2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)

Global Path Service: a New Inter-domain Routing Scheme

Wei Zhang
Depart.of CST, Tsinghua University
Wuhan commanding comunications Academy
Beijing, China
E-mail: zw@netarchlab.tsinghua.edu.cn

Jun Bi
Depart. of CST, Tsinghua University
Beijing, China
E-mail: Junbi@cernet.edu.cn

Jianping Wu
Depart. of CST, Tsinghua University
Beijing, China
E-mail: jianping@cernet.edu.cn

Hongcheng Tian
Depart. of CST, Tsinghua University
Beijing, China
E-mail: tianhc08@mails.tsinghua.edu.cn

***Abstract*—This paper introduces a new inter-domain routing scheme based on a centralized routing service: Global Path Service (GPS). This routing service provides alternate inter-AS paths different to ordinary BGP routes. This scheme facilitates diversified inter-AS forwarding paths through GPS-enabled ASes offering their inter-domain transit tunnels (e.g. MPLS). A GPS center concatenates those transit tunnels into available GPS paths. This routing scheme facilitates a kind of diversified source routing at the AS-level.**

When a packet needs to be forwarded, the local router queries the GPS routing service to obtain a Path ID which will serve as a kind of source routing primitive. The packet will be routed based on this path ID from AS to AS, using tunnels to cross intermediate ASes on the given path. This paper illustrates a framework to implement the infrastructure of diversified routing. This infrastructure should allow for novel inter-domain routing services, such as inter-domain Quality of Service (QoS) or multipath routing. We evaluate the availability, scalability and path diversity of the GPS service by simulations.

Keywords- Inter-domai;, multipat;, Routing; BGP

I. INTRODUCTION

Internet routing can be understood as an end to end path exploration process. The path travels within and/or between administrative domains (Autonomous Systems/ASes) with corresponding routing protocols. The inter-domain routing infrastructure is highly distributed. BGP is the only de facto inter-domain routing protocol which supports routing policies of every individual AS. Each AS makes independent routing decisions. This scheme makes it difficult to manipulate inter-domain paths, because nothing is for sure about whether the next hop AS will take an expecting route or not. There is no existing coordinating mechanism to facilitate diversified or determined inter-domain routing in BGP. BGP guarantees reacheability but no more admirable routing versatilities.

However, versatile inter-domain routing is valuable and appealing functioning [1] when we put an eye on the new requirements of the future Internet applications. End users are expecting multipath routing services [14] to provide resilient/high throughput inter-domain connections, and some sites may need determined inter-AS path services to provide secured ephemeral routes between them. QoS [12, 13] and multicast are also desirable routing functioning; however they are still difficult to be deployed in an inter-domain scenario. If it is hard enough to design versatile routing algorithms to substitute BGP, it will be admirable to have alternate inter-domain routing schemes/services in addition to BGP routing.

Motivated by these considerations, we propose a new inter-domain routing scheme: Global Path Service (GPS) as an alternate inter-domain routing scheme which focuses on routing over diversified AS level paths. GPS will facilitate a centralized policy coordinating mechanism when each GPS-enabled AS provides policy compliant transit tunnels to the GPS center. In our proposal, BGP guarantees reacheability and GPS enhances enriched routing services (e.g. multipath routing, QoS routing) at the AS level.

The rest of the paper proceeds as follows. In section II, we illustrate an overview of the GPS. In section III, we describe our basic design conceptually. In section IV, we evaluate the feasibility of this scheme with simulations. In section V we review related works. At last we draw a conclusion in section VI.

II. GPS ARCHITECTURE

We define the Global Path Service (GPS) as an internet-wide infrastructure which will provide concentrated inter-AS path calculation service to facilitate specific inter-domain IP routing. In this framework, a GPS center will be deployed as a focal contact point and centralized computing server. Each GPS-enabled AS must at least have a GPS agent which is a local contact point. We can expect that the GPS center has powerful computing capacity and collects available AS transit information from all GPS agents. Hopefully a collective topology can be generated.

