Verspätung in Packet Voice Networks

Inhalt

Einführung
Grundlegender Sprachdatenfluss
Funktionsweise der Sprachkomprimierung
Standards für Verzögerungstoleranzen
Quellen der Verzögerung
Coder (Processing)-Verzögerung
Verzögerung der Packetisierung
Verzögerung der Serialisierung
Verzögerung bei Warteschlangen/Pufferung
Verzögerung des Netzwerk-Switching
Jitter-Verzögerung
Erstellen des Verzögerungsbudgets
Single-Hop-Verbindung
Zwei Hops in einem öffentlichen Netzwerk mit einem C7200, der als Tandem-Switch fungiert
Zwei-Hop-Verbindung über ein öffentliches Netzwerk mit einem Tandem-PBX-Switch
Zwei-Hop-Verbindung über ein privates Netzwerk mit einem Tandem-PBX-Switch
Auswirkungen mehrerer Komprimierungszyklen
Überlegungen für Verbindungen mit hoher Verzögerung
Zugehörige Informationen

Einführung

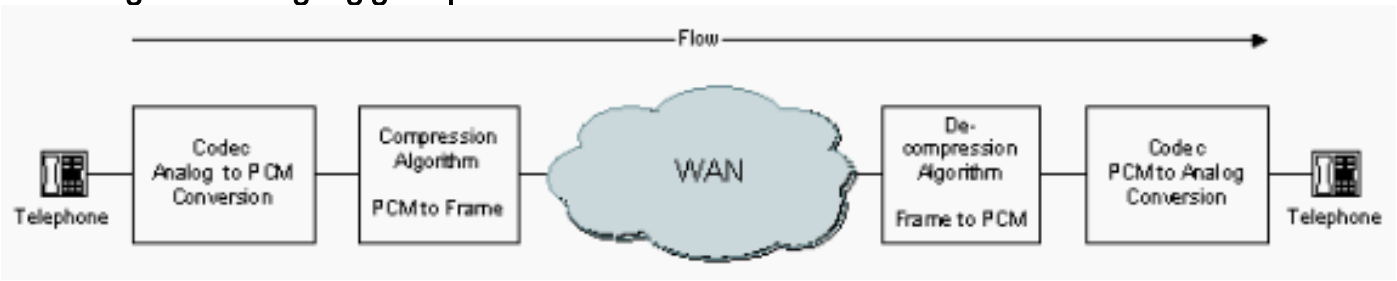
Wenn Sie Netzwerke entwerfen, die Sprache über Paket-, Frame- oder Zellinfrastrukturen übertragen, ist es wichtig, die Verzögerungskomponenten im Netzwerk zu verstehen und zu berücksichtigen. Wenn Sie alle potenziellen Verzögerungen korrekt berücksichtigen, wird sichergestellt, dass die Netzwerkleistung insgesamt akzeptabel ist. Die allgemeine Sprachqualität ist eine Funktion vieler Faktoren, darunter der Komprimierungsalgorithmus, Fehler und Frame-Verlust, die Echounterdrückung und die Verzögerung. In diesem Whitepaper werden die Ursachen für Verzögerungen bei der Verwendung von Cisco Routern/Gateways über Paketnetzwerke erläutert. Obwohl die Beispiele auf Frame Relay ausgerichtet sind, gelten die Konzepte auch für VoIP- (Voice over IP) und VoATM-Netzwerke (Voice over ATM).

Grundlegender Sprachdatenfluss

Der Fluss einer komprimierten Sprachschaltung wird in diesem Diagramm dargestellt. Das analoge Signal vom Telefon wird vom Sprach-Decoder (Codec) in Pulscode Modulation (PCM)-Signale digitalisiert. Die PCM-Stichproben werden dann an den Komprimierungsalgorithmus weitergeleitet, der die Sprache in ein Paketformat für die Übertragung über das WAN komprimiert. Auf der anderen Seite der Cloud werden die gleichen Funktionen in umgekehrter Reihenfolge

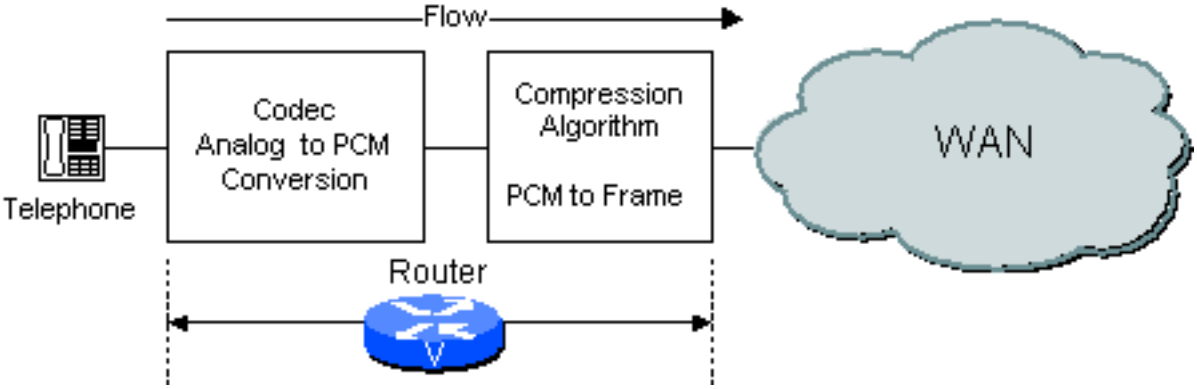
ausgeführt. Der gesamte Datenfluss ist in Abbildung 2-1 dargestellt.

