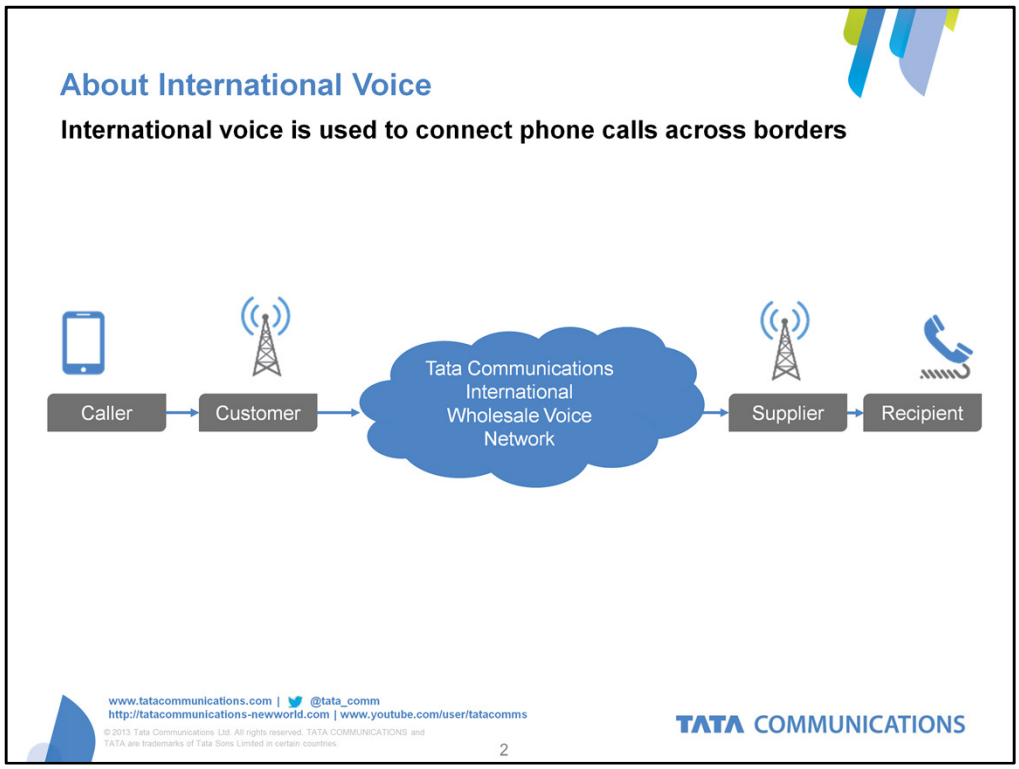
TATA COMMUNICATIONS

Understanding our Data Sets

DataHack14

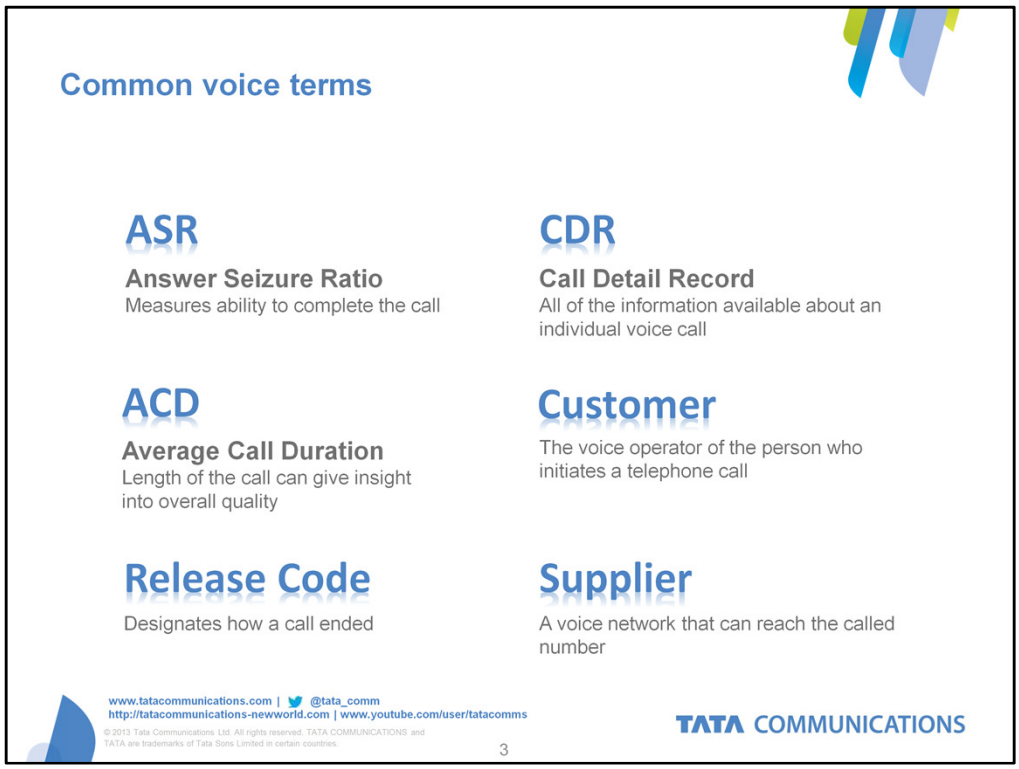
www.tatacommunications.com | @tata_comm
http://tatacommunications-newworld.com | www.youtube.com/user/tatacomms



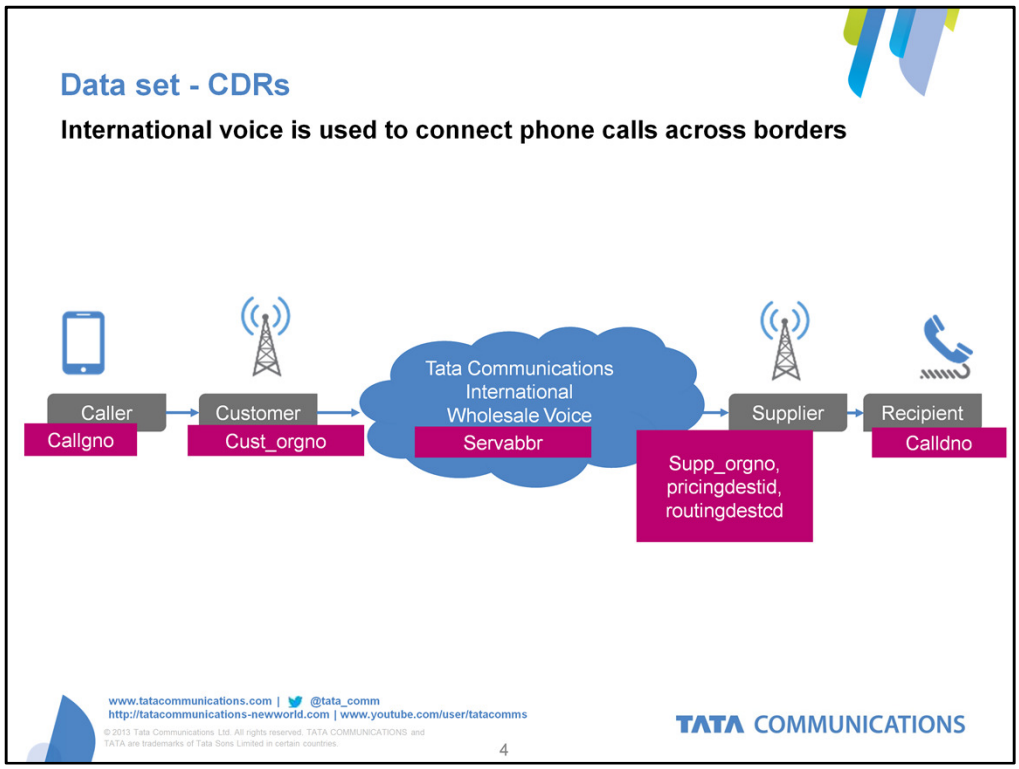
The main area where we have supplied data is around international voice calling. This service carries voice calls, from both landline and mobile phones, as well as IP telephony services such as Vonage, Skype or a cable company calling package, between countries.

