Internet Engineering Task Force
Internet-Draft
Obsoletes: 5136 (if approved)
Intended status: Informational
Expires: April 19, 2011

X. Cui
Huawei

October 16, 2010

Defining Network Capacity
draft-cui-ippm-rfc5136bis-00

Abstract

This document defines the metric of network capacity, including link capacity aspect, router capacity aspect and path capacity aspect. RFC5136 has defined link capacity and path capacity, where the router impact is implicitly considered in link capacity. However, in this document, router capacity is considerred as a separate factor of path capacity, no longer a factor of link capacity. This document explicitly describes the router capacity and its impact to network capacity, e.g. how to evaluate path capacity.

This document is derived from RFC5136 and obsoletes RFC5136.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 19, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of

publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction

The IPPM working group has defined a framework for IP Performance Metrics [RFC2330] and a set of IP Performance Metrics, such as One-way Delay Metric [RFC2679], Packet Delay Variation Metric [RFC3393] and Network Capacity Metric [RFC5136].

Network capacity, which is defined in [RFC5136], is one of the most important IP Performance Metrics in internet. In [RFC5136], network capacity consists of link capacity, path capacity, link usage, link utilization, available link capacity and available path capacity. [RFC5136] also introduces the definitions, measurement and calculation methods and some important formulas.

As stated in [RFC5136], "measuring the capacity of a link or network path is a task that sounds simple, but in reality can be quite complex". There are so many factors and so complicated coupling (between these factors) that the factor of router capacity is not explicitly stated in [RFC5136]. Router is an important element of internet and it is also an essential component of path. In [RFC5136] router impact is implicitly considered in link capacity, but it should be considered in path and path capacity instead, because router is a part of path while not a part of link.

This memo explicitly presents that the router factor should be considered in path, path capacity and related metrics (e.g. available path capacity). For the integrity of network capacity metrics, this memo additionally defines router capacity, router usage, router utilization and available router capacity.

This memo is the latest development based on [RFC5136] and draws heavily from it.

The remainder of this memo is structured as follows.

Section 2.1 contains component definitions and explanations (node, host, router, link, path, etc.)

Section 2.2 contains nominal physical capacity and explanations of link and router.

Section 2.3 give IP-layer capacity definitions and explanstions. It is structured in 11 subsections:

- IP-layer Bits (section 2.3.1)
- IP-type-P Link Capacity (section 2.3.2)
- IP-type-P Link Usage (section 2.3.3)
- IP-type-P Link Utilization (section 2.3.4)

- IP-type-P Available Link Capacity (section 2.3.5)
- IP-type-P Router Capacity (section 2.3.6)
- IP-type-P Router Usage (section 2.3.7)
- IP-type-P Router Utilization (section 2.3.8)
- IP-type-P Available Router Capacity (section 2.3.9)
- IP-type-P Path Capacity (section 2.3.10)
- IP-type-P Available Path Capacity (section 2.3.11)

Section 3 describes changes from [RFC5136]. Section 4 gives some complementary discussion. Section 5 gives discussion conclusion.

1.1. Overview of Capacity

Any physical medium requires that information be encoded and, depending on the medium, there are various schemes to convert information into a sequence of signals that are transmitted physically from one location to another.

While on some media, the maximum frequency of these signals can be thought of as "capacity", on other media, the signal transmission frequency and the information capacity of the medium (channel) may be quite different. For example, a satellite channel may have a carrier frequency of a few gigahertz, but an information-carrying capacity of only a few hundred kilobits per second. Often similar or identical terms are used to refer to these different applications of capacity, adding to the ambiguity and confusion, and the lack of a unified nomenclature makes it difficult to properly build, test, and use various techniques and tools.

We are interested in information-carrying capacity, but even this is not straightforward. Each of the layers, depending on the medium, adds overhead to the task of carrying information. The wired Ethernet uses Manchester coding or 4/5 coding, which cuts down considerably on the "theoretical" capacity. Similarly, RF (radio frequency) communications will often add redundancy to the coding scheme to implement forward error correction because the physical medium (air) is lossy. This can further decrease the information capacity.

In addition to coding schemes, usually the physical layer and the link layer add framing bits for multiplexing and control purposes. For example, on SONET there is physical-layer framing and typically also some layer-2 framing such as High-Level Data Link Control (HDLC), PPP, or ATM.

Aside from questions of coding efficiency, there are issues of how access to the channel is controlled, which also may affect the

capacity. For example, a multiple-access medium with collision detection, avoidance, and recovery mechanisms has a varying capacity from the point of view of the users. This varying capacity depends upon the total number of users contending for the medium, how busy the users are, and bounds resulting from the mechanisms themselves. RF channels may also vary in capacity, depending on range, environmental conditions, mobility, shadowing, etc.

The important points to derive from this discussion are these: First, capacity is only meaningful when defined relative to a given protocol layer in the network. It is meaningless to speak of "link" capacity without qualifying exactly what is meant. Second, capacity is not necessarily fixed, and consequently, a single measure of capacity at any layer may in fact provide a skewed picture (either optimistic or pessimistic) of what is actually available.

1.2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Definitions

In this section, we specify component definitions and capacity definitions.

2.1. Component Definitions

In this section, we specify component definitions for network. We define "node", "Non-IP-Node", "host", "router", "link" and "path" clearly in this section, then we define capacity of network in next section.

