Relation between Bandwidth measurement methodology and BGP routing traffic engineering processes .

RED-SEA
Communications
Services

Introduction

I would like first to sincerely thank the AIS-14's committee, the Ministry of telecommunications in Djibouti, as well as Djibouti-Telecom to allow me to present this subject.

- The purpose of this presentation is to put in relation the common Internet measurement bandwidth techniques with the routing processes of an ISP's Internet Backbone and its peers.

- It is very common that the end-users and corporates test their bandwidth, however results can be unexpected and drive major complains to the ISP.

  Many Africans ISPs are connected to Tier1 or Tier2 ISPs in Europe. With the development of the submarine cables, major Tier1 and Tier2 ISPs setup more and IP PoPs with CDN servers in Africa.

  Still, the majority of the IP routes and contents are sourced from the USA and Europe. As a result latencies and congestions issues are still a reality for all end-users, which can impact the bandwidth performances results.

- Bandwidth speed measurement tools can be very useful if the methodology and the interpretation of the results are coherent.
  It can improve a lot the work of IT-engineers regarding routing and traffic engineering processes, but also developers of netcode applications and optimize new IT's services trend like VOD, cloud services, voip etc.

➢ AFNOG summit is great event where many Telco stakeholders can share their knowledge and improve the experience of the Internet in Africa.

➢ **The final aim of this document is to make Telco professionals participate in this research community, knowing that this subject can be benefic for all Internet end-users.**

I) Different Metrics of network bandwidth from Internet measurement speed literature.

- **Regarding the terminology of this particular thread, should we commonly speak about Internet speed or Internet bandwidth?**

- **"Internet bandwidth measure" terms will be preferred.**
  In the context of data networks, the term "bandwidth" generally quantifies the data rate at which a network link or an end-to-end network path can flow.

  The perception of bandwidth is central to digital communications related to the Internet, specifically to packet networks as it points to the amount of data a link or a network-path can deliver per unit of time.

  The term bandwidth can define the net bit rate, the channel capacity, or the maximum throughput of a logical or physical communication path in a digital communication system.

- It is important to **differentiate** the **bandwidth measure** on a **single link** or on an **end-to-end network path** :

  - For instance, an IT-engineer designing a corporate network will mainly use bandwidth measurements tools on single links like the one of its switches, routers and servers.

  - In the case of designing an ISP network, since the Internet is an interconnected network of operators, the notion of bandwidth related to an end-to-end path will be additionally deeply considered.
    This time the bandwidth performance will not only be affected by its own backbone, but also by the networks of its Internet peers connected. The traffic engineering and equipment specs (cpu, interfaces, protocol …) of an ISP generally differ from one to another one.

  - The degree of impact is predisposed mainly by the size of the ISP, for instance a Tier-1's ISP bottleneck issue will have a bigger influence on the Internet than a Tier-3's ISP congestion.

I) Different Metrics of network bandwidth from Internet measurement speed literature.

- **Internet speed measures can be globally defined into 3 different metrics, where each one can be used in order to highlight specifics performance traits of your Internet connectivity:**

➢ **Capacity is the maximum bandwidth that a single link or an end-to-end network-path can deliver.**

The Capacity is generally defined as the maximum number of bits per second ("bps") that a network element can transfer. **Regarding an end-to-end network path, the capacity is fixed by the slowest or thinnest network element along the path.**

Capacity on a single link is generally accurate and is the most common metric that everyone is aware. For an ADSL end-user, it will be the local- loop between the DSLAM and its ADSL modem, for a corporate it will be the Fiber2Office interface speed between the access network and its router …

➢ **Available bandwidth is the residual capacity of a link. It is simply the capacity minus its utilization.**

This metric is quite interesting since it allows many developers to estimate how much residual bandwidth their application can use on the link or a network-path.

Available bandwidth is strongly related to the oversubscription rate, but also to the behavior for the end-users. It is clear that the available bandwidth at 2am is higher than at 7pm for an end-user.

➢ **Achievable Throughput is the amount of data that is successfully sent from a source to a destination via an end-to-end network path. Every component (logical or physical) along the path can influence this metric. This metric reflects very well the Internet behavior, since it also introduces strongly the notion of latency or round-time-trip between a source and its destination on the Internet.**

Achievable throughput considers the network-path but also additional factors like network protocols, host and destination speeds, TCP buffers, CPUs performances; whereas Capacity and available bandwidth just consider the network path. These factors are very hard to estimate and can be assimilated as the noise of your bandwidth performance.

But the most important factors are the TCP size and the RTT.
Considering that default "Windows" TCP size is 64Kbytes, if an end-user have a 100Mbps, and want to download a file from a server that he can ping with a RTT of 100ms, we will have a:
"TCP achievable throughput = (64000 x 8) / 0.1 = 5.12 Mbps"

I) Different Metrics of network bandwidth from Internet measurement speed literature.

- As we can see, the results between the capacity and the achievable throughput do not reflect the same performance. In order to maximize its download speed, the end-user can use a download software accelerator in order to download multiple TCP connections from the server. The total speed can neighbor 100Mbps, but this performance does not interest an end-user since commercially he expects to subscribe a capacity which allows him to download with this rate from a source.

