

White Paper on Transactions and Valuation Associated with Inter-Carrier Routing of Internet Protocol Traffic

- or -

BGP for Bankers

Version 0.2bis
August, 2000
Bill Woodcock
Packet Clearing House

Preface

This document is an artifact of its time, and "Internet time" moves quickly. The telecommunications investment bubble that began in the late 1990s and ended abruptly in 2001 is, as I write this, only five years in the past, but already seems a very different world. One which in retrospect seems heady, naïve, crass, and indulgent. But things are cyclic, and it seems appropriate to publish this now, with the benefit of some retrospection, at a time when it can be seen free of the hyperbolic enthusiasms of an investment boom.

I wrote this paper through the summer of 2000, as a consequence of several years of badgering from investment bankers, all eager to capitalize upon what they saw as a fast-moving stream of monetizable commercial meta-data. I was quite skeptical of this notion, but simultaneously curious enough about the workings of the world to investigate it, hoping to discover whether my knee-jerk skepticism was the rational response of a sober person confronted by foolishness, or the irrational defensiveness of an engineer reacting to incursion by those who would re-shape his working environment in new and less-pleasant, but inevitable and logical, ways.

Late in the summer of 2000, after informally releasing several partial drafts of this paper on PCH's web site, I found myself scheduled to speak on other topics at a meeting of Reseau IP Europeen, or RIPE, the European regional Internet resource registry and operators' forum, in Prague. Although I'd found myself engaged in hallway discussions for most of the lunch-hour, I had a few scant minutes to eat before taking the podium, and so when I entered the conference center's dining hall purposely looked for a table at which I knew no one, and took a seat at the far end of just such a sparsely-populated table. After a few minutes, one of the three young men at the other end of the table, oddly out-of-place in their neckties, leaned over, having seen the PCH logo on my shirt, and asked whether, by chance, I knew Bill Woodcock. I admitted to, in fact, being Bill Woodcock, whereupon they revealed that they were members of the now-infamous Enron's just-created "bandwidth trading unit" and that the drafts of my paper were the required reading upon which they and their cohort had been depending. Although this was prior to some of Enron's greatest crimes, and many still considered them to be some form of legitimate, if unusually rapacious, business concern, I abandoned further work on this paper in disgust, and withdrew it from PCH's web site. I think that both community morals and legislative action in the years since have condemned the forms of use to which some few individuals and organizations like Enron put economic information

It should be understood by the reader that my background at the time I wrote this was as an engineer, an entrepreneur, and a businessperson, but notably a neophyte amateur economist rather than a trained academic. And I had the fondness for barely-grasped polysyllabic words which adheres to those still in their twenties, for which I beg your indulgence now, in the interest of historical verisimilitude.

Throughout this document, I've distinguished 2006 commentary from 2000 original text by the use of italics.

– Bill Woodcock, November 2006, Berkeley California

The nature of the transactions by which Internet traffic is exchanged between carriers has of late become a matter of some interest to the financial community. Whereas until this point such transactions have been relatively unexamined and occurred directly between two principal parties within the industry either as a future or less commonly a simple good, it is now the case that third parties are showing an interest in basing derivative financial instruments upon these transactions.

Inasmuch as a commodity market is necessary to support derivative transactions and a commodity market is not possible without a great degree of financial transparency, a good first step in that direction is to establish a clear common understanding of the services being exchanged and the nature of the transactions by which this occurs. This paper attempts to describe the means by which Internet traffic is exchanged and catalog the various transactions by which the costs of doing so are borne.

Overview of the Field and Terminology

Internet traffic occurs as bi-directional conversations, typically but not necessarily between a pair of endpoints. In a typical conversation, such as that required to view a web page, the "client" computer will send a "packet" to the "server" on which the desired web page is housed, seeking to initiate a connection. The server will acknowledge the packet with a packet of its own, indicating a readiness to communicate. A "session" or "flow" having been established, the client will generate another packet requesting the web page itself. The server will respond with as many packets as necessary to convey the requested information. The client will acknowledge the transfer, and one more pair of packets will be exchanged in agreeing to tear down the session. During the course of this conversation, the client will have sent a relatively small number of very small packets to the server, while the server will have sent a larger number of larger packets back to the client. The vast majority of flows across the Internet are characterized by a volumetric asymmetry of this sort, although the asymmetry may be such that the client transmits more than the server, or the server originates a data transfer, or any of a number of other variations.

These conversations are nearly always between machines owned by organizations which would be characterized as end-user customers relative to the Internet infrastructure. That is, very few conversations are originated by or directed to computers owned by Internet carriers, or which comprise the infrastructure of the Internet itself.