Abbildung 2-1 Durchgängiger Sprachdatenfluss



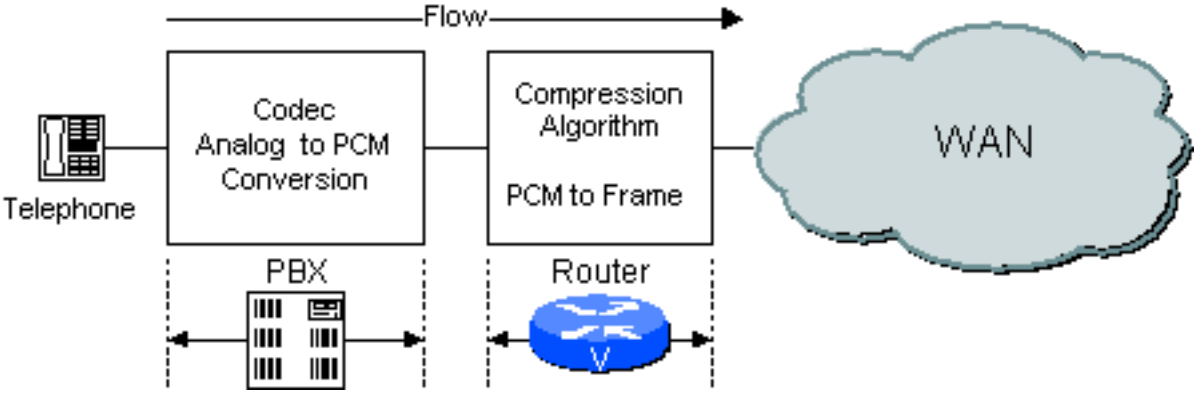
Je nach Konfiguration des Netzwerks kann der Router bzw. das Gateway sowohl Codec- als auch Komprimierungsfunktionen oder nur eine davon ausführen. Wenn beispielsweise ein analoges Sprachsystem verwendet wird, führt der Router/Gateway die CODEC-Funktion und die Komprimierungsfunktion aus, wie in Abbildung 2-2 dargestellt.

Abbildung 2-2 Codec-Funktion im Router/Gateway



Wenn ein digitales PBX-System verwendet wird, führt das PBX-System die Codec-Funktion aus, und der Router verarbeitet die PCM-Beispiele, die ihm vom PBX-System übergeben werden. Ein Beispiel ist in Abbildung 2-3 dargestellt.