2.1.1. Node

IPv6 Specification [RFC2460] defines node is a device that implements IPv6. Framework for IP Performance Metrics [RFC2330] defines host is a computer capable of communicating using the Internet protocols; includes "routers". The notion of host from [RFC2330] is equal to the notion of node from RFC2460. In this document, a node is a computer that implements IP protocol.

Note in this document any node without specicial statement is an IP node.

2.1.2. Non-IP-Node

In this document, a Non-IP-Node is a device that can transmit, receive or forward bit flow, but doesn't implement IP protocol. The examples of Non-IP-Node are ethernet switch and hub.

Note the Non-IP-Node may be part of link and impact the link capacity, for example, consider an ethernet switch that can operate ports at different speeds.

2.1.3. Host

IPv6 Specification [RFC2460] defines a host as any node that is not a router. This document adopts this definition, and the notion of host in this document dosn't includes "routers".

2.1.4. Router

[RFC2460] defines a router is a node that forwards IP packets not explicitly addressed to itself. This document adopts this definition.

2.1.5. Link

[RFC2460] defines link is a communication facility or medium over which nodes can communicate at the link layer, i.e., the layer immediately below IPv6. Examples are Ethernets (simple or bridged); PPP links; X.25, Frame Relay, or ATM networks; and internet (or higher) layer "tunnels", such as tunnels over IPv4 or IPv6 itself. [RFC2330] defines link is a single link-level connection between two (or more) hosts; includes leased lines, ethernets, frame relay clouds, etc. This document adopts the definition from [RFC2460].

Note that link is a bidirectional concept, link terminal and link-layer middle-box are included in link.

2.1.6. Path

As defined in [RFC2330], a path of length n is a sequence of the form (N0, L1, N1, ..., Ln, Nn), where n >= 0, each Ni is a node, each Li is a link between Ni-1 and Ni, each N1...Nn-1 is a router. A pair (Li, Ni) is termed a 'hop'. In an appropriate operational configuration, the links and routers in the path facilitate network-layer communication of packets from N0 to Nn.

Note that path is a unidirectional concept and a path of length one is not equal to the corresponding link. In this case, the Link (i.e., L1) is a part of the path, i.e., the sequence of (N0, L1, N1).

2.2. Definition: Nominal Physical Capacity

Nominal physical link capacity, NomCap(L), is the theoretical maximum amount of data that the link L can support. For example, an OC-3 link would be capable of 155.520 Mbit/s. We stress that this is a measurement at the physical layer and not the network IP layer, which we will define separately. While NomCap(L) is typically constant over time, there are links whose characteristics may allow otherwise, such as the dynamic activation of additional transponders for a satellite link.

Note when we define nominal physical capacity of link, link terminals are considered while the nodes (host or router) which are connected by the link are not gathered. This is because the link terminal (e.g., network interface card) is not integrant of computer, it is only an accessory of the computer. However, there may be some Non-IP-Node in the link, such as an ethereal switch. The physical link capacity is affected by the switch's ability to process and forward information bits for the given link.

The nominal physical link capacity is provided as a means to help distinguish between the commonly used link-layer capacities and the remaining definitions for IP-layer capacity. The nominal physical capacity provides an upper bound on link capaciy of both IP-layer and link-layer.

However, it is difficult to define the nominal physical capacity of a router. The routers are designed under many limitation, such as physical bound of CPU, memory and system bus. We usually use a pair of common principle to estimate a router: packet per second and bit per second. These two principles are coupled together, and in general, we almost can not correctly estimate a router by either of them. So we don't define nominal physical router capacity in this document.

2.3. Capacity at the IP Layer

There are many factors that can reduce the IP information carrying capacity of the link. However, the goal of this document is not to become an exhaustive list of such factors. Rather, we outline some of the major examples in the following section, thus providing food for thought to those implementing the algorithms or tools that attempt to measure capacity accurately.

The remaining definitions are all given in terms of "IP-layer bits" in order to distinguish these definitions from the nominal physical capacity of the link.

2.3.1. Definition: IP-layer Bits

IP-layer bits are defined as eight (8) times the number of octets in all IP packets received, from the first octet of the IP header to the last octet of the IP packet payload, inclusive.

IP-layer bits are recorded at the destination D beginning at time T and ending at a time T+I. Since the definitions are based on averages, the two time parameters, T and I, must accompany any report or estimate of the following values in order for them to remain meaningful. It is not required that the interval boundary points fall between packet arrivals at D. However, boundaries that fall within a packet will invalidate the packets on which they fall. Specifically, the data from the partial packet that is contained within the interval will not be counted. This may artificially bias some of the values, depending on the length of the interval and the amount of data received during that interval. We elaborate on what constitutes correctly received data in the next section.

2.3.1.1. Standard or Correctly Formed Packets

The definitions in this document specify that IP packets must be received correctly. The IPPM framework recommends a set of criteria for such standard-formed packets in Section 15 of [RFC2330]. However, it is inadequate for use with this document. Thus, we outline our own criteria below while pointing out any variations or similarities to [RFC2330].