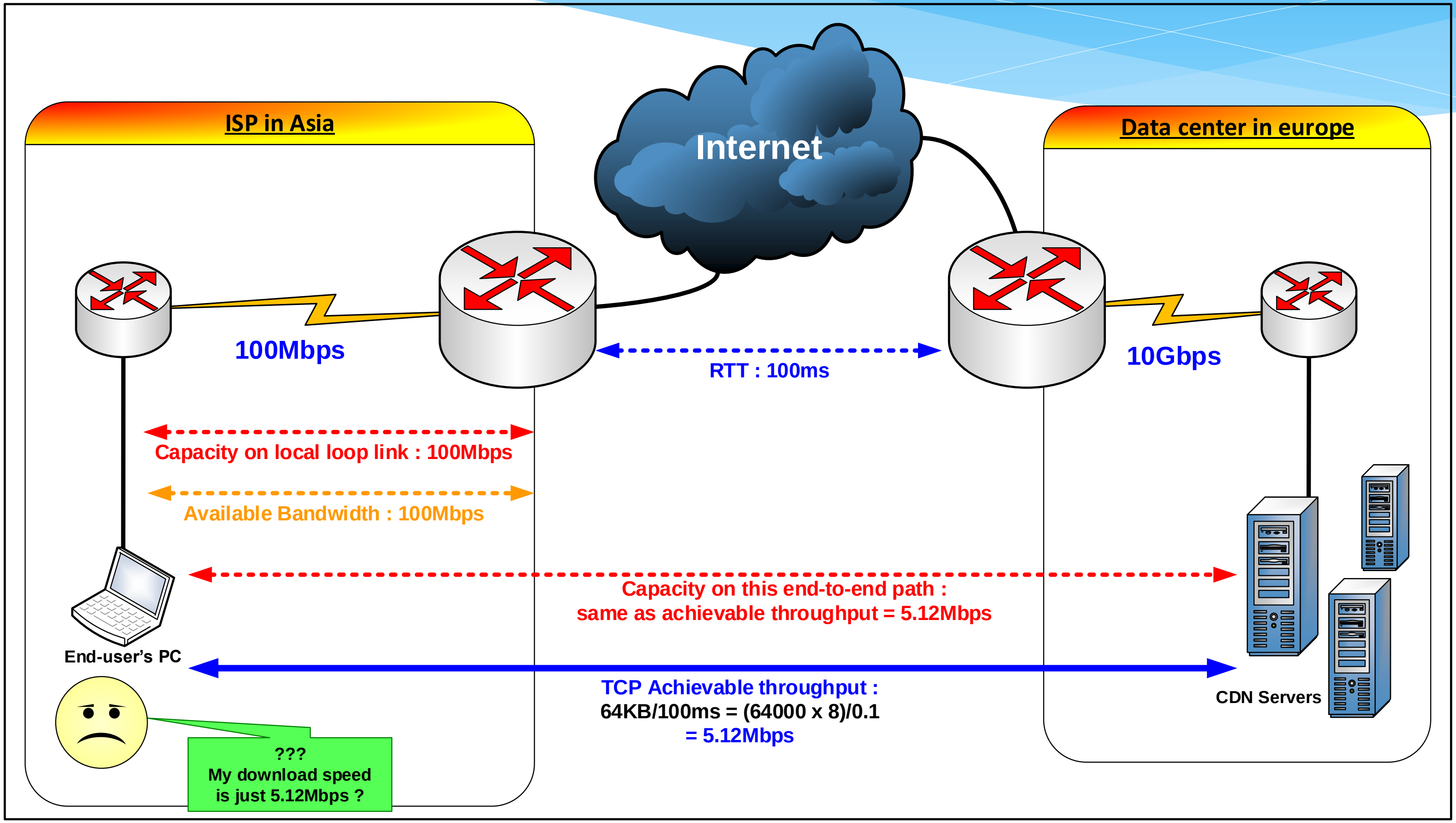

ISP in Asia
Internet
Data center in europe
100Mbps
RTT : 100ms
10Gbps
Capacity on local loop link : 100Mbps
Available Bandwidth : 100Mbps
Capacity on this end-to-end path :
same as achievable throughput = 5.12Mbps
End-user's PC
TCP Achievable throughput :
64KB/100ms = (64000 x 8)/0.1
= 5.12Mbps
CDN Servers
???
My download speed
is just 5.12Mbps ?

Fig A.1 : (TCP) Achievable Throughput vs Capacity purchased

II) Association of different bandwidth Metrics and its terminology

- **"Internet Bandwidth estimation" is a subject related to the Internet, since there is not any official standard methodology and the fact that algorithms are related to various research communities, it can result in some misunderstanding regarding the use of the metrics.**

➢ **The Internet bandwidth measurement literature however shows a common terminology, which can result a shared accepted practice in the use of the metrics.**

- Ultimately all bandwidth measurement tools attempt to identify the performances of a network, but it not always clear how to associate this vague notion of bandwidth towards specific performance metrics.

  And in the worst case scenario, it is not clear if a particular methodology actually measures the bandwidth metric it says to measure. It also happens that tools claiming to use the same metrics give different results without necessarily meaning that the results are incoherent.

