

VOIP EN REDES CORPORATIVAS

Adriana Ferreira

Operación y Mantenimiento
ANTELDATA
Montevideo, Uruguay
aferreira@anteldata.com.uy

Marcelo Pepe

Area Técnica
ANTELDATA
Montevideo, Uruguay
mpepe@anteldata.com.uy

Fernando Lopez

Operación y Mantenimiento
ANTELDATA
Montevideo, Uruguay
flopez@antel.net.uy

Julio Guani

Area Técnica
ANTELDATA
Montevideo, Uruguay
jguani@adinet.com.uy

Resúmen

Se analizan varios escenarios de implementaciones de Voz Sobre IP (Voice Over IP - VoIP) corporativas desde la perspectiva del proveedor de servicios Anteldata. Se comparan estas diferentes alternativas tanto a nivel de estandarización y crecimiento como a nivel técnico, comparando la calidad de la voz y los retardos. Se concluye con la solución más conveniente para los intereses de la empresa. Este trabajo recoge la experiencia de 3 años de Anteldata brindando soluciones de VoIP corporativas.

Palabras claves

Redes Corporativas, VPN, VoIP, calidad de servicio, retardo de jitter, retardo punta a punta.

INTRODUCCION

Anteldata es la Unidad de Negocios de Datos de Antel y dentro de su portafolio de servicios se encuentran las redes privadas ruteadas (de capa 3) con telefonía corporativa.

El objetivo es presentar un análisis de 3 implementaciones de este servicio:

1. La voz se transmite sobre el mismo enlace virtual (Private Virtual Circuit - PVC) que los datos, utilizando fragmentación, intercalado, compresión de encabezados, y priorizando el tráfico de voz frente al de datos.
2. La voz se transmite sobre el mismo PVC que los datos, pero solamente se utiliza priorización del tráfico de voz sobre el de datos.
3. La voz y los datos se transmiten sobre 2 PVCs distintos.

Los primeros servicios brindados por Antel se realizaron de acuerdo a la implementación 1, con velocidad de acceso en las sucursales de 64 Kbps y 2 canales de voz.

Se mencionarán los hechos que servirán de argumento a la conclusión de que esta implementación no es generalizada en el resto del mundo y es muy compleja de estandarizar, resultando muy poco escalable para un proveedor de servicios como Antel.

Esta situación motivó un estudio de alternativas más sencillas surgiendo entonces las alternativas 2 y 3.

Se presentarán los resultados del estudio de las 3 implementaciones, comparando los parámetros de calidad de la llamada, ancho de banda de acceso, calidad de

servicio (Quality Of Service – QoS) ATM, etc, y se compartirán las conclusiones que se desprenden de este estudio, incluso desenmascarando algunos “mitos” de la voz sobre IP como ser el ancho de banda real utilizado. Cabe acotar que el análisis se realiza sobre equipos Cisco^R y que las tecnologías utilizadas en capa 2 son Frame Relay y ATM.

VOZ SOBRE IP - VOIP

Tendencia

Por diversas razones, no solo tecnológicas, las empresas de telecomunicaciones se encuentra en un fuerte proceso de integración de aplicaciones de tiempo-real sobre Redes IP, en particular Voz (Telefonía), Fax, video, etc.

Telefonía Corporativa

En la figura 1 se muestra el escenario clásico sobre el cual se maneja la telefonía y las redes de datos a nivel corporativo.

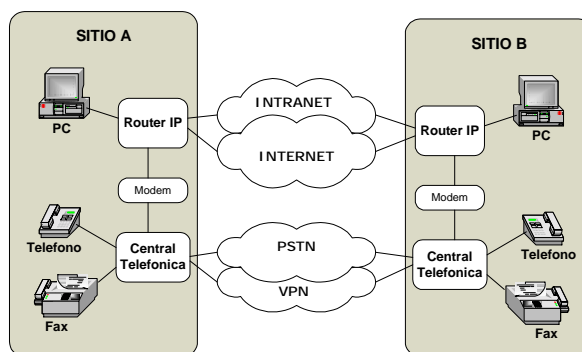


Figura 1. Arquitectura corporativa clásica.

Está planteado un horizonte de solución tecnológica con teléfonos IP y prescindiendo de centrales de conmutación de circuitos, pero en la actualidad hay pocas implementaciones de este modelo y se utiliza la tecnología de voz sobre IP (VoIP – Voice over IP), principalmente para las conexiones troncales entre centrales telefónicas (PBX) como se muestra en la figura 2:

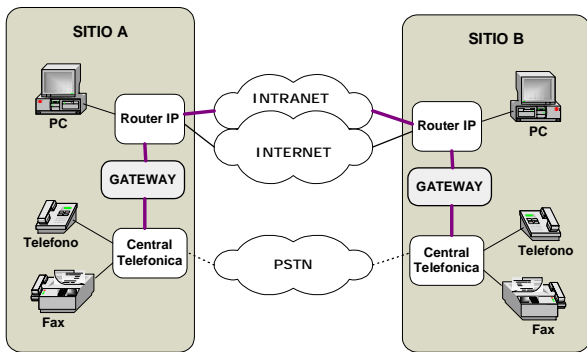


Figura 2. Arquitectura VoIP más difundida.

El "Gateway" de Voz/IP es el componente clave de una solución de voz sobre IP al facilitar la conversión de las llamadas telefónicas convencionales al mundo de IP. Normalmente, tienen interfaces analógicas o digitales a la red telefónica, y disponen de interfaces Ethernet, Frame Relay o ATM hacia la red IP

El llamado Gatekeeper es un punto central de control en una red H.323 [1] proporcionando servicios de control de llamada, traducción de direcciones y control de admisión. Además facilita el control del ancho de banda utilizado y localiza los distintos gateways.

El principal beneficio que motiva esta convergencia es la reducción de costos, que debería ser analizada en cada caso. En este análisis, además de la inversión en equipamiento específico de VoIP (Gatekeeper, terminales, Gateways), hay que considerar los costos asociados a los nuevos requerimientos de la red IP y el costo de un soporte técnico más especializado.

Calidad de la Voz

Es un concepto subjetivo que mide la fidelidad e inteligibilidad de la conversación percibida por quienes experimentan la misma.

En el caso de redes que utilizan tecnología VoIP, los principales elementos que afectan la calidad de voz son Claridad y Retardo punta-a-punta.

Claridad

La medida más popular de claridad es el índice Mean Opinión Score (MOS) según especificación UIT-P.800 [2], que se implementa a través de aplicación de test predeterminados a un grupo significativo de personas. Otras formas de medir la claridad son el PSQM y el PAMS, todos manejan la misma escala de 1 a 5 como se puede observar en la Tabla 1.

La claridad depende fundamentalmente del Codificador-Decodificador (CODEC) y de la pérdida de paquetes en la red. También influyen el ruido, el Voice Activity Detection (VAD), que suprime paquetes de silencio, y el eco.