The topology of the Internet is a very loose hierarchy, rooted at many "Internet exchange points" around the world. Large Internet carriers, originally termed "Network Service Providers" or NSPs under the United States government's National Information Infrastructure plan of the early 1990s but now more commonly called "backbone providers", connect to many geographically-distributed Internet exchange points, and are able to carry traffic long distances on behalf of their customers. Somewhat smaller Internet carriers, originally described by the phrase "Internet Service Providers" or ISPs, which has since fallen into looser usage, but now more specifically described as "regional providers", typically connect primarily to multiple backbone providers, and perhaps supplement this with connections to a small number of Internet exchange points within their service area. Large corporate end-users, and carriers who serve individual consumers and were originally called "Internet Access Providers" or IAPs, are typically tied to the Internet through a single connection to an ISP or NSP, and do not generally participate in any Internet exchange points.

A conversation between two computers owned by different end-user organizations will typically traverse the client's ISP, the client's ISP's NSP, a Internet exchange point, the server's ISP's NSP, and the server's ISP, before reaching the network which actually houses the server. Note that there's a symmetry displayed here, in the traversal of customer, ISP, and NSP toward the Internet exchange point, and NSP, ISP, and customer away from the Internet exchange point. Obviously

replies from the server to the client take roughly the reverse path, although it's not likely to consist of exactly the same specific entities in each role in the hierarchy. That is, the roles within the hierarchy are symmetric, but individual queries and responses within a conversation typically do not take a symmetric path through the same specific carriers. This property will be discussed at length later.

There are a number of terms used to describe directionality of traffic propagation and direction relative to the top of the hierarchy. From any point in the hierarchy, any entity on a connected path between the point under discussion and the root of the hierarchy (the Internet exchange point) is referred to as "upstream" whereas anything distal to the point under discussion, but on a connected path, is referred to as "downstream." The host from which a packet is transmitted is referred to the "originator," "sender" or "source," while the addressee is referred to as the "destination," "recipient" or "sink."

A "query" is an unsolicited packet which begins a conversation while a "response" is the reply which it solicits. Together, these may form a whole conversation or flow, or a small conversation within a larger series of related ones. This is more commonly referred to as a "transaction," but since I intend to discuss financial matters within this paper, I'll avoid this usage in order to prevent semantic overloading and distinguish between protocol conversations and financial transactions.

Note that all of these terms with the exception of "client" and "server" are agnostic with respect to the end-user's idea of the direction of the conversation. They're applied to the direction of individual packets, rather than the direction of the causal intentionality of the conversation as a whole. On the other hand, although both endpoints source and sink packets in a conversation, the "client" is typically but not always that which causes a new connection to be originated, and a "server" is typically passive and responds primarily to externally-originated stimulus.

Each of the terms mentioned so far apply generally to packets and conversations and the data which is carried in them. One specific subset of data is the routing protocol which establishes the paths which specific data flows will take through the hierarchy of connected facilities. This protocol is called BGP, or Border Gateway Protocol, since it's used primarily at the borders of carriers' networks, that is, at the gateways between a carrier's network and those other networks which are tangent to it. Unlike many environments, the Internet relies upon in-band transmission of routing data. The routing data is data like any other data, and by design nearly always takes the same paths, even though it's the data which is being used to establish those paths in the first place. This is a somewhat recursive concept, but does solve problems as well as creating them. The principal merit of this system is that it avoids the introduction of new false-positive routing information. That is, if a link must exist and be traversable by data in order for routing data to cross it and announce its existence to the other side, and all this must occur prior to the link's use for general non-routing data, then a non-existent link cannot be introduced into the data-transmission topology. If routing data were carried on a separate out-of-band network, a synchronization method would have to be used to insure that this problem did not occur. A corresponding drawback of in-band transmission of routing data is that it encourages the holding of false-positive information which was previously valid but has become false. That is, when a link becomes inoperative, the routing protocol cannot discover this critical information independently, except through the very slow process of observing that several sequential "keepalive" or "tickle" status packets which it expects to receive from a neighbor router have gone missing. Thus most implementations prefer to rely upon information which the equipment at each end of the link injects into the routing protocol based upon its knowledge of the physical state of the link.

Thus there are two principal types of information carried within routing protocols: announcements and withdrawals. Announcements are series of packets which contain lists of new destinations which are reachable via a link, typically the link which the announcement packets are traversing.