In this case, the network from which the call is coming is the customer – they pay us – while the network where the call is going is the supplier – we pay them to complete the call to the dialed number.

While in most cases the caller is on the customer network and the called party is on our supplier network, this is not always true. There can be intermediary providers that might aggregate calls from small regional players, for example, or other wholesale providers that might be sending traffic to our network where we are the supplier. Similarly, suppliers may not always be the network that the called party gets service from.



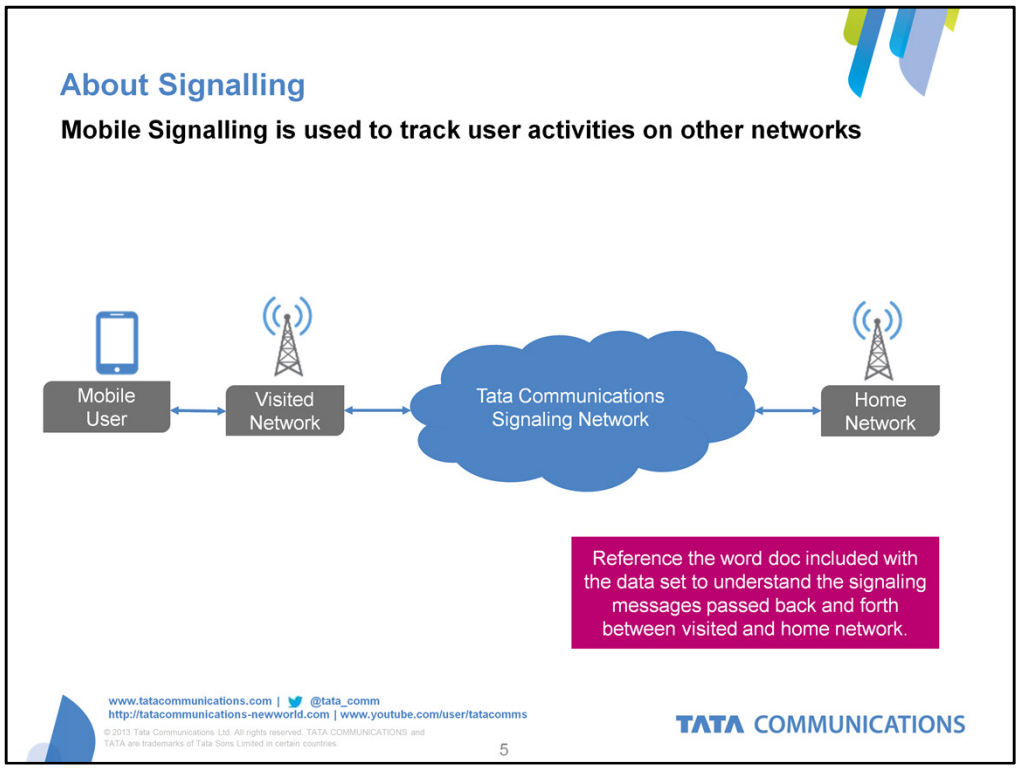
Some of the terms that will help you understand the international voice data set include CDR, or the total record of one voice call, and the release code, which designates how a call ended – no answer, bad number, lost connection, hang up, etc. Average call duration and Answer Seizure ratio are both used to measure quality – a good connection with no static or noise results in longer calls, while answer seizure ratio refers to the number of calls that are answered versus the number of calls that don't complete (i.e. are seized).