En la tabla 1 se muestran algunas características de los codecs más comunes.

Codecs	Código	Bit rate en Kbps	MIPS	MOS
G.711	MIC comp.	64	<<1	4,0
G.722	ADPCM	32-64	~1	3,9
G.728	LD-CELP	16	5	3,6
G.729	CS-ACELP	8	20	3,9
G.723,1	MPC,MCQ y ACELP	6,3-5,3	18	3,9-3,6

Tabla 1. Características de los diferentes codecs.

Retardo punta-a-punta

Es el tiempo que demora la señal de voz desde el que habla (emisor) hasta el que escucha (receptor). El mismo se compone fundamentalmente del retardo de codificación/decodificación, retardo de paquetización, el retardo de jitter y del retardo en la red.



Figura 3. Principales componentes del retardo.

Codecs	frames en bytes	Retardo del algoritmo	Retardo de procesamiento
G.711	1	~0	~0
G.728	10	0,6 ms	~0
G.729	10	15 ms	5 ms
G.723,1	20-24	37,5 ms	5 ms

Tabla 2. Retardo de los diferentes codecs.

El retardo de paquetización depende de cuantos frames se agrupan en un paquete IP, en el caso de que cada paquete IP se forme con un frame este retardo es nulo.

El retardo del buffer de jitter se implementa para absorber las variaciones del retardo en la red (jitter).

Conclusión

Para transportar Voz sobre una red IP es necesario que esta cumpla requerimientos de retardo, variación del retardo y pérdida de paquetes, que no eran necesarios para el transporte puro de datos (aplicaciones que no sean en tiempo-real).

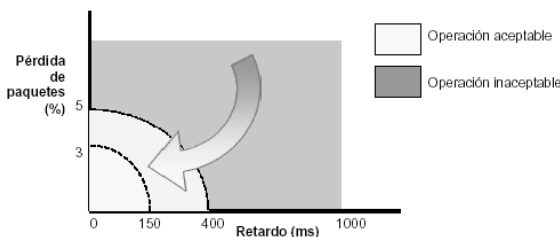


Figura 4. Requerimientos de la Red según ITU G.114 [2]

SITUACIÓN ACTUAL DE ANTELDATA

Descripción de la situación

Actualmente existe gran demanda por soluciones integradas para las redes privadas corporativas – de capa 3 - en las cuales se incluya el servicio de voz. ANTELDATA cuenta para ello con un soporte a nivel de capa 2, que básicamente se da con ATM [3], enlaces sincrónicos, Frame-Relay (FR)[4] y Asynchronous Digital Line Subscriber (ADSL), además de una capa de red integrada por distintos modelos de equipos Cisco, en su mayoría. Por una cuestión de costos, las redes corporativas se implementan principalmente con tecnologías XDSL.

En este escenario, y teniendo en cuenta el uso generalizado de IP, la decisión fue brindar voz sobre IP.

Los clientes de esta solución (VoIP), reclaman la misma calidad de voz que ellos perciben con servicios de telefonía básica. Por tratarse de una aplicación de tiempo real, ésta aplicación es extremadamente sensible al retardo. Por tal motivo desde un principio se enfocaron los esfuerzos en pos de brindar la mejor calidad de servicio en lo referente a la voz, y para esto se hizo uso de características y facilidades a configurar en los equipos de capa 3 utilizados. En dichos equipos, los procesadores digitales de señal (DSPs) segmentan la señal de voz en marcos y los almacenan en paquetes de voz. Estos paquetes son transportados vía IP conforme con la especificación H323 de la ITU-T[2], que especifica el envío de multimedia (voz, video y datos) a través de una red.

Funcionamiento y funcionalidades

Funcionamiento

El servicio que brinda ANTELDATA de Redes Privadas Virtuales (VPN - Virtual Private Network) con Telefonía Corporativa tiene alcance nacional y en general es contratado por empresas que tienen sus sucursales distribuidas en el territorio nacional.

La topología de las VPNs es en estrella, considerando al centro de la misma como el sitio Central y al resto como Sucursales. El sitio central se une a una sucursal mediante un único camino virtual, por el cual se trafica conjuntamente la voz y el resto de la información (que de aquí en más se nombrará simplemente como “datos”). El acceso a las sucursales es ADSL (ATM) y Frame Relay

(FR) para aquellos puntos sin cobertura ADSL (5% aproximadamente). En el punto central el acceso es FR.

El esquema es el siguiente:

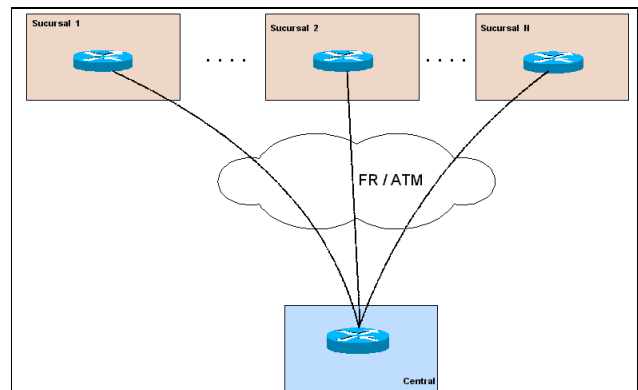


Figura 5. Topología de las redes privadas

Funcionalidades

Para que los usuarios puedan percibir un nivel de calidad de voz aceptable, se debe garantizar para este tráfico ciertos requerimientos de ancho de banda, retardo y variación del retardo.

Para asegurar calidad, se buscan las siguientes características:

- Mantener un ancho de banda mínimo garantizado para la voz
- Reducir la variación del retardo

A continuación se detallan las prestaciones utilizadas para cumplir con las características antes mencionadas:

ancho de banda

Los enlaces mas comunes son de 64 Kbps y este ancho de banda es compartido entre la voz y los datos. Como consecuencia se decidió garantizarle a los paquetes de voz un cierto porcentaje del total del ancho de banda y darle determinada prioridad.

Reducción de ancho de banda - cRTP

El protocolo de transporte de tiempo real (RTP- Real Time Transport Protocol), provee funciones de red de punta a punta y de servicios para datos de tiempo real sensibles al retardo, como ser la voz o el video. RTP trabaja con encolamiento para dar prioridad al tráfico de voz sobre otro tráfico, los servicios incluyen: identificación del tipo de carga, numeración de secuencia, colocación de marcas de tiempo, entre otros. Un encabezado RTP contiene una marca de tiempo y un número de secuencia, permitiendo al dispositivo receptor encolar tantos como sea necesario para remover la variación del retardo y el retardo mediante sincronización de paquetes, todo esto para obtener un flujo continuo de sonido.

Para evitar el consumo innecesario del ancho de banda disponible, se comprime el encabezado RTP (cRTP). Así, en todo un tramo, se comprime los encabezados de los paquetes de voz, incluyendo el de IP/UDP/RTP logrando reducir los 40 bytes de encabezados a sólo entre 2 y 4 bytes. Existe la recomendación de usar cRTP en enlaces menores a los 768kbps [5].