First, data that is in error at layers below IP and cannot be properly passed to the IP layer must not be counted. For example, wireless media often have a considerably larger error rate than wired media, resulting in a reduction in IP link capacity. In accordance with the IPPM framework, packets that fail validation of the IP header must be discarded. Specifically, the requirements in [RFC1812], Section 5.2.2, on IP header validation must be checked, which includes a valid length, checksum, and version field.

The IPPM framework specifies further restrictions, requiring that any transport header be checked for correctness and that any packets with IP options be ignored. However, the definitions in this document are concerned with the traversal of IP-layer bits. As a result, data from the higher layers is not required to be valid or understood as that data is simply regarded as part of the IP packet. The same holds true for IP options. Valid IP fragments must also be counted as they expend the resources of a link even though assembly of the full packet may not be possible. The IPPM framework differs in this area, discarding IP fragments.

For a discussion of duplicates, please see Section 4.2.

In summary, any IP packet that can be properly processed must be included in these calculations.

2.3.1.2. Type P Packets

The definitions in this document refer to "Type P" packets to designate a particular type of flow or sets of flows. As defined in [RFC2330], Section 13, "Type P" is a placeholder for what may be an explicit specification of the packet flows referenced by the metric, or it may be a very loose specification encompassing aggregates. We use the "Type P" designation in these definitions in order to emphasize two things: First, that the value of the capacity measurement depends on the types of flows referenced in the definition. This is because networks may treat packets differently (in terms of queuing and scheduling) based on their markings and classification. Networks may also arbitrarily decide to flow-balance based on the packet type or flow type and thereby affect capacity measurements. Second, the measurement of capacity depends not only on the type of the reference packets, but also on the types of the packets in the "population" with which the flows of interest share the links in the path.

All of this indicates two different approaches to measuring: One is to measure capacity using a broad spectrum of packet types, suggesting that "Type P" should be set as generic as possible. The second is to focus narrowly on the types of flows of particular interest, which suggests that "Type P" should be very specific and narrowly defined. The first approach is likely to be of interest to providers, the second to application users.

As a practical matter, it should be noted that some providers may treat packets with certain characteristics differently than other packets. For example, access control lists, routing policies, and other mechanisms may be used to filter ICMP packets or forward packets with certain IP options through different routes. If a capacity-measurement tool uses these special packets and they are included in the "Type P" designation, the tool may not be measuring the path that it was intended to measure. Tool authors, as well as users, may wish to check this point with their service providers.

2.3.2. Definition: IP-type-P Link Capacity

We define the IP-layer link capacity, C(L,T,I), to be the maximum number of IP-layer bits that can be transmitted from the source S and correctly received by the destination D over the link L during the interval [T, T+I], divided by I. The "maximum" means that IP-type-P

link capacity is the capacity representation when the link is fully utilized (i.e., nominal physical link capacity is fully used.)

In theory, IP-layer link capacity may be calculated out from nominal physical link capacity. Usually, for any link whose link protocol is given, we can know well the encapsulation, overhead and overtail of the link layer protocol. In these cases, for Type P Packets, whose length is Lp, we can get IP-layer link capacity as:

C(L,T,I)
= [Lp/(Lh + Lp + Lt)] * [1 - BER(T, T+I)] * [1 - BDR(T, T+I)] * P(L)

In this formula,
- Lp denotes type P packet length (in IP layer),
- Lh denotes link layer protocol overhead length,
- Lt denotes link layer protocol overtail length,
- BER(T, T+I) denotes Block Error or Lost Rate during the interval [T, T+I],
- BDR(T, T+I) denotes Block Duplication Rate during the interval [T, T+I],
- P(L) denotes nominal physical link capacity of the given link.

Like nominal physical link capacity, IP-type-P link capacity is also a theoretical maximum value. But IP-type-P link capacity is not constant over time, becauese there are many types of link layer protocol and BER and BDR (e.g., BER/BDR of radio channel) may vary in different period.

As defined in section 2.1.5, link is the layer 2 connection between nodes, so the nodes which are connected by the link are not part of the given link. However, there may be some Non-IP-Node in the link, such as an ethereal switch. The IP-type-P link capacity is affected by the switch's ability to process and forward IP packets for the given link.

IP-type-P link capacity is affected by on-way Non-IP-Node but not affected by the nodes which are connected by the IP link. This means, the injecting node may affect how many packets are transpored between the source S and the destination D during the interval [T, T+I], and the incepting node may also affect how many packets are correctly received in the destination D, but these factors do not affect IP-type-P link capacity, because the capacity is the maximum value can be represented by the link.

IP-type-P link capacity is similar to IP-type-P link usage in some percent. The comparison is described in the next section.

2.3.3. Definition: IP-type-P Link Usage

The average usage of a link L, Used(L,T,I), is the actual number of IP-layer bits from any source, correctly received over link L during the interval [T, T+I], divided by I.

An important distinction between usage and capacity is that the capacity is a theoretical value (constant number) while the usage is a factually represented value (variable number). This is to say, Used(L,T,I) is not the maximum number, but rather, the actual average rate that IP bits are correctly received.

The information transmitted across the link can be generated by any source, including those sources that may not be directly attached to either side of the link. In addition, each information flow from these sources may share any number (from one to n) of links in the overall path between S and D.

2.3.4. Definition: IP-type-P Link Utilization

We express usage as a fraction of the overall IP-layer link capacity.