Withdrawals are the converse, lists of destinations which may no longer be reached via a link. Announcements, which are also often called "advertisements," can best be understood as solicitations for the transmission of data. They are an offer made by an entity to another entity to accept packets on behalf of some destination. Although they contain no information about price or guarantee of performance, they do state the path by which they intend to convey the data if they receive it. Each unique entity capable of routing traffic on the Internet is called an "autonomous system" and is identified by a unique integer between 1 and 65535. <<<actually highest non-reserved>>> The path information contained within a routing announcement is in the form of a list of autonomous system numbers, or ASNs. This is called the "AS path." The form of an announcement is a list of "prefixes" which are the names of destinations for Internet traffic, and the path by which traffic would travel to that prefix if given to the announcing party for retransmission. To give a concrete example, the following is a hypothetical routing announcement:

Prefix	AS Path
10.0.1.0/24	1 20 18 5
10.0.2.0/24	1 20 18 12
192.168.0.0/16	1 25

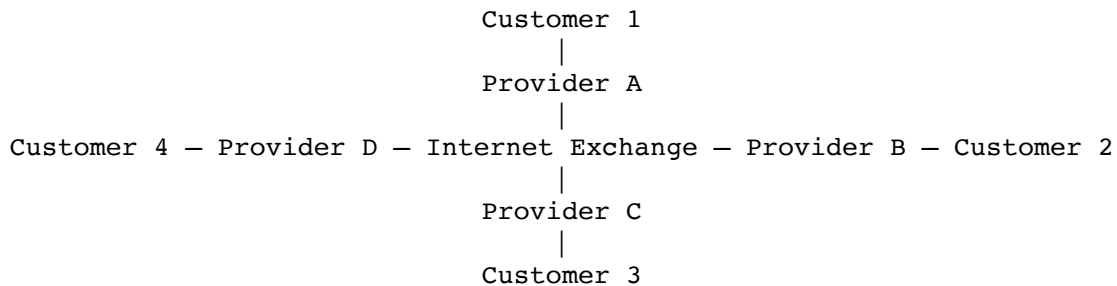
Just by examining this simple table, an experienced Internet provider can derive quite a bit of knowledge about the relationships between the autonomous systems mentioned in the three AS paths. First, since each of the three paths begins with the ASN "1," that the announcement is coming across a link which connects the point from which we're observing to the organization which is identified by the ASN 1. So we (the hypothetical "we" who are receiving this announcement) must have some sort of adjacency relationship with AS 1. Internet exchange points are not visible in AS paths, so we can't derive positional information within the hierarchy just from these three paths, but experience would allow a provider to deduce that information from previous knowledge about the ASes involved.

Next, we know that if we were to create, or receive for retransmission, a packet addressed to a host within the prefix 10.0.1.0/24 (for instance, at the IP address 10.0.1.25), we understand that AS 1 would forward that packet to AS 20, based upon AS 20's advertisement to AS 1 that AS 20 has reachability of 10.0.1.0/24 via AS 18. AS 5 is the originating AS for the 10.0.1.0/24 prefix, so we know that the traffic will terminate at a machine which is within AS 5's network. AS paths are built and depend upon this sort of transitive trust, and in the business world, this trust between adjacent ASes is encoded in contractual relationships between the corporations which they represent. One's trust that if one delivers a packet addressed to 10.0.1.25 to AS 1 that it won't be malevolently misdirected is based upon one's knowledge that AS 1 has a contract with AS 20 which mutually assures them of the legitimacy of announcements between each other, that AS 20 has a similar contract with AS 18, and that AS 18 has a similar contract with AS 5. By the transitive property, one thus has indirect but no less valid assurance that AS 5 is the proper destination for a packet to 10.0.1.25.

As of 2006, there have been competing efforts for several years to establish a standard for cryptographic signature of either the AS which originates a prefix, or an entire AS path and each of the adjacencies within it. Thus far, heavyweight proposals have met with little enthusiasm from the operations community which they purport to serve, but the ever-increasing problem of spam originating through falsified prefix originations continues to increase the cost of failure to arrive at a solution of some sort to this problem, whether it be cryptographic or procedural.