Selección del CODEC

En VoIP se usan CODECs de baja velocidad de bit. A pesar de que el G.723 usa menor ancho de banda que el G.729 (5.3kbps y 8kbps respectivamente), en la práctica se ha percibido una mejora notoria en cuanto a calidad de la voz con el uso del último. Ver tablas 1 y 2.

Nota: Se configuran los equipos de modo de transportar tonos DTMF fuera de la banda separadamente del flujo de la voz codificada, para evitar que sean distorsionados debido a la optimización de los codecs de baja velocidad.

Habilitación de la detección de actividad de voz (VAD)

Una conversación típica contiene entre el 35% y 50% de silencios. Usando la característica que permite detectar la actividad de voz VAD, los paquetes de silencio son suprimidos. Esto se traduce en una reducción del 35% en el ancho de banda en VoIP [6].

Clasificación y marcado

La clasificación identifica la clase o grupo al cual un paquete pertenece. Los routers usan varios criterios de correspondencia para situar al tráfico en un cierto número de clases, como ser listas de acceso, puertos de entrada, marcado de los bits del campo Type Of Service (ToS) del paquete IP, para identificar tráfico como de voz.

Las VPNs de Anteldata, clasifican los paquetes de voz (incluyendo a los de señalización de voz) mediante el marcado de DSCP, contribuyendo con la baja pérdida, bajo retardo, baja variación del retardo, y asegurando buen servicio en el ancho de banda [7].

El método de clasificación y marcado de paquetes, recomendado por Cisco, es la llamada “Modular QoS Command-Line Interface” (MQC) , y consiste en separar la clasificación de la política de tráfico (conocida como “policy”), permitiendo múltiples características de calidad de servicio (QoS - Quality of Service) configuradas todas juntas para múltiples clases. Se usa un mapa de clase para clasificar el tráfico basado en varios criterios de correspondencia y un mapa de policy para determinar qué debería suceder con cada clase. Luego se aplica la policy al tráfico de entrada o de salida de una interfaz.

Priorización del tráfico de voz

Low Latency Queueing

Luego de que todo el tráfico haya sido ubicado en clases con determinada QoS basándose en sus requerimientos de calidad, se garantiza un ancho de banda y una prioridad de servicio por medio de un mecanismo de encolamiento de salida inteligente.

Para VoIP es necesario una cola de prioridad . Se puede usar cualquier tipo de encolamiento para darle a los paquetes de voz alta prioridad, pero se recomienda usar la llamada Low Latency Queueing (LLQ) por ser flexible y fácil de configurar.

LLQ usa el método de configuración MQC para dar prioridad a una clase y reservar un mínimo ancho de banda para otras clases de tráfico. Hay también una clase por defecto que es usada para determinar el tratamiento de todo el tráfico no clasificado, en donde cada flujo compartirá aproximadamente igual cantidad del restante ancho de banda disponible [6].

Con LLQ, los datos sensibles al retardo, como la voz, son colocados en la cola de mayor prioridad y son los primeros en ser enviados. Esto es así a menos que este tráfico exceda el ancho de banda de prioridad configurado para dicha cola y que ese ancho de banda lo necesite alguna de las colas reservadas (es decir, cuando existe congestión). Cuando una interfaz se congestiona, la cola de mayor prioridad es atendida hasta que la carga alcance el valor de Kbps configurado, asignado a ese tráfico. El tráfico excedente es descartado.

Las clases reservadas son atendidas de acuerdo a su ancho de banda configurado, el cual es usado para calcular un peso. El peso es usado para determinar con qué frecuencia es atendida una cola reservada y cuántos bytes son enviados a la vez – esto está basado en el algoritmo conocido como Weighted Fair Queueing (WFQ) [5].

La Figura 6, muestra cómo trabaja LLQ.



Figura 6. Funcionamiento del encolamiento LLQ

Lo que sigue es el procedimiento a llevar a cabo para configurar LLQ:

- Crear un mapa de clases para el tráfico de VoIP y definir criterios de correspondencia (por puertos UDP, puertos TCP de señalización de VoIP, etc).
- Crear un mapa de política de tráfico (policy map) y asociarlo a los mapas de clases de VoIP.
- Habilitar LLQ: esto se logra aplicando el policy map a la interfaz de salida WAN .

Fragmentación e Intercalado

Puesto que el tráfico de voz debe compartir el enlace con el resto de los datos, y dado que el primero es sensible al retardo, es necesario intercalar o insertar los paquetes de voz entre los paquetes de datos previamente fragmentados. Aún con una buena estrategia de colas y priorizando el tráfico de voz, hay veces que la cola con prioridad está vacía y un paquete de otra clase es entonces atendido. Si un paquete priorizado de voz alcanza la cola de salida mientras se está atendiendo dicho paquete, deberá esperar una cantidad de tiempo considerable antes de ser enviado. Si se asume que el paquete VoIP debe esperar detrás de un paquete de datos, y que tal paquete puede ser como máximo igual en tamaño a la MTU (1500 bytes para una interfaz), es posible calcular el tiempo de espera tomando en cuenta la velocidad del enlace [5].

Para un enlace de 64 kbps y una MTU de 1500 bytes, se tiene el siguiente retardo de serialización:

$$\text{Retardo serialización} = (1500\text{bytes} * 8\text{bits/byte}) / (64000 \text{ bits/s}) = 187.5 \text{ ms}$$

Por lo tanto, en un enlace de 64 kbps un paquete VoIP necesitaría esperar hasta 187.5 ms antes de que pueda ser enviado. Tomando en cuenta que la especificación de ITU G.114 [2] recomienda un retardo punta a punta menor a 150 ms para una buena calidad de voz, ese valor de retardo no es aceptable.

Es entonces necesario fragmentar los paquetes largos de datos de modo de reducir el tiempo que toma enviar cada fragmento y así bajar el retardo de los paquetes de voz. Si bien se recomienda un tamaño de fragmento tal que el tiempo de serialización sea de 10 ms, en la práctica se vio necesario ajustar este valor, cosa no muy fácil de lograr. El tamaño de fragmento, el tiempo de envío del mismo y la velocidad del enlace se relacionan de la siguiente manera:

$$\text{Tamaño fragmento} = (t \text{ envío [s]} * BW \text{ [bps]}) / (8 \text{ [bits/byte]})$$

No basta con fragmentar únicamente, pues un paquete de voz podría llegar a tener que esperar por todos los fragmentos de un mismo paquete de datos, por tanto es imprescindible intercalar paquetes de voz entre los fragmentos de datos.