➢ **Using the correct terminology allows to understand the methodology, the correct use of the metrics and its associated algorithms and the results.**

II) Association of different bandwidth Metrics and its terminology

- The performance of the bandwidth can be measured on a **"single link"** or an **"end-to-end network path"** which is a sequential association of single links.

- On an end-to-end network path, we can have:
  - A **"physical-transmission-link"** (fiber, copper, microwave) which is generally logically transparent.
  - A **"data-layer-link"** (layer2) which is always supported by a physical-transmission-link, and appears also logically transparent on an IP end-to-end path.

- At this point, we can qualify a physical-transmission-link or a data-layer-link as a **"segment"**.

  It is well-known that the term segment is always referenced to **"physical point-to-point link"**, a **"virtual circuit"**, or a **"shared access local area network"** (Ethernet collision domain, fiber distributed data interface, ring).

- On IP point of view, on an end-to-end network path, we can have:
  - A "transport-link", commonly called an **"IP-layer-link"** which is a logical link clearly appearing in a network path. An IP-layer-link is generally supported and transmitted by a segment.
  - The sequence of IP-layer-links is delimited by hops. On most networks, including the Internet, packets typically need to pass through several routers before they reach their final destination. Each time a packet is forwarded to the next router, a **"hop"** occurs. The hop number is the sum of hops that a packet has taken towards its destination.
  - A hop can be clustered into a sequence of segments (for instance two routers connected by two switched will show one hop but the transmission will be supported on three segments).

- **As result, we can define an "end-to-end network path" as a sequence of IP hops which delimit the logical path from a source to its destination.**

