A novel design for maximum use of public IP Space by ISPs – one IP per customer

6/20/2018

Jim McNally, James Lopeman - Plusten
Mark Steckel - Citywisper

Abstract

This paper outlines a new design for ISP networks which results in a nearly one to one customer to IP ratio for public IP address space. It takes advantage of several properties of IP networking, using them in a novel way. Doing this allows up to nearly a 4 to 1 increase in utilization of public IP address space versus current methods. It also allows complete or partial conversion of an existing ISP Network to use this design without affecting current customers.

We first review IP addressing and routing in general, look at how it is applied in most ISP networks today, and then present the updated approach. We currently use this design in our own ISP Networks.

Background

Computer networks are broken into layers which are both logical and physical. For IP Networks the classical way to talk about this is Layer 2 and Layer 3 functionality. Layer 2 is generally seen as the physical layer, although in a complete OSI reference model the physical layer is actually Layer 1. However most people do not directly use layer 1 for any purpose themselves. Layer 2 in an Ethernet Network utilizes MAC addressing to define end points of a layer 2 Network. MAC addresses uniquely identify computer interfaces on Ethernet networks and are generally assigned by the factory when the network interface is manufactured. They are supposed to be globally unique from all other NICs in existence, although there are usually ways to set them post-manufacture if desired.

IP addressing works at Layer 3. Layer3 includes IP to IP packet handling as well as routing between networks. It also defines the networks themselves both logically and physically from the computer’s point of view.

In order for two computer devices to send data to each other on the local layer2 network they must know each other's MAC address. A data packet contains a header which includes the MAC address from which the packet originated and the MAC address of the destination device. This allows two-way Communications between two computer devices since once a packet is received by a device it automatically knows how to send data back to the originating device because its MAC address is included in the header of the packet. Since MAC addresses are generally fixed permanently to a given interface, and IP addressing is often defined dynamically, there needs to be a way to find the MAC address of a LAN device where the data is being sent from the IP address. This happens when layer 2 and layer 3 are correlated in the Address Resolution Protocol.

Layer3 is defined by IP addresses. These take the form of 4 hexadecimal octets (in IPV4) which are dynamically assigned either through a local area network protocol such as DHCP or statically by an administrator.

In order to function on the LAN there has to be a way to translate the IP information to Mac addresses, since computers on a local network can only talk to each other via Mac address. This is generally done through the Address Resolution Protocol or ARP. When one device needs to send data to a second device on a local network but only knows that second device's IP address, it uses ARP to determine the second device's MAC address. The originating device sends broadcast ARP request which propagates throughout the local network to all the connected devices specifically asking for the MAC address of the device with the requested IP address. When the device with that IP address receives the query, it replies, effectively saying “I am here at this MAC address”. Since now both the originating and the receiving device know each other's MAC addresses, communication can now successfully take place.

In order for this to occur an IP network must be defined. Any given local network (or subnet) is assumed to have two or more devices talking to each other. Because the network address and broadcast address have a specific place in a network, the starting IP address and the size of the network must be known.

For example, if we define a logical network starting at the IP address 2.0.1.0 and encompassing 256 IP addresses, known as a /24 network, the lowest and highest addresses are reserved for special use within the network. The lowest address, 2.0.1.0, is called the network address and logically defines (with the “/24” netmask value) the overall network itself. The highest network address, 2.0.1.255, is reserved as the broadcast address. Packets destined for all devices on the local network will be sent to the broadcast address. All the devices on the local network must listen for packets with the broadcast address so they can determine if the ARP packets (or any other broadcast protocol) are intended for them or not.

Only the network definition (ex: 2.0.1.0/24), the starting boundary of the network and its size, is needed in order to define everything necessary for devices to communicate with each other via IP on a local network. Using the starting IP address of the network and its size (the number of IP addresses it contains), that single piece of information allows any two devices can find each other via ARP and then communicate.

Devices on the same subnet can communicate directly with only the local network definition, however devices on different subnets require something more.

In order to make internet connections function, we need to be able to send data from the local network (LAN) out to other networks where it will be received by devices on those networks. All of these networks use the same layer-3 definitions and IP address functions like the broadcast address within their local scope. When we need to talk outside of the local network, we need to add another piece of information and device: the Gateway router IP address.

A router is a device which physically connects more than one local (to the router) network, and knows how to direct traffic between the networks. When a router receives a packet from one network which is destined to another network (which it knows about), it sends the packet to the second Network. The simplest router has only two IP addresses each on a different subnet (ex: 2.0.1.0/24 and 3.0.1.0/24), however routers can connect to many networks simultaneously.