Existen tres estrategias para la fragmentación e intercalado, una de las cuales es válida sólo para enlaces puramente FR (y por tanto no se aplica a enlaces FR-XDSL). De las dos restantes, se ha descartado la utilización de la que propone cambiar el tamaño de la MTU, ya que aunque es simple exige que quien se encargue del reensamblado sea la aplicación final y depende además de que los paquetes IP no estén marcados como de no-fragmentación. Por tales motivos, se usa la Fragmentación y el intercalado del protocolo Multilink Point-to-Point (MLPPP); tal protocolo configura enlaces PPP tanto sobre FR como sobre ATM. En este caso se determina el tamaño de los fragmentos y se habilita el intercalado en la interfaz. Los paquetes son fragmentados en un extremo del enlace y reensamblados en el otro [5].

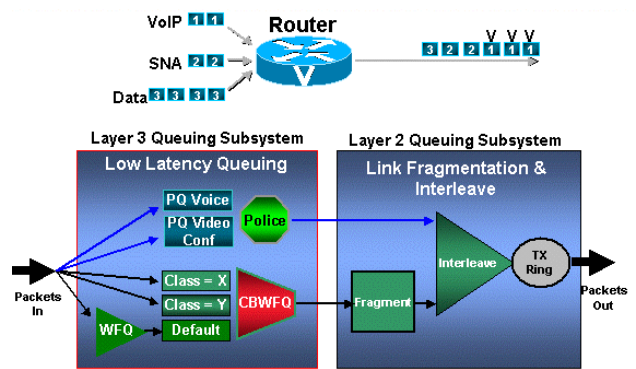


Figura 7. Estrategia de encolamiento y Fragmentación e intercalado.

Modelado del tráfico

En el sitio central de una VPN se utiliza FR. Esto hace que se deba modelar el tráfico pues el reloj de la interfaz es mayor que el ancho de banda garantizado (Committed Information Rate – CIR) para el enlace que tiene con cada sucursal. Mediante el llamado “Traffic shaping” se envían ráfagas pequeñas de tráfico.

Sin el modelado el sitio central intentaría enviar a velocidades mayores que las que los sitios remotos podrían recibir, causando descarte de tráfico.

Resultados y evaluación

La experiencia en la operación de la red fue mala. Se descubrieron varios problemas que afectaban básicamente a la voz, estos en su gran mayoría aparecen debido a problemas, tanto con el propio sistema operativo de los equipos Cisco (IOS) como por incompatibilidades de versiones de estos IOS.

Estos problemas de IOS se debían básicamente a “bugs” relativos a MLPPP, LLQ y al protocolo que se utiliza entre

gateway y gatekeeper para el mapeo dinámico de los números telefónicos y las direcciones IP correspondientes.

El esquema a configurar no es sencillo, ya que requiere de muchos parámetros y la configuración de múltiples interfaces virtuales por enlace en el sitio central, esto trae consigo un problema de escalabilidad ya que el número total de interfaces está limitado en la mayoría de los IOS.

La incompatibilidad entre versiones de IOS hace muchas veces necesaria la ampliación de memoria para algunos equipos de cliente aumentando su costo.

Para solucionar los problemas encontrados se tuvo que homologar con precisión las configuraciones y los IOS de las distintas plataformas. Cabe notar el problema de tener que realizar una homologación cada vez que sea necesario adquirir un router nuevo (que se da, por ejemplo, en caso de discontinuidad de un equipo), puesto que contará con diferente IOS. Esto también trae consigo problemas de escalabilidad.

ANÁLISIS DE OTRAS ALTERNATIVAS

Descripción de alternativas

Debido a los problemas de escalamiento y estandarización de la solución de redes corporativas mencionada en el punto anterior y que actualmente está brindando Anteldata, se resolvió estudiar alternativas más simples, que permitan dar el mismo tipo de servicio y que impliquen una configuración más sencilla a nivel de los equipos de capa 3 (routers). Cabe notar que esta era una de las limitantes en el escalamiento y estandarización en la solución anterior ya que cualquier modificación o inclusión de nuevos comandos en el router podía traer consigo tanto un cambio de sistema operativo del router, como el agregado de más memoria o incluso la imposibilidad de utilizar algún equipo ya adquirido para dar este servicio.

A partir del estudio de las alternativas posibles se encontraron dos implementaciones que podían cumplir con los requerimientos buscados.

Estas alternativas son las siguientes:

- La voz se transmite sobre el mismo PVC que los datos, pero solamente se utiliza priorización del tráfico de voz frente al de datos.
- La voz y los datos se transfieren sobre dos PVCs distintos.

A continuación se describe cada una de estas dos implementaciones y se muestran los resultados y conclusiones obtenidas del estudio de cada una de ellas, así como también algunos resultados generales interesantes que se obtuvieron.

Análisis de alternativa 1 PVC

A su vez dentro de esta alternativa se estudian 2 posibles implementaciones:

- Se limitan los datos mediante un mecanismo llamado Committed Access Rate (CAR) [8] propio de los equipos Cisco.
- Se prioriza la voz mediante el mecanismo LLQ

Alternativa 1 PVC con CAR

Descripción

Esta alternativa transmite voz y datos para un mismo PVC, utilizando la funcionalidad llamada Committed Access Rate (CAR) para limitar los datos, y de esta forma dejar un ancho de banda suficiente para la voz.

La topología utilizada es en estrella como ya fue expuesto.

Se comenzó el estudio de esta alternativa conectando únicamente dos routers, para luego simular una red con más sucursales.

El router central tiene enlace FR [4] mientras que el router sucursal enlace ATM, en particular ADSL.

Los equipos utilizados fueron, para las sucursales, Cisco 827-4V y para el central un equipo Cisco 1751. El Cisco 827-4V tiene una interface Ethernet, una ATM y cuatro puertos de voz. El 1751 contiene tarjetas VIC-2FXS (con 2 puertos de voz), WIC-1T (con interface serial), así como también una interface FastEthernet.

Las velocidades (o access-rate) involucradas en cada una de las interfaces son, de 128 Kbps para el lado ATM mientras que para FR es de 512 Kbps. El PVC entre los equipos se crea de 128 Kbps (ver figura 8).

Se configuró CAR en ambas puntas. Con esta prestación se puede definir límites de velocidad para determinado tráfico de paquetes (en este caso particular: datos), a partir de los cuales se define qué paquetes cumplen o exceden las condiciones preestablecidas, las cuales se definen con los siguientes parámetros:

- Average rate. Determina la velocidad de transmisión en promedio. El tráfico que está por debajo de esta velocidad cumple con las condiciones.
- Normal burst: Es el tamaño de ráfaga máximo.
- Exceed burst: Cuando el tamaño del exceed burst supera al del normal burst, se permite una ráfaga adicional cuyo tamaño es la diferencia entre el exceed y el normal burst.

CAR aplica un criterio dependiendo de si las condiciones preestablecidas se cumplen (Conform) o no (Exceed Action). En este caso particular los paquetes clasificados como conform son transmitidos, mientras que los restantes se descartan.