Util(L,T,I) = (Used(L,T,I) / C(L,T,I))

Thus, the utilization now represents the fraction of the capacity that is being used and is a value between zero (meaning nothing is used) and one (meaning the link is fully saturated). Multiplying the utilization by 100 yields the percent utilization of the link. By using the above, we can now define the available capacity over the link.

2.3.5. Definition: IP-type-P Available Link Capacity

We can now determine the amount of available capacity on a congested link by multiplying the IP-layer link capacity with the complement of the IP-layer link utilization. Thus, the IP-layer available link capacity becomes:

AvailCap(L,T,I) = C(L,T,I) * (1 - Util(L,T,I))

2.3.6. Definition: IP-type-P Router Capacity

As mentioned in section 2.2, we don't define nominal physical router capacity in this document, we only discuss the router IP capaicy for given packet type.

We define the IP-type-P router capacity, C(R,T,I), to be the maximum number of IP-layer bits (in the formation of type P packet) that can

be correctly transfered from the ingress interfaces to the egress interfaces during the interval [T, T+I], divided by I. Like nominal physical link capacity and IP-layer link capacity, IP-type-P router capacity is also a theoretical maximum value and typically constant over time.

Note this is only a nominal value or an approximation, because the accurate IP layer router capacity depends on many factors. Any router faces the common challenge, its capacity representation depends on its atchitect design, memory (e.g. queueing) management, interface deployment and other implementation issues. for example, a router can support 1000 interfaces and the capacity of 1T bps for IP type P at best. When we configure this router with 100 interfaces/ links, we can get this capacity value (i.e., 1T bps). But if the router is configured with only one ingress interface/link and one egress interface/link, maybe the maximum capacity value this router can present is less than 1T bps, because of its internal bus structure factors, even each link has the IP layer capacity of 2T bps.

On the other hand, as link capacity is node-independent, router capacity is not dependent on bits injection. The ingress link (i.e., the link which is attached to the ingress interface) may affect how many packets are injected to the router and the egress link (i.e., the link which is attached to the egress interface) may affect how many packets are forwarded to the next hop, but note the router capacity is the maximum number that we can get in all cases, for the given type P packets.

2.3.7. Definition: IP-type-P Router Usage

The average usage of a Router R, Used(R,T,I), is the actual number of IP-layer bits (in the formation of type P packet) correctly transfered from any ingress interface to the right engree interface during the interval [T, T+I], divided by I.

An important distinction between usage and capacity is that Used(R,T,I) is not the maximum number, but rather, the actual number of IP bits that are correctly transfered.

The information forwarded through the router can be generated by any source, including those sources that are not directly attached to the router. In addition, each information flow from these sources may share the router in their respective path.

2.3.8. Definition: IP-type-P Router Utilization

We express usage as a fraction of the overall IP-layer router capacity.

Util(R,T,I) = (Used(R,T,I) / C(R,T,I))

Thus, the utilization now represents the fraction of the capacity that is being used and is a value between zero (meaning nothing is used) and one (meaning the router is fully saturated). Multiplying the utilization by 100 yields the percent utilization of the router. By using the above, we can now define the capacity available throuth the router.

2.3.9. Definition: IP-type-P Available Router Capacity

We can now determine the amount of available capacity on a congested router by multiplying the IP-layer router capacity with the complement of the IP-layer router utilization. Thus, the IP-layer available router capacity becomes:

AvailCap(R,T,I) = C(R,T,I) * (1 - Util(R,T,I))

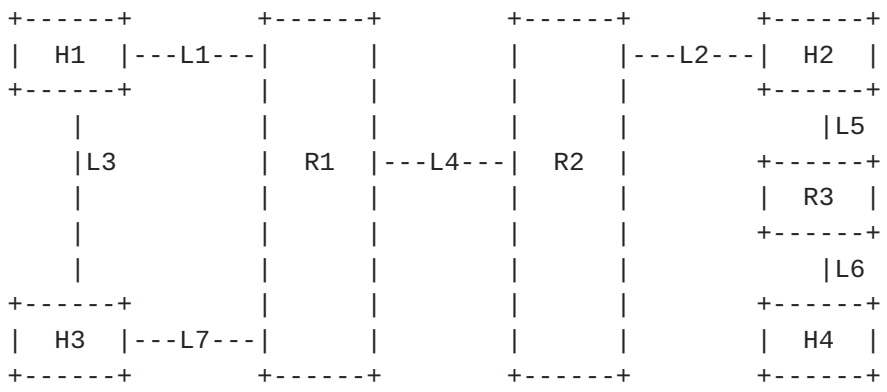
As mentioned in router capacity section, AvailCap(R,T,I) is only an approximation, because the accurate available router capacity depends on many internal factors.

2.3.10. Definition: IP-type-P Path Capacity

Using our definition for IP-layer link capacity and IP-layer router capacity, we can then extend these notions to an entire path.

We define the IP-type-P IP-layer path capacity, C(P,T,I), to be the maximum number of IP-layer bits (in the formation of type P packet) that can be correctly transfered from the source to the destination during the interval [T, T+I], divided by I. Like link capacity and router capacity, path capacity is also a theoretical number.

