPs is desirable, where locality promotion is one of the possible approaches. In this paper, we propose a locality promotion mechanism based on BGP routing information of an ISP and show by simulations that it can reduce inter-domain traffic, prefers shorter connections and peering links over transit links while P2P applications can achieve a better performance as well.**

I. Introduction

Network traffic generated by Peer-to-Peer (P2P) applications, like file sharing and video streaming, constitutes a large portion of the overall Internet traffic [1]. Since P2P applications usually do not know the underlying network topology, peers select the source of their downloads based on overlay metrics or even randomly. Due to this fact, P2P traffic uses more network resources than necessary and causes high load and congestion on particular network links. Moreover, traditional traffic engineering approaches are very difficult to apply and the large amount of traffic generates high costs for Internet Service Providers (ISP), especially because of the expensive interconnection links between ISPs [2]. To reduce inter-domain traffic, some ISPs filter P2P traffic and throttle the bandwidth for P2P applications [3]. This bandwidth limitation together with the suboptimal peer selection of P2P applications can result in decreased application performance.

To overcome these problems, the collaboration of P2P applications and ISPs would be beneficial for both parties, where P2P applications could achieve a better performance and ISPs would be able to manage P2P traffic and reduce operational costs. A possible solution for collaboration is that ISPs provide some network information (e.g., topology) and P2P applications adapt their behavior based on this information. There have been several approaches proposed for such a collaboration [3], [4], [5], [6] and the IETF has also recently established a working group on application layer traffic optimization (ALTO) [7]. A prominent approach is locality promotion, which aims at localizing P2P traffic in the network of an ISP, thus reducing costly inter-domain traffic and increasing application performance.

This paper addresses the problem of locality promotion and proposes 1) an algorithm to calculate locality ratings of peers based on Border Gateway Protocol (BGP) routing information and 2) shows how to apply this rating in a BitTorrent-based file sharing application by using Biased Neighbor Selection (BNS), Biased Unchoking (BU), or their combination. This work has been developed in the SmoothIT (Simple Economic Management Approaches of Overlay Traffic in Heterogeneous Internet Topologies) project [8] that investigates Economic Traffic Management (ETM) mechanisms, leading to a triple-win situation, where ISPs, overlay providers, and users all benefit from a collaboration. ETM mechanisms intend to provide appropriate incentives to all players to achieve the management of overlay traffic. This paper focuses on BGP-based locality promotion – one of the mechanisms beside others [9] developed by SmoothIT –, where locality is expressed based on BGP attributes. BGP is the routing protocol between Autonomous Systems (AS) and reflects a.o. ISP preferences and AS hop counts in its route selection. Based on such attributes the BGP-based locality promotion calculates a peer rating that is suitable to rate peers outside the AS of an ISP. Peers can retrieve ratings from their ISPs via the SmoothIT Information Service (SIS) and adapt their peer selection accordingly (i.e. BNS, BU).

The remainder of the paper is structured as follows. Section II gives an overview on existing approaches to promote locality in P2P overlays. Section III introduces briefly the SmoothIT architecture, while the BGP-based locality promotion mechanism is presented in Section IV. BitTorrent and the use of locality information by peers are described in Section V. Section VI contains the simulation setup and the performance evaluation results. Finally, Section VII concludes this paper.

II. Related Work

Several approaches have been recently proposed which try to facilitate the collaboration between P2P applications and ISPs. A set of approaches propose to deploy caches in the network [10] that store popular content and serve only peers in the local network, which results in less traffic on inter-domain links. However, caches have to support multiple different P2P protocols in order to be able to serve users using different applications. Furthermore, legal issues may arise when copyright protected content is cached by an ISP.

Other approaches foresee some sort of collaboration between overlay application and ISP, where locality promotion

978-1-4244-7116-4/10/$26.00 ©2010 IEEE

is a dominant approach. It aims at better utilizing the local network resources of an ISP, while reducing the amount of traffic and associated costs on interconnection links between ISPs. The P4P project proposes an ISP-controlled traffic localization, where an information server, called iTracker, provides locality information to the overlay [6]. The iTracker may communicate directly with peers or application trackers such as the BitTorrent tracker. A similar approach, called Oracle service, is proposed in [5]. Peers query the Oracle service to find local peers from a list of potential neighbors. Since the Oracle service knows about network information, it can respond with an optimized choice. Another approach, called Ono [11], does not rely on a service provided by an ISP as the previous approaches do, but it exploits Content Distribution Network (CDN) servers, like Akamai servers, as landmarks to gain proximity information. It performs periodical DNS lookups of CDN servers in order to build and maintain proximity maps. The ALTO IETF Working Group [7] is also addressing the P2P traffic management problem [2] and defines the ALTO Service that provides network related information to overlays. Currently, the ALTO protocol is under development that shall be used to exchange information between overlay and network. The BGP-based locality promotion proposed in this paper is a possible ALTO service.

To use locality information in BitTorrent, applications can apply the concept of BNS [4] and BU [12]. In BNS the neighbor set of peers is adjusted so that it contains a larger portion of local peers. BU adjusts the optimistic unchoking process so that peers prefer local neighbors to which they start uploading instead of randomly selecting a neighbor.