The router keeps an internal routing table that tells it how to route packets. The routing happens because the table has a list of destination networks, and a router IP address associated with each entry (a gateway). When a packet arrives, the router looks in the table to see if it has an entry for the network that IP is on, and if it does, sends the packet on to the device at the associated IP address in the table. The routing table can be statically defined or maintained and updated by a routing protocol such as RIP , OSPF or BGP.

At a minimum an IP router will have in its routing table its own interfaces for the networks that it connects to directly and therefore will automatically know how to route packets from one network on one side of itself to a network on the other side.

This solves the case of two networks with a router separating them and how packets flow between those two networks from the point of view of the router itself. However if we need to send a packet from a device on one network to a device on a network through the router, or beyond what the router is directly connected to, we need to add a new element to the design: a network Gateway. All the other devices on the local network need to know the router's IP address ahead of time, so when the device they need to talk to is not on the local network (to the extent of which they all already understand), they know to send the packet to the Gateway router.

Gateway router IPs are defined by specific entries in the routing table.   But there are literally hundreds of millions of networks across the Internet, and no routing table could hold them all. So for the case where there is not a specific entry in the routing table, there is a default gateway entry, which says “if I don't know where else to send this packet, send it to the default gateway IP”. With the assumption that the default gateway has a connection to a router that can reach other routers that can in turn reach the destination of the packet, this is how routing of packets across large networks or the Internet occurs. These table entries are usually set statically if there are a small number of entries, or they can be set by a Routing Protocol like OSPF or BGP to allow the table to be populated automatically.

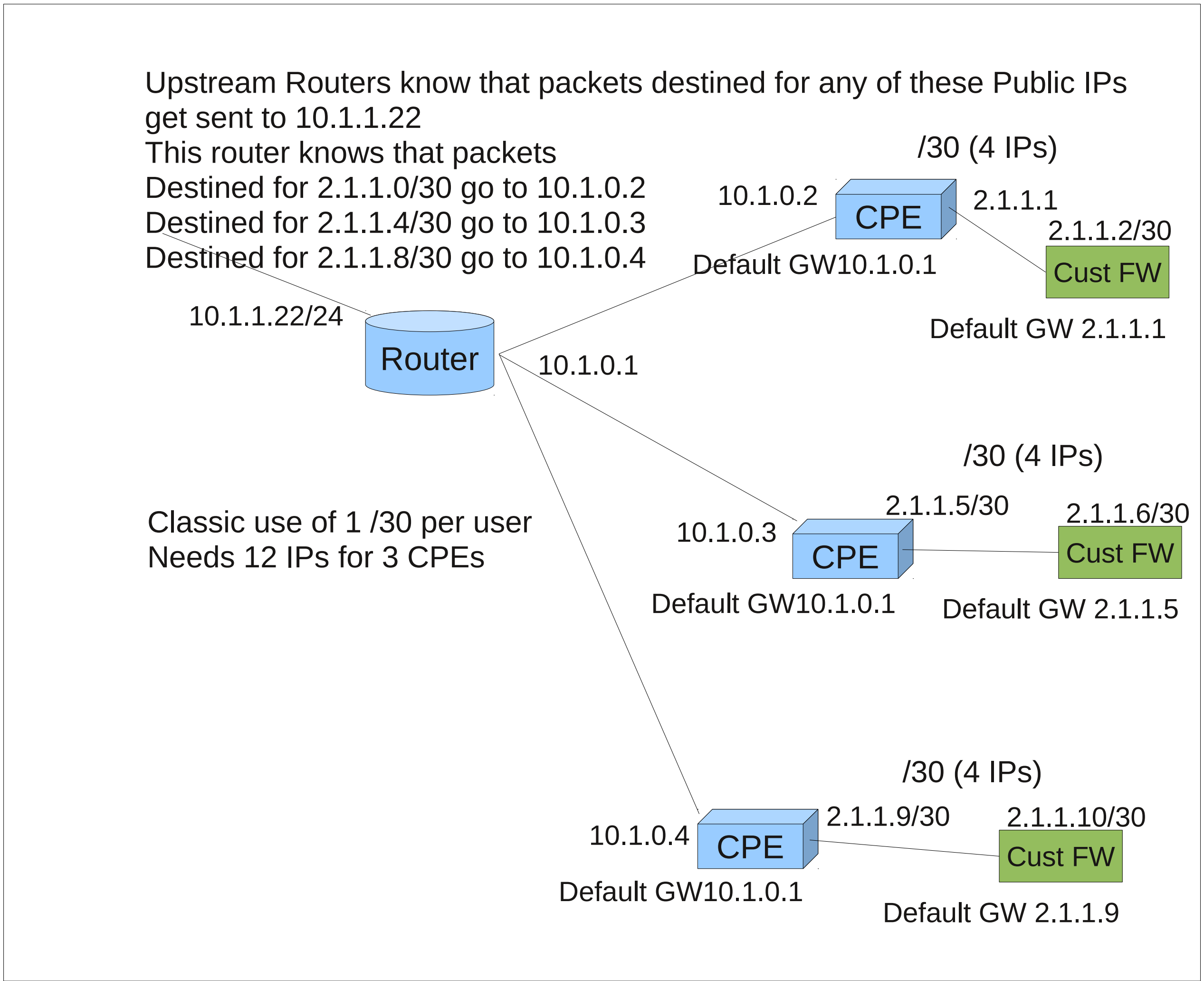
So in order to make networking function we have the 2 specific elements we need - the local network definition including the size and broadcast address of the local network, and a default gateway where the router will send packets that are not on its local network. It is assumed that the routing table in that default gateway device knows where to send packets on to their correct destination, or to the next router in the chain eventually leading to their destination.