As mentioned earlier, the path of length n is a sequence of the form (N0, L1, N1, ..., Ln, Nn) and N1, N2, ..., Nn-1 are all routers and part of the path. But these links and routers may be part of one or multiple paths, for example, in the following scenario:



Host, Router, Link and Path

Figure 1

There are multiple paths in this network, such as:

Path P1 (from H1 to H2) -- (H1, L1, R1, L4, R2, L2, H2);
Path P2 (from H1 to H3) -- (H1, L3, H3);
Path P3 (from H1 to H3) -- (H1, L1, R1, L7, H3);
Path P4 (from H2 to H1) -- (H2, L2, R2, L4, R1, L1, H1);
Path P5 (from H2 to H3) -- (H2, L2, R2, L4, R1, L7, H3); and,
Path P6 (from H2 to H4) -- (H2, L5, R3, L6, H4).

Note this is not an exhaustive list. There are many other paths in this network, e.g., (R1, L4, R2).

In this scenario, the path (H1, L3, H4) and the path (H2, L5, R3, L6, H4) are exclusive path. The IP-layer capacity of an exclusive path may be calculated by:

$$C(P,T,I) = \min_{\{1..n\}} \{C(Ln,T,I), C(Rn,T,I)\}$$

we can also find that the link of L1, L2, L4 are all shared by multiple paths and the router of R1 and R2 are the same. Because of the capacity sharing, path capacity rather depends on the capacity contribution from the links and the routers than the IP-layer capacity of themselves. So for any given path whose link or router overlaps with other path, the IP-layer path capacity becomes more complex, it depends on not only the IP-layer capacity of the links and the routers but also the "competitive" traffic (also in formation of type P packet) of other paths, which have overlap segment with the given path. This means the capacity of non-exclusive path is a variable, is external situation dependent.

It is very difficult to calculate IP-type-P path capacity of non-exclusive path in general but we can get out the maximum number of path capacity from links and routers, to indicate the upper bound on path capacity.

The maximum number of IP-layer capacity of non-exclusive path may be calculated by:

Cmax(P,T,I) = min {1..n} {C(Ln,T,I), C(Rn,T,I)}

2.3.11. Definition: IP-type-P Available Path Capacity

Using our definition for IP-layer available link capacity and IP-layer available router capacity, we can then extend these notions to an entire path, such that the IP-layer available path capacity simply becomes that of the link and router with the smallest available capacity along that path.

AvailCap(P,T,I) = min {1..n} {AvailCap(Ln,T,I), AvailCap(Rn,T,I)}

Since measurements of available capacity are more volatile than that of link capacity, we stress the importance that both the time and interval be specified as their values have a great deal of influence on the results. In addition, a sequence of measurements may be beneficial in offsetting the volatility when attempting to characterize available capacity.

3. Changes from RFC5136

In general, this document clarifies some definitions (e.g., path) and expounds that the capacity metrics (e.g., IP-type-P link capacity) are theoretical number. In addition, usage metrics (e.g., IP-type-P link usage) are very different from capacity metrics because they are actual number represented in measurement cases.

3.1. Node Definition

Section 2.1 from [RFC5136] has been changed from:

We define nodes as hosts, routers, Ethernet switches, or any other device where the input and output links can have different characteristics.

to Section 2.1.1 of this memo:

Node is a computer that implements IP protocol.

Reason/summarization:

The reason for this modification is to follow the most idiomatic definition. Non-IP device is excluded in the notion of node.

3.2. Link Definition

Section 2.1 from [RFC5136] has been changed from:

A link is a connection between two of these network devices or nodes.

to Section 2.1.5 of this memo:

Link is a communication facility or medium over which nodes can communicate at the link layer, i.e., the layer immediately below IPv6. Examples are Ethernets (simple or bridged); PPP links; X.25, Frame Relay, or ATM networks; and internet (or higher) layer "tunnels", such as tunnels over IPv4 or IPv6 itself.

Reason/summarization:

The reason for this modification is to clarify the notion. The connection between layer1 or layer2 devices is not an absolute link, but only a segment of link.

3.3. Path Definition

Section 2.1 from [RFC5136] has been changed from:

We then define a path P of length n as a series of links (L1, L2, ..., Ln) connecting a sequence of nodes (N1, N2, ..., Nn+1). A source S and destination D reside at N1 and Nn+1, respectively.

to Section 2.1.6 of this memo:

A path of length n is a sequence of the form (N0, L1, N1, ..., Ln, Nn), where n >= 0, each Ni is a node, each Li is a link between Ni-1 and Ni, each N1...Nn-1 is a router. A pair (Li, Ni) is termed a 'hop'. In an appropriate operational configuration, the links and routers in the path facilitate network-layer communication of packets from N0 to Nn.

Reason/summarization:

The reason for this modification is to emphasize that routers in the path are essential component of the path.

3.4. Definition: Nominal Physical Capacity

Section 2.2 from [RFC5136] has been changed from "Definition: Nominal Physical Link Capacity" to "Definition: Nominal Physical Capacity". And some statement are added, including:

Note when we define nominal physical capacity of link, link terminals are considered while the nodes (host or router) which are connected by the link are not gathered. This is because the link terminal (e.g., network interface card) is not integrant of computer, it is only an accessory of the computer. However, there may be some non-IP-node in the link, such as the ethereal switch. The physical link capacity is affected by the switch's ability to process and forward information bits for the given link.

and,

However, it is difficult to define the nominal physical capacity of a router. The routers are designed under many limitation, such as physical bound of CPU, memory and system bus. We usually use a pair of common principle to estimate a router: packet per second and bit per second. These two principles are coupled together, and in general, we almost can not correctly estimate a router by either of them. So we don't define nominal physical router capacity in this document.