En la práctica se configuraron los siguientes parámetros: normal burst size de 1500 bytes, un exceed burst size de 2000 bytes (estos son los valores mínimos), y en un comienzo se configuró un average rate de 32 Kbps, el cual luego se fue modificando. En lo referente al burst y al exceed burst size, se tomaron dichos valores porque aseguran que sólo un paquete de datos se puede anteponer a un paquete de voz. Si se envían dos paquetes de datos de 1500 bytes consecutivos, el segundo será descartado ya que 3000 bytes exceden el extended burst size (2000 bytes).

En cuanto a la codificación y las prestaciones utilizadas para la voz se utilizó el codec g729r8 que realiza compresión CELP (Code Excited Linear Predictor) y que permite que la voz sea codificada en tramas de 8 Kbps. Este codec generalmente provee una calidad de voz tan buena como Adaptive Differential Pulse Code Modulation (ADPCM) de 32 Kbps. No fue utilizada la prestación de VAD, dado que su inclusión provocaría en determinadas ocasiones una sensación de entrecorte de la llamada, lo que haría descender el MOS.

Se configuró un tamaño de 40 bytes para el payload de los paquetes de voz pues esto disminuye el ancho de banda ocupado por los mismos. La explicación de este punto se encuentra en la sesión “Análisis alternativa 2” (Resultados).

Para realizar pruebas en esta solución y con el fin de generar el tráfico suficiente, se realizan flujos de pings desde una computadora conectada en la LAN del sitio sucursal (827), hacia otra conectada en la LAN del sitio central (1751). También se envían pings de 6000 bytes espaciados, ya que esto provoca el envío de cuatro paquetes de 1500 bytes y luego una pequeña pausa. Esta última situación es la más desfavorable para el buffer de jitter dinámico, pues le es difícil adaptarse a estas situaciones cambiantes.

Las pruebas de voz consisten en establecer dos llamadas cruzadas simultáneas. Estas llamadas se realizaban conjuntamente con los pings especificados en el párrafo anterior. Luego se analizaban los resultados tanto a nivel del escucha como a través de algunos valores que proporciona el router.

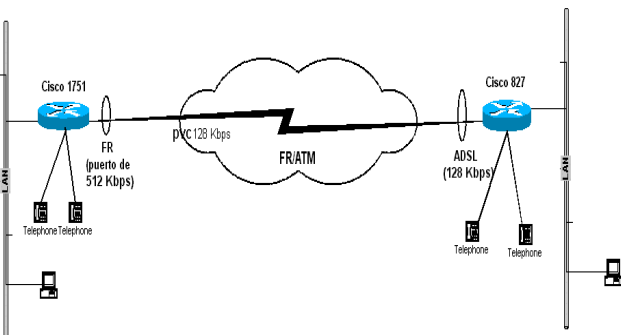


Figura 8. Alternativa 1 PVC con CAR

Resultados

Con los parámetros de CAR del punto anterior, se procedió a realizar las pruebas antes mencionadas, encontrándose una calidad de voz aceptable aunque un mayor retardo del esperado. Teóricamente el retardo de jitter máximo era de un paquete de datos, es decir, $1500(\text{bytes}) * 8(\text{bits/bytes}) / 128\text{kbps} = 93,75 \text{ ms}$. Sin embargo, el retardo obtenido en la práctica es de aproximadamente 200ms, lo cual es comparable al retardo provocado por dos paquetes de datos.

Luego, se procedió a subir el average rate de 32 kbps a 64 kbps y a 96 kbps, encontrándose que tanto el retardo como la calidad de voz perceptibles se mantenían constantes.

Esto sí concuerda con lo esperado dado que el exceed burst size es de 2000 bytes para todos los casos, por lo que no debería anteponerse más de un paquete de datos a los de voz.

Finalmente se probó agregando 2 sucursales adicionales y enviando tráfico por todas las sucursales simultáneamente. Nuevamente en esta oportunidad, no empeoró la situación, esto se debe a que el sector más comprometido es del lado ATM que para todas sus sucursales el retardo teórico continúa siendo de $1500 * 8 / 128000 = 93,7\text{ms}$, mientras que para el lado FR ahora se pueden anteponer tres paquetes de datos (uno por cada subinterfaz), por lo que el retardo de jitter máximo teórico en este caso es de $1500 * 3 * 8 / 512000 = 70,3\text{ms}$.

Conclusión

En esta solución se logró una calidad de voz aceptable, con retardos de aproximadamente 200ms, sensiblemente superior al recomendado (menor a 150ms).

Es importante remarcar que los datos prácticos no concordaron con los teóricos.

Alternativa 1 PVC con priorización LLQ para voz.

Descripción

Esta alternativa utiliza un mismo PVC para voz y datos, pero en esta oportunidad se utiliza prioridad LLQ para preferir el tráfico de voz respecto al de datos.

La topología utilizada es en estrella, idéntica al caso anterior, con el mismo ancho de banda para el PVC (128Kbps) e iguales velocidades de serialización (512kbps router central y 128Kbps router sucursal).

Para lograr aplicar prioridad LLQ en la punta ATM (827) se debió configurar calidad de servicio ATM Variable Bit Rate – No Real Time (vbr-nrt) aplicando traffic-shaping, de lo contrario, no era posible aplicar la clase para la prioridad LLQ en la interfaz. En la punta FR (1751) el traffic-shaping no era un requisito necesario por lo cual no fue configurado. Los parámetros de shaping configurados en primera instancia fueron: Peak Cell Rate (PCR) = 128kbps, Sustainable Cell Rate (SCR) = 64kbps y un Burst = 2 celdas [3].

Se empleó el codec g729r8 y 40 bytes de payload como en los casos anteriores (ver Análisis alternativa 2 PVCs - resultados).

Para realizar pruebas en esta solución y con el fin de generar el tráfico suficiente, se realizan flujo de pings desde una computadora conectada en la LAN del sitio sucursal (827), hacia otra conectada en la LAN del sitio central (1751). También se envían, al igual que en la implementación anterior, pings de 6000 bytes espaciados, para provocar el envío de cuatro paquetes de 1500 bytes y luego una pequeña pausa.

Las pruebas de voz consistían en establecer dos llamadas cruzadas simultáneas. Estas llamadas se realizaban conjuntamente con los pings especificados en el párrafo anterior. Luego se analizaban los resultados tanto a nivel de apreciación auditiva como a través de algunos valores que proporciona el router.

Resultados

Con los parámetros de shaping especificados en el punto anterior se procedió a realizar las pruebas antes mencionadas, encontrándose una calidad de voz sumamente deteriorada. Una explicación de este efecto es que para cumplir con los parámetros de shaping es necesario limitar el tráfico a la salida y de esta forma cumplir con el contrato establecido, esto hace que paquetes de voz y datos indistintamente sean retardados a la salida, aumentando particularmente el retardo en los paquetes de voz.

Por lo antes expuesto, se procedió a cambiar los parámetros de shaping de forma tal de cancelar su efecto PCR=128kbps, SCR=128kbps y un Burst=33celdas.