In our scheme, Internet Service Providers (ISPs) need to establish transit tunnels (e.g. MPLS) between their GPS-enabled AS Border Routers (ASBRs) if they want to make a profit on providing GPS services. Of course these transit tunnels are compliant with their routing policies. Their local

978-1-4244-6542-2/$26.00© 2010 IEEE

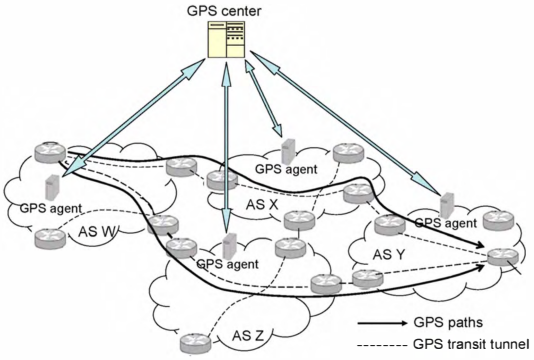

Figure 1 GPS framework

GPS agent will report these transit tunnels to the GPS center afterwards.

The path service is built by calculating inter-AS paths based on a transit aware AS-level topology. Each AS path will be assigned a Path ID (PID). Each PID needs to be disseminated to the corresponding GPS agents on the path that the PID represents. The GPS agents will then bind the PID with a local transit tunnel. Consequently GPS-enabled ASBRs may tunnel packet as long as a valid PID embedded in the IP header. How to use the PID will be illustrated in Section III. Figure 1 illustrates a conceptual framework of GPS.

In GPS, the path exploration is the most important functioning. However, in this paper we focus on the architecture design of an inter-domain routing infrastructure instead of discussing how to implement various inter-domain path services. GPS centre calculates inter-AS path by concatenating available transit tunnels provided by intermediate GPS-enabled ASes. We believe that there exist various heuristic multipath exploration algorithms to facilitate routing diversity, but this topic is beyond the realm of this paper. Another related issue is the consideration of applying generalized policy expressing mechanism, such as Routing Policy Specification Language (RPSL[2]) between the GPS agents and center. Under a centralized control it will be straightforward to detect policy conflicts. Neither this consideration will be included in this paper.

In our scheme, the intra-domain routing is unchanged. Inter-domain routing now has two versions, either goes with BGP or with GPS. BGP and GPS serve as symbiotic inter-domain routing schemes. Our intention is to enrich inter-domain routing services and BGP functioning will not be affected. The detailed GPS design is illustrated next.

III. GPS DESIGN

A. GPS components

1) GPS center

GPS center is the core element of the infrastructure. It has powerful computing capacity as we have mentioned before. It can be an Internet Data Center (IDC) or other similar facilities. GPS providers can be ISPs or telecommunication operators or other Internet practitioners. A GPS provider may have a disseminated GPS center if its services cover a large range of the Internet. However, the GPS center should operate under a logically centralized control. A backup scheme will be necessary in case the GPS center becomes a single failure point.

2) GPS agent

GPS agents should be deployed in each participating ASes and under the full administration of the GPS center. GPS agents collect necessary AS-wide routing information, which includes IP prefixes assigned/administrated within the local AS, neighboring ASes and available AS transit tunnels etc. An AS may have multiple GPS agents which can be aggregated at a rendezvous point before they reach the GPS center. Each ISP may use GPS agent to express their specific local transit policies in a generalized format. The GPS centers distribute PIDs for each AS on the path via GPS agents. GPS agents will bind these PIDs with valid tunnels on the corresponding ASBRs. GPS agents also maintain IP prefix/ASN mapping information received from GPS center. There will be special protocols designed to fulfill these tasks.

3) GPS-enabled ASBR

Some ASBRs must be upgraded to support GPS functioning. The GPS-enabled ASBRs should establish transit tunnels to each neighbor ASes, and bind PIDs with these tunnels under the instructions of GPS agents. For each incoming packet, GPS ASBR checks the optional fields in the IP header to decide either GPS forwarding or BGP forwarding. For each outbound packet originated from local AS, GPS ASBR consults the GPS agent for destination ASN and PID which will be stamped afterwards. Of course a cache mechanism can be helpful to improve the performance of PID stamping and hopefully the startup time of a flow will be reduced considerably.

4) Host upgrade & GPS service proxy

Host/end system upgrade is optional in the GPS scheme. The GPS service can be absolutely transparent to the users' applications/end systems, because ISPs may deploy GPS proxies here and there within their ASes. These proxies are GPS-aware network devices which will initiate GPS service request by adding optional field in the IP header and forward the packets to a GPS-enabled ASBR when necessary. If the host/application requires GPS services, it might use a proxy or optionally upgrade its APIs to support GPS. How to initiate a GPS packet will be illustrated in the following subsection where we describe the format of the optional field in the IP header.