3.5. IP-type-P Link Capacity

Section 2.3.2 from [RFC5136] has been changed from:

We define the IP-layer link capacity, C(L,T,I), to be the maximum number of IP-layer bits that can be transmitted from the source S and correctly received by the destination D over the link L during the interval [T, T+I], divided by I.

As mentioned earlier, this definition is affected by many factors that may change over time. For example, a device's ability to process and forward IP packets for a particular link may have varying effect on capacity, depending on the amount or type of traffic being processed.

to Section 2.3.2 of this memo:

We define the IP-layer link capacity, C(L,T,I), to be the maximum number of IP-layer bits that can be transmitted from the source S and correctly received by the destination D over the link L during the interval [T, T+I], divided by I. The "maximum" means that IP-type-P link capacity is the capacity representation when the link is fully

utilized (i.e., nominal physical link capacity is fully used.)

In theory, IP-layer link capacity may be calculated out from nominal physical link capacity. Usually, for any link whose link protocol is given, we can know well the encapsulation, overhead and overtail of the link layer protocol. In these cases, for Type P Packets, whose length is Lp, we can get IP-layer link capacity as:

C(L,T,I)
= [Lp/(Lh + Lp + Lt)] * [1 - BER(T, T+I)] * [1 - BDR(T, T+I)] * P(L)

In this formula,
- Lp denotes type P packet length (in IP layer),
- Lh denotes link layer protocol overhead length,
- Lt denotes link layer protocol overtail length,
- BER(T, T+I) denotes Block Error or Lost Rate during the interval [T, T+I],
- BDR(T, T+I) denotes Block Duplication Rate during the interval [T, T+I],
- P(L) denotes nominal physical link capacity of the given link.

Like nominal physical link capacity, IP-type-P link capacity is also a theoretical maximum value. But IP-type-P link capacity is not constant over time, becauese there are many types of link layer protocol and BER and BDR (e.g., BER/BDR of radio channel) may vary in different period.

As defined in section 2.1.5, link is the layer 2 connection between nodes, so the nodes which are connected by the link are not part of the given link. However, there may be some Non-IP-Node in the link, such as the ethereal switch. The IP-type-P link capacity is affected by the switch's ability to process and forward IP packets for the given link.

IP-type-P link capacity is affected by on-way Non-IP-Node but not affected by the nodes which are connected by the IP link. This means, the injecting node may affect how many packets are transpored between the source S and the destination D during the interval [T, T+I], and the incepting node may also affect how many packets are correctly received in the destination D, but these factors do not affect IP-type-P link capacity, because the capacity is the maximum value can be represented by the link.

IP-type-P link capacity is similar to IP-type-P link usage in some percent. The comparison is described in the next section.

Reason/summarization:

This modification clarifies the definition and calculation of link capacity and explicitly indicates that node doesn't affect link capacity but the Non-IP-Node which is part of link does.

3.6. IP-type-P Router Capacity

Section 2.3.6 is newly added in this memo, as:

As mentioned in section 2.2, we don't define nominal physical router capacity in this document, we only discuss the router IP capaicy for given packet type.

We define the IP-type-P router capacity, C(R,T,I), to be the maximum number of IP-layer bits (in the formation of type P packet) that can be correctly transfered from the ingress interfaces to the egress interfaces during the interval [T, T+I], divided by I. Like nominal physical link capacity and IP-layer link capacity, IP-type-P router capacity is also a theoretical maximum value and typically constant over time

Note this is only a nominal value or an approximation, because the accurate IP layer router capacity depends on many factors. Any router faces the common challenge, its capacity representation depends on its atchitect design, memory (e.g. queueing) management, interface deployment and other implementation issues. for example, a router can support 1000 interfaces and the capacity of 1T bps for IP type P at best. When we configure this router with 100 interfaces/ links, we can get this capacity value (i.e., 1T bps). But if the router is configured with only one ingress interface/link and one egress interface/link, maybe the maximum capacity value this router can present is less than 1T bps, because of its internal bus structure factors, even each link has the IP layer capacity of 2T bps.

On the other hand, as link capacity is node-independent, router capacity is not dependent on bits injection. The ingress link (i.e., the link which is attached to the ingress interface) may affect how many packets are injected to the router and the egress link (i.e., the link which is attached to the egress interface) may affect how many packets are forwarded to the next hop, but note the router capacity is the maximum number that we can get in all cases, for the given type P packets.

Reason/summarization:

This modification defines IP-layer router capacity aspect from network capacity.

3.7. IP-type-P Router Usage

Section 2.3.7 is newly added in this memo, as:

The average usage of a Router R, Used(R,T,I), is the actual number of IP-layer bits (in the formation of type P packet) correctly transfered from any ingress interface to the right engree interface during the interval [T, T+I], divided by I.

An important distinction between usage and capacity is that Used(R,T,I) is not the maximum number, but rather, the actual number of IP bits that are correctly transfered.

The information forwarded through the router can be generated by any source, including those sources that are not directly attached to the router. In addition, each information flow from these sources may share the router in their respective path.

Reason/summarization:

This modification defines IP-layer router usage aspect from network capacity.