The advertisement of a willingness to receive traffic on behalf of a destination and retransmit it to that destination is an advertisement of value being offered, and this is what's at the heart of the contracts between ASes. In the case of customer-provider relationships, the contract specifies an exchange which is primarily one of transitive routing announcements from provider to customer in

exchange for money from customer to provider. In addition, the customer is also almost certainly announcing some or all of their own local (not transitive from another AS) routes to the provider, and both customer and provider exchange assurances of a variety of responsible technical and business practices surrounding the essential transaction. In the case of peer-to-peer relationships, the contract specifies a bi-directional exchange of non-transitive routing information between two Internet service providers. Thus the two types of relationship are called "transit" and "peering." It's very important to note that both of these terms have been very badly semantically overloaded. The word "peering" is used in a routing-protocol level sense to refer to any exchange of routing information between adjacent parties ("peers") as opposed to specifically one of local routes between a pair of Internet providers. Both terms are also very broadly used in both qualified and unqualified forms by marketing personnel who attach no specific meaning to either one. I will attempt to use both terms in their strict sense within this document: peering referring to a relationship between two Internet providers in which both exchange some or all of their customer routes sans monetary exchange, and transit referring to a relationship between a provider and a customer in which the provider advertises a full set of transitive routes to a customer in exchange for money. Stated in another way, if someone sells transit, they're offering their customer an advertisement of reachability for the whole Internet, that is, they're soliciting traffic bound for any destination on the Internet. If someone forms a peering relationship with another party, each advertises reachability of only those networks which are downstream from them, that is, they solicit traffic on behalf of only their customers. The distinction between transit and peering is the key to understanding the economics of Internet traffic. None of the details and refinements which we will discuss in this paper will be intelligible or useful without a fundamental understanding of this basic concept upon which they're founded.



In this simplistic example, Customer 1 would have a transit agreement with Provider A, by which it would receive routes to customers 2,3, and 4. Customer 2 would have a transit agreement with Provider B, by which it would receive routes to customers 1, 3, and 4. Customer 3 would have a transit agreement with Provider C, by which it would receive routes to customers 1, 2, and 4. Customer 4 would have a transit agreement with Provider D by which it would receive routes to customers 1, 2, and 3.

Provider A would have a peering relationship with Provider B (across the Internet exchange point, which is why Internet exchange points are sometimes referred to as "peering points") by which it would receive a route to Customer 2 only, and advertise a route to Customer 1 only. Provider A would have a peering relationship with Provider C by which it would receive a route to Customer 3 only, and advertise a route to Customer 1 only. Provider A would have a peering relationship with Provider D by which it would receive a route to Customer 4 only, and advertise a route to Customer 1 only. Providers B, C, and D would similarly have full-meshes of peering relationships by which they could assemble full sets of routes to all four destination customers.

The function of a provider, then, is to advertise reachability of their own customer to the world via peering relationships, and to aggregate the world's announcements together and pass them on to their customer via a down-stream transit relationship.

<<<Diagram flow-of-data versus flow-of-money>>>

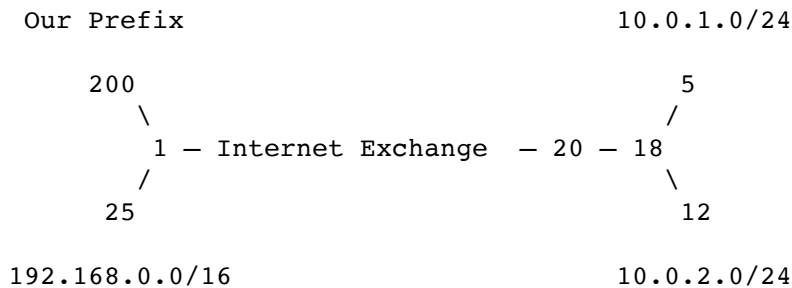
Note that each customer is paying their own provider, and the providers jointly finance the operation of the Internet exchange point, but no provider pays any other provider in this example. Thus money flows upstream within the hierarchy, eventually reaching direction-free equilibrium at the Internet exchange point, regardless of the direction of packet propagation. Thus within this system and all else being equal, a packet from Customer 1 to Customer 2 has the same value and cost as a packet from Customer 2 to Customer 1. The flow of money is independent of, and unrelated to either the direction of the net flow of the bulk of the data, or the customer's perception of who originated the conversation. All customers pay in roughly equal measure to talk to one another; there's no such thing as a "more valuable" customer in a technical or business sense, only in a marketing sense. This is a point of understanding which people without technical literacy often fail to reach; one often hears talk of Microsoft being a "valuable customer" because many individual consumers want to reach their content, or of AOL being a "valuable customer" because many content-providers want to reach their individual consumers' eyeballs. Obviously these are cross-canceling assertions, yet there exists a ceaseless stream of naive entrants into the Internet routing market, so one hears them with unfortunate frequency.

This is not to say that all customers pay exactly the same amount or that they all cost the same amount to service. There are more and less profitable customers, but that only affects value within the customer-provider pair, and isn't a measure of the value of a connection between that provider and any other provider or customer.