Abbildung 2-3 Codec-Funktion in PBX

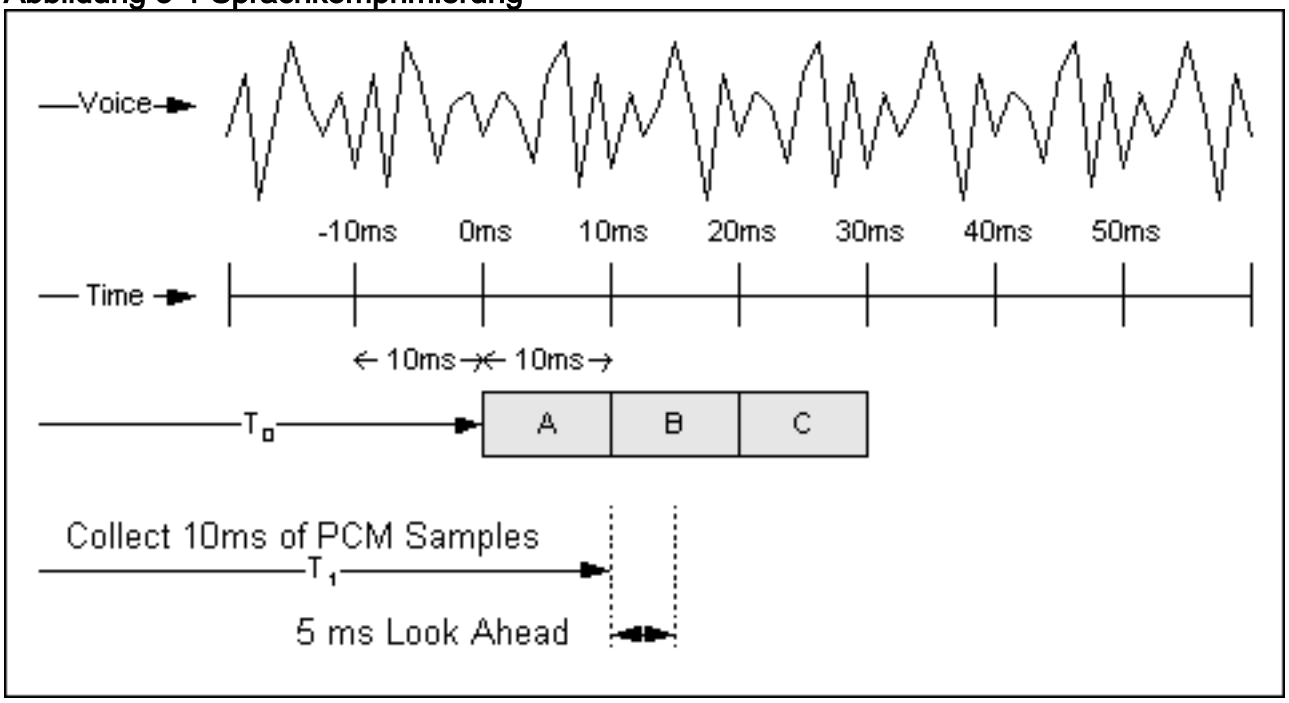


Funktionsweise der Sprachkomprimierung

Die in Cisco Router/Gateways verwendeten Komprimierungsalgorithmen mit hoher Komplexität analysieren einen Block von PCM-Stichproben, die vom Voice-Codec bereitgestellt werden. Die Länge dieser Blöcke hängt vom Codierer ab. Beispielsweise beträgt die von einem G.729-Algorithmus verwendete Blockgröße 10 ms, die von den G.723.1-Algorithmen verwendete Blockgröße 30 ms. Ein Beispiel für die Funktionsweise eines G.729-Komprimierungssystems ist in

Abbildung 3-1 dargestellt.

Abbildung 3-1 Sprachkomprimierung



Der analoge Sprachstream wird in PCM-Stichproben digitalisiert und in Schritten von 10 ms an den Komprimierungsalgorithmus übermittelt. Der Blick nach vorne wird in Algorithmic Delay behandelt.

Standards für Verzögerungstoleranzen

Die Internationale Fernmeldeunion (ITU) berücksichtigt Netzverzögerungen für Sprachanwendungen in Empfehlung G.114. In dieser Empfehlung werden drei Frequenzbänder für einmalige Verzögerungen definiert (siehe Tabelle 4.1).

Tabelle 4.1 Verzögerte Spezifikationen

Bereich in Millisekunden	Beschreibung
0-150	Für die meisten Benutzeranwendungen geeignet.
150-400	Akzeptabel, vorausgesetzt, Administratoren sind sich der Übertragungszeit und der Auswirkungen auf die Übertragungsqualität von Benutzeranwendungen bewusst.
Über 400	Nicht akzeptabel für die allgemeine Netzwerkplanung. Es wird jedoch anerkannt, dass diese Obergrenze in einigen Ausnahmefällen überschritten wird.

Hinweis: Diese Empfehlungen gelten für Verbindungen mit Echo, die angemessen kontrolliert werden. Dies impliziert, dass Echounterdrücker verwendet werden. Eine Echounterdrückung ist

erforderlich, wenn die Verzögerung in eine Richtung 25 ms (G.131) überschreitet.

Diese Empfehlungen richten sich an die nationalen Telekommunikationsverwaltungen. Daher sind diese strenger als in privaten Sprachnetzwerken üblich. Wenn dem Netzwerkdesigner Standort und geschäftliche Anforderungen von Endbenutzern bekannt sind, können sich weitere Verzögerungen als akzeptabel erweisen. Bei privaten Netzwerken sind 200 ms Verzögerung ein vernünftiges Ziel und 250 ms ein Limit. Alle Netzwerke müssen so konzipiert sein, dass die maximal erwartete Verzögerung der Sprachverbindung bekannt ist und minimiert wird.