Aquí el problema pareció mejorar considerablemente. Sin embargo, los cortes que se producían en las llamadas de voz aún eran excesivos.

El problema nuevamente se encontraba del lado del equipo Cisco 827. La explicación al problema es la siguiente: para transmitir un paquete primero se acumula en la cola de capa 3, donde actúa la priorización (LLQ). Luego, el 827 toma paquetes de esta cola y los coloca en un buffer circular, el "transmit ring" o "tx-ring"[9]. Luego el hardware/software de Segmentation And Reassembly (SAR) toma a su vez paquetes desde el tx-ring y los fragmenta en celdas aplicándose luego el tráfico shaping a la salida de ATM.

Por tanto, si llega un paquete de voz, la priorización (LLQ) lo pone delante de la cola de capa 3. Sin embargo para ser transmitido debe esperar que los paquetes que se encuentran en el tx-ring sean transmitidos, lo cual adiciona un retardo.

Por defecto, el tamaño de esta cola circular es de 16 paquetes, por lo que se llevó este valor al mínimo configurable de 2 paquetes.

No se puede disminuir el valor de la cola a un paquete. Esto abre la posibilidad que en el peor caso, un paquete de voz

tenga que esperar para ser transmitido el envío de dos paquetes IP que se encuentren en el tx-ring produciendo la variación del retardo correspondiente.

Por lo antes expuesto el retardo esperado en este caso será:
 $2 * 1500 * 8 / 128 = 187.5ms$

En la práctica el retardo obtenido fue de aproximadamente 200ms lo cual concuerda con lo esperado.

Por último se probó disminuir el tamaño de la Maximum Transfer Unit (MTU) de los paquetes a 540 bytes. Esto bajó el retardo a poco menos de 100ms pues, si bien la cola circular puede contener como máximo dos paquetes, ahora estos son de 540 bytes.

Conclusión

En esta oportunidad los datos prácticos se acercaron a los teóricos, obteniendo una calidad de voz aceptable, aunque aquí también con retardo excesivo.

En lo referente a las pruebas realizadas con MTU=540 bytes si bien los resultados fueron muy superiores, el inconveniente es que la fragmentación IP puede producir problemas a nivel aplicaciones IP, el más conocido de ellos es el relacionado con "Path MTU Discovery"[10] por lo que esta opción fue descartada

Análisis alternativa 2 PVCs

Descripción

Esta alternativa se basa en la creación de dos PVCs por cada conexión entre routers, uno exclusivamente para la voz y otro exclusivamente para los datos.

Al igual que en las implementaciones anteriores la topología utilizada es en estrella en donde uno de los routers funciona como router Central y es a él que llegan todos los enlaces de las sucursales.

Se comenzó el estudio de esta alternativa conectando únicamente dos routers, para luego simular una red corporativa completa.

Uno de los equipos tiene enlace Frame Relay (FR) mientras que el otro tiene enlace ATM, en particular ADSL.

Los equipos utilizados fueron, Cisco 827-4V (lado ATM) y Cisco 1751 (lado FR).

Las velocidades (o access-rate) involucrados en cada uno de los enlaces son, 64 Kbps del lado ATM mientras que del lado FR es de 512 Kbps. El access-rate elevado del lado FR se debe a que el tamaño de los frames en FR es de 1500 bytes y, al no utilizarse fragmentación, estos deben ser sacados de la interface a una alta velocidad para que no influya demasiado sobre los paquetes de voz que no deben experimentar un retardo elevado.

Se crean dos PVCs entre los equipos, uno para datos y otro para voz. Estos se crean con calidad de servicio ATM

Unspecified Bit Rate (UBR) para el pvc de datos y Variable Bit Rate – No Real Time (VBR-NRT) para el pvc de voz [3], priorizando por lo tanto el PVC de voz a nivel de capa 2. De esta forma el retardo que sufre un paquete de voz es como máximo el que genera una celda ATM, $53 \times 8 / 64000 = 6.63$ ms. Notamos que en la configuración del servicio vbr-nrt se deben configurar los parámetros de traffic-shaping, haciendo que el enlace de datos tenga PCR = 64 y el de voz PCR = 64 y SCR = 64[3].

Del lado FR en principio no se prioriza ninguno de los PVCs pero se configura traffic-shaping en ambos. En este caso los valores de traffic-shaping configurados fueron los siguientes Committed Information Rate (CIR) = 64000, Committed Burst Size (Bc) = 640 y Excess Burst size (Be) = 0 para los datos y CIR = 64000, Bc = 15000 y Be = 0 para la voz.

A nivel de capa 2 se mantienen los parámetros de CoS ATM y traffic-shaping a lo largo de toda la red para el PVC que corresponda, incluso se debe configurar en el último tramo del lado ATM, tramo DSLAM (DSL Access Multiplexer) – 827, de lo contrario se producen pérdidas de paquetes de voz en el sentido FR - ATM.

En la figura 9 se muestra un esquema de la configuración.

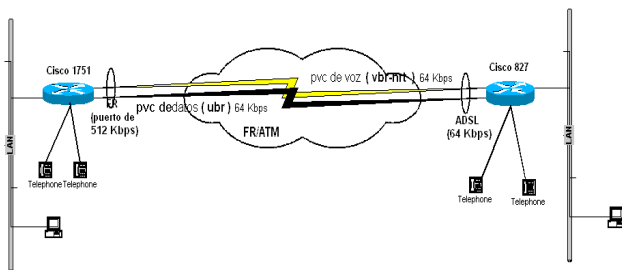


Figura 9. Esquema de prueba con 2 PVCs.

En cuanto a la codificación y las prestaciones utilizadas para la voz podemos notar que se utilizó el codec g729r8 y no se utilizó la prestación de VAD, por las mismas razones que no fue utilizado en las pruebas anteriores.

Las prestaciones de calidad de servicio, como ser fragmentación, interleaving o compresión de encabezado, tampoco fueron utilizadas.

Es posible también configurar el tamaño del payload de los paquetes de voz a utilizar, en una primera instancia se resolvió utilizar el valor que traen por defecto los equipos Cisco, 20 bytes. Luego se verá que este valor tuvo que ser modificado debido a algunos resultados obtenidos.

Para realizar el testeado de este tipo de solución se realizan las mismas pruebas que para las alternativas anteriores, generando tráfico suficiente como para saturar los enlaces.

Una vez que los enlaces se saturan se realizan llamadas telefónicas y se observan los resultados tanto a nivel del

escucha como a través de algunos valores que proporciona el router.

Resultados

Con un alto nivel de tráfico, una llamada de voz presenta una buena calidad, presentando un retardo de jitter de 70 ms, este valor de retardo de jitter se obtuvo mediante el comando “show call active voice brief” [11] que entre los varios parámetros que presenta, muestra el tamaño buffer de jitter (en ms) del equipo Cisco. Este parámetro es representativo de la variación del retardo en la red y al sumarle el retardo de la red, el retardo del codec, y el retardo de paquetización, se obtiene el retardo total punta a punta de la conversación telefónica.