With this deeper understanding, then, let's take another look at our previous example of a received routing announcement, with the assumption that we're looking (receiving it) from the point of view of a customer, which we'll call AS 200:

Prefix	AS Path
10.0.1.0/24	1 20 18 5
10.0.2.0/24	1 20 18 12
192.168.0.0/16	1 25

If we diagram the topology which we're able to discern from this, and we know that ASes 1 and 20 are major providers which peer with each other, then this is the result:



From this routing table, we know AS 18 to be an ISP (since they're selling service to ASes 5 and 12), and if we know for certain that AS 1 is a peer of AS 20, then we also know AS 18 to be a customer in a transit relationship with AS 20, since if ASes 18 and 20 were peers at another Internet exchange point, we would not have visibility of any of AS 18, 5, or 12, since there wouldn't be transitive communication of their existence between ASes 1 and 20.

A Closer Examination of the Transit Transaction

What exactly is being sold in a transit relationship? Although anyone possessed of a rudimentary understanding of Internet economics can easily define the transaction, unfortunately few people make the necessary study to gain that understanding. However the transaction occurs many times a day, and people have widely divergent impressions, so it's not uncommon that both parties to the transaction fail to understand what they're buying and selling. *In fact, in a survey conducted by Gordon Cook and myself in 2001, fewer than a half-dozen Internet service providers globally understood what measurements of their networks would be necessary to allow them to directly observe their cost of goods.* If both parties misunderstand the nature of the transaction they're entering into in the same way, it's probably fair to say that they're not actually engaging in the transaction that we believe them to be, but are instead fulfilling a common fantasy which has little to do with business; nonetheless, money changes hands and has an unfortunate effect on the market, so it's useful to understand not only the transaction, but also common modes of misunderstanding it.

As currently practiced, the transit transaction is most simply formulated as an exchange of money from the customer for the provider's willingness to receive from the customer and make a best-effort attempt to retransmit packets bound for any destination on the Internet, and to make a best-effort attempt to readvertise any valid routes heard from the customer to all the other providers of which the Internet is composed, thereby soliciting traffic on behalf of the customer. Obviously that best-effort is frequently quantified, since a provider which advertised universal reachability but discarded all packets would not in fact be providing a useful service. The efforts aimed at such quantification are generally lumped under the rubric of QoS, or "Quality of Service," and the agreements which attempt to contractually specify them are generally called SLAs, or "Service Level Agreements."

The transit transaction is often bundled with other indirectly-related services which are also thought to be useful to the customer, and easily provided by most Internet providers. Most common are domain name service, Usenet news feed or reading service, mail exchange host service, configuration of equipment, and access to the provider's security incident response team. Occasionally tangible products are bundled as well, most commonly the routing hardware which is necessary at the customer's end of the communications facility. All of these products and services are usually handled internally by larger or more competent customers. There are also a set of assumptions regarding the relationship which would not be categorized as either products or services, but are more closely related to security, privacy, and intellectual property protection. These would typically include a degree of privacy protection, such that neither aggregated nor specific information about a customer's utilization of the network be distributed to parties outside of the transaction, information about the nature of any attacks being conducted against the customer be withheld from anyone without a warrant or subpoena, and the expectation that some forms of attack by or against the customer be blocked within the provider's infrastructure rather than passed through to the customer or the world. These assumptions regarding the service would be made by any informed customer regardless of their size, and are not beneficially cleaved from the transit service.

Quality of Service

Most discussion of QoS focuses upon three properties: loss, latency, and jitter. Loss and latency are independent characteristics, and jitter refers to variability in the values of each of the other two.

Loss, or more specifically, the probability with which any individual future packet will be discarded en route prior to arriving at its destination, and the fraction of the total number of historically transmitted packets which have been discarded en route, is essentially a measurement of

efficiency, since all Internet protocols expect and deal with even fairly high rates of loss through detection and retransmission, by throttling the rate at which they transmit, or in the worst case by simply evidencing qualitatively worse service. Most commonly, applications which use TCP, Transmission Control Protocol, for the transfer of data combine the first two methods. TCP provides a reliable service for data transfer, that is, it attempts to guarantee eventual transmission to the application above it by monitoring loss, retransmitting any lost packets, and providing a complete reassembled data-stream to the application on the receiving side. Meanwhile, upon encountering loss, it throttles down its transmission rate until the loss becomes less evident. This is called "flow control" and is a very beneficial property in network traffic, since it allows many conversations to coexist simultaneously on the network without the degree of negative interaction that they would display if each application were greedy and attempted to use bandwidth as quickly as possible. Since by limiting their transmission rates, applications are able to reduce the fraction of lost packets, they also reduce the necessity of retransmitting, which is lost efficiency: data being carried across the network twice. Thus they all come out ahead by behaving in an altruistic manner rather than pursuing self-interest in the area of transmission rate.