III. Architecture

The BGP-based locality promotion mechanism is part of an overall architecture developed in the SmoothIT project [8]. In the architecture peers use the SmoothIT Information Service (SIS) to interact with the ISP and receive topology information [9] (cf. Fig. 1). In case of the BGP-based locality promotion, peers receive SIS ratings from the SIS server, where the SIS rating is calculated based on BGP routing information and it reflects the preference of a candidate peer. A peer will prefer to contact peers with higher SIS rating by using BNS, BU, or their combination. ISPs deploy the SIS server in their own domain and it serves only peers that are connected to the network of the ISP who runs the SIS server.

The components of the SmoothIT architecture can be mapped to components developed by the ALTO Working Group [7]. In this context, the SIS server corresponds to the *ALTO server* and the BGP-based locality promotion service is a possible *ALTO service* [13]. The routers in an ISP's network are the basic *source of the topology information* aggregated in the SIS of the ISP. The locality information can in general be utilized by peers participating in an overlay network, acting as both *resource providers* and *consumers*. The specific example evaluated in this work is a BitTorrent overlay which also features a tracker, or a *resource directory* in ALTO terms. Depending on the client solution, the tracker and/or the peers can contact the SIS server directly. We will only consider

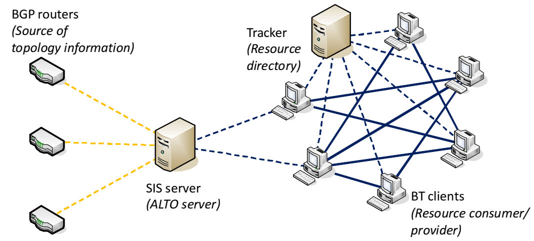

Fig. 1: Basic SmoothIT architecture.

solutions where P2P clients contact the SIS server, since in this case the users can decide whether they want to support locality or not. A query to the SIS server consists of a list of peer addresses to be rated in relation to the querying peer. A response consists of the same list of peers, with the SIS rating value attached to each entry in the list. This rating value is computed from BGP routing entries as described in Section IV. The different ways the clients can use this rating value are described in Section V.

IV. Peer Rating based on BGP Routing Information

The BGP-based locality promotion service assists P2P applications by providing the SIS rating value for each potential overlay neighbor. The peer rating value is calculated based on the BGP attributes associated with the route from the local peer to each potential neighbor peer. The idea here is to re-use topology information and ISP preferences existing in routers to support locality-awareness.

The BGP routing protocol [14] is used to exchange routing information between ISPs and it determines routes for inter-domain routing between ASes. BGP can consider ISP policies in the routing decision, so that ISPs can influence the route selection according to their preferences, e.g., costs or connection types. Because of this, BGP routing information can be used to assist peer selection in P2P applications in order to reduce inter-domain traffic and to achieve a more efficient inter-domain traffic pattern between ISPs. This is an incentive for ISPs to use their BGP routing information to provide a service for P2P applications. P2P applications in turn will follow the implicit peer recommendations of the ISP if they can improve their performance this way. We will show scenarios where this is the case in Section VI. The BGP-based locality promotion is only effective for P2P traffic that crosses ISP domains and the traffic management of P2P traffic within the domain of a single ISP is not considered here.

In the following, we first review relevant BGP attributes and the BGP route selection process based on [14], then we present the BGP-based peer rating calculation algorithm.

A. Relevant BGP Attributes

The peer rating calculation algorithm uses the Local Preference, the AS Path, and the MED BGP attributes.

1) Local Preference: It is used to prefer a route to reach a certain destination if there are multiple alternative exit routes from the local AS to the given destination. Routers select the route with the highest local preference value from all alternative routes (cf. Section IV-B).

In inter-domain routing ISPs prefer to select next hops in a way that considers the business relations and agreements established with their neighbors. In general the following relations can be distinguished: customer, provider, peering, and backup [15]. The local preference attribute can be used to prefer a route over another one to the same destination. ISPs often use the local preference value to implement their policies that reflect their business relations with their neighbors [15].

2) AS Path: It contains a sequence of AS path segments. An AS path segment is either an AS_SEQUENCE or an AS_SET. An AS_SEQUENCE contains an ordered list of AS numbers that a route has traversed, while an AS_SET is an unordered set of AS numbers. AS_SETs are used in case of route aggregation and they reduce the size of the AS path information. However, by using AS_SETs the information about the exact AS numbers a route traversed is also lost.

From the AS path attribute it is possible to calculate the AS path length that defines how many ASes a path traverses and it approximates relative distance. The AS path length is calculated by counting the number of AS numbers in AS_SEQUENCEs and AS_SETs, where an AS_SET counts as 1 independent of how many ASs are in the set. Routers select the route with shorter AS path from all alternative routes (cf. Section IV-B).

3) MED: By using the multi-exit discriminator (MED) attribute, an AS can advertise to external ASs its preference for incoming routes into its own AS. However, an external AS might select another route and it does not have to follow the suggestion. Routers select the route with the lowest MED value from all alternative routes (cf. Section IV-B).

B. BGP Route Selection

BGP can receive multiple alternative routes to the same destination from different BGP speakers. BGP will select only one route as the best path from all alternatives. Considering only the local preference, AS path, and MED BGP attributes, the BGP route selection takes the following steps [14]:

- Prefer the route with the largest local preference value.
- If the alternative routes have the same local preference value, prefer the route with the shortest AS path.
- If the alternative routes have the same AS path length, prefer the route with the lowest MED value.

The peer rating calculation algorithm works similar to BGP and it preserves the relative importance of the attributes mentioned above. However, while BGP compares alternative routes to the same destination, the peer rating algorithm compares alternatives to different destinations (i.e. to candidate peers).