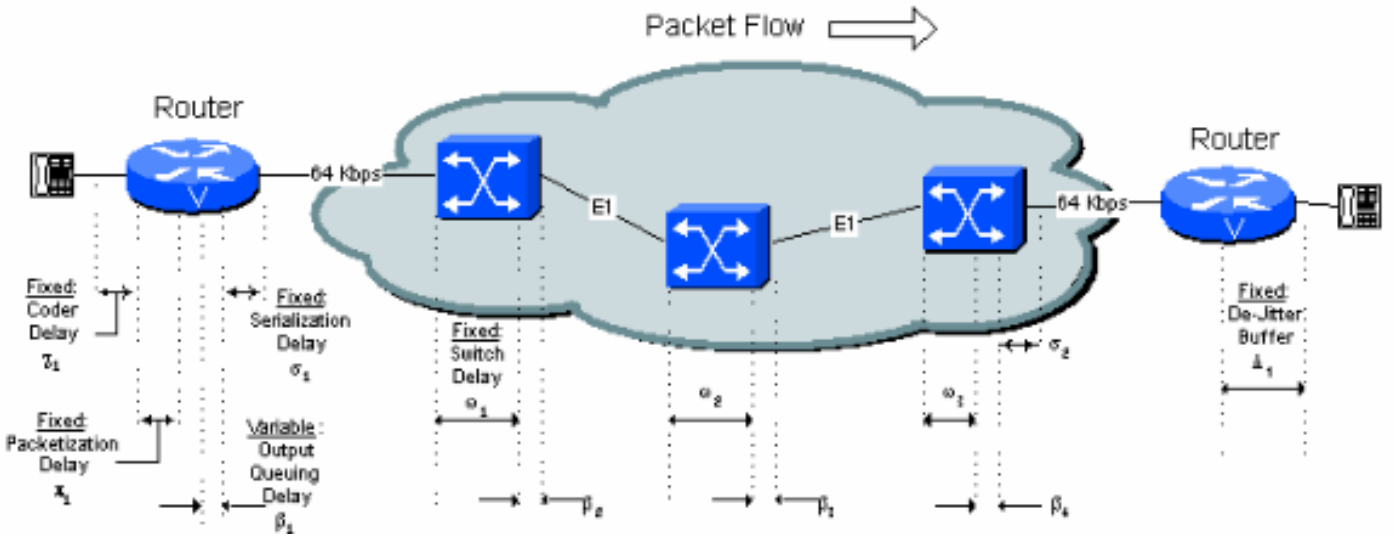
Quellen der Verzögerung

Es gibt zwei verschiedene Verzögerungsarten, die als fixed und variable bezeichnet werden.

- Komponenten mit fester Verzögerung vergrößern die Gesamtverzögerung der Verbindung direkt.
- Veränderliche Verzögerungen entstehen durch Warteschlangenverzögerungen bei den Ausgangs-Trunk-Puffern auf dem mit dem WAN verbundenen seriellen Port. Diese Puffer führen im gesamten Netzwerk zu variablen Verzögerungen, die als Jitter bezeichnet werden. Variable Verzögerungen werden über den De-Jitter-Puffer am empfangenden Router/Gateway gehandhabt. Der De-Jitter-Puffer wird im Abschnitt "De-jitter Delay ($_{IGN}$)" dieses Dokuments beschrieben.

In Abbildung 5-1 werden alle Quellen für Verzögerungen bei fester und variabler Verzögerung im Netzwerk angegeben. Jede Quelle wird in diesem Dokument ausführlich beschrieben.

Abbildung 5-1: Verzögerungsquellen



Coder (Processing)-Verzögerung

Die Taktverzögerung ist die Zeit, die der digitale Signalprozessor (DSP) benötigt, um einen Block von PCM-Proben zu komprimieren. Dies wird auch als Verarbeitungsverzögerung ($_{Hinzufügen\ von}$ Prozessen) bezeichnet. Diese Verzögerung variiert je nach verwendetem Sprachcoder und Prozessorgeschwindigkeit. So analysieren algebraischer Code beispielsweise ACELP-Algorithmen (Linear Prediction) einen 10-ms-Block von PCM-Stichproben und komprimieren diese anschließend.

Die Komprimierungszeit für einen Prozess mit konjugierter Struktur-Algebraischer Code für

exzessive lineare Prognose (CS-ACELP) liegt je nach Laden des DSP-Prozessors zwischen 2,5 ms und 10 ms. Wenn der DSP vollständig mit vier Sprachkanälen geladen ist, beträgt die Coder-Verzögerung 10 ms. Wenn der DSP mit nur einem Sprachkanal geladen wird, beträgt die Coder-Verzögerung 2,5 ms. Verwenden Sie für Designzwecke die Worst-Case-Zeit von 10 ms.

Die Dekomprimierungszeit beträgt etwa zehn Prozent der Komprimierungszeit für jeden Block. Die Dekomprimierungszeit ist jedoch proportional zur Anzahl der Stichproben pro Frame, da mehrere Stichproben vorhanden sind. Die schlimmste Dekomprimierungszeit für einen Frame mit drei Stichproben beträgt 3 x 1 ms oder 3 ms. Normalerweise werden zwei oder drei Blöcke komprimierter G.729-Ausgabe in einem Frame abgelegt, während eine komprimierte G.723.1-Ausgabe in einem Frame gesendet wird.

Tabelle 5.1 enthält Angaben zu Verzögerungen bei Codierungen im besten und schlechtesten Fall.

Tabelle 5.1 Verzögerung bei der Bearbeitung von Best/Worst-Case-Tests

Cursor	Rate	Erforderlicher Beispielblock	Coder-Verzögerung für Best Case-Szenario	Coder-Verzögerung bei Worst Case
ADPCM, G.726	32 Kbit/s	10 ms	2,5 ms	10 ms
CS-ACELP, G.729A	8,0 Kbit/s	10 ms	2,5 ms	10 ms
MP-MLQ, G.723.1	6,3 Kbit/s	30 ms	5 ms	20 ms
MP-ACELP, G.723.1	5,3 Kbit/s	30 ms	5 ms	20 ms

Algorithmische Verzögerung

Der Komprimierungsalgorithmus stützt sich auf bekannte Sprachmerkmale, um den Stichprobenblock N richtig zu verarbeiten. Der Algorithmus muss einige Kenntnisse über die Bestandteile von Block N+1 aufweisen, damit der Stichprobenblock N präzise reproduziert werden kann. Dieser Blick nach vorn, der wirklich eine zusätzliche Verzögerung ist, wird als algorithmische Verzögerung bezeichnet. Dadurch wird die Länge des Komprimierungsblocks effektiv erhöht.