Con un retardo de jitter de 70 ms, el retardo total es de aproximadamente 115 ms (retardo intrínseco del codec = 15ms + retardo de la red = 30 ms + retardo de paquetización = 0ms), que cumple con la recomendación de la ITU-T G.114 de que el retardo punta a punta debe ser menor a 150 ms para tener una calidad aceptable para la mayoría de las aplicaciones [2].

Al realizar la segunda llamada se observa que ésta presenta entrecortes y por lo tanto una mala calidad de la voz, incluso la llamada que se encontraba en curso se degrada. Este entrecorte y degradación de la voz se debe a que el ancho de banda que ocupa cada llamada de voz es muy alto para los anchos de banda de línea definidos en el PVC. Cabe notar que este bandwidth elevado se debe al overhead ATM y al tamaño del payload de los paquetes de voz utilizado. A continuación se muestra el ancho de banda que ocupa cada llamada.

Cada paquete de voz, a ser encapsulado en celdas ATM, tiene el siguiente largo en bytes:

Headers AAL5	8 bytes
DSAP/SSAP (0xAAAA)	2 bytes
Campo Control SNAP (0x03)	1 byte
OUI SNAP (0x000000)	3 bytes
Ethertype (0x800)	2 bytes
IP/RTP/UDP	40 bytes
Datos Codec G.729	20 bytes
Total:	76 bytes

Por lo tanto este paquete debe ser encapsulado en dos celdas ATM, con un tamaño total de $53 \times 2 = 106$ bytes, lo que indica que para transmitir 20 bytes de un paquete de voz con el codec g.729, el equipo en realidad transmite 106 bytes y por lo tanto una sesión g.729 de voz utilizará $8 \text{ Kbps} \times (106/20) = 42.4 \text{ Kbps}$, lo que indica que dos llamadas de voz ocupan 84.8 Kbps.

Estos datos fueron corroborados prácticamente graficando el tráfico en el PVC de voz a nivel de los equipos de capa

2. En la figura 10 se muestran estas gráficas. Notamos que para poder realizarlas sin que se produjera descarte de paquetes hubo que aumentar el ancho de banda del PVC de voz a 128 Kbps.

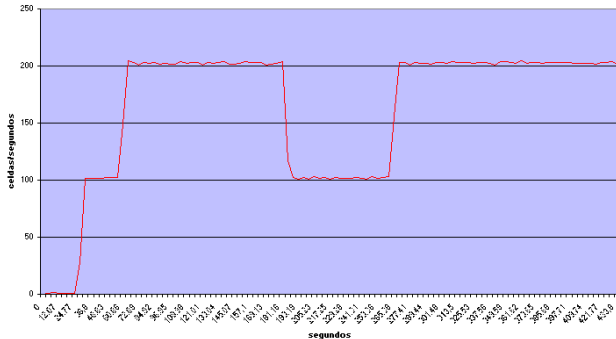


Figura 10. Ancho de banda de voz con una y dos llamadas

En esta gráfica se ve claramente el ancho de banda utilizado por dos llamadas de voz, así como también el descenso a la mitad de éste cuando se cursa una única llamada. Se corrobora el ancho de banda teórico calculado que, para dos llamadas, es de 200 celdas/segundo y por lo tanto el bandwidth es $200 * 53 * 8 = 84800$ bps.

Debido a este ancho de banda involucrado y con la intención de bajarlo y poder utilizar velocidades de 64 Kbps en los enlaces, se aumenta el payload de los paquetes de voz resultando en una reducción del bandwidth utilizado por cada llamada de voz.

Aumentando entonces el payload del paquete de voz a 40 bytes resulta lo siguiente:

Cada paquete de voz, a ser encapsulado en celdas ATM, tiene el siguiente largo:

Headers AAL5	8 bytes
DSAP/SSAP (0xAAAA)	2 bytes
Campo Control SNAP (0x03)	1 byte
OUI SNAP (0x000000)	3 bytes
Ethertype (0x800)	2 bytes
IP/RTP/UDP	40 bytes
Datos Codec G.729	40 bytes

Total: 96 bytes

Por lo tanto estos 96 bytes ocupan exactamente el payload de dos celdas ATM con un tamaño total de $53 * 2 = 106$ bytes. Una sesión de voz con el codec g.729 queda entonces de $8 \text{ Kbps} * (106/40) = 21.2 \text{ Kbps}$, y por consiguiente las dos sesiones de voz ocuparán 42.4 Kbps.

Notamos que este valor es exactamente la mitad del que se obtenía con paquetes de voz de 20 bytes.

Nuevamente este dato teórico se corroboró prácticamente graficando el tráfico a nivel de capa 2.

De esta forma con el puerto ADSL en 64 Kbps se pueden cursar las dos llamadas telefónicas sin inconvenientes.

Las pruebas se realizaron con esta configuración y haciendo que los enlaces se saturen mediante floodings de pings y transferencia de archivos vía ftp. Los resultados obtenidos para dos llamadas telefónicas simultáneas fueron muy buenos, en este caso el retardo de jitter obtenido mediante el comando "show call active voice brief" fue de 69 ms en ambas puntas, lo que se traduce en un retardo punta a punta de aproximadamente 115 ms.

Para testear el funcionamiento de este tipo de solución en una situación más realista, se realizó una red corporativa de prueba en la cual se tienen varias sucursales, cada una con equipos Cisco 827-4V y un router central Cisco 1751 al que llegan todos los PVCs (ver figura 11). Nuevamente se crean PVCs distintos para voz y para datos, cada uno de 64 Kbps, el access-rate del lado central (FR) es de 512 Kbps mientras que el de las sucursales es de 64 Kbps (ADSL). En lo que refiere a CoS, traffic-shaping, codecs y prestaciones de la voz, las configuraciones fueron análogas a la prueba anterior.

En este caso se configura del lado FR una facilidad que implementa Cisco llamada PVC Interface Priority Queuing (PIPQ) que permite realizar priorización por PVC. Existen cuatro niveles de priorización: high, médium, normal y low. El frame FR es examinado en la interface para obtener el valor del Data-Link Connection Identifier (DLCI), luego es mandado a la respectiva cola de prioridad basado en el nivel de prioridad configurado para ese DLCI.

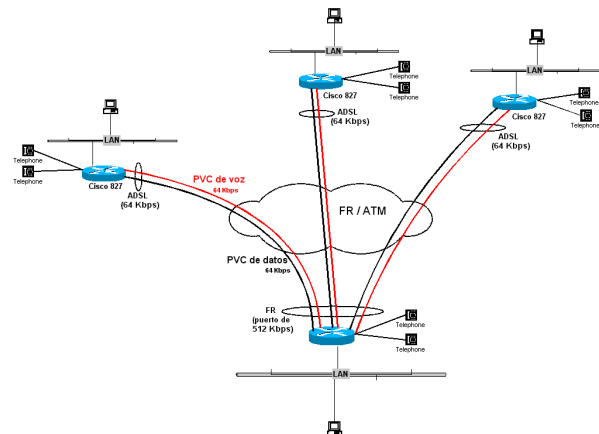


Figura 11. Red corporativa de prueba (caso real).