B. GPS functioning

1) Global AS topology generating

The GPS center collects regional AS topology from the agents deployed in each ISP. It might obtain a comprehensive Internet topology if globally deployed. Ideally the quality of the global AS topology can be much better than a BGP's view, since many "invisible" peering AS links in [3] can be explicitly detected in GPS. An improved topology view is very helpful to enhance the routing performance.

2) IP Address and AS numbers mapping

Since GPS path will be built on the granularity of AS level, we need to map the source/destination IP address to the corresponding source/destination ASN before an AS path can be assigned. The GPS agent will play an important role in this function.

Each GPS agent registers local IP prefixes (IP addresses assigned to the local ISP or IP addresses within the administrative of the local AS) with the ASN of the ISP, the GPS agent collects this mapping information and report it to the GPS center. Note that each IP prefix in the registration can only be set uniformly in /24 format in case of overlapping. The GPS center will disseminate the collective mapping information to every GPS agent periodically. This

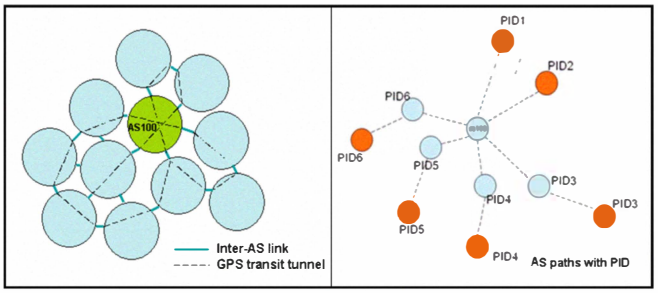

Figure2 Path generation and PID assignment. A spanning tree

mechanism guarantees GPS agents have global mapping information for all GPS-enabled ASes. When ASBRs need this mapping information they query local GPS agents. This mapping information can be cached in the ASBR for a while when routing performance being taken into consideration.

3) Path generation and PID distribution

This part involves the major functioning of a GPS center. As we have mentioned the algorithms of path generation can be very tricky, therefore we simply introduce a heuristic method that only makes the scheme feasible and convenient to discuss the mechanism of PID distribution.

In our algorithm, the abstraction of a GPS AS is a "switch", and its neighboring ASes determine its switch matrix. With the assumption that a GPS center has got the transit tunnels (candidate links of a path) provided by some ISPs, it can explore AS paths between any given ASes on this topology. We generate a spanning tree rooted from a given AS with Breadth First Search (BFS) algorithm. For each branch of this spanning tree (from the root to a leaf) we assign a 32-bit Path ID. Figure 2 illustrates this method of path generation and PID assignment. Based on the transit allowance topology, a BFS tree rooted from AS 100 will be generated and the PIDs will be assigned to each ASes on its branches. These PIDs can be disseminated to each corresponding GPS agents. GPS center will alter the root to do the same thing and generate enough PIDs to ensure each pair of ASes having at least one path.

4) PID and local transit tunnels binding

When a GPS agent received a PID update message from the GPS center, it will update the bindings between the PIDs and the local AS transit tunnels. Each ASBR must promptly update the binding and send back acknowledgement to the GPS agent. If a transit tunnel becomes invalid, the binding

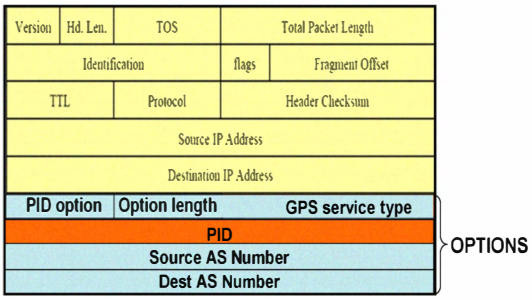

Figure 3 PID is wedged in IPv4 header's optional field

fails. This failure will be reported to the GPS center by GPS agents explicitly and the path will be recalculated to bypass the failed AS.

5) PID stamping and forwarding

In this scheme, PID is used as a source route primitive and should be embedded in the header of each packet. We give an example to embed GPS information in the IPv4 header's optional field as shown in Figure 3. In the optional field, PID, source ASN and destination ASN will take 4 Bytes respectively. IPv6 header has more flexible options therefore

PID embedding will be feasible as well. Of course, this example is not the only way to wedge a PID into an IP packet and there might be better solutions than our preliminary design.