C. Peer Rating Calculation

The peer rating algorithm assigns SIS ratings to all IP address ranges. A peer represented by its address will receive the SIS rating that is assigned to its address. The algorithm differentiates between local and remote address ranges. Local address ranges are in the AS of the ISP who operates the SIS server; remote ones are outside. The peer rating algorithm assigns the highest SIS rating to local address ranges. This ensures that first local peers will be selected by P2P applications if they follow the recommendations reflected by the peer rating values. Thus, if a certain content is available on peers in the local AS, a peer will prefer to download it from such peers. Regarding remote address ranges, the peer rating algorithm assigns SIS rating values according to the preference of the BGP routing protocol. This ensures that peers outside of the AS will be selected by P2P applications according to the ISP's routing preference, e.g., a peer will prefer to download from peers that are located in a peering AS over those located in a provider AS.

```
for each IP range that is in the network of the ISP {
  /* Assign the highest peer rating value */
  if(MED_flag == true) {
    peer_rating = (MAXPREF+1)*(MAXAS+1)*(MAXMED+1)
  } else {
    peer_rating = (MAXPREF+1)*(MAXAS+1)
  }
}
for each BGP routing entry that is selected as a
best path by BGP {
  /* Assign the peer rating value based on the BGP
     attributes */
  if(MED_flag == true) {
    peer_rating = local_pref*(MAXAS+1)*(MAXMED+1) +
                 (MAXAS-as_hops)*(MAXMED+1) +
                  MAXMED-med
  } else {
    peer_rating = local_pref*(MAXAS+1) +
                  MAXAS-as_hops
  }
}
```

Fig. 2: Peer rating calculation algorithm

Following the logic of the BGP route selection described above, the peer rating calculation algorithm prefers first peers that are reachable via routes with the largest local preference, then peers via routes with the shortest AS path length, and finally peers via routes with the lowest MED value.

The peer rating calculation algorithm is shown by a pseudo code in Fig. 2. In the first for-loop the algorithm assigns the highest peer rating to each local address range. In the second for-loop the algorithm takes only the BGP routing entries that have been selected as the best path by the BGP protocol and assigns to the corresponding IP ranges a peer rating based on their BGP attributes (i.e., local_pref, as_hops, med). The MED attribute is only taken into account if the MED flag is set (see details below). The algorithm uses the following three parameters that are calculated as the maximum values of the respective BGP attributes:

- MAXPREF: It is equal to or greater than the largest local preference value used by BGP speakers in the network of the ISP.
- MAXAS: It is equal to or greater than the largest AS path length used by BGP speakers in the network of the ISP.
- MAXMED: It is equal to or greater than the largest MED value used by BGP speakers in the network of the ISP.

Similar to the BGP route selection algorithm, the local preference attribute is considered the most important value in the peer rating calculation. The algorithm assumes that ISPs assign a non-overlapping range of local preference values to each type of business relation, which allows the comparison of local preference values assigned to routes to different destinations. Such local preference assignment is often the case [15]; for example, local preference values in the range 90-99 might be used for customers, 80-89 for peers, 70-79 for providers, and 60-69 for backup links [15]. If an ISP does not assign local preferences in this way, an additional mapping shall be applied, which maps IP ranges to local preference values.

The AS path attribute represents the distance to a destination. The peer rating calculation prefers peers with shorter AS path length, again similar to BGP. This is because if P2P applications establish connections to closer peers, the application will most probably reach a better performance and the overall traffic will be reduced, since the traffic traverses less ASes.

The MED attribute represents a suggestion of route preference from neighboring ISPs. The peer rating calculation tries to follow this suggestion, if the routes to two different peers have the same local preference and AS path length. Since MED values have a local scope with respect to an inter-connection between two ISPs and MED values from different neighboring ISPs cannot be compared in general, the peer rating calculation takes MED values into account only if neighboring ISPs set the same range of MED values in a common way. Otherwise a mapping of MED values could be applied. The use of MED in the peer rating calculation is optional, which is reflected by the MED flag in the algorithm.

BGP selects the best route from alternative routes to the same destination, while the peer rating calculation algorithm compares routes to different destinations based on their BGP attributes. Therefore, as discussed above, the following prerequisites need to be fulfilled to be able to use the algorithm: 1) Non-overlapping ranges of local preference values are assigned to routes according to the business relations an ISP has with its neighbors. Otherwise a mapping needs to be applied; 2) The MED value is only considered in the algorithm if all neighbors of the ISP use the same range of values in a common way. Otherwise a mapping is applied or the MED value is not taken into account in the SIS rating calculation; 3) The Cisco-defined weight BGP attribute is either not used or is set to the same value for all routes. Otherwise, the algorithm has to be extended to take the weight attribute into account.

The SIS server uses the peer rating algorithm to calculate SIS ratings and it assigns a SIS rating to each IP address range. At the end of the algorithm the server has a list of IP ranges with the SIS rating assigned to each of them. For the rating calculation the server has to know the local IP ranges and has to have access to the BGP routing table of the ISP. The local IP ranges are either configured on the SIS server or the server retrieves them from routers running intra-domain routing protocols, e.g., OSPF. To access the BGP routing table, the SIS server reads it from a BGP speaker in the local AS using standard interfaces. The SIS server can read the routing table either over SNMP or BGP and the routers do not need any modification. The SIS rating calculation shall be repeated periodically, e.g., once a day, in order to reflect possible changes in local IP ranges or in the BGP routing.