The magenta boxes show where in the call process the various data points come from:

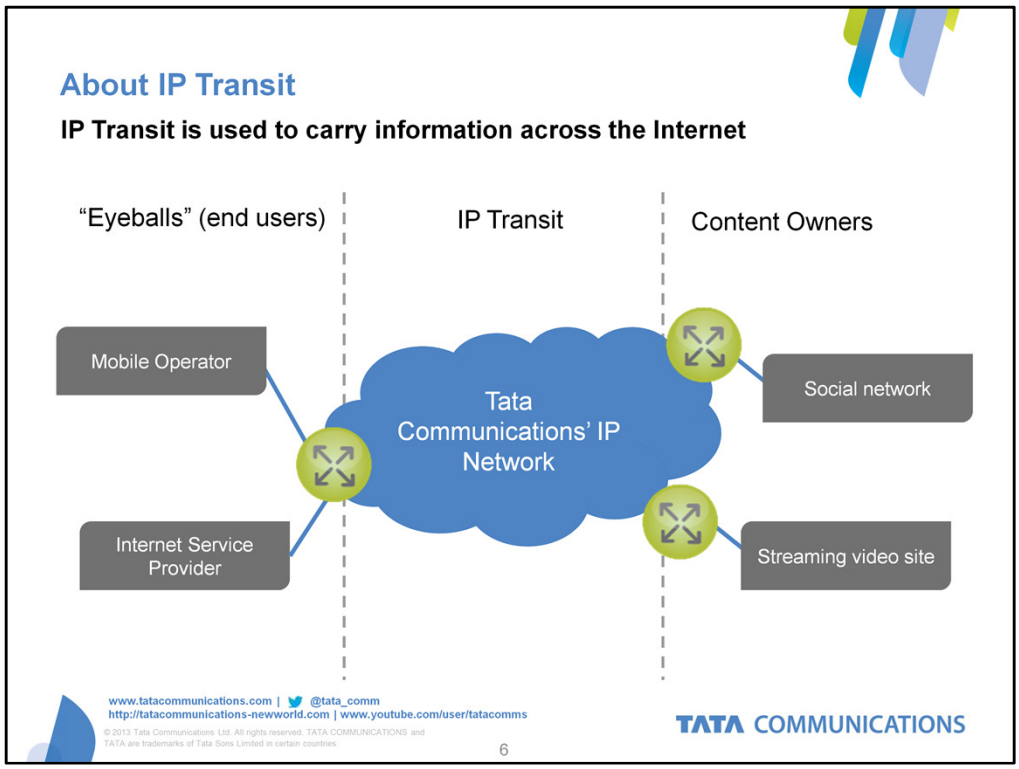
Calligno= calling number
Cust_orgno = customer ID number
Servabbr = the service level we use to terminate the call, or take it to its destination via our network of suppliers
Supp_orgno = the supplier id number
Pricingdestid = the ID of the pricing for the call
Routingdestcd = tells us where the call is going
Calldno = the called number

Calling and called numbers have been scrambled to preserve customer anonymity, so please don't try calling them! We have however preserved correlations, so a number is always scrambled the same way each time it appears.