Dies geschieht wiederholt, sodass Block N+1 in Block N+2 usw. schaut. Der Nettoeffekt erhöht die Gesamtverzögerung für die Verbindung um 5 ms. Dies bedeutet, dass die gesamte für die Verarbeitung eines Datenblocks benötigte Zeit 10 m bei einem konstanten Overhead-Faktor von 5 ms beträgt. Siehe Abbildung 3-1: Sprachkomprimierung.

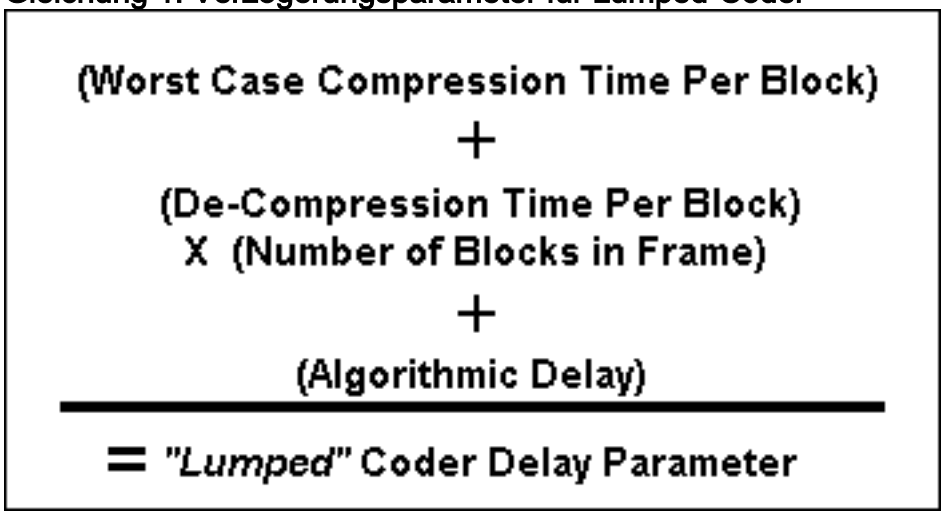
- Die algorithmische Verzögerung für G.726-Coder beträgt 0 ms.
- Die algorithmische Verzögerung für G.729-Coder beträgt 5 ms.
- Die algorithmische Verzögerung für G.723.1-Coder beträgt 7,5 ms.

Bei den Beispielen im verbleibenden Teil dieses Dokuments wird von einer G.729-Komprimierung

mit einer Payload von 30 ms/30 Byte ausgegangen. Um das Design zu vereinfachen und einen konservativen Ansatz zu verfolgen, gehen die Tabellen im verbleibenden Teil dieses Dokuments von der schlimmsten Fall-Codierungsverzögerung aus. Die Codierungsverzögerung, die Dekomprimierungsverzögerung und die algorithmische Verzögerung werden in einem Faktor zusammengefasst, der als Codierungsverzögerung bezeichnet wird.

Die Gleichung zum Generieren des gespalteten Coder-Delay-Parameters lautet:

Gleichung 1: Verzögerungsparameter für Lumped-Coder



Die für das restliche Dokument verwendete Verzögerung bei der Suche nach G.729 ist:

Komprimierungszeit für Worst Case pro Block: 10 ms

Dekomprimierungszeit pro Block x 3 Blöcke: 3 ms

Algorithmische Verzögerung 5 ms —

Gesamt (Reichweite) 18 ms

Verzögerung der Packetisierung

Packetization Delay ($_{Sample}$) ist die Zeit, die zum Füllen einer Paket-Payload mit kodierter/komprimierter Sprache benötigt wird. Diese Verzögerung ist abhängig von der vom Vocoder benötigten Blockgröße und der Anzahl der Blöcke, die in einem Frame platziert werden. Die Packetisierungsverzögerung kann auch als Akkumulierungsverzögerung bezeichnet werden, da sich die Sprachmuster in einem Puffer ansammeln, bevor sie freigegeben werden.

In der Regel müssen Sie sich um eine Paketverzögerung von maximal 30 ms bemühen. Bei den Cisco Routern/Gateways müssen Sie die folgenden Zahlen aus Tabelle 5.2 verwenden, die auf der konfigurierten Nutzlastgröße basieren:

Tabelle 5.2: Häufig

Cursor		Nutzlastgröße (Byte)	Packetisierungsverzögerung (ms)	Nutzlastgröße (Byte)	Packetisierungsverzögerung (ms)
PCM,	6	160	20	240	30