When a packet needs to be routed to a remote AS, The GPS-enabled host or a proxy may fill in the GPS service type into the optional field and the PID, source and destination ASN bits with padding of all "1". Once a GPS-enabled ASBR received this packet, it needs to fill in the padding bits with correct GPS information. It will check the cache first. If there is a miss in the cache, it will query the GPS agent. The GPS agent will promptly responds either the available GPS information (PID, source and destination ASN) or a "service not available" message (the destination IP does not have a mapping registration or no PID available). If the ASBR successfully retrieved necessary GPS information, it will fill in the PID and source and destination ASN accordingly, otherwise it eliminates the optional field and forwards the packets as per ordinary BGP routing.

Since we have a profile of the GPS scheme, the following will further introduce its functioning on the control plane and data plane.

C. GPS control plane

The control plane has four processes: firstly, GPS information collection; secondly, path generation and PID distribution; thirdly, IP address/ASN mapping information dissemination; fourthly, PID request and response.

The GPS information collection will be conducted by GPS agents. The information includes: local dwelling IP prefixes, the neighboring ASNs and local transit tunnels.

The path generation and PID distribution involves the GPS center, agents and corresponding ASBRs as we have described previously. The consequence of this process is to bind PIDs with available transit tunnels.

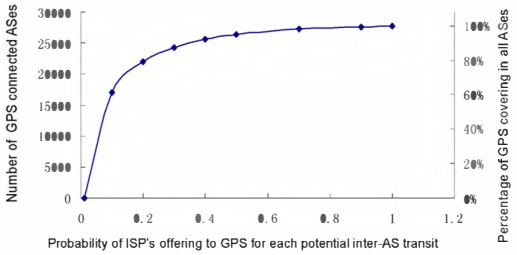

Figure 4 Availability of GPS with different probability that ISP will offer their potential transit tunnels.

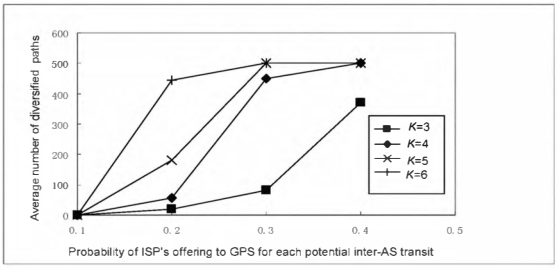

Figure 5 The average number of diversified paths increases with higher probability of GPS provisioning and longer path length of *k* hops

The GPS center will disseminate IP prefixes/ASN mapping information to each GPS agent. This information can be updated from time to time. All prefixes are in the form of a /24 mask length which is normally the finest granularity of the BGP advertisement. After this process, we can expect a GPS agent to maintain necessary information for IP prefix/ASN mapping and AS paths to remote ASes.

The last process will happen when an ASBR needs to know how to fill in the optional fields in a GPS packet. It will send a PID query to the GPS agent. The GPS agent checks the destination ASN and looks up a PID directed to the destination AS before a response is composed. This information can be cached in the ASBR afterwards.

D. GPS data plane

GPS data plane is straightforward as we assume the transit tunnels are ready and all PID bindings are valid. When a GPS-enabled host (or a proxy) needs to forward a packet to another AS, the outbound packet will be routed to the first GPS-enabled ASBR, and then this ASBR will stamp source and destination ASNs and PID into the IP header of the packet after querying a GPS agent or referring to its cache. The following AS transit procedure are similar: ASBRs check the IP header to see if the local AS is the destination AS. The intermediate ASes will tunnel the packet to the next AS on the path as per the PID in the header of the packet until it reaches the destination AS. Although we can set complicated scenarios for this trip, more complex failover mechanism can be implemented in the control plane. Anyway, in case that the intermediate ASBR fails to resolve the PID to a valid transit tunnel, the last resort is to forward the packet in BGP routing scheme. Since BGP guarantees the reacheability of this packet, the best we can expect is that all ASes supporting GPS can enjoy the merits of the new service without any risks.

IV. EVALUATION

The GPS relies heavily on the availability of transit tunnels in participating ASes. To understand to what extend the GPS service can be practically used, we simulated the incremental process of GPS provisioning. Based on different probability of GPS provisioning, we measured the potential of GPS routing diversities. The last concern is the scalability of the scheme. We put GPS into an Internet-wide scenario to verify its scalability.