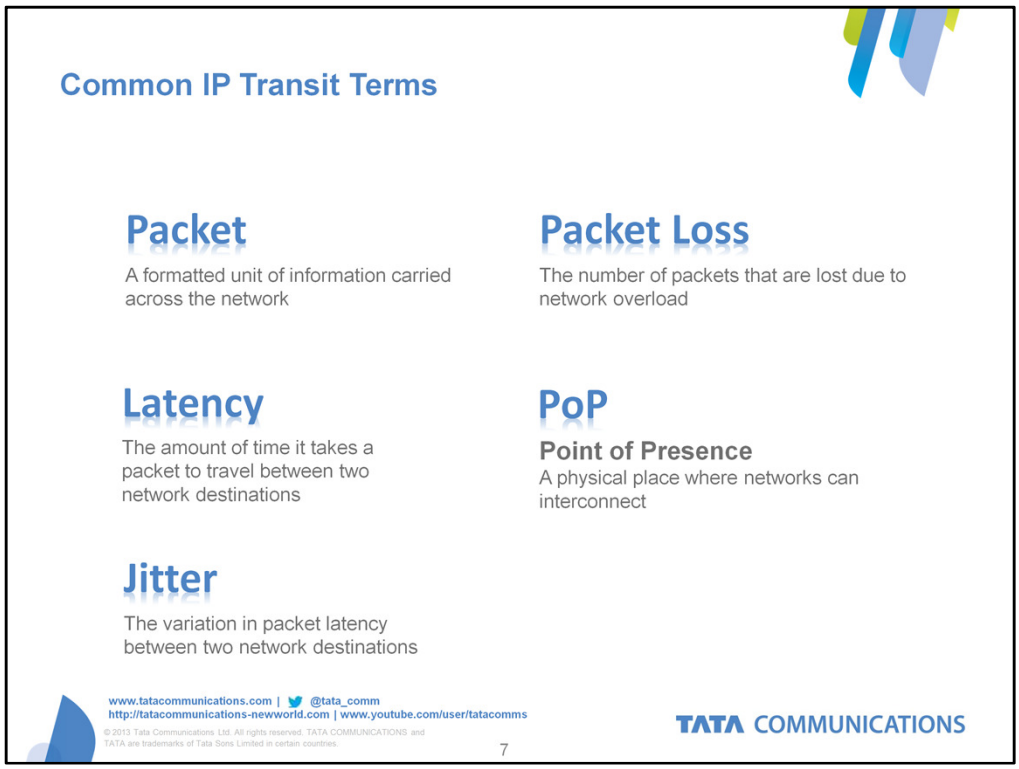
Our mobile signalling service is used to keep track of mobile user activity while they are roaming. This allows the home network to bill the user appropriately, and pay the visited network for the activity that takes place on their network.

Along with the data set is a detailed word document that describes the signalling messages that pass across our network. Signalling messages can go both directions between the home and visited network – in our data set they are described by the origination and destination of the individual message, and not necessarily consistently in reference to the subscriber.



IP transit is a service that carries information between networks connected to the Internet. Typically, end users request information from content or application providers – companies such as social networks, search engines, news or video sites, or any of the hundreds of other websites you might visit every day. User networks, such as ISPs, mobile broadband providers, or large enterprises, connect to our Tier 1 IP backbone as customers to reach the other networks and content sources that make up the Internet, while content providers connect to our network to reach the end users (known in the industry as "eyeballs") that consume their services.

A customer may take a circuit of any number of types, from a local loop from a regional provider to a piece of fibre in a data center, to connect to via a dedicated port to one of our routers at any of our Points of Presence, or physical places where we have located network equipment.



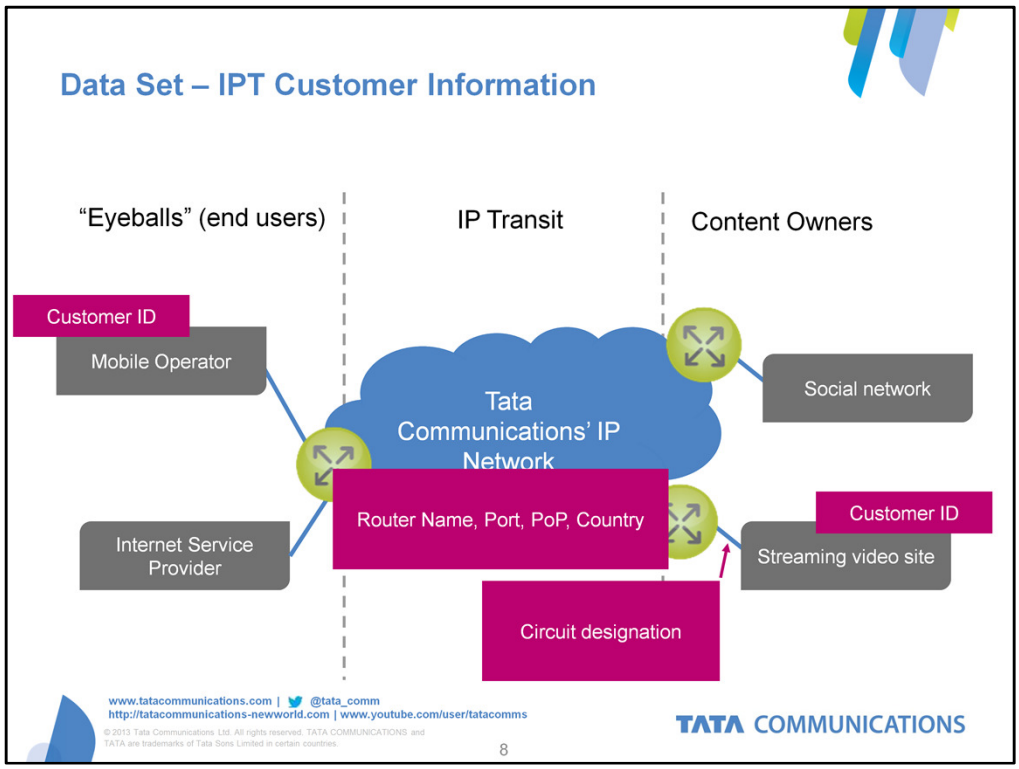
These are some of the common terms to use to understand IP transit.

