How we fixed Latency Problems over Satellite links

Summary

Latency used to be a huge problem when using Satcom links for modern internet browsing. This is because modern web design and the number of objects per page has exponentially increased as the increasing bandwidth speeds with 4G or fibre optic networks has also increased. Satellite links previously were unable to deal with the increasing number of objects as a small link (for example over 512Kbps) as it would create a traffic jam through the satellite pipe when using transmission control protocol (TCP) as an example.

Latency cannot be sped up when using satellite networks due to the speed of light and the requirement to link to geosynchronous satellites. We have however fixed all of the problems associated with browsing and running ERP or Cloud based software solutions experienced with a higher latency through satellite through the following technology solutions. In essence we cant speed up the speed up light but we can optimise traffic to stop a traffic jam normally experienced with satellite connectivity.

The areas that we provide a technology solution to fix a problem caused by latency are:

- Dynamic Network Management
- HTTP Acceleration
- Internet page acceleration (IPA)
- Pre-fetching
- TCP Acceleration

Dynamic Network Management

Dynamic network data management is critical for a large enterprise managing multiple sites in PNG. Our network architecture requires the entire network to be managed behind the satellite and cable infrastructure to enable optimal browsing and internet usage and to deploy our technology fixes for issues that suffer under latency. Our network has been designed this way because of the low amount of internet infrastructure in PNG. It is optimal for managing many sites across rural and urban Pacific locations.

In essence all our sites have a base 42/8 download/upload network speed and our contracts for enterprises are set as a single bandwidth allocation (Critical Information Rate- CIR). The only additional cost is managed by a single connection fee per site. We can then dynamically manage your traffic flow based on your individual priority needs and flexibly respond to your enterprise's usage. This allows us to open or restrict all the site's connections to small and large links at a rate of 30 times per second. This enables us to respond to large packet requests managed under IPA pre-fetching and also allocate unused bandwidth to each site based on all our customer's usage. So why is this better:

- Most of the problems experienced with latency occur because rural bandwidth links are restricted to a set rate which causes data backing up behind infrastructure links

(512Kbps). By dynamically managing the links we can respond to large packet requests by users across your network regardless of where they are located. Dynamic management is critical to enable our additional support systems under TCP Acceleration, HTTP Acceleration as well as Internet page acceleration (IPA) Pre-fetching.