II) Association of different bandwidth Metrics and its terminology

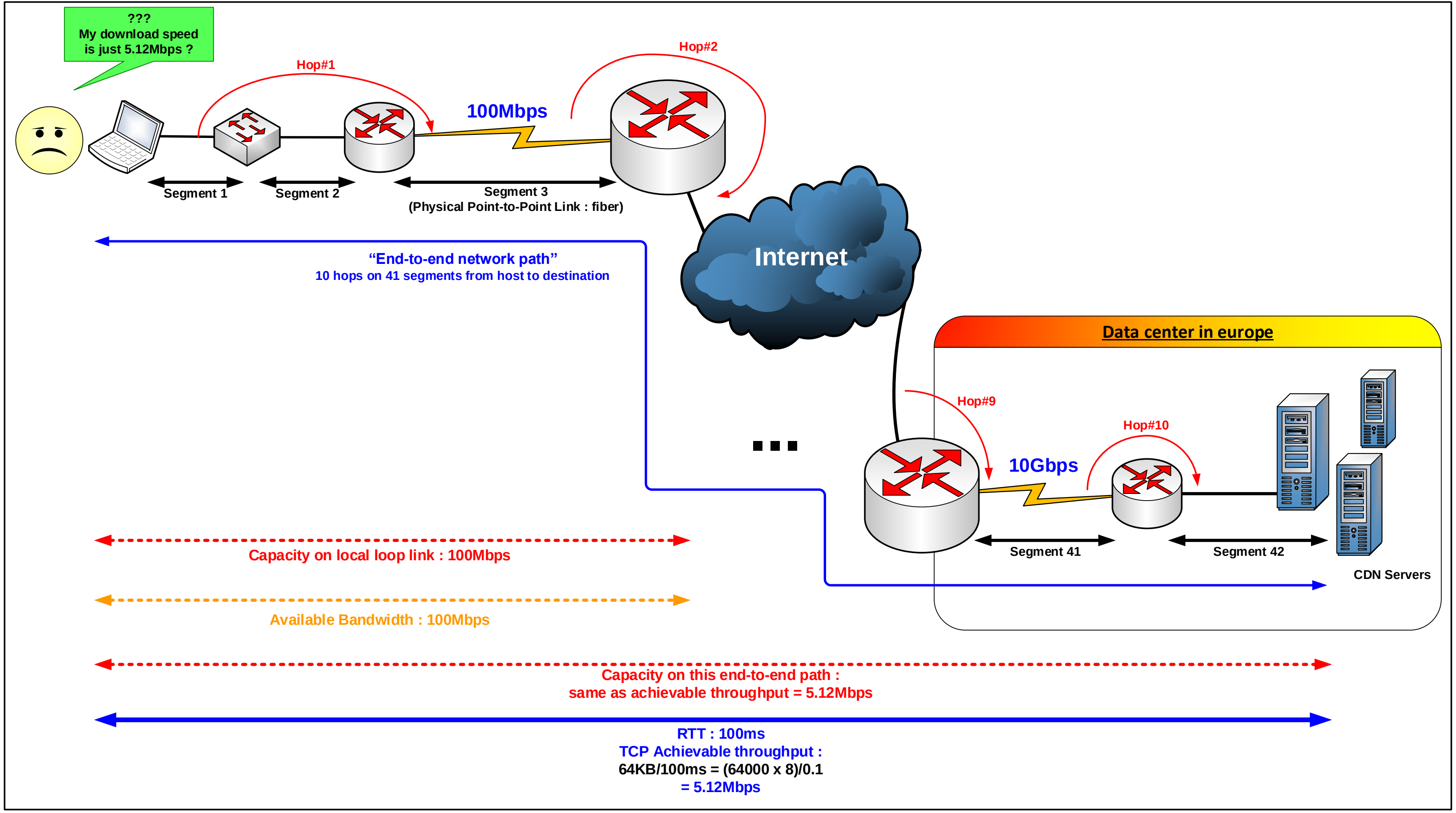

???
My download speed
is just 5.12Mbps ?
Hop#1
Hop#2
100Mbps
Segment 1
Segment 2
Segment 3
(Physical Point-to-Point Link : fiber)
Internet
"End-to-end network path"
10 hops on 41 segments from host to destination
Data center in europe
Hop#9
Hop#10
10Gbps
...
Segment 41
Segment 42
CDN Servers
Capacity on local loop link : 100Mbps
Available Bandwidth : 100Mbps
Capacity on this end-to-end path :
same as achievable throughput = 5.12Mbps
RTT : 100ms
TCP Achievable throughput :
64KB/100ms = (64000 x 8)/0.1
= 5.12Mbps

Fig A.2 : (TCP) Achievable Throughput vs Capacity purchased on an end-to-end network path

- With these terms, we can apprehend that the notion of bandwidth is related to the specs of a link regarding the data-rate transmission, and has "spacio-temporal" relation with the topology of an end-to-end network path.
- The "spacio-temporal" characteristic can be referred such as the available bandwidth on each particular link of and end-to-end path on a particular time-period, thus allowing to localize thin links and bottlenecks.

II) Association of different bandwidth Metrics and its terminology

Internet bandwidth measurement is generally used through two different techniques:

- **"Flooding-based" *** algorithms which use multiples parallels TCP connections and sum the combined rate of all connections. **Multiple TCP connections are preferred over a single connection because it can quickly estimate the "capacity" of the end-user on the network-path.**

- The use of **"optimizing-probing tools" * allows generally to estimate the available bandwidth on different cross-traffic network paths**. Estimating available bandwidth is a very important notion for applications which need strong Internet stability like adaptive video streaming, gaming, voip etc. Two common methodologies are used:

  - The **"Self Loading periodic Streaming"** ("SLOPS") is an algorithm which make a source sends a periodic packet stream of equal-sized packets to a destination at a certain rate. If the stream rate is greater than the path's available bandwidth, the stream will cause a short-term overload in the queue of the thinnest link of the end-to-end network path. One-way delays of the probing packets will keep increasing as each packet of the stream queues up at the narrow link. Oppositely if the stream rate is lower than the available bandwidth, the probing packets will go through the path without causing increasing buildup in the narrowed link, thus not increasing their one-way delays.

  - The **"Train of packet pairs"** ("TOPP") is a measurement technique which send a sequence of ICMP pair with the Timestamp and the Timestamp reply message, where the header carries 3 temporal informations which are the time when packet leaves the source, arrives at the destination and then leave the destination. Thanks to the timestamps, the transmission rate is calculated through the window error between the estimated delay and the real one.

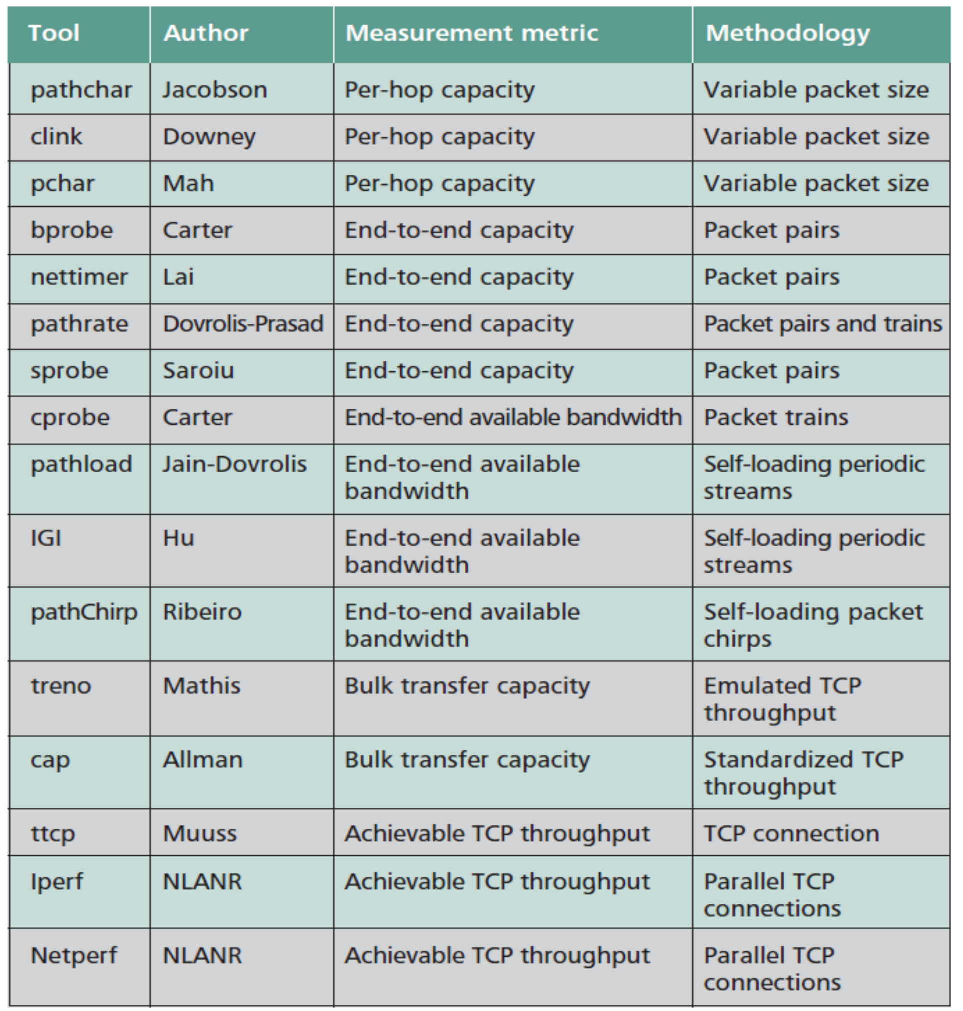
- **For that point, Internet measurement literature often refers a key concept called the "Maximum Burst Size" which is the maximum number of bytes that can be sent contiguously from a source host to a destination host across the network during a certain period of time without dropping a packet.**