Data is sent across the Internet in units called packets, which are formatted pieces of data that carry their origination and destination information, as well as a data packet.

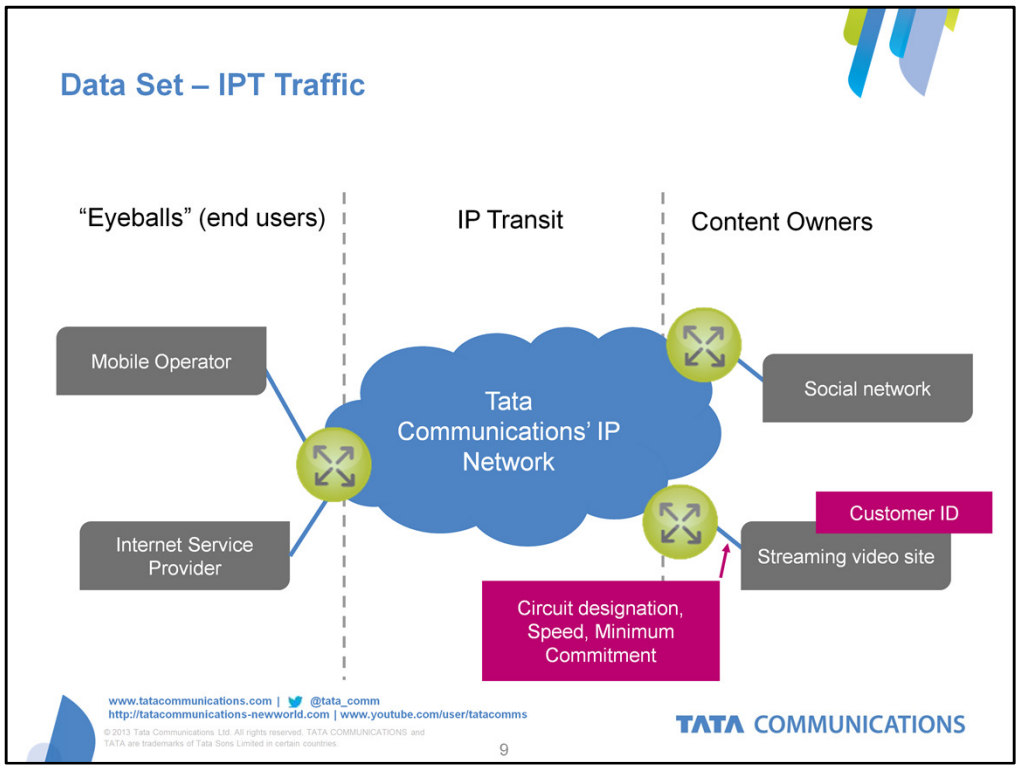
Latency, jitter and packet loss are the standard measures of network quality. Latency is the amount of time it takes the packet to pass through the various routers and fibre optic or copper wires between its origination and destination. Latency is typically measured in milliseconds – the latency between the East and West Coasts of the United States is somewhere around 75 milliseconds.