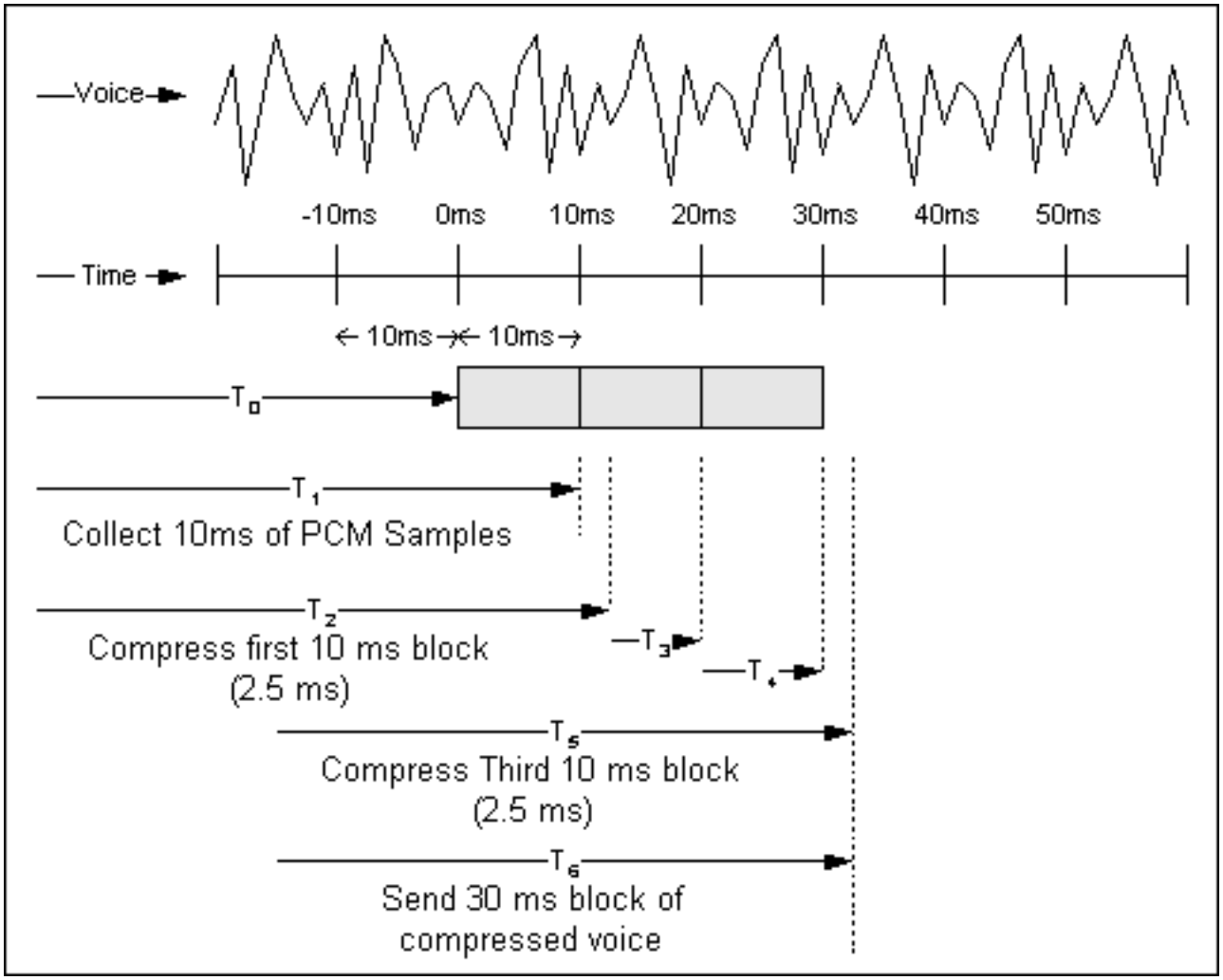
G.711	64 Kbit/s				
ADPCM, G.726	32 Kbit/s	80	20	120	30
CS-ACELP, G.729	8,0 Kbit/s	20	20	30	30
MP-MLQ, G.723.1	6,3 Kbit/s	24	24	60	48
MP-ACELP, G.723.1	5,3 Kbit/s	20	30	60	60

Sie müssen die Paketverzögerung gegen die CPU-Last abwägen. Je geringer die Verzögerung, desto höher die Frame-Rate und desto höher die Last auf der CPU. Auf einigen älteren Plattformen können 20 ms-Payloads die Haupt-CPU belasten.

Verzögerung der Pipeline im Packetingprozess

Obwohl bei jedem Sprachmuster sowohl algorithmische Verzögerungen als auch Paketverzögerungen auftreten, überschneiden sich die Prozesse, und diese Pipeline bietet einen Nettovorteil. Betrachten Sie das Beispiel in Abbildung 2-1.

Abbildung 5-2: Pipelining und Packetisierung



Die obere Linie der Abbildung zeigt ein Beispiel für eine Sprachwellenform. Die zweite Zeile ist eine Zeitskala in Schritten von 10 ms. Bei T_0 beginnt der CS-ACELP-Algorithmus, PCM-Stichproben vom Codec zu sammeln. Bei T_1 hat der Algorithmus seinen ersten 10-ms-Samplerblock gesammelt und beginnt, ihn zu komprimieren. Beim T_2 wurde der erste Block komprimiert. In diesem Beispiel beträgt die Komprimierungszeit 2,5 ms, wie durch T_2-T_1 angegeben.

Der zweite und dritte Block werden bei T_3 und T_4 gesammelt. Der dritte Block wird mit T_5 komprimiert. Das Paket wird mit T_6 assembliert und gesendet (es wird angenommen, dass es sich um ein sofortiges Paket handelt. Aufgrund der Pipelinecharakter der Komprimierungs- und Packetisierungsprozesse beträgt die Verzögerung vom Beginn des Prozesses bis zum Absenden des Sprach-Frames T_6-T_0 oder ca. 32,5 ms.