During the operation of the system, peers that want to support locality, query the SIS server of their ISP with a list of IP addresses (or ranges) for SIS ratings. These IP addresses are the addresses of candidate peers that the querying peer wants to connect to. The SIS server looks up the SIS rating entry corresponding to each address in the list according to longest prefix match. Then the server returns the list of IP addresses with their SIS ratings to the peer that uses this rating information as explained in the next section.

V. Modifications of P2P Clients

In this section, we describe possible modifications of the P2P client software which permit to include the SIS ratings in their peer selection process. We start with a brief description of BitTorrent. Then, we give an overview of the two modifications we use for our performance evaluation in Sect. VI, i.e., Biased Neighbor Selection and Biased Unchoking.

A. BitTorrent

The BitTorrent protocol forms a mesh-based overlay and utilizes multi-source download to distribute content. For each shared file, one overlay is formed, a so-called *swarm*. To facilitate the multi-source download, a shared file is split into smaller pieces, called *chunks*, which are in turn again separated into sub-pieces or *blocks* [16], [17]. In the following, we will focus on the description of the relevant mechanisms of BitTorrent that are utilized for locality promotion.

Each peer has only a limited number of other peers it has direct contact with in the swarm. These neighbors know about each other's download progress, i.e., which chunks the other has already downloaded. This enables a peer A to signal its interest in downloading chunks to a neighbor B holding chunks that the local peer still misses. We say that peer A is *interested* in peer B.

A peer joining a swarm typically initializes its neighbor set by contacting a *tracker*, i.e., an index server with global information about the peer population of a swarm. A standard tracker responds with a random subset of all peers in the swarm. Once a peer A has received a list of contacts in the swarm, it tries to establish connections to them. If it is successful, the according remote peer B is added to A's neighbor set and vice versa.

Every 10 seconds, a peer decides to which of its interested neighbors it will upload data to. These peers are called *unchoked*, the rest is *choked*. In standard BitTorrent, there are 3 regular unchoke slots which are awarded to the peers that offer the currently highest upload rate to the local peer. This strategy is called *tit-for-tat* and provides an incentive for peers to contribute upload bandwidth to the swarm. If the local peer has already downloaded the complete file, i.e., it is a *seeder*, the slots are given to all interested neighbors in a round-robin fashion. Additionally, every 30 seconds a random peer not currently unchoked is selected for *optimistic unchoking* for the next 30 seconds. This allows a peer to discover new mutually beneficial data exchange connections.

B. Biased Neighbor Selection

The specific implementation of BNS used in our experiments leaves the responsibility of choosing close neighbors with the peers and not with the tracker. It is therefore a form of peer-based BNS. Other implementations of BNS have been proposed in literature [4], [5], [6]. In our implementation, the clients request a much larger number of contacts than usual from the tracker (1000 instead of 50), either by repeating the query several times or by specifying a larger number of contacts to be returned. This is possible with most of the common tracker implementations. After this sufficiently large set of candidate peers has been received by the local peer, all of them are rated by the SIS server. Then, the local peer tries to establish connections to the closest ones first until it has $l \cdot N_{min}$ neighbors, where $0 < l \leq 1$ is the targeted share of close neighbors and N_{min} is the minimum number of neighbors a peer tries to have in any case (typically 40). After that, it adds $(1-l) \cdot N_{min}$ neighbors randomly from the remainder of the candidates. In our evaluation, we set $l = 0.9$ so that 90% of the neighborset consists of the peers with the best ratings. Similar values for l have been used in [4] and [3]. However, these studies differentiate only between local and remote peers while our implementation is able to cope with a continuous range of peer ratings containing more information than whether another peer is in the same AS or not.

C. Biased Unchoking

The mechanism of BU evaluated here is an expansion of the unchoking implementation used in [12]. It influences the choice of peers for optimistic unchoking, so that closer peers are preferred instead of randomly picking an interested neighbor. Specifically, it unchokes one of the interested neighbors with the highest rating value that are not yet unchoked. Within this set, the choice is random. The rating value is retrieved from the SIS server.

VI. Performance Evaluation

In this section we use simulations to assess the impact of BGP-based locality promotion on the inter-domain traffic and the application performance. We follow the methodology and use the same scenarios as we did in [12].

However, in [12], we used a metric that only discerned between neighbors being in the same AS and neighbors outside the local peer's AS. This allowed for an abstract modeling of the underlay topology. In contrast to [12], we use a more complex network topology here to model the comparably wider range of rating values returned by the BGP-based locality promotion.

A. Simulator

The simulation experiments have been performed using the ProtoPeer framework [18] which is intended for P2P simulations and prototyping. This framework already has different network models from which we chose the flow-based one for our study. The flow-based network model simulates the property of TCP that the bandwidth of a physical network link is shared among all data connections using that particular link. Furthermore, a constant startup delay of 10 ms is added to every connection to simulate the TCP handshake.