  The maximum burst size is determined by the queue size of the bottleneck router and by the current network traffic at that router. Note that other factors such as traffic shaping, policing, or QoS may cause such queuing behavior. This can restrict the effective bandwidth when it is small. So, MBS is also called as effective queue size.

  Not exceeding the maximum burst size is the key to obtaining good TCP achievable throughput.

II) Association of different bandwidth Metrics and its terminology

- A non-exhaustive list of common bandwidth measurement tools with their respective methodology :



Tool	Author	Measurement metric	Methodology
pathchar	Jacobson	Per-hop capacity	Variable packet size
clink	Downey	Per-hop capacity	Variable packet size
pchar	Mah	Per-hop capacity	Variable packet size
bprobe	Carter	End-to-end capacity	Packet pairs
nettimer	Lai	End-to-end capacity	Packet pairs
pathrate	Dovrolis-Prasad	End-to-end capacity	Packet pairs and trains
sprobe	Saroiu	End-to-end capacity	Packet pairs
cprobe	Carter	End-to-end available bandwidth	Packet trains
pathload	Jain-Dovrolis	End-to-end available bandwidth	Self-loading periodic streams
IGI	Hu	End-to-end available bandwidth	Self-loading periodic streams
pathChirp	Ribeiro	End-to-end available bandwidth	Self-loading packet chirps
treno	Mathis	Bulk transfer capacity	Emulated TCP throughput
cap	Allman	Bulk transfer capacity	Standardized TCP throughput
ttcp	Muuss	Achievable TCP throughput	TCP connection
Iperf	NLANR	Achievable TCP throughput	Parallel TCP connections
Netperf	NLANR	Achievable TCP throughput	Parallel TCP connections

III) Routing and traffic engineering influence in bandwidth measures.

As a reminder, the following points are essential, and need to be taken into consideration strongly, even for the most basic design of a network running BGP within an ISP. These notions are useful when traffic engineering using BGP is implemented.

1. **When a customer advertises its prefixes to an ISP:**
   a. The customer will allow IP-packets flows from the ISP (or the Internet) to its networks.
      We can say that the customer downloads IP-packets from the ISP (and reciprocally we can say that the ISP uploads IP-packets to the customer).
   b. By advertising its prefixes, a customer controls its inbound/ingress/download IP-packets traffic.

2. **When a customer receives routes from its ISP:**
   a. The customer will be able to send IP-packets from its network to the ISP (and so the Internet).
      So the customer uploads IP-packets to the ISP (and reciprocally it is the same as the ISP downloads IP-packets from the customer).
   b. By receiving prefixes or routes, a customer controls its outbound/egress/upload IP-packets traffic.

3. **Let's simply memorize that routes announcements allow IP-packets stream/flow from networks in the opposite way:**
   a. advertising routes means receiving/download IP-packets,
   b. receiving prefixes or routes allows sending/upload IP-packets.

4. To influence upload-paths for Internet routes, the BGP local-preference attribute will be manipulated at the gateway router of the given traffic class. A higher local-preference is preferred. The default value local-preference being 100, the local-preference of routes coming in from the less preferred neighbors will be set to less than 100(e.g. 90).

5. Policy Based routing could be also used to influence upload/outbound traffic. "IP next-hop" instructions could be required to set more granular traffic engineering when it is required.

6. To influence the download path for Internet routes, the BGP as-path attribute can be manipulated at the gateway of the traffic class of interest.
   A shorted as-path is preferred over longer ones. As-prepend operations could come handy for that.
   As such, ISP-A's as-number shall be prepended for updates of the given class to all neighbors apart from the one through which traffic download is preferred

III) Routing and traffic engineering influence in bandwidth measures.

- BGP routing and traffic engineering processes focus mainly on prefixes announcement.

However an ISP wants always to maximize its revenue as much as possible, the priority to use the IP transit connectivity to its maximum efficiency is fundamental.
Still common best practices of BGP multihoming do not take care much about additional latencies which can lower achievable throughput performances.