One more thing that needs to be understood about the routing table in any router is that when there are overlapping network definitions in the table, the most specific route always wins. This means if we had a table entry for a /24 network and a second entry for a /30 network contained within that /24 network, the /30 definition in the table will take precedence over the larger /24 table entry. In this way a hierarchy of routing definitions can be created which allow for things such as route aggregation or internal routing of smaller networks to specific destinations.   This means that smaller networks can be “carved out” by the routing table even if they are logically contained in the larger “local” network.

An important thing to note for our purposes here is that even though both local and non-local networks are defined within the routing table, local networks behave differently than non local networks because they **are** local. ie they do not need a gateway entry to be reached by the devices within that local network. If a destination device is within the scope of the local network, the packets are just sent to the local device – Layer 3 routing is not involved.

IP allocation efficiency

Since public IP space, specifically IPv4 space, is extremely limited in most cases, it is important to use it as efficiently as possible. For an ISP handing out IPv4 space to its customers, this is often done by using a small network for each customer which the ISP routes to that customer's CPE device.   So in this design case the more efficiently you can give out those networks to customers, the more efficient the use of the IP space.



If we look at a /30 network, it is made up of four IP addresses. For instance in the network 2.1.1.0/30, the first IP is the network address, 2. 1. 1. 0; the 4th is reserved as the broadcast address, 2.1.1.3., This leaves only the middle two addresses to be used by two devices. One of these needs to be the default gateway, ie the router where the customer's device sends data destined for the outside world, and only the second can be used by the customer for their own device, generally a firewall. What this structure does, however, is make very inefficient use of the IP space available, since at a minimum 50% of the IP addresses are now unusable by specific customer devices since the network and broadcast addresses need to be reserved, and in fact the default gateway address is also effectively unusable by a customer since it needs to be kept for the ISP router connecting to that customer. Larger networks for, instance a /29 network with 8 total IP addresses, is slightly more efficient, since only three of the eight addresses cannot be used by the customer, but this is still inefficient. If a customer needs exactly 2 static IP addresses for two devices, an entire /29 must be used for them, effectively throwing away 3 customer-usable IP addresses (not including the network, broadcast and gateway IPs which take together would mean only 2 of the 8 addresses are actually used by the customer).