Dieses Beispiel basiert beispielsweise auf der optimalen Verzögerung. Wenn die Worst-Case-Verzögerung verwendet wird, beträgt die Zahl 40 ms, 10 ms für Coder-Verzögerung und 30 ms für Packetisierungsverzögerung.

Beachten Sie, dass in diesen Beispielen algorithmische Verzögerungen nicht berücksichtigt werden.

Verzögerung der Serialisierung

Die Serialisierungsverzögerung ($_n$) ist die feste Verzögerung, die erforderlich ist, um einen Sprach- oder Daten-Frame auf die Netzwerkschnittstelle zu übertragen. Sie steht in direktem Zusammenhang mit der Taktrate auf dem Trunk. Bei niedrigen Taktraten und kleinen Frame-

Größen ist das zusätzliche Flag erforderlich, um Frames zu trennen, erheblich.

Tabelle 5.3 zeigt die erforderliche Serialisierungsverzögerung für unterschiedliche Frame-Größen bei unterschiedlichen Leitungsgeschwindigkeiten. Diese Tabelle verwendet zur Berechnung die Gesamtgröße des Frames, nicht die Payload-Größe.

Tabelle 5.3: Verzögerung der Serialisierung in Millisekunden für unterschiedliche Frame-Größen

Frame-Größe (Byte)	Leitungsgeschwindigkeit (Kbit/s)										
	19.2	56	64	128	256	384	512	768	1024	1544	2048
38	15.83	5.43	4.75	2.38	1.19	0.79	0.59	0.40	0.30	0.20	0.15
48	20.00	6.86	6.00	3.00	1.50	1.00	0.75	0.50	0.38	0.25	0.19
64	26.67	9.14	8.00	4.00	2.00	1.33	1.00	0.67	0.50	0.33	0.25
128	53.33	18.29	16.00	8.00	4.00	2.67	2.00	1.33	1.00	0.66	0.50
256	106.67	36.57	32.00	16.00	8.00	5.33	4.00	2.67	2.00	1.33	1.00
512	213.33	73.14	64.00	32.00	16.00	10.67	8.00	5.33	4.00	2.65	2.00
1024	426.67	149.29	128.00	64.00	32.00	21.33	16.00	10.67	8.00	5.31	4.00
1500	625.00	214.29	187.50	93.75	46.88	31.25	23.44	15.63	11.72	7.77	5.86
2048	853.33	292.57	256.00	128.00	64.00	42.67	32.00	21.33	16.00	10.61	8.00

In der Tabelle beträgt die Serialisierungsverzögerung eines CS-ACELP-Sprachrahmens mit einer Länge von 38 Byte (37+1-Flag) für eine 64-Kbit/s-Leitung 4,75 ms.

Hinweis: Die Serialisierungsverzögerung für eine 53-Byte-ATM-Zelle (T1: 0,275 ms, E1: 0,207 ms) ist aufgrund der hohen Leitungsgeschwindigkeit und der kleinen Zellengröße zu vernachlässigen.

Verzögerung bei Warteschlangen/Pufferung

Nachdem die komprimierte Sprach-Nutzlast erstellt wurde, wird ein Header hinzugefügt, und der Frame wird in die Warteschlange für die Übertragung über die Netzwerkverbindung gestellt. Sprache muss im Router/Gateway absolute Priorität haben. Aus diesem Grund darf ein Sprach-Frame nur auf einen bereits abgespielten Daten-Frame oder auf andere Sprach-Frames vor diesem Frame warten. Im Wesentlichen wartet der Sprach-Frame auf die Serialisierungsverzögerung aller vorherigen Frames in der Ausgabewarteschlange. Die Warteschlangenverzögerung ($_{ßn}$) ist eine variable Verzögerung und hängt von der Trunk-Geschwindigkeit und dem Status der Warteschlange ab. Mit der Warteschlangenverzögerung sind

zufällige Elemente verknüpft.

Nehmen Sie beispielsweise an, Sie befinden sich in einer 64-Kbit/s-Leitung und werden hinter einem Daten-Frame (48 Byte) und einem Sprach-Frame (42 Byte) in die Warteschlange gestellt. Da es eine zufällige Art gibt, wie viel von dem 48-Byte-Frame abgespielt wurde, können Sie sicher davon ausgehen, dass im Durchschnitt die Hälfte des Datenrahmens ausgegeben wurde. Basierend auf den Daten aus der Serialisierungstabelle beträgt Ihre Datenrahmenkomponente 6 ms * 0,5 = 3 ms. Wenn Sie die Zeit für einen weiteren Sprach-Frame in der Warteschlange (5,25 ms) hinzufügen, ergibt dies eine Gesamtzeit von 8,25 ms Warteschlangenverzögerung.

Wie die Warteschlangenverzögerung charakterisiert wird, liegt beim Netzwerktechniker. Im Allgemeinen muss die Leistung nach der Installation des Netzwerks angepasst werden, um den schlimmsten Fall zu bewältigen. Je mehr Sprachleitungen den Benutzern zur Verfügung stehen, desto höher ist die Wahrscheinlichkeit, dass das durchschnittliche Sprachpaket in der Warteschlange wartet. Der Sprach-Frame wartet aufgrund der Prioritätsstruktur nie hinter mehr als einem Datenrahmen.