A. Methodology

We generated an AS-level topology by using BGP RIB and update messages measured by Routeviews [4]. For each AS, we put all its neighboring ASes into a switch matrix which represents potential GPS transit tunnels. Then we chose a certain probability that the AS would take to offer each potential transit tunnel to GPS. As the probability increases, more and more transit tunnels appeared in the GPS topology.

We applied the path generation algorithm illustrated in section III to see how many ASes can be connected by GPS paths and measured the ratio of the GPS coverage in the Internet. Without lose of generality, in our algorithm we use a metric of "hops" as the same in BGP, and all paths generated in our simulations are "simple path" (cycle free).

In order to justify the diversity of the GPS paths, we changed the root of the spanning tree from one AS to another and obtained more paths between two ASes. We also changed the search order of the BFS algorithm, e.g. in ascending/descending order of the ASN, consequently the shape of the spanning tree changed and generated more different paths. We measured an average outcome of this algorithm on routing diversity.

At last, we calculated the number of PIDs generated by our algorithm. This number, together with a theoretical analysis, gave us a rough scale of magnitude of all possible inter-domain paths. We used this estimation to verify the scalability of the GPS in an Internet-wide background.

B. Results

Availability Figure 4 shows the availability of GPS with different probability that ISP will offer potential transit tunnels. Note that even with the probability of 0.1, half of the ASes will be in a connected sub-graph. If the probability be raised to 0.2, there will be more than 20000 ASes being connected by the GPS routing scheme. If the probability reached 0.5, nearly 92% ASes in the Internet will be covered by GPS routing. In our simulation the transit is randomly chosen and only tuned by the probability. Practical GPS providers might purchase transit tunnels purposely and the availability will improve accordingly.

Diversity The diversity of inter-AS paths can not be exhausted at a large scale. Moreover, there are too many heuristic path generation algorithms which cannot be included

in our preliminary design. To evaluate the potential to support multipath routing, we chose two measurements to reflect the diversity of routing provided by GPS.

The first measurement is the average number of possible shortest paths between any pair of ASes. This measurement reflects the potential of multipath exploration on greedy routing policy. Another measurement is the average number of potential paths within a given distance. This measurement is defined as *k*-hops average multipath availability, which accounts for all simple paths within the length of *k* hops from the source to a destination. Both of these two measurements are averaged in our simulations.

Our experiment shows that the shortest path will not have many alternate equal cost paths. Even on the transit provisioning probability of 1, the average number of shortest paths is only 1.008. If we allow for some diversified paths a little bit longer than the shortest path, the average diversity of routing increases considerably.

Figure 5 shows the result of routing diversity evaluation under various GPS transit provisioning probabilities. With the probability of 0.2 and k>4, the connected ASes will have ideal average routing diversities (more than 150 diversified paths).

The above experiment shows that the GPS center can be extremely effective on multipath routing because it has a more complete view of routing state than any distributed routing schemes. All paths are compliant with routing policies, which is critical in the case of inter-domain routing.

Scalability We calculated the number of PID generated by GPS with the maximum transit probability. In our topology there were 27737 ASes. There were approximately less than 25000 PIDs generated from any single spanning tree rooted on a given AS. Even if we generate PIDs by taking each AS as the root of a spanning tree, there will be no more than 27737*25000 PIDs. For each AS, it only needs to handle maximally about 27736+25000 PIDs. Theoretically, given n ASes, the total number of PID maintained by GPS center will scale with n^2, and each GPS agent has to handle no more than $2n$ PIDs.

However, we don't need that many PIDs. Actually there will be many overlapped/attached paths which can be identified by the same PID. With the help of more efficient path generation algorithm, the number of PID will reduce considerably. After all, if compared with the size of BGP RIB, GPS PID registration in each ASBR only accounts for trivial resources.

V. Related Works

There are some proposals, e.g. R-BGP [5], route splicing [6], MIRO [7] and YAMR [8] also allow for discovery of additional interdomain routes besides BGP. All require establishing additional states on routers and might introduce heavy burden to BGP routers on maintaining multi-topology routing. In these schemes there is no centralized routing policy coordinating mechanism. The flexibility of inter-domain routing services can hardly be implemented on the basis of bilateral negotiation between neighboring ASes.