Jitter is the variation in latency for packets – this is important because high variation can result in packets arriving out of order, which affects the quality of applications such as voice calls or streaming video.

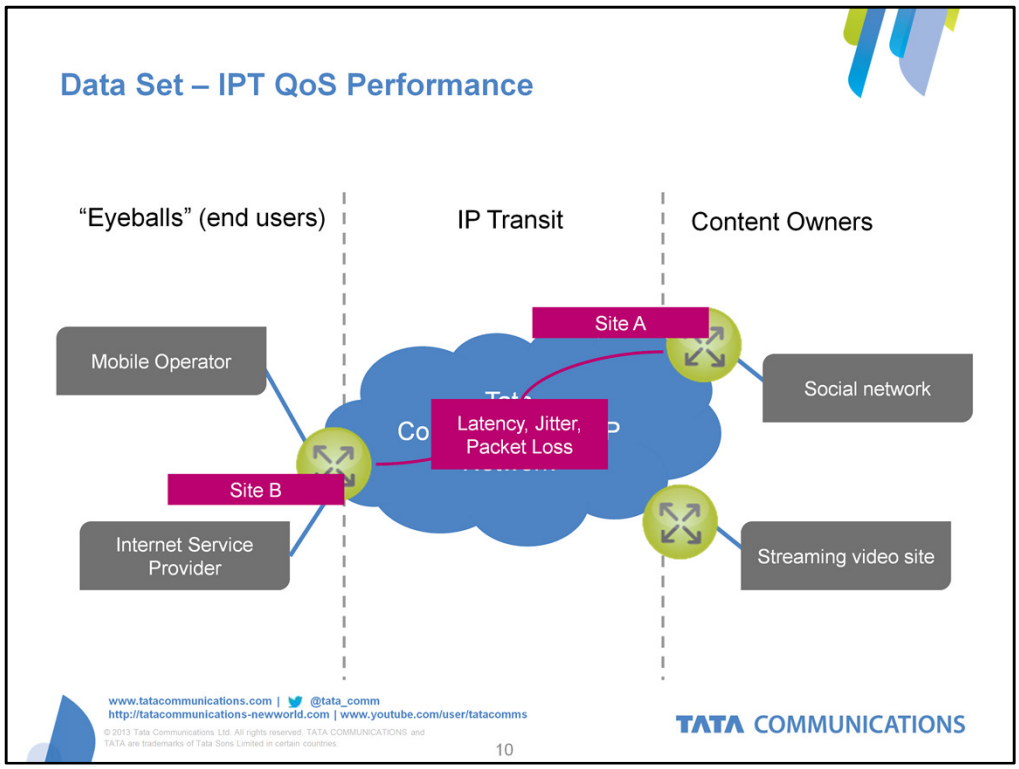
Finally, packet loss refers to packets that are lost due to being dropped from the router buffer when the network is overloaded.



There are three data sets that reference IP transit. The first focuses on the route that the packet takes from the customer to our network. Remember that customers can be both content and eyeball networks. Each magenta box shows where in the process each data point references.



The second data set focuses on the connection between our network and the customer. Speed refers to the amount of data flowing across that connection, while minimum commitment references the minimum capacity that we've promised to provide to that customer.



The third set of data focuses on what happens to packets as they cross our network, measuring latency, jitter and packet loss as they move between two endpoints on our network.

TATA COMMUNICATIONS

Thank you

Visit tatacommunications.com
Follow us on Twitter, our New World Blog and YouTube:

@tata_comm

http://tatacommunications-newworld.com

www.youtube.com/user/tatacomms