Common case of manipulation of communities and local preference :

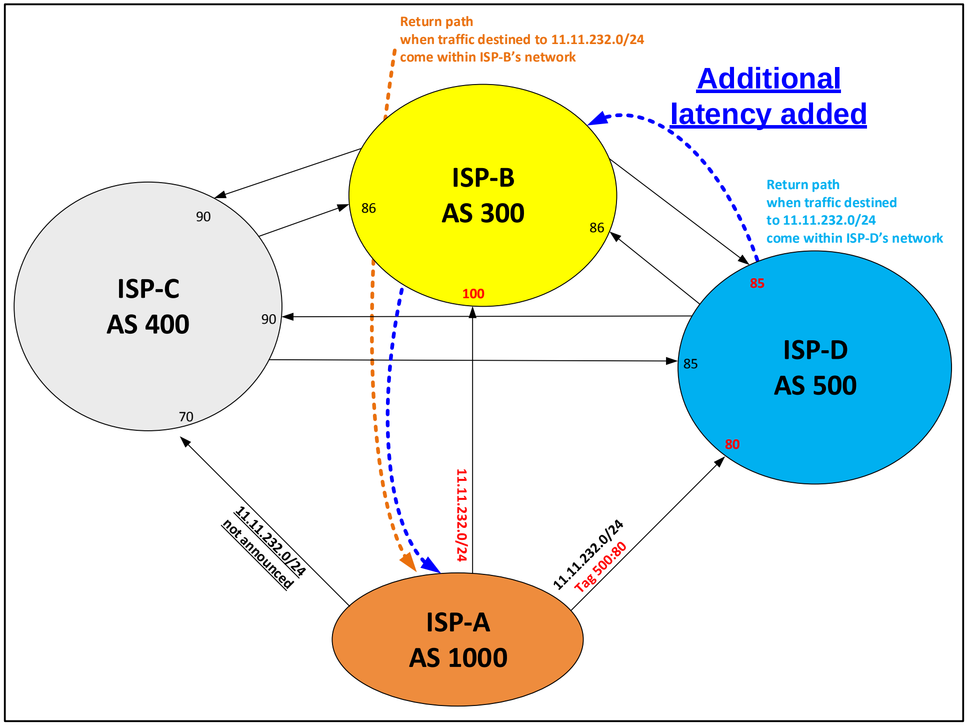


Figure A.3: Return traffic path for prefixA 11.11.232.0/24
where community values are set to 300:100 and 500:80