3.8. IP-type-P Router Utilization

Section 2.3.8 is newly added in this memo, as:

We express usage as a fraction of the overall IP-layer router capacity.

Util(R,T,I) = (Used(R,T,I) / C(R,T,I))

Thus, the utilization now represents the fraction of the capacity that is being used and is a value between zero (meaning nothing is used) and one (meaning the router is fully saturated). Multiplying the utilization by 100 yields the percent utilization of the router. By using the above, we can now define the capacity available throuth the router.

Reason/summarization:

This modification defines IP-layer router utilization aspect from network capacity.

3.9. IP-type-P Available Router Capacity

Section 2.3.9 is newly added in this memo, as:

We can now determine the amount of available capacity on a congested router by multiplying the IP-layer router capacity with the complement of the IP-layer router utilization. Thus, the IP-layer available router capacity becomes:

AvailCap(R,T,I) = C(R,T,I) * (1 - Util(R,T,I))

As mentioned in router capacity section, AvailCap(R,T,I) is only an approximation, because the accurate available router capacity depends on many internal factors.

Reason/summarization:

This modification defines IP-layer available router capacity aspect from network capacity.

3.10. IP-type-P Path Capacity

Section 2.3.3 from [RFC5136] has been changed from:

Using our definition for IP-layer link capacity, we can then extend this notion to an entire path, such that the IP-layer path capacity simply becomes that of the link with the smallest capacity along that path.

C(P,T,I) = min {1..n} {C(Ln,T,I)}

The previous definitions specify the number of IP-layer bits that can be transmitted across a link or path should the resource be free of any congestion. It represents the full capacity available for traffic between the source and destination. Determining how much capacity is available for use on a congested link is potentially much more useful. However, in order to define the available capacity, we must first specify how much is being used.

to Section 2.3.10 of this memo:

We define the IP-type-P IP-layer path capacity, C(P,T,I), to be the maximum number of IP-layer bits (in the formation of type P packet) that can be correctly transfered from the source to the destination during the interval [T, T+I], divided by I. Like link capacity and router capacity, path capacity is also a theoretical number.

The IP-layer capacity of an exclusive path may be calculated by:

C(P,T,I) = min {1..n} {C(Ln,T,I), C(Rn,T,I)}

It is very difficult to calculate IP-type-P path capacity of non-

exclusive path in general but we can get out the maximum number of path capacity from links and routers, to indicate the upper bound on path capacity.

The maximum number of IP-layer capacity of non-exclusive path may be calculated by:

Cmax(P,T,I) = min {1..n} {C(Ln,T,I), C(Rn,T,I)}

Reason/summarization:

This modification clarifies how to correctly evaluate path capacity. Router capacity is considered for path capacity.

3.11. IP-type-P Available Path Capacity

Section 2.3.7 from [RFC5136] has been changed from:

Using our definition for IP-layer available link capacity, we can then extend this notion to an entire path, such that the IP-layer available path capacity simply becomes that of the link with the smallest available capacity along that path.

AvailCap(P,T,I) = min {1..n} {AvailCap(Ln,T,I)}

to Section 2.3.11 of this memo:

Using our definition for IP-layer available link capacity and IP-layer available router capacity, we can then extend these notions to an entire path, such that the IP-layer available path capacity simply becomes that of the link and router with the smallest available capacity along that path.

AvailCap(P,T,I) = min {1..n} {AvailCap(Ln,T,I), AvailCap(Rn,T,I)}

Reason/summarization:

This modification clarifies how to correctly evaluate available path capacity. Available router capacity is considered for available path capacity.

4. Discussion

4.1. Time and Sampling

We must emphasize the importance of time in the basic definitions of these quantities. We know that traffic on the Internet is highly

variable across all time scales. This argues that the time and length of measurements are critical variables in reporting available capacity measurements and must be reported when using these definitions.

The closer to "instantaneous" a metric is, the more important it is to have a plan for sampling the metric over a time period that is sufficiently large. By doing so, we allow valid statistical inferences to be made from the measurements. An obvious pitfall here is sampling in a way that causes bias. For example, a situation where the sampling frequency is a multiple of the frequency of an underlying condition.

4.2. Hardware Duplicates

We briefly consider the effects of paths where hardware duplication of packets may occur. In such an environment, a node in the network path may duplicate packets, and the destination may receive multiple, identical copies of these packets. Both the original packet and the duplicates can be properly received and appear to be originating from the sender. Thus, in the most generic form, duplicate IP packets are counted in these definitions. However, hardware duplication can affect these definitions depending on the use of "Type P" to add additional restrictions on packet reception. For instance, a restriction only to count uniquely-sent packets may be more useful to users concerned with capacity for meaningful data. In contrast, the more general, unrestricted metric may be suitable for a user who is concerned with raw capacity. Thus, it is up to the user to properly scope and interpret results in situations where hardware duplicates may be prevalent.

4.3. Other Potential Factors

IP encapsulation does not affect the definitions as all IP header and payload bits must be counted regardless of content. However, IP packets of different sizes can lead to a variation in the amount of overhead needed at the lower layers to transmit the data, thus altering the overall IP link-layer capacity.

Should the link happen to employ a compression scheme such as RObust Header Compression (ROHC) [RFC3095] or V.44 [V44], some of the original bits are not transmitted across the link. However, the inflated (not compressed) number of IP-layer bits should be counted.