En este caso se le asigna prioridad máxima a los DLCIs correspondientes a los PVCs de voz. De esta forma el retardo máximo que puede llegar a experimentar un paquete

de voz es el debido a un frame FR de 1500 bytes, siendo el retardo de $1500 \cdot 8 / 512000 = 23.4$ ms. Notamos que del lado sucursal la priorización de la voz está dada con la configuración de la calidad de servicio ATM configurada en capa 2 y en el router.

Las pruebas se realizaron, al igual que las anteriores, haciendo cursar un elevado tráfico de datos en el punto central (enlace FR), nuevamente se utilizó floodings de pings y transferencia de archivos vía ftp entre sucursales y central.

Se realizaron cuatro llamadas simultáneas entre sucursales, y entre sucursales y central, obteniendo una muy buena calidad en la voz, comparable con la obtenida en las pruebas anteriores. El valor del retardo de jitter obtenido para este caso fue de 70 ms.

Conclusión

Como conclusión para este tipo de implementación con dos PVCs, se puede decir que es una solución muy robusta, de simple configuración, escalable y de buen funcionamiento tanto a nivel de voz como de datos.

Cabe notar que para el buen funcionamiento se debe configurar la calidad de servicio en toda la red de capa 2,

incluyendo los tramos DSLAM – 827. Este esquema permite además la utilización de access-rates bajos (64 Kbps) en las sucursales.

Es importante destacar que el jitter de 70 ms que se obtiene en las conversaciones de voz es debido al buffer de jitter que implementa Cisco y que este jitter es posible de reducir hasta un valor tan pequeño como el retardo que genera la serialización de una celda ATM, o sea $53 \cdot 8 / 64000 = 6.6$ ms. El funcionamiento del buffer de jitter se explica más adelante.

Como posible desventaja se observa que se deben operar y mantener dos PVCs por cada una de las sucursales, lo que puede eventualmente generar inconvenientes de administración si ocurriese alguna caída del servicio.

La tendencia de Anteldata para los próximos años es aumentar presencia de ADSL y generalizar el acceso GHDSL. Con esta nueva tecnología, será posible cambiar el puerto central de FR a GHDSL, permitiendo por lo tanto la reducción del access-rate. Esto es posible gracias a que la priorización de los paquetes de voz en el puerto central se hará análogamente al de las sucursales, utilizando la calidad de servicio ATM.

Tabla 3. Comparación de las distintas implementaciones.

Alternativas	Claridad	Retardo punta a punta	Aprovechamiento de ancho de banda	Control de ancho de banda de datos	Recursos de ancho de banda	Provisionamiento y Gestión ATM	Costo de los equipos	Complejidad de la solución	Estandarización
1 PVC (solución actual de AntelData)	buena	bajo	bueno	si	Bajos	bajo	alto	alta	no
1 PVC:									
con CAR	buena	medio	bueno	si	Altos	bajo	medio	baja	no
con priorización	buena	medio	muy bueno	no	Altos	bajo	medio	media	no
2 PVCs	muy buena	bajo	muy bueno	si	Bajos	medio	medio	baja	si

Resultados Interesantes

Es interesante observar que asociado al concepto de la tecnología VoIP está muy arraigada la idea de bajo ancho de banda utilizado. Lo cual en gran parte de los casos no es así. Es cierto que una llamada de VoIP generalmente utiliza menos de 60 Kbps, que comparte el ancho de banda con otras aplicaciones y que en redes gestionadas hay técnicas para optimizar los recursos insumidos. A modo de ejemplo una comunicación de VoIP con G.729 y la configuración por defecto de un Gateway DSL utiliza 43 Kbps en vez de los 12.5 Kbps que se suele creer.

Los equipos Cisco manejan un buffer llamado "jitter buffer"[11]. Este buffer recibe los paquetes de voz de la red IP a intervalos irregulares y estos a su vez pueden llegar fuera de secuencia. El buffer mantiene los paquetes por un período corto, los reordena si es necesario, y después los envía a intervalos de tiempo equiespaciados hacia el decodificador en el Digital Signal Processor (DSP) en el gateway o router.

Algoritmos en el DSP determinan el tamaño y comportamiento del buffer de jitter, basado en la configuración del usuario y las condiciones de jitter actuales de la red para maximizar el número de paquetes enviados correctamente y minimizar el retardo.

El tamaño del buffer de jitter y por ende el retardo es configurable por el usuario.

Este parámetro fue modificado durante el estudio de la nueva implementación con un PVC, comprobando su funcionamiento y obteniendo al modificarlo una mejora en la calidad de la voz.

COMPARACIÓN DE LAS IMPLEMENTACIONES

En la tabla 3 se muestra una comparación de las implementaciones vistas.

CONCLUSIONES

De todo lo antes mencionado se obtiene que la solución de 2PVCs resulta más adecuada para la situación actual de Antel, dado que es escalable, de simple configuración y presenta una muy buena calidad en lo que refiere a la voz. Adicionalmente, dado que muchas de las redes corporativas presentan sucursales en el interior, y en muchos casos con enlaces troncales con alta probabilidad de congestión,

cualquiera de las soluciones con 1 PVC sería inaceptable para el servicio de voz.

Y la principal debilidad detectada de esta solución es la dependencia con la tecnología ATM, que muy probablemente será cuestión de análisis en el futuro.

También notamos que los resultados de este estudio son basados, en su mayoría, en la utilización de equipamiento Cisco.

REFERENCIAS

- [1] V. Kumar, M. Korpi, S. Sengodan, *IP Telephony with H.323*, Wiley (2001)
- [2] International Communication Union, <http://www.itu.int>
- [3] D. McDysan, D. Spohn, *ATM Theory and Applications*, McGraw-Hill, 1999
- [4] *Frame Relay Networking Guide*, Frame RelayForum, www.frforum.com
- [5] *Quality of Service for Voice over IP*, Cisco Systems, 2002
- [6] *VoIP over Frame Relay with Quality of Service (Fragmentation, Traffic Shaping, LLQ / IP RTP Priority)*, Cisco Systems, 2003
- [7] *Classifying VoIP Signaling and Media with DSCP for QoS*, Cisco Systems, 2002
- [8] *Committed Access Rate*, Cisco System, <http://www.cisco.com/>
- [9] *Understanding and Tuning the tx-ring-limit Value*, Cisco Systems, <http://www.cisco.com/>
- [10] *IP Fragmentation and PMTUD*, Cisco Systems, <http://www.cisco.com/>
- [11] *Playout Delay Enhancements for Voice over IP*, Cisco Systems, <http://www.cisco.com/>