The other frequently-used data transmission protocol is UDP, User Datagram Protocol, which is an "unreliable" service. This is not a qualitative statement, but rather a value-neutral descriptive one. UDP is used when guaranteed retransmission of lost packets would be disadvantageous. This is the case any time "late" is worse than "not at all" and the value of the bandwidth to subsequent data is higher than the value of the lost data.

Understanding the impact of loss requires an awareness of the response which applications will have to the effects of loss on their attempted transmissions. In the simplest case, for example the one-way transmission of a stream of live video, lost packets are not retransmitted, since by the time they were retransmitted, they would have arrived either too late to be useful, or out-of-order and thus not useful. In this case, loss is unnoticed by the sender, and correlates directly with the quality of the video as perceived by the receiving viewer. In the more common case of a reliable transfer, as of components of a web page, loss has a more dramatic effect. If 10% loss is encountered, transmission takes more than 10% longer, since the lost packets still occupied transmission time, the loss must be detected, retransmission negotiated, and each lost packet retransmitted. The loss rate affects the retransmitted packets as well, so if a link is capable of transmitting 1,000 packets per second with a 10% loss rate, it will actually take more than 1.12 seconds to transmit 1,000 packets. This would mean that the "goodput" or portion of the throughput which actually consists of data arriving intact and on-time rather than lost, errored, or control packets, will be somewhat less than 90%. In theory, since the same piece of data may get hit by loss each time it's retransmitted, the 1,000 packets might never actually all be received; however that's statistically improbable in the absence of an algorithmic problem with the method of the data encoding or the retransmission method. Note that a subset of loss is the introduction of errors, which are similarly detected and corrected, and have much the same effect on the network as loss, although slightly worse since they consume transmission resources all the way to the destination, rather than being lost somewhere en-route and consuming no further resources beyond the point of loss.

Latency is the measure of the delay between the insertion of a packet into the network and its reception at the destination. Latency can be most usefully measured as "one-way delay," that is, the delay which occurs between the transmission and reception of a single packet. However, precise measurement of one-way delay requires special equipment, since it depends upon the presence of a high-resolution common time-base between the sender and receiver which is not present in the general case. Much more common but less useful is measurement of round-trip delay, that is, the amount of time it takes for an "echo request" packet to be inserted into the network, propagated to its destination, received, an "echo reply" packet generated at the receiving end, that reply transmitted and propagated back to the first machine, and an elapsed time measurement performed. This obviously is subject to delay at the receiving end, where the

host may have higher priorities than generating echo replies for unknown parties, and it also measures only the sum of the unrelated forward and reverse paths through the Internet. It's thus a marginally useful tool for evaluating the network's handling of bidirectional transactions, but only a very crude tool for diagnosing problems in the network itself. Bidirectional delay measurement is most frequently performed using a tool called "ping" which is available on most all computing platforms, and "ping" has entered the lexicon as a verb.

High latency can have a variety of pathological effects, ranging from simply making transactional behaviors marginally slower to complete, to making voice interactions unusable. High latency is typically of greatest detriment when combined with a high loss or error rate, since it compounds the problem of loss and error detection and correction, and when combined with high rates of throughput. Under low latency, low throughput, low loss, low error rate situations, the requirement that a transmitting device wait for acknowledgement of receipt of a block of data isn't very onerous... The data can be transmitted in relatively large blocks, and an acknowledgement of receipt will come back quickly, so transmission can begin again without much delay. If the round-trip delay is long, it takes longer for acknowledgements to come back, which means that either the transmitter sits idle for long periods between transmissions, or the transmitter can continuously transmit, but then must keep a copy of each block of data locally until it's been acknowledged. That seems like a reasonable solution until the throughput gets high... under continuous transmission on a 622mbit link with 1500ms bidirectional latency, at least 117 megabytes of RAM will be occupied by already-transmitted data at any time. In the worst case, an acknowledgement packet is lost, and despite the fact that the transmitted data has arrived intact a timeout period must be endured and then a new transaction must be completed to get data flowing again. All of this is hampered by high latency.

High latency may also make some kinds of time-sensitive data useless. The best example of this is voice telephony, which becomes essentially useless once a human-perceptible delay has been introduced into the conversation.