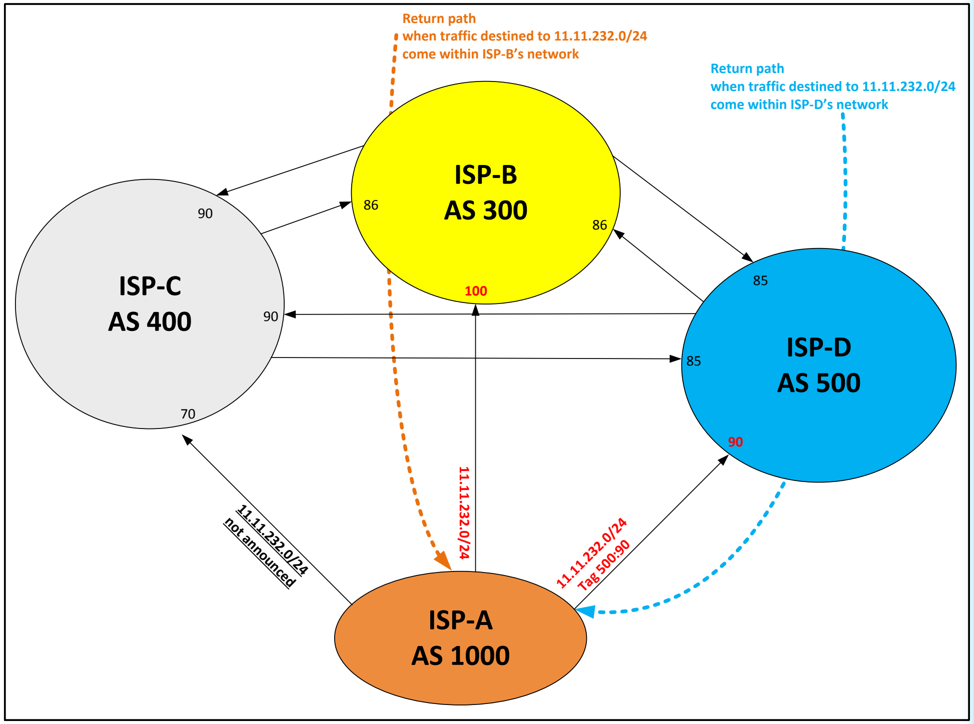
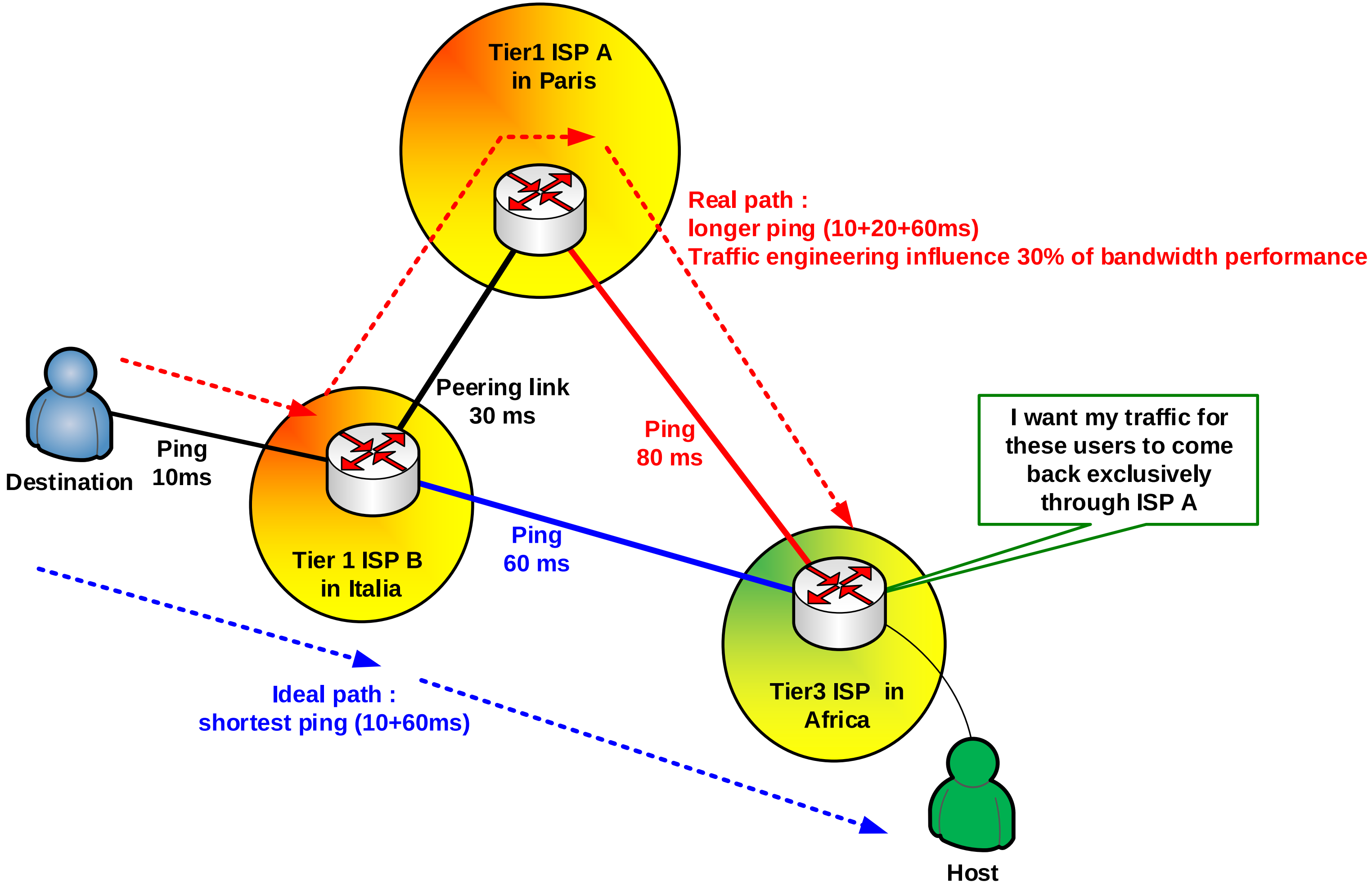


Figure A.4: Return traffic path for prefixA 11.11.232.0/24
where community values are set to 300:100 and 500:90

III) Routing and traffic engineering influence in bandwidth measures.

- Common BGP case which higher the latency for a Tier3 ISP.
  IP transit Link utilization from ISP-A is maximized.
  However achievable throughput performance has significantly decreased.



Tier1 ISP A in Paris
Real path : longer ping (10+20+60ms)
Traffic engineering influence 30% of bandwidth performance
Peering link 30 ms
Ping 80 ms
Destination
Ping 10ms
Tier 1 ISP B in Italia
Ping 60 ms
I want my traffic for these users to come back exclusively through ISP A
Tier3 ISP in Africa
Ideal path : shortest ping (10+60ms)
Host

III) Routing and traffic engineering influence in bandwidth measures.