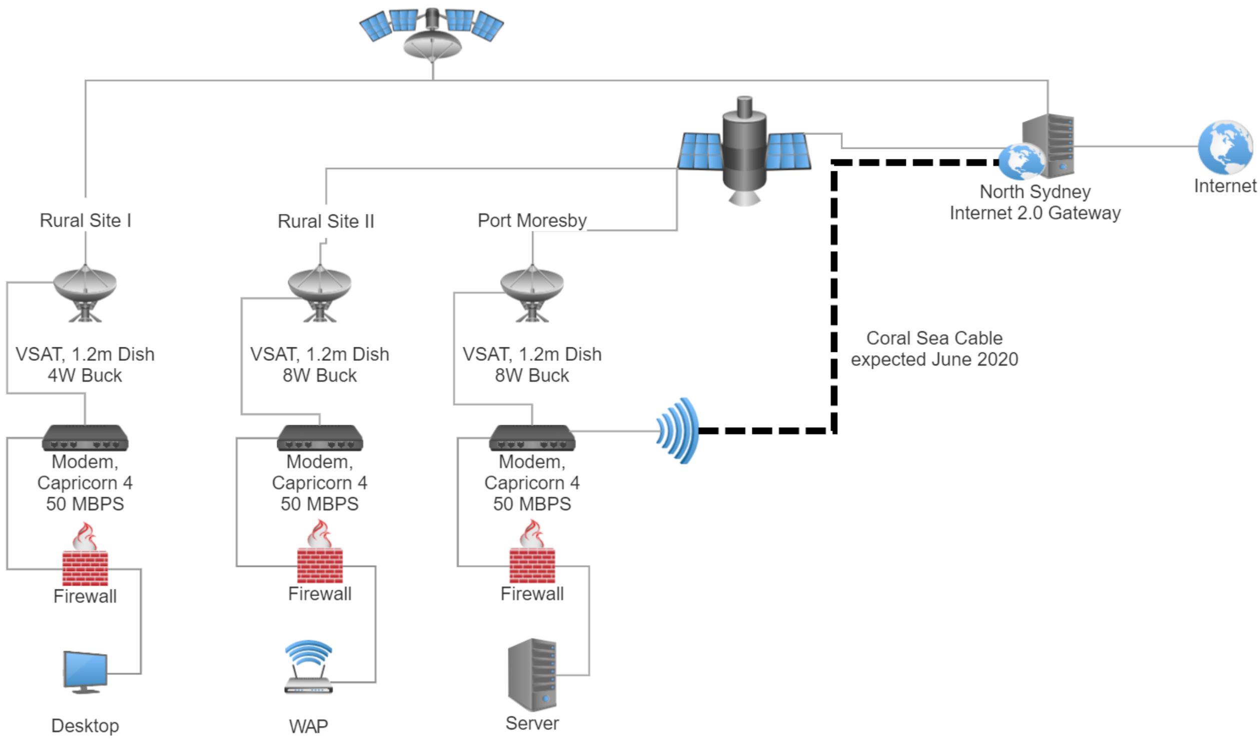

Figure 1

Figure 1 is a basic network diagram that explains how you connect as an example with Astrolab. It is based on one customer running multiple rural sites in the Pacific and their Port Moresby site has an impending connection to the Coral Sea Cable with Astrolab.

So how do we prioritise your traffic in a dynamically managed network?

There are many ways to mange your individual needs to ensure you are receiving the best customer experience and your usage is optimally using your bandwidth. During onboarding we will request a network priority list and discuss in detail which solution we apply to ensure your needs are met. The methods we use to adapt to your usage requirements are: Traffic Management and Priority lists.

Traffic management

Figure 2 is an example of managing different types of traffic from the same network. If you have multiple software systems, a VOIP server or an important ERP system we can prioritise these over normal web browsing and email to ensure a better user experience.

Traffic Management by VSAT

Figure 2

Prioritising by grouping

We can create a grouping tiered system on your contract up to 4 layers per group. This means that you can build a group of sites and allocate a CIR to that group and this group shares the CIR allocated to it prioritising the VSAT based on your tiered allocation of your bandwidth. Figure 3 is an example of how you could allocate a large contract across your multiple sites under your purchased Bandwidth using two groups and splitting them into two CIR pipes. We are very flexible in this prioritisation method so the options will be tailored to your individual usage requirements.