Jitter, or variability in loss and latency, also compounds existing problems, as well as creating a few new ones of its own. Many applications and protocols seek to estimate network conditions and optimize their transmissions to encounter the fewest problems possible. Jitter, or continuously-fluctuating conditions, makes this a moving target, and fine-tuning for the much-different conditions of half a second ago may exacerbate current conditions. In a continuously varying network topology, particularly one in which one-way delay is changing from moment to moment, it's probable that packets will be delivered out-of-order. This can pose some problems for traditional TCP file transfer applications, which may have to endure a timeout period in order to understand that a packet from the middle of a block of data doesn't need to be retransmitted, it's just late. Where it's really problematic, however, is in real-time data streams like voice and video, in which a packet which is delivered subsequent to ones which contain later data which has already been displayed, is worse than loss, since it must be discarded rather than displayed, and it occupies network bandwidth which could have been used for subsequent data. Jitter also makes it difficult to synchronize a time-base across the network, and impossible to use pairs of round-trip echoes to estimate one-way delay in each direction.

Each of these QoS characteristics can be quantified and evaluated by end-users, although not so easily by providers unless they have access to test facilities at the customer premises. Transit contracts have begun to occasionally include rebate structures related to QoS targets, such that a customer which can demonstrate that they have received service which falls below some threshold may receive a pro-rated rebate for that portion of the service or, extremely rarely, may even receive additional compensation from the provider. Note that such compensation is always tied to quantifiable network performance characteristics, and never to business losses on the part of the customer, which are nearly infinitely variable and outside the realm of influence of the Internet provider. Total rebates are also universally contractually limited such that in no period do

they exceed the amount paid by the customer to the provider. That is, SLAs preclude a net payment from provider to customer.

Costs of Delivery

Reachability of fundamentally the entire Internet is a baseline requirement for all customer-serving portions of an Internet provider's network, and QoS characteristics are byproducts of good network engineering, so the qualities that customers are looking for from providers aren't actually what's in the forefront of providers' attention. Instead, properties of customers' utilization of the network which directly affect cost and performance tend to occupy providers' minds.

...

Charging Schemes

Relative to the age of the industry, service providers have experimented with a surprisingly small variety of billing schemes. The vast majority of service is sold under one of three prevailing schemes: flat-rate, tiered, or 95th percentile.

Flat-rate is the most common charging scheme. Under a flat-rate scheme, the customer pays a fixed amount prior to the beginning of each period, typically one month, purchasing the option to send and receive data at a rate which cannot exceed a cap, for the period. The price is dependent upon the capped maximum rate of transmission and reception, which are typically symmetric. The periods are typically contracted for in terms of 12 or 24 months at a rate which is fixed for the duration of the contract. Carriers typically expect customers to actually send or receive some 1% - 70% of the number of bits which they're contractually permitted to, and find their profit in aggregating many such customers together and hoping that they don't all attempt to utilize the service simultaneously. Since the timeframe of the utilization is quite small, this is a business model which is beneficial to all concerned, although it's sometimes spun as "overselling" by promoters of fear, uncertainty, and doubt. All carriers "oversell," and any carrier which didn't would be shunned by customers due to the astronomical price which would result. Flat rate transit transactions are essentially a bet by the provider that the customer's utilization will be low, and a bet by the customer that their utilization will be high.

Tiered is the next most common transaction. A tiered model is essentially one in which the maximum rate of transmission and reception is capped as in a flat-rate transit transaction, but at a higher price, and the customer receives discounts if their average (or occasionally a percentile) utilization stays below a set of thresholds. Viewed from the opposite perspective, for each utilization threshold the customer surpasses, the rate at which they pay increases, until they finally hit the overall rate cap, at a much higher cost than if they'd signed up for the same cap as a flat-rate transaction. This is also a bet, but reverses the assumption of the flat-rate model: the customer bets that their utilization will be low, while the provider bets that the customer's utilization will be high.

While neither flat-rate nor tiered transactions are simple fees-for-service, both are at least legitimate risks openly undertaken between customer and provider. The third most common model, however, is more accurately viewed as a scam, a means of concealing billing irregularities or falsifications. Under the 95th percentile scheme, the provider samples the customer's rate of utilization periodically, takes the 95th percentile measurement, and bills for the month based upon that rate. The most fundamental problem with this is that the calculation of a percentile is a unidirectional equation; that is, it's unauditible. Whereas an average can be multiplied by the