We extended ProtoPeer with an implementation of the BitTorrent protocol according to the specifications in [16]

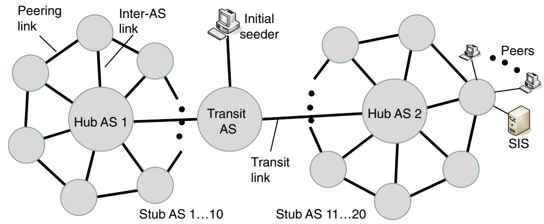

Fig. 3: The simulated topology.

and [17]. It comprises all key mechanisms of BitTorrent: the rarest-first piece-selection, the tit-for-tat peer selection (choke algorithm), and the neighbor set management. The latter two mechanisms play a major role in this study because we modify them so that they can make use of the information provided by the SIS server. The code was also used for the study in [12].

B. Simulation Model

The underlay network in the simulation has a multi-AS topology depicted in Fig. 3. It consists of 20 Tier 3 stub-ASes (if not mentioned otherwise) where the peers and the SIS servers are located. Each stub-AS is directly connected to two other stub-ASes, depicting peering relationships. Furthermore, each stub-AS is connected to a Tier 2 hub-AS. There are two hub-ASes connected to half of the stub-ASes each. The two hub-ASes in turn are interconnected via a Tier 1 transit-AS. We denote the direct links between stub-ASes as peering links, the links between stub-ASes and hub-ASes as inter-AS links and the links between hub-ASes and the transit-AS as transit links. We use the same notation for the traffic flowing over these links. The access links of peers have a speed of 16 Mbit/s downlink and 1 Mbit/s uplink (a typical ADSL connection). The initial seed has a symmetric connection with 10 Mbit/s both in downlink and uplink. The access links of peers and the inter-AS links can constitute network bottlenecks. If not stated otherwise, we model the peering, inter-AS and transit links as well dimensioned. We keep the network topology as simple as possible in order to understand the basic behavior of the mechanisms. Still, the topology is complex enough to allow to observe the effects of different types of AS relations (peering, customer-to-provider, and transit) and of longer AS level routes on the locality promotion.

The overlay contains one BitTorrent swarm with a file of size 154.6 MB generated from an example TV show. Chunks are 512 KB and blocks are 16 KB large. At the beginning of each run, the swarm contains only the initial seed with the complete file. The initial seed and the tracker are placed in the transit-AS. Peers are only connected to stub-ASes. When new peers arrive, they join one of the stub-ASes randomly, following an uniform distribution. New peers join the swarm with an exponentially distributed inter-arrival time with a mean value of 10 s. Peers stay online until they downloaded the entire file plus an additional, exponentially distributed seeding time with a mean value of 10 minutes. The mean number of concurrently online peers is between 120 and 200 depending on the scenario (peers have no offline times). Each run simulates the swarm for 6.5 hours and consequently contains about 2300 downloads. Since the initial warm-up phase took about 1.5 hours in all runs, we discard this phase for the evaluation.

The chosen parameters for our simulation study are very similar to the ones used in [12]. They are motivated by a measurement campaign of BitTorrent swarms where we found that the fraction of a swarm that is located in a given AS is below 10% in almost all cases. Furthermore, the number of peers in the wide majority of swarms was below 200. The only exceptions are swarms distributing popular movie files which are considerably larger. Details can be found in [12].

C. Evaluation of the BGP-based Locality Promotion

We compare the performance of regular BitTorrent with several variants implementing BGP-based locality promotion. In all experiments, we compare 4 different peer behaviors: (1) regular BitTorrent (regBT), (2) BitTorrent with Biased Unchoking (BU), (3) BitTorrent with Biased Neighbor Selection (BNS), and (4) BitTorrent with both BNS and BU (BNS&BU).

For the rating calculation algorithm (cf. Fig. 2) we used the following parameters. We set MED_flag = false and MAXAS = MAXPREF = 100. The value for local_pref = 80 by default and local_pref = 90 for peers in peering ASes. The parameters for BNS are set as described in Section V-B. The implementation of BU as described in Section V-C does not require parameter settings.

In order to assess the performance from an ISP's perspective, we consider the amount of traffic on the different link types. This traffic was measured in intervals of one minute during the whole simulation and then averaged over one simulation run. If the source and the destination of a data transfer is in the same AS, the traffic is considered as intra-AS traffic. Otherwise, it contributes to the total traffic on peering links, inter-AS links or transit links. For these types of traffic, we sum up the data flowing over all links of one type, i.e., a connection spanning four links generates four times the traffic in the statistic as a connection over one link. To judge the overlay performance from the user's point of view, we consider the download times of the peer. This is the time when a peer issues its first block request until it has completely received the file. Here, we average the download times of all peers in one simulation run. For each parameter setting we run 10 simulations and show average values over all runs for all observed variables. The confidence intervals for a confidence level of 95% are calculated and shown for all results.

To evaluate the performance of the BGP-based locality, we use four different scenarios as described in the following.

1) Experiment "Load": In this experiment, we evaluate the influence of load conditions on the effectiveness of BGP-based locality promotion. Load here means the fraction of leechers in the swarm. To generate different load scenarios, we vary the mean seeding time of the peers from 5 to 30 minutes.