Overlay networks can provide diversified paths above the network layer [9, 10, 11], but the overlay routing may not be efficient because of the lack of the underlay topology. The "topology proximity problem" is the major problem encountered by overlay applications. Moreover, ISPs dislike P2P overlay since the traffic is out of their control.

MP-TCP [15] studied a TCP congestion control over multi-path routing schemes, but not focused on the routing scheme which mainly deals with network layer implementations. In [15], the multiple path routing implementation is assumed, either in a scenario that source routing between peering routers is allowed (overlay routing) or that multihoming hosts can choose different ISP initially.

LISP [16] proposed by CISCO also implies diversified inter-domain routes. Their scheme is based on Core/Edge Split (CES) routing architecture and similar to multihoming scenarios.

To our best knowledge, none of above can provide diversified inter-domain routing which obviously requires centralized policy coordination and management.

VI. Conclusion

GPS provides an alternate inter-domain routing scheme which will compensate the inherited architectural deficiency of BGP. Although this proposal is still in its preliminary stage, it tries to address the problem of routing versatility. Based on a centralized routing service, GPS introduces a scheme to facilitate diversified inter-domain routing service, which may bring many possibilities to the future Internet.

Acknowledgment

This research is supported by the Key Project of Chinese National Programs for Fundamental Research and Development (973 program) No. 2009CB320501 and Specialized Research Fund for the Doctoral Program of Higher Education (SRFDP) No 200800030034.

The authors would like to thank the participants for their contributions: Yangyang Wang, Hongyu Hu, Baobao Zhang and Guang Yao.

References

[1] Israel Cidon, et al. "Analysis of Multi-Path Routing", IEEE/ACM TON 1999 885.
[2] RPSL. http://tools.ietf.org/html/rfc2622
[3] Ricardo Oliveira et al. "Quantifying the Completeness of the Observed Internet AS-level Structure", to appear in UCLA Computer Science Department - Technical Report TR-080026-2008.
[4] Routeviews, http://www.routeviews.org.
[5] N. Kushman et al. R-BGP: Staying connected in a connected world. In Proc. 4th USENIX NSDI,Cambridge, MA, Apr. 2007. SIGCOMM '04
[6] Murtaza Motiwala et al., Path Splicing. SIGCOMM'08, August 17–22, 2008, Seattle, Washington, USA.
[7] W. Xu and J. Rexford. "MIRO: Multi-path Interdomain Routing". In Proc. ACM SIGCOMM, Pisa, Italy, Aug. 2006.
[8] Igor Anatolyevich et al. "YAMR: Yet Another Multipath Routing Protocol", Technical Report No. UCB/EECS-2009-150,Oct. 2009
[9] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris. "Resilient Overlay Networks". In Proc. 18th ACM Symposium on Operating Systems Principles (SOSP), pages 131–145, Banff, Canada, Oct. 2001.

[10] D. G. Andersen, A. C. Snoeren, and H. Balakrishnan. "Best-path vs. multi-path overlay routing". In *Proc. ACM SIGCOMM Interne Measurement Conference*, Miami, FL, Oct. 2003.

[11] K. P. Gummadi, H. V. Madhyastha, S. D. Gribble, H. M. Levy, and D. Wetherall. "Improving the reliability of Internet paths with one-hop source routing". In *Proc. 6th USENIX OSDI*, San Francisco, CA, Dec.2004.

[12] Rui Prior.et al. "Towards Inter-Domain QoS Control", ISCC Proceedings of the 11th IEEE Symposium on Computers and Communications Pages: 731 - 739 ,2006

[13] Cao Yuanming et al. Initiator-Domain-Based SLA Negotiation for Inter-domain QoS-Service Provisioning, ICNS 2008

[14] Dan Li et al. Multipath Inter-domain Routing Based on Subdividing Locators, APCIP 2009

[15] Huaizhong Han, Srinivas Shakkottai, C. V. Hollot, R. Srikant, Don Towsley. "Multi-path TCP: a joint congestion control and routing scheme to exploit path diversity in the internet", IEEE/ACM Transactions on Networking (TON) archive Volume 14 , Issue 6 (December 2006) Pages: 1260 - 1271

[16] FARINACCI, D., FULLER, V., ORAN, D., MEYER, D., AND BRIM, S. Locator/ID Separation Protocol (LISP). Internet draft, draft-ietf-lisp-04.txt, 2009.