number of averaging periods within the billing period to rearrive at the total number of bits sent, a verifiable figure, any percentile measurement depends upon the granularity, frequency, and alignment of the samples within the window of the calculation. Since utilization is actually binary (either a bit is being transmitted or it's not, at any instant in time), continuous instantaneous measurement of any connection at a frequency equal to the maximum transmission rate of the line would yield a 95th percentile measurement of zero on any line which was in use less than 5% of the time, and equal to the line capacity on any line which was in use more than 5% of the time. In more concrete terms, actual utilization of 6% will always yield a 95th percentile measurement of 100% of the maximum transmission capacity of the line, if the sampling frequency is high enough. Indeed, exhaustive analysis of actual customer data shows that the 95th percentile algorithm maps much more closely to the maximum transmission rate of the pipe than to the amount of data put through the pipe (Odlyzko, 1999). Under cover of providing customers with a "high burstable capacity," providers which use the 95th percentile scheme invariably use a very large-capacity port facing the customer, and many ensure that the pipe stays at least 6% full by propagating LAN broadcasts, active measurements, and other unsolicited traffic toward the customer. By varying the frequency and alignment of the sampling window, they can vary the percentile result between the pipe size and the average utilization. When customers complain about high bills, 95th percentile providers usually explain that any previous providers the customer was used to had so constricted traffic to the customer's web site that the customer didn't realize the web site's latent popularity, which the 95th percentile provider has now unleashed. Customers fall for this with surprising regularity.

Billing for actual traffic, that is, a charge to the customer which is in any way based upon the number of bits sent or received, is unfortunately vanishingly uncommon. The obvious algorithm here would be for the customer to pay the fixed cost of maintaining their interconnection to the provider, plus a negotiated rate per bit transmitted or received, plus a surcharge for any additional services used. Note that paying for the actual number of bits transmitted or received is functionally equivalent to paying for the average utilization, since the latter is simply a function of the former and the billing period.

Unfinished Notes

A future is simply a firm agreement to buy something at a certain date at a certain price. It's nearly the same as buying the object itself.

An option is paying a small amount now to reserve the right to buy something at a certain price (the "exercise price") at a date in the future.

Peering routers versus POP routers, peering/local routes only, input filtering to keep peers from using you to transit to other peers.

95th percentile versus average versus actual versus peak capacity.

pay per packet, pay by average (same), pay flat rate, pay "tiered", pay by 95th percentile, pay by blocks. Distinguish futures from goods.

Examine customers-can't-be-peers misassumption. Works until two peering customer talk to each other.

"Paid peering" Bits paid for once, twice, or never.

Probability of loss (retries, "goodput")

Latency (one-way, two-way, out of order delivery)

Ratio of peaks to valleys (burstability)

Per-prefix cost of delivery ("distance")

Time of day scarcity/demand
unpredictability of future costs (temporal distance)
Jitter
Anonymity

QoS: Not an issue if receiver pays. Predictive in nature. Is it necessarily true that the provider can predict demand better than the customer? QoS mostly treated in bulk, rather than per-packet. Classes of service not currently used much.

v4 multicast, v6, v6 multicast.

Anycast and caching as a service, or a means of transparently reducing cost of delivery.

Sender pays/recipient pays
Hot-potato/cold-potato
Full advertisements/regional advertisements/MEDs

bilat/multilat/mmlpa

"private peering" with half-circuits.

Old issue of whether customers are more valuable to web sites, or web sites are more valuable than customers. Customers vs. content.

Acknowledgments:

Thanks to Geoff Huston and Bill Manning for critical contributions to this document.

Afterward

By far the greatest actor upon communications pricing in the five years which have elapsed since I wrote the above has been the displacement of the legacy Time Domain Multiplexed (TDM) Public Switched Telephone Network (PSTN) by the packet-switched Internet for the carriage of essentially all of the world's communications traffic. The small-but-highly-visible portion of that traffic which consists of real-time person-to-person voice telephony had also comprised the bulk of legacy carriers' profits, and some of those carriers proved very slow (in Internet terms) to bring their revenue and accounting models up to the standards of the more modern (and lower margin) network of which they found themselves to be a small subset.

This created a tension between, on the one hand, the reactionary impulses of many incumbent telephone companies, which sought to employ regulatory barriers to innovation and competitive pressures, and on the other hand, the general public, who desired faster, cheaper, more reliable communications services, the more forward-looking regulators and communications ministries, which valued growth and progress over retrenchment, and the Internet service providers who built the Internet and served the demands of the market.

As I said, five years is a long time, at the Internet's pace of development. For a primer on the communications economics as they now obtain in the post-transition, 21st century network, see my more recent White Paper on the Economics of Multi-Party Interconnection for the Production of Internet Value, or Peering for Politicians. But to adapt the famous words of Abbie Hoffman: Never trust a white paper over thirty days old.

Author's Contact Information:

Bill Woodcock
Research Director
Packet Clearing House
572 B Ruger Street, Box 29920
The Presidio of San Francisco
San Francisco, CA 94129-0920 USA
+1 415 BGP PEER
woody@pch.net

Footnotes

Smartest guys in the room