Fig. 4 shows the mean value of the total utilized bandwidth for the different mechanisms and load scenarios. The scenario with 5 minutes mean seeding time is the one with the highest load. To judge the distribution of the total traffic among the different link types, the share of the total traffic is shown in different colors for each traffic type (inter-AS, transit, peering

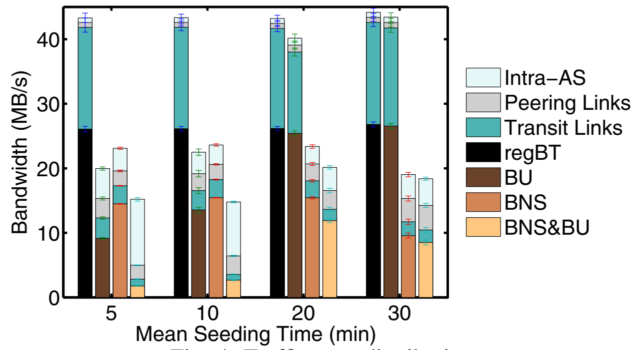

Fig. 4: Traffic sum distribution.

and intra-AS, from bottom to top). Thus, the complete bar is the total average bandwidth utilized. The color of the inter-AS traffic share denotes the type of BitTorrent variant used in the experiment.

We can observe that the total traffic is reduced by as much as 60% by the BGP-based locality if BNS and BU are used in conjunction at the client side. This traffic reduction is due to the fact that data takes shorter routes when the peers are locality-aware, thus consuming bandwidth on less links than in the regBT case. The larger share of intra-AS and peering traffic leads to a lower traffic demand in the topology as a whole. This change in the traffic distribution is again most prominent for the BNS&BU case, where the clients utilize best the locality information provided by the BGP-based algorithm.

The results also show that the locality-awareness has a higher effect on the traffic distribution when the load in the swarm is higher. While the BNS and BNS&BU variants still save a large amount of traffic for average seeding times of 20 and 30 minutes, the share of intra-AS traffic drops in these scenarios. This effect was already explained in [12], where it was shown that the lower number of local interested neighbors in swarms with low load strongly reduce the effectiveness of locality promotion. Especially Biased Unchoking suffers from this, as can also be concluded from the near-similar behavior of regBT and BU in the scenario with a mean seeding time of 30 minutes.

Finally, we observe no large impact of the evaluated mechanisms on the mean download times of the file. These are 14.6, 9.9, 2.6, and 1.7 minutes in the scenarios with 5, 10, 20, and 30 minutes mean seeding time, respectively. They do not differ significantly (below 10s) among the investigated mechanisms. Therefore, we argue that a user will not see a big difference in the performance of the application, while the gains for an ISP are potentially large.

2) Experiment "Swarm Distribution": Here, the number of peers in each AS is varied to evaluate the impact of the swarm distribution on the locality-promotion mechanism. We do this by simulating a topology with 10, 20 and 40 stub-ASes, which leads to 10%, 5% and 2.5% of the swarm per AS on average, due to the uniform distribution of newly arriving peers among the ASes. A lower number of peers in one AS means less opportunity for locality-promotion.

Taking a look at the resulting traffic distribution again, we can see this effect in the total traffic consumption as well as in the share of intra-AS traffic for all mechanisms, cf. Fig. 5. The

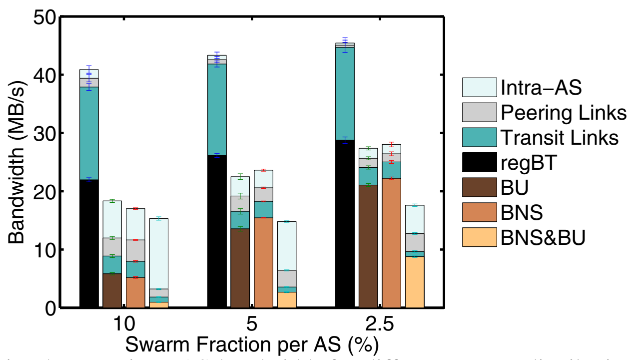

Fig. 5: Mean inter-AS bandwidth for different swarm distributions.

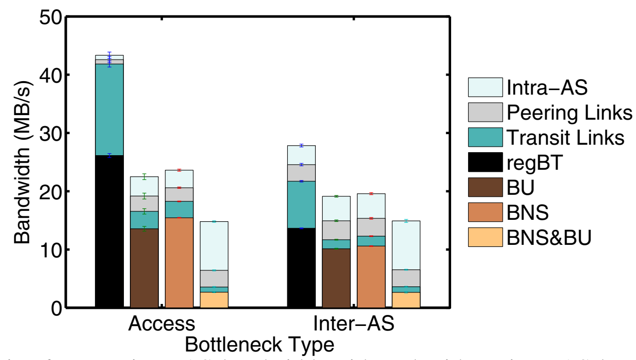

Fig. 6: Mean inter-AS bandwidth with and without inter-AS bottlenecks.

amount of intra-AS and peering traffic is much larger in the scenario with 10% of the peers per AS on average. BGP-based locality promotion still reduces the overall traffic significantly and especially saves costs by reducing the share of inter-AS and transit traffic if BNS and BU is used by the clients.

However, since the number of peers in the same AS and in peering ASes is reduced, there are simply less overlay neighbors with a good locality rating to be preferred by the clients. Therefore, the potential for keeping traffic local or between peering ASes is much lower. BGP-based locality promotion still manages to prefer shorter connections over longer ones, which can also be observed in the reduced amount of transit traffic in comparison to regBT.

Since we again have no bottleneck in the network, the location of neighbors does not have an effect on the utilized download bandwidth per peer. As a consequence, the download times are not affected by the number of stub-ASes nor by the different mechanisms. For all configurations, the mean download times are slightly below 10 minutes.