Verzögerung des Netzwerk-Switching

Das öffentliche Frame-Relay oder ATM-Netzwerk, das die Endpunktstandorte miteinander verbindet, ist die Ursache der größten Verzögerungen bei Sprachverbindungen. Die Netzwerk-Switching-Verzögerungen (im $_{Stadium}$) sind ebenfalls am schwierigsten zu quantifizieren.

Wenn die Weitverkehrsverbindung von Cisco oder einem anderen privaten Netzwerk bereitgestellt wird, können die einzelnen Komponenten der Verzögerung identifiziert werden. Im Allgemeinen sind die festen Komponenten von Übertragungsverzögerungen auf den Trunks im Netzwerk und variable Verzögerungen entstehen durch Warteschlangenverzögerungen, die Frames in und außerhalb von zwischengeschalteten Switches verzögern. Zur Schätzung der Verbreitungsverzögerung wird häufig eine Schätzung von 10 Mikrosekunden/Meile oder 6 Mikrosekunden/km (G.114) verwendet. Intermediär-Multiplexing-Geräte, Backhauling, Mikrowellenverbindungen und andere Faktoren, die in Betreibernetzwerken vorkommen, stellen jedoch viele Ausnahmen dar.

Die andere wichtige Komponente der Verzögerung ist die Warteschlange innerhalb des Wide Area Network. In einem privaten Netzwerk können bestehende Warteschlangenverzögerungen gemessen oder ein Hop-basiertes Budget im WAN geschätzt werden.

Typische Carrier-Verzögerungen für Frame-Relay-Verbindungen in den USA sind 40 ms fest und variabel 25 ms, was einer Gesamtverzögerung im schlimmsten Fall von 65 ms entspricht. Aus Gründen der Einfachheit, in den Beispielen 6-1, 6-2 und 6-3, sind alle Verzögerungen bei der Serialisierung mit niedriger Geschwindigkeit in der festen Verzögerung von 40 ms enthalten.

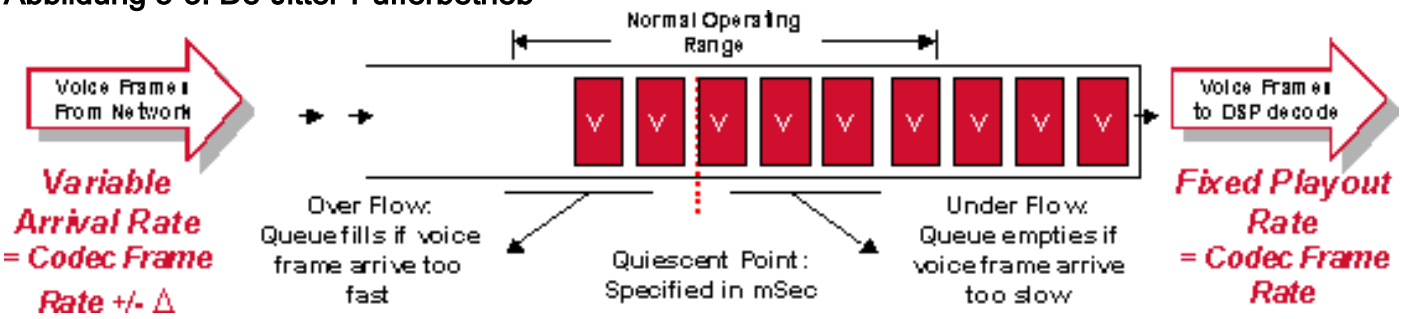
Diese Zahlen wurden von US Frame Relay Carriern veröffentlicht, um überall in den USA eine Abdeckung zu ermöglichen. Es ist zu erwarten, dass zwei Standorte, die geografisch näher sind als der schlechteste Fall, eine bessere Verzögerung der Leistung, aber Beförderer in der Regel nur den schlimmsten Fall dokumentieren.

Frame Relay Carrier bieten manchmal Premium Services an. Diese Services sind in der Regel für Sprach- oder Systems Network Architecture (SNA)-Datenverkehr bestimmt, bei dem die Netzwerkverzögerung garantiert wird und der unter dem Standard-Servicelevel liegt. So kündigte ein US-Carrier kürzlich einen solchen Service mit einer Gesamtverzögerungsgrenze von 50 ms an, nicht aber mit 65 ms für den Standarddienst.

Jitter-Verzögerung

Da Sprache ein Dienst mit konstanter Bitrate ist, muss der Jitter aus allen variablen Verzögerungen entfernt werden, bevor das Signal das Netzwerk verlässt. Bei Cisco Routern/Gateways erfolgt dies über einen De-Jitter (z. B.)-Puffer am Remote-Router/Gateway (Receiver). Der De-Jitter-Puffer wandelt die variable Verzögerung in eine feste Verzögerung um. Es enthält die erste erhaltene Probe für einen Zeitraum, bevor es sie ausgibt. Diese Haltezeit wird als "ursprüngliche Play Out Delay"-Verzögerung bezeichnet.

Abbildung 5-3: De-Jitter-Pufferbetrieb