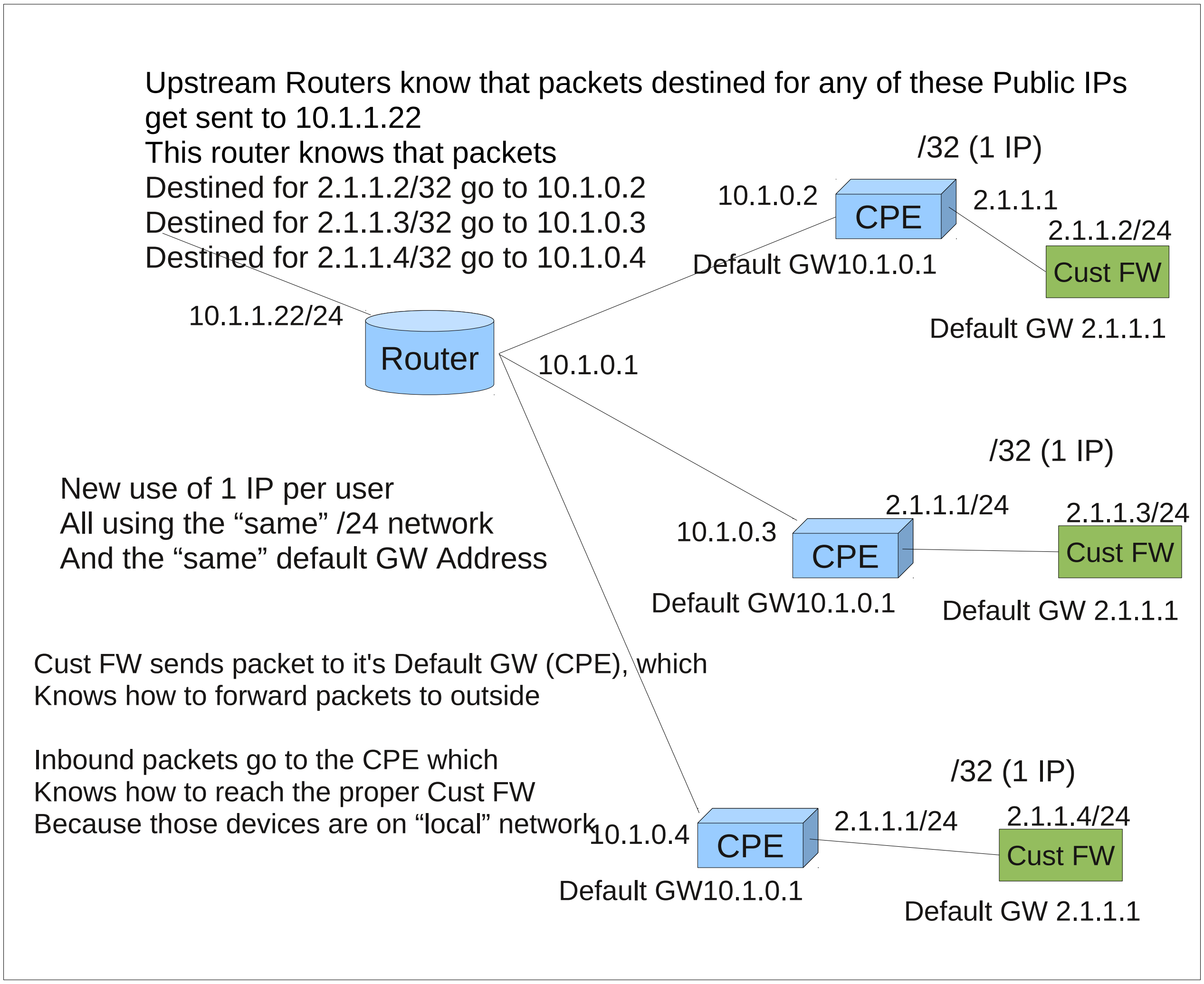
It is possible to define networks that are smaller than a /30. The network and broadcast address in /30 or any other network larger than /30 are technically only needed to determine where the beginning IP and the broadcast IP are within the network. A /31 network only has two IP addresses in it, therefore there can only be two devices attached to it, which means that a packet sent across that network must by definition reach the other device on the end of the cable. If all customers who only need a single public routable IP were handled with /31 Networks, this would double the efficiency of IP allocation versus using /30 Networks. However it turns out that many end-user devices like firewalls do not correctly understand /31 Networks or only do so under certain circumstances. We have run into this issue on numerous occasions since using /31 Networks was one of the early things we attempted in order to increase the efficiency of our use of Public IP space.

Technically a slash 32 network exists and contains only a single IP address. However as an actual **local** network this is of limited usefulness since it only defines a single device which could send data to itself – an example would be the loopback address in a device. However a /32 entry in the routing table is perfectly logical and is understood by routers - it says any packet destined for this single IP address needs to be sent to this gateway which will itself know what to do with that packet. Our new design makes extensive use of this functionality.

New design using one IP per customer

If we allocate a single /32 IP address to each customer who only needs a single public IP address this would dramatically increase the efficiency of allocation of the public IP space which we have, nearing 1:1 IP to customer ratio (the ideal). A packet sent to that customer through the ISP Network will find the customer device at that IP address through the normal routing mechanisms as long as there is a routing table entry in the system's routers for that /

32. The difficulty is finding an IP address for the ISP’s gateway device which connects to the customers port where that /32 exists.   We can’t have a second IP on that network for the gateway since there is only a single IP in the network by definition.



Upstream Routers know that packets destined for any of these Public IPs get sent to 10.1.1.22
This router knows that packets
Destined for 2.1.1.2/32 go to 10.1.0.2
Destined for 2.1.1.3/32 go to 10.1.0.3
Destined for 2.1.1.4/32 go to 10.1.0.4

New use of 1 IP per user
All using the “same” /24 network
And the “same” default GW Address

Cust FW sends packet to it's Default GW (CPE), which
Knows how to forward packets to outside

Inbound packets go to the CPE which
Knows how to reach the proper Cust FW
Because those devices are on “local” network

To make this work we take advantage of both the route table specificity and the feature of the “router” next to the customer connection to “block” the view of it’s “local network” to the rest of the ISP network.   We set the customer and the CPE radio or router to be on a larger network than they really are so there **is** a default gateway and broadcast IP defined, and then only use the /32 IP (but with a bigger netmask) on the customer side so that the customer firewall is happy.   We then set up routing across the entire system for the /32 as pointing to the CPE on it’s ISP-side interface, and when the packets destined for the customer arrive, the CPE router knows where to send them (to the customer IP which our CPE sees as a larger than /32 network).   To send out packets, the customer device sends them to it’s “default gateway” in the local CPE network, and the CPE and ISP router sends them on from there.

This decoupling of the inbound-to-the-customer /32 from the outbound-from-the-customer packets on the larger “local LAN” is what makes this work.   The most unusual (and baffling to uninformed eyes) feature if this is that all the CPE devices used in this way which are using the same large “fake LAN” will have the same public IP address. Since this is not actually used for any part of Layer 3 routing it doesn't matter to the overall ISP system – inbound routing to the customer is done through the ISP network. When the packets reach the local customer gateway device (usually the CPE radio or router) they transition to a Layer2 connection to the customer device. And once the outgoing packets from the customer hit the CPE device they are routed out via the ISP network with only the /32 IP still in the packet header as the from-address, which isn't normally used again until the far-end destination somewhere on the Internet needs to send packets back to the customer.

Two limitations to this design which may be good or bad for ISPs depending on their specific needs is that customer’s will not be able to see each other directly if the second customer’s IP is in the local “network” defined at the CPE. This is because the CPE device will (incorrectly in reality but correctly by protocol) assume that that second customer is on the local network even though it’s not. In most cases this will not be a problem - many ISPs implement a client isolation system to prevent clients from directly talking anyway. But if this functionality is needed, either the two customers need to be in different “fake” LANs or given their own /30 networks so that standard routing can prevail. It also means that the CPE devices are not reachable by their public IP, since all the CPE/gateway devices have the same IP address on their public side so routing in the ISP network could not discriminate between them in any event. This may make monitoring more difficult, but they are all still reachable (obviously) by their internal addresses in the ISP network. Some customers may not like not being able to ping their gateway from the outside world, but frankly this automatically adds a layer of security since these CPE/gateway devices are now all blocked from Internet access directly. We currently firewall all of these Gateway devices individually in our ISP, and the need to do that is removed in this new design by the function of the design itself.

In addition, since the specificity of the /32 routes for the IPs wins out against all the other routes in the network, we can actually make use of these IPs anywhere in the network – they do not have to be used at specific sites only. It also allows the use of currently delegated Network and Broadcast IPs on public subnets already in use by customers without affecting the customer's operation – the Network and Broadcast IPs aren't actually used for packets outside of the Layer2 local network, so using them in the Layer 3 routed network is fine.

Conclusion

The effect of using this new design for ISP networks is primarily to drastically increase the efficiency of public IP use while leaving normal operations as seen by the customers unaffected – they still think they are connected to a normal network. The ISP does have slightly more work to do in the case where a customer needs more than one public IP, since individual /32 entries are needed in the routing table(s) for each IP being used, but this is a very fair tradeoff for being able to use all the currently unusable IPs in the public space. Using OSPF or some other routing protocol which handles this table population behind the scenes makes this relatively simple. In our case we are now able to increase the amount of Public IPs for our customers to use by a factor of nearly 4 without actually needing to acquire more. And client isolation is inherent in the design, which is good for most customers.

The “gateway” (ie the CPE) is automatically blocked from the Internet, which is good from a security point of view, but does mean that all management must be done in the interior network instead of the public IP on the device.   In our case this is fine, since all our management is done on the interior RFC-1918 IP space.   And in cases where CPE devices do need to be able to directly contact each other, they will have to be handled in the classic manner, but they should be the exception rather than the rule.