3) Experiment "Inter-AS Bottlenecks": Up until now, the BGP-based locality promotion had no positive effect for the end user, since the underlay connection length had no effect on the available bandwidth of an overlay connection. However, in reality, longer connections may experience a lower throughput. Additionally, providers may throttle the bandwidth of P2P connections leaving their network [11]. Therefore, we introduce bottlenecks in the inter-AS links only (labeled 'Inter-AS' in Fig. 6 and 7) by limiting the speed to 3072 kbit/s, i.e., three times the upload capacity of one peer. We compare the results for this scenario with the corresponding results for a topology with only the access links as bottlenecks (labeled 'Access').

Since we reduced the bandwidth available to some connections in the network, the total amount of traffic is reduced for all mechanisms in the scenario with inter-AS bottlenecks (cf. Fig. 6). Also, the share of intra-AS and peering traffic is enlarged, even in the regBT case. This can be attributed to the tit-for-tat policy of BitTorrent, which prefers connections with a higher throughput, in this case all connections that do not utilize an inter-AS link. Still, the BGP-based locality promotion manages to additionally lower the total traffic and increase the share of intra-AS and peering traffic in comparison to regBT in case of inter-AS bottlenecks. Also, the efficiency of BNS and BU is increased in comparison to the access link bottleneck case, since the effect of these algorithms is reinforced by the tit-for-tat behavior of BitTorrent. Only in the case of BNS&BU the potential for traffic savings seems to be already fully exploited, so that there is no additional decrease of traffic in the inter-AS bottleneck case.

In contrast to the previous scenarios, the download times in the inter-AS bottleneck scenario now heavily depend on the used overlay algorithms (cf. Fig. 7). This is due to the fact that the throughput and therefore the download speed between two overlay neighbors depends on their position in the topology. The locality promotion mechanisms prefer peers that are not reached via the bandwidth-limited inter-AS links, i.e., peers in the same AS or in peering ASes. Thus, they also lead to shorter download times in comparison to regBT in the inter-AS bottleneck case. Especially the utilization of the BGP locality information with both BNS and BU leads to download times comparable to the access link bottleneck scenario. This should provide an incentive for end users to support this locality-promotion scheme, at least in cases with bottlenecks within the network core.

4) Experiment "Fraction of Locality-Aware Peers": Here, inter-AS links are again bottlenecks, but we now vary the share of peers that actively promote locality based on the BGP-based rating. Peers that do not participate in the locality promotion use the regBT client. We vary the share of peers that utilize a locality-aware mechanism from 0% (corresponding to the regBT case) to 100% (corresponding to the previous results).

We can observe that even if only a small share of the peers promote locality, the overall traffic is slightly reduced in comparison to regBT (cf. Fig. 8). The amount of traffic saved increases with the share of peers utilizing the BGP-based rating, until finally almost half of the traffic can be saved if all peers implement BNS and BU. From the remaining

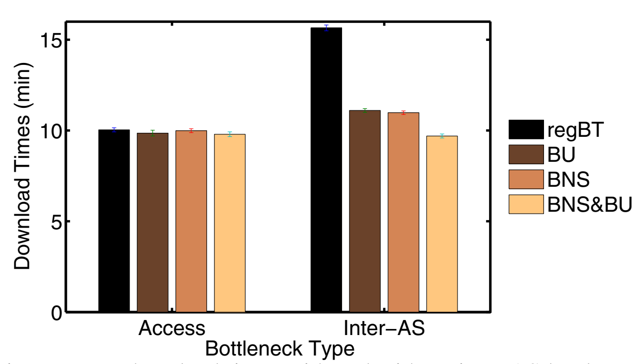

Fig. 7: Mean download times with and without inter-AS bottlenecks.

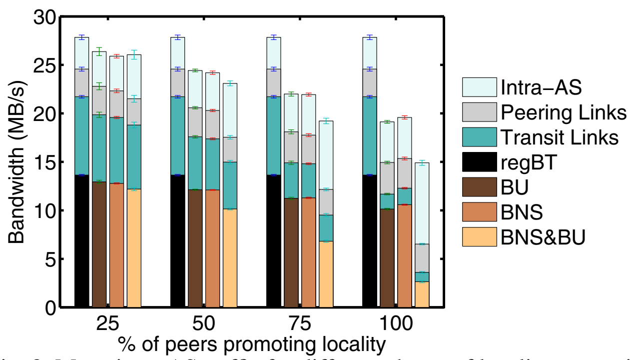

Fig. 8: Mean inter-AS traffic for different shares of locality promoting peers in the swarm.

TABLE I: Mean download times (in minutes) of locality promoting peers and non-locality promoting peers (in brackets).

Share (%)	25	50	75	100
regBT	- (29.50)	- (29.50)	- (29.50)	- (29.50)
BU	22.95 (24.85)	20.24 (21.41)	18.54 (18.94)	16.65 (-)
BNS	13.04 (22.25)	13.59 (18.43)	13.52 (16.41)	13.40 (-)
BNS&BU	10.90 (20.33)	11.39 (15.89)	11.14 (14.00)	11.09 (-)

traffic, again a major share is intra-AS and peering traffic, which causes the stub-AS providers much less costs than traffic forwarded to tier-2 ISPs.

Since we still have inter-AS bottlenecks, the download times are influenced by the use of locality-aware mechanisms. We discern between the two classes of peers, the ones that promote locality and the ones that do not, cf. Table I. In general, BGP-based locality-awareness shortens the download times for both groups, but the group of peers actively promoting locality profit more. Still, the positive effect of locality promotion is enjoyed also by regular peers, since a local neighbor may be preferred over a remote one by a locality-aware peer even if it is not aware of locality itself.