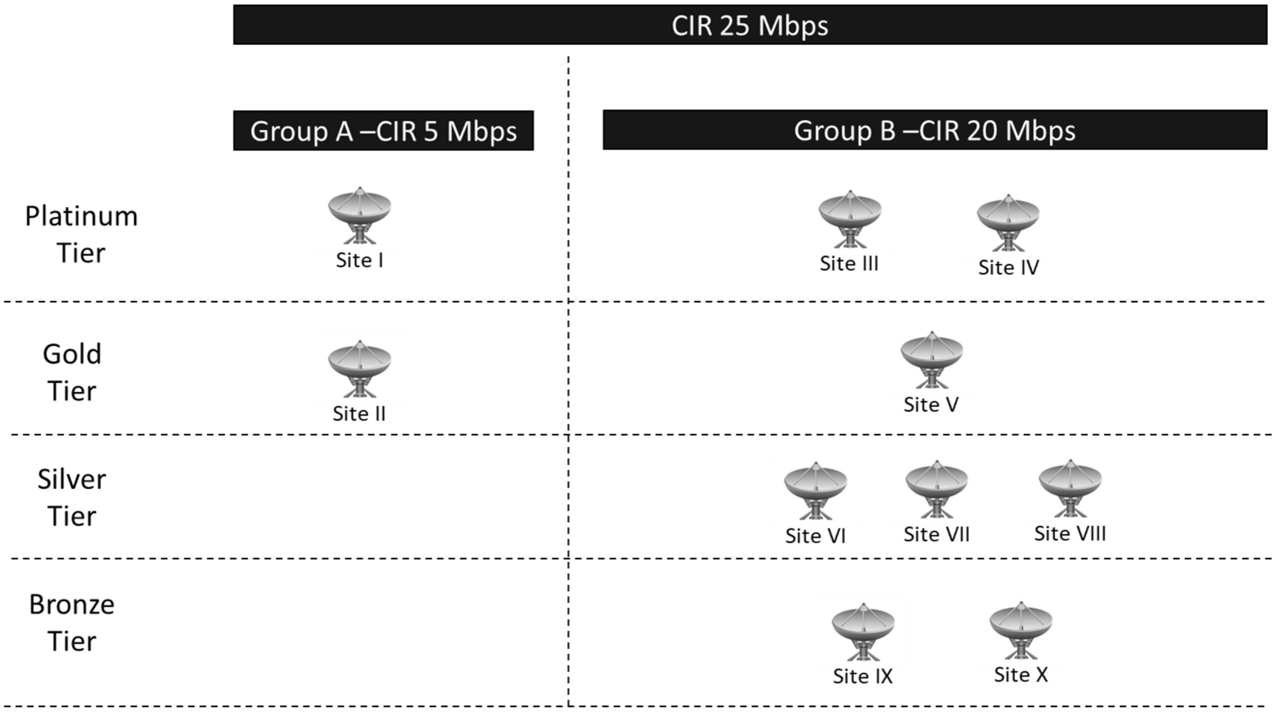

Figure 3

TCP Acceleration

TCP performance over satellite without special treatment is impacted due to the latency introduced by the distance between the Earth and the satellite. Typically, this distance translates to a latency between 240 to 280 msec, depending on where the sending and receiving sites are in the satellite footprint. This makes the round-trip time due to propagation delay at least 480 milliseconds.

During a connection TCP employs four congestion control mechanisms to avoid generating an inappropriate amount of network traffic for a certain network condition. These algorithms are:

- Slow start
- Congestion avoidance
- Fast retransmit
- Fast recovery

These algorithms are used to adjust the amount of unacknowledged data that can be injected into the network and to retransmit segments dropped by the network. The effective performance then changes between different operating systems as the TCP stack is implemented differently.

TCP traffic without acceleration

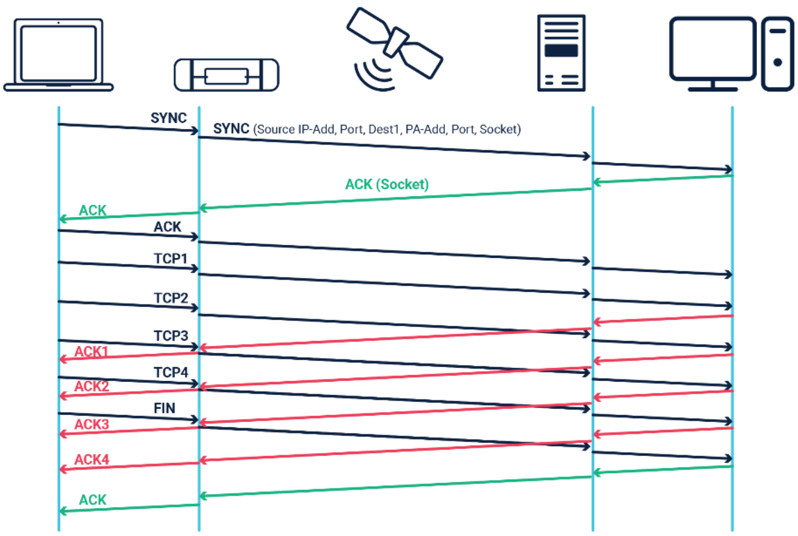

Figure 4

Astrolab includes a special mechanism to overcome the limitations described above and boost TCP performance while maintaining the reliability of the transmission. This Performance Enhancement Proxy (PEP) mechanism is referred to as "TCP Acceleration". This protocol is used between the Gateway and VSATs to carry the traffic. The protocol provides reliability and congestion control. The mechanism employed is transparent to the end user and application – there is no need to change any configuration or install any client on the TCP end device. The mechanism works as follows:

- Selective retransmits and larger windows are used to improve the performance of the TCP sessions

- TCP data segments received by the VSAT are locally acknowledged by the VSAT (acting as the remote side of the PEP) on one side and the DPS on the other side (as recommended in RFC 3135). This speeds up TCP slow start and allows the sending TCP host to quickly open its congestion window.
- NAKs - Local Negative Acknowledgments are employed to trigger local (and faster) error recovery when significant error rates occur (spike noises).
- Retransmission – The VSAT and DPS locally retransmit data segments lost on the satellite link, thus aiming at faster recovery from lost data. In order to achieve this, the VSAT and DPS uses acknowledgments arriving from the end system that receives the TCP data segments, along with appropriate timeouts, to determine when to locally retransmit lost data.
- Inbound and outbound TCP-IP headers are compressed.

TCP traffic with acceleration

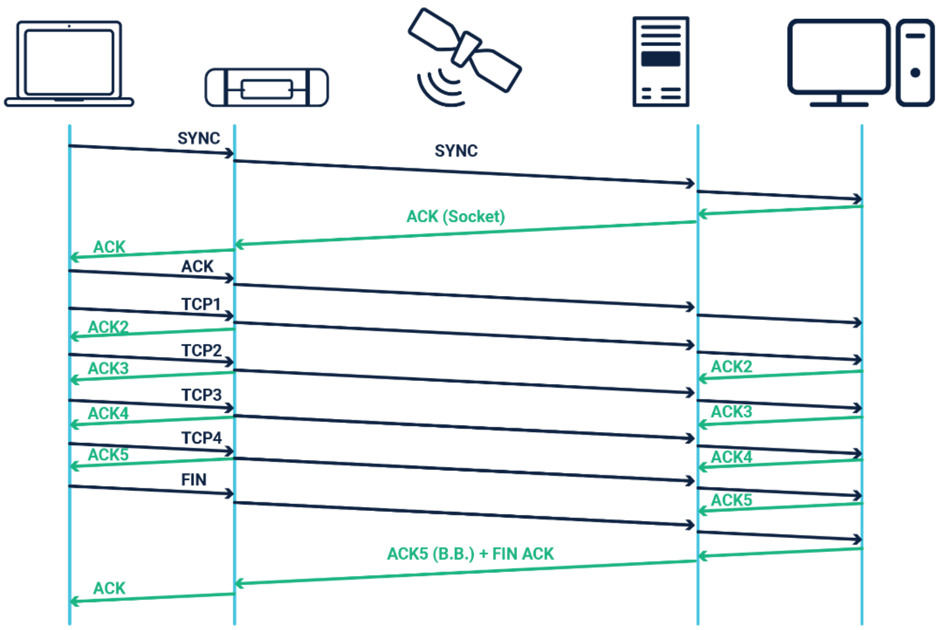

Figure 5

HTTP Acceleration

In a satellite environment, HTTP performance suffers from latency even when the TCP layer is accelerated. Dealing specifically with the HTTP traffic further improves the bandwidth usage and browsing speed.

Our Internet Page Acceleration (IPA) improves the Internet browsing experience while minimizing traffic on the inbound as well as the outbound channels of the satellite network. This means Astrolab can provide a higher service experience on the same bandwidth.

The number of inbound packets is reduced by 35% on average. The number of outbound packets drops by an average of 30%. The total accumulated time to load all pages is 40% below a comparable network without the IPA. Although performance varies from site to site, these figures represent the capabilities in general.

HTTP traffic without pre-fetching

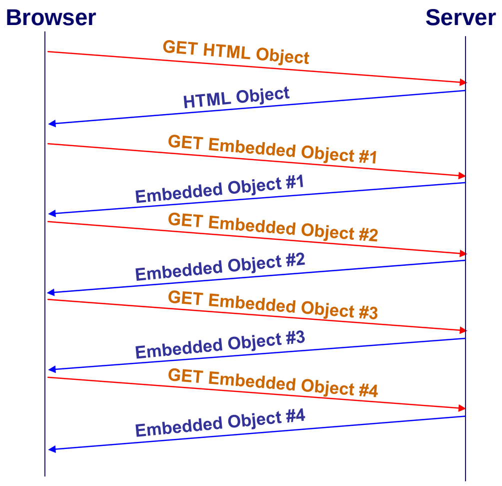

Figure 6

The IPA solution is based on a Hub Page Accelerator (HPA) collocated with the hub and a Remote Page Accelerator (RPA) residing in the VSATs. The mechanisms employed are as follows:

Pre-fetching

Fetching an HTML page requires many TCP connections, one for each embedded object (images, applets, frames, etc.). Sometimes one object must be completely loaded before embedded objects are fetched. For example, an HTML frame set must be completely loaded before the enclosed frames are fetched, which in turn must be completely loaded before the embedded images are fetched.

With IPA, the entire page is provided as a single element to the user browser. As a result, the complete web page is available faster. The IPA pre-fetches the entire web page from the Internet at the first GET request, and then automatically pushes it to the remote site without the need to send additional GET requests for each object. This saves space segment (bandwidth) by eliminating the GET requests on the inbound, and reduces the satellite delay. Consequently, the user does not wait for each component object to be downloaded and displayed.

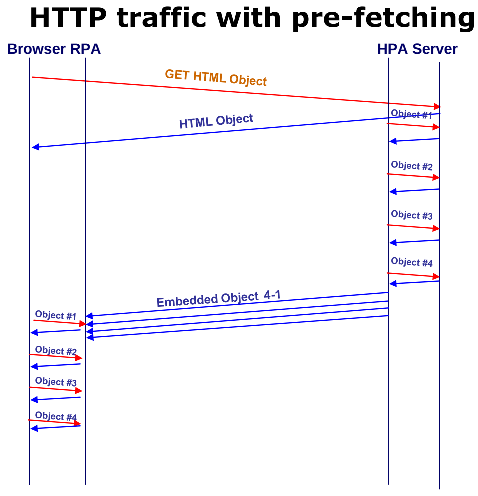

Figure 7

Persistent TCP connections

Instead of opening a new TCP connection for every HTTP transaction, the RPA opens a single, persistent connection to the HPA server after the first object request is received. The browser at the customer's host and the RPA at the VSAT will establish a standard HTTP dialog, with a new TCP connection for every transaction. However, these connections are established locally and do not suffer from long delays. The RPA will aggregate all the HTTP information into the single TCP link with the HPA. The CPE will then handle it as another TCP connection, using spoofing and encapsulating it in the "Backbone" protocol.

This implementation saves the latency and protocol overhead involved in the multiple TCP connections required to request a complete web page. This reduction in latency is mostly associated with avoiding the TCP 3-way handshake required for opening and closing connections.

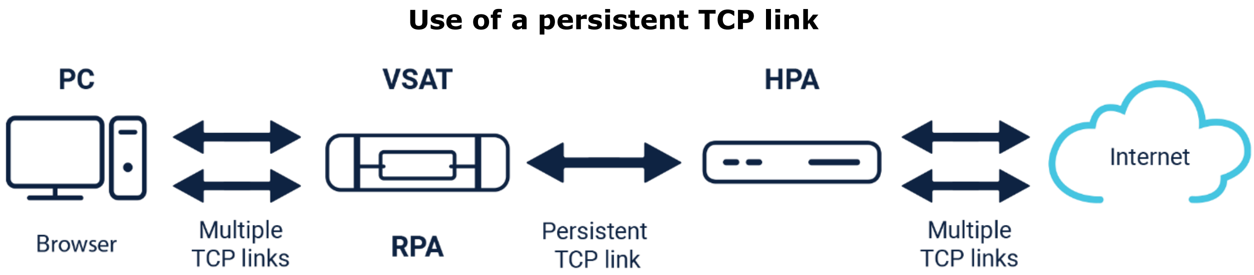

Figure 8

Header compression

Inbound reduction of bandwidth is performed by the RPA by compressing the HTTP header of inbound packets. This reduces the amount of traffic flowing in the inbound and increases the efficiency of the space segment usage.

DNS Caching

The Modem sites support DNS caching. This feature reduces repeated transactions to DNS servers from different PCs connected to the same VSAT and adds to the link efficiency.