4.4. Common Terminology in Literature

Certain terms are often used to characterize specific aspects of the presented definitions. The link with the smallest capacity is

commonly referred to as the "narrow link" of a path. Also, the link with the smallest available capacity is often referred to as the "tight link" within a path. So, while a given link may have a very large capacity, the overall congestion level on the link makes it the likely bottleneck of a connection. Conversely, a link that has the smallest capacity may not be the bottleneck should it be lightly loaded in relation to the rest of the path.

Also, literature often overloads the term "bandwidth" to refer to what we have described as capacity in this document. For example, when inquiring about the bandwidth of a 802.11b link, a network engineer will likely answer with 11 Mbit/s. However, an electrical engineer may answer with 25 MHz, and an end user may tell you that his observed bandwidth is 8 Mbit/s. In contrast, the term "capacity" is not quite as overloaded and is an appropriate term that better reflects what is actually being measured.

4.5. Comparison to Bulk Transfer Capacity (BTC)

Bulk Transfer Capacity (BTC) [RFC3148] provides a distinct perspective on path capacity that differs from the definitions in this document in several fundamental ways. First, BTC operates at the transport layer, gauging the amount of capacity available to an application that wishes to send data. Only unique data is measured, meaning header and retransmitted data are not included in the calculation. In contrast, IP-layer link capacity includes the IP header and is indifferent to the uniqueness of the data contained within the packet payload. (Hardware duplication of packets is an anomaly addressed in a previous section.) Second, BTC utilizes a single congestion-aware transport connection, such as TCP, to obtain measurements. As a result, BTC implementations react strongly to different path characteristics, topologies, and distances. Since these differences can affect the control loop (propagation delays, segment reordering, etc.), the reaction is further dependent on the algorithms being employed for the measurements. For example, consider a single event where a link suffers a large duration of bit errors. The event could cause IP-layer packets to be discarded, and the lost packets would reduce the IP-layer link capacity. However, the same event and subsequent losses would trigger loss recovery for a BTC measurement resulting in the retransmission of data and a potentially reduced sending rate. Thus, a measurement of BTC does not correspond to any of the definitions in this document. Both techniques are useful in exploring the characteristics of a network path, but from different perspectives.

5. Conclusion

In this document, we have defined a set of quantities related to the capacity of links, routers and paths in an IP network. In these definitions, we have tried to be as clear as possible and take into account various characteristics that links, routers and paths can have. The goal of these definitions is to enable researchers who propose capacity metrics to relate those metrics to these definitions and to evaluate those metrics with respect to how well they approximate these quantities.

In addition, we have pointed out some key auxiliary parameters and opened a discussion of issues related to valid inferences from available capacity metrics.

6. Security Considerations

This document specifies definitions regarding IP traffic traveling between a source and destination in an IP network. These definitions do not raise any security issues and do not have a direct impact on the networking protocol suite.

Tools that attempt to implement these definitions may introduce security issues specific to each implementation. Both active and passive measurement techniques can be abused, impacting the security, privacy, and performance of the network. Any measurement techniques based upon these definitions must include a discussion of the techniques needed to protect the network on which the measurements are being performed.

7. IANA Considerations

This document has no actions for IANA.

8. Acknowledgments

The author would especially like to acknowledge Phil Chimento and Joseph Ishac for their great contribution on the item of network capacity. The author would like to acknowledge Mark Allman, Patrik Arlos, Matt Mathis, Al Morton, Stanislav Shalunov, and Matt Zekauskas for their contribution on [RFC5136], which is the basis of this document.

The author would also like to acknowledge Brian E Carpenter, Adrian Farrel, Spencer Dawkins, David Harrington and Barry Leiba for their

review and discussion in the early stage of this document.

9. References

9.1. Normative References

[RFC1812] Baker, F., "Requirements for IP Version 4 Routers", RFC 1812, June 1995.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, "Framework for IP Performance Metrics", RFC 2330, May 1998.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

9.2. Informative References

[RFC2679] Almes, G., Kalidindi, S., and M. Zekauskas, "A One-way Delay Metric for IPPM", RFC 2679, September 1999.

[RFC3095] Bormann, C., Burmeister, C., Degermark, M., Fukushima, H., Hannu, H., Jonsson, L-E., Hakenberg, R., Koren, T., Le, K., Liu, Z., Martensson, A., Miyazaki, A., Svanbro, K., Wiebke, T., Yoshimura, T., and H. Zheng, "RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed", RFC 3095, July 2001.

[RFC3148] Mathis, M. and M. Allman, "A Framework for Defining Empirical Bulk Transfer Capacity Metrics", RFC 3148, July 2001.

[RFC3393] Demichelis, C. and P. Chimento, "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)", RFC 3393, November 2002.

[RFC5136] Chimento, P. and J. Ishac, "Defining Network Capacity", RFC 5136, February 2008.

[V44] ITU Telecommunication Standardization Sector (ITU-T) Recommendation V.44, "Data Compression Procedures", November 2000.

Author's Address

Xiangsong Cui (editor)
Huawei
KuiKe Bld., No.9 Xinxi Rd., Shang-Di Information Industry Base
Beijing, 100085
P.R. China

Phone:
Email: Xiangsong.Cui@huawei.com