VII. Conclusion

The BGP-based locality promotion mechanism, proposed in this paper, provides a possible collaboration of ISPs and P2P applications in order to manage P2P traffic. It uses existing BGP routing information of an ISP and assigns a rating value to peers in the overlay network according to this routing information. P2P applications can take this rating value into consideration in their peer selection mechanism by using Biased Neighbor Selection (BNS), Biased Unchoking (BU), or their combination. By applying the locality promotion mechanism, the ISP can reduce the amount of costly inter-domain traffic and P2P applications can achieve a better performance in the form of shorter download time.

According to the simulation results the total traffic is reduced by up to 60% if peers act according to the rating value. If peers are locality-aware, P2P traffic takes shorter routes and crosses less network links in total, which results in less bandwidth consumption. The best result for this traffic reduction is achieved if peers apply both BNS and BU together. The BGP-based locality promotion manages to prefer peering traffic over transit traffic as well as shorter connections (i.e. crossing less ASs) over longer ones. By localizing P2P traffic in the network of an ISP, the intra-domain traffic is increased, but the costly inter-domain transit traffic is reduced. Additionally, if inter-domain links represent a bottleneck, locality-aware peers achieve a shorter download time compared to regular BitTorrent. Thus, BGP-based locality promotion together with the combined BNS and BU is beneficial for both ISPs and P2P applications.

As future work it is planned to extend the simulation study with additional swarm and network parameters as well as investigate additional network topologies and possible cooperation of ISPs. Furthermore, in the SmoothIT project the mechanism together with other approaches will be evaluated in a field trial in an ISP network.

Acknowledgments

This work has been performed in the framework of the EU ICT Project SmoothIT (FP7-2007-ICT-216259). The authors would like to thank all SmoothIT partners for useful discussions on the subject of the paper.

References

[1] Cisco Systems, "Cisco Visual Networking Index: Forecast and Methodology, 2008-2013," White Paper, June 2009.

[2] J. Seedorf and E. Burger, "Application-Layer Traffic Optimization (ALTO) Problem Statement," IETF RFC 5693, October 2009.

[3] S. Le Blond, A. Legout, and W. Dabbous, "Pushing BitTorrent Locality to the Limit," Tech. Rep., Dec 2008.

[4] R. Bindal, P. Cao, W. Chan, J. Medval, G. Suwala, T. Bates, and A. Zhang, "Improving Traffic Locality in BitTorrent via Biased Neighbor Selection," in *Proc. of the 26th IEEE International Conference on Distributed Computing Systems*, 2006.

[5] V. Aggarwal, A. Feldmann, and C. Scheideler, "Can ISPs and P2P systems co-operate for improved performance?," *ACM SIGCOMM Computer Comm. Rev. (CCR)*, vol. 37, no. 3, pp. 29–40, July 2007.

[6] H. Xie, R. Y. Yang, A. Krishnamurthy, Y. G. Liu, and A. Silberschatz, "P4P: Provider portal for applications," *SIGCOMM Computer Comm. Rev. (CCR)*, vol. 38, no. 4, pp. 351–362, 2008.

[7] "Application-layer traffic optimization (alto)," IETF Working Group, http://www.ietf.org/html.charters/alto-charter.html, 2009.

[8] The SmoothIT project, "http://smoothit.org/," 2008.

[9] P. Racz, S. Soursos, M. Á. C. Rodriguez, S. Spirou, F. Hecht, I. Papafili, G. D. Stamoulis, Hasan, and B. Stiller, "Economic traffic management for overlay networks," in *ICT-MobileSummit 2009*, June 2009.

[10] Osama Saleh and Mohamed Hefeeda, "Modeling and Caching of Peer-to-Peer Traffic," in *Proc. of the IEEE International Conference on Network Protocols*, Washington, DC, USA, 2006.

[11] D. R. Choffnes and F. E. Bustamante, "Taming the torrent: a practical approach to reducing cross-ISP traffic in Peer-to-Peer systems," *SIGCOMM Computer Comm. Rev.*, vol. 38, no. 4, pp. 363–374, 2008.

[12] S. Oechsner, F. Lehrieder, T. Hoßfeld, F. Metzger, K. Pussep, and D. Staehle, "Pushing the performance of biased neighbor selection through biased unchoking," in *9th International Conference on Peer-to-Peer Computing*, Seattle, USA, Sep 2009.

[13] P. Racz and Z. Despotovic, "An ALTO Service based on BGP Routing Information," Internet Draft, June 2009.

[14] Y. Rekhter, T. Li, and S. Hares, "A Border Gateway Protocol 4 (BGP-4)," IETF RFC 4271, January 2006.

[15] M. Caesar and J. Rexford, "BGP routing policies in ISP networks," *IEEE Network*, vol. 19, no. 6, 2005.

[16] A. Legout, G. Urvoy-Keller, and P. Michiardi, "Rarest first and choke algorithms are enough," in *Proc. of the 6th ACM SIGCOMM Conference on Internet Measurement*, Rio de Janeiro, Brazil, 2006.

[17] "Bittorrent specification," http://wiki.theory.org/BitTorrentSpecification, Last accessed: February 2010.

[18] "Protopeer," http://protopeer.epfl.ch/index.html, Last accessed: February 2010.