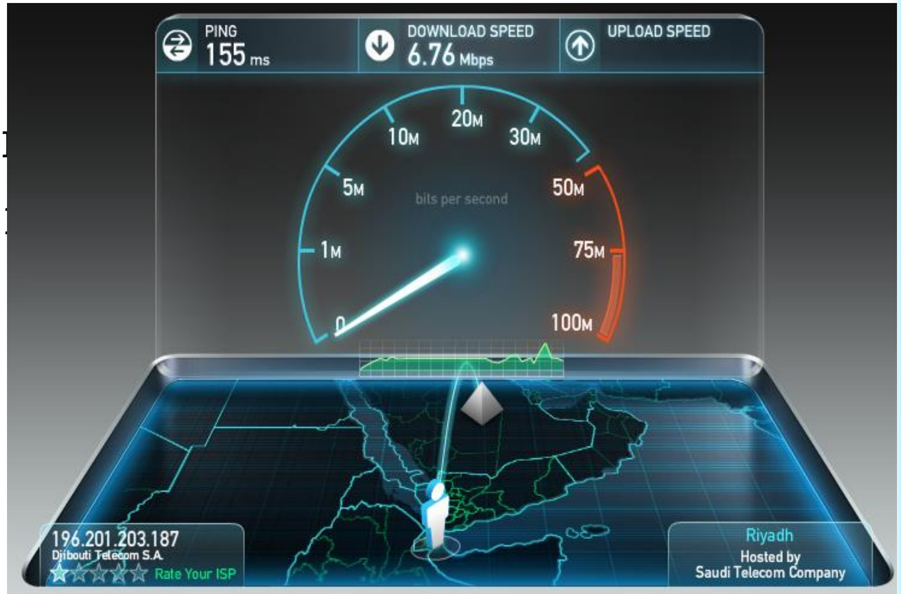
Concrete example of flooding tools :

- By increasing the TCP windows size and doing parallel transfer files, SpeedTest.net give a result approaching a capacity or an achievable bandwidth result.
  However this result consider a network path but not some important additional factors like network protocols, host and destination speeds, TCP buffers etc. Now on that point the end-user can be sometimes surprised since some application sensible to the achievable throughput will not reflect the same great results as speedtest.net.

- In our case, we can see that the notion of the nearest server by Speedtest.net is based on a simple geographic information but does not consider the as-path neither the latency.
  We are not here to criticize SpeedTest.net, however it is important to see how far these results (which are true but not detailed in which way they should be interpreted) can misinform end-users on a large-scale if they are not correctly aware about the bandwidth metrics and their characteristics.

```
tracert www.speedtest.net
Tracing route to cs62.adn.edgecastcdn.net [93.184.219.82] over a maximum
of 30 hops:
  5     9 ms     9 ms     9 ms  access1.intnet.dj [193.251.143.7]
  6    14 ms    11 ms    12 ms  41.189.226.33
  7    88 ms    88 ms    90 ms  so-6-1-1.edge1.Djibouti1.Level3.net [212.73.206.237]
  8    89 ms    90 ms    90 ms  ae-11-11.ebr3.Frankfurt1.Level3.net [4.69.200.230]
  9    88 ms    90 ms    89 ms  ae-93-93.csw4.Frankfurt1.Level3.net [4.69.163.14]
 10    88 ms    89 ms    96 ms  ae-4-90.edge3.Frankfurt1.Level3.net [4.69.154.199]
 11    88 ms    88 ms    88 ms  195.22.214.52
 12    89 ms    99 ms    90 ms  edgecast.franco31.fra.seabone.net [89.221.34.93]
 13    89 ms    89 ms    89 ms  93.184.219.82
Trace complete.


BGP routing table entry for 196.201.203.0/24, version 290099639
Paths: (2 available, best #1, table default)
  Advertised to update-groups:
     1062
  174 30990
    195.66.226.76 from 195.66.226.76 (38.28.1.241)
      Origin incomplete, metric 20032, localpref 100, valid,
                  external, best
      Community: 174:21101 174:22008 5459:1 5459:60
```



PING 155 ms

DOWNLOAD SPEED 6.76 Mbps

UPLOAD SPEED

196.201.203.187
Djibouti Telecom S.A.

Riyadh
Hosted by
Saudi Telecom Company

III) Routing and traffic engineering influence in bandwidth measures.

- Majority of tools focus on giving bandwidth measures results on an end-to-end path, but few of them are able to localize the thinnest link on an end-to-end path.
- If we can have statistics information on thinnest links associated with net-flow analyzer, IT-engineers could readapt the traffic engineering more accurately.
- Same can be said for developers : information of thinnest links will allow to give more accurate control on adaptive mechanisms bandwidth related to TCP traffic.
- Few tools provide localization information on bottleneck links. However some of them need to be readapted, for instance “pipechar” need ICMP requests from routers, but this feature is sometimes disabled by administrators. “Netest” is no more developed. “STAB” gives promising results.
- Moreover the Internet evolves with many new constraints : Poison traffic (DDOS), lack of adaptability of firewall to quickly identify bandwidth optimized-probing-tools.
  These new factors complicate more and more the correct usage of bandwidth metric.

Conclusion

Conclusion :

- Internet bandwidth measurement literature is quite developed and advanced.

  However the Internet evolution requests the most-used bandwidth metrics to be constantly readapted and tested.

- Returned shared experience from operators and Telco stakeholders could improve the development of these tools and the quality of service of the end-users.

- Developing a community who will at least establish and share a common knowledge around the use of these tools and the optimization of the Internet traffic, could be very benefic.

Thank you for sharing your time and attention,

RED-SEA Communications Services

Thanks

- I would like to thank the following supports who contribute to this subject :
  - Philip Smith
  - Bramwel Wakachunga - CCIE# 38370 (R&S)
  - Moustapha Mohamed Ismael
  - Khaled Naguib

Contact :

- abas_ikbal@hotmail.com or ikbalabas@gmail.com
- http://www.linkedin.com/in/ikbalabas
- http://networktalk.wordpress.com/

RED-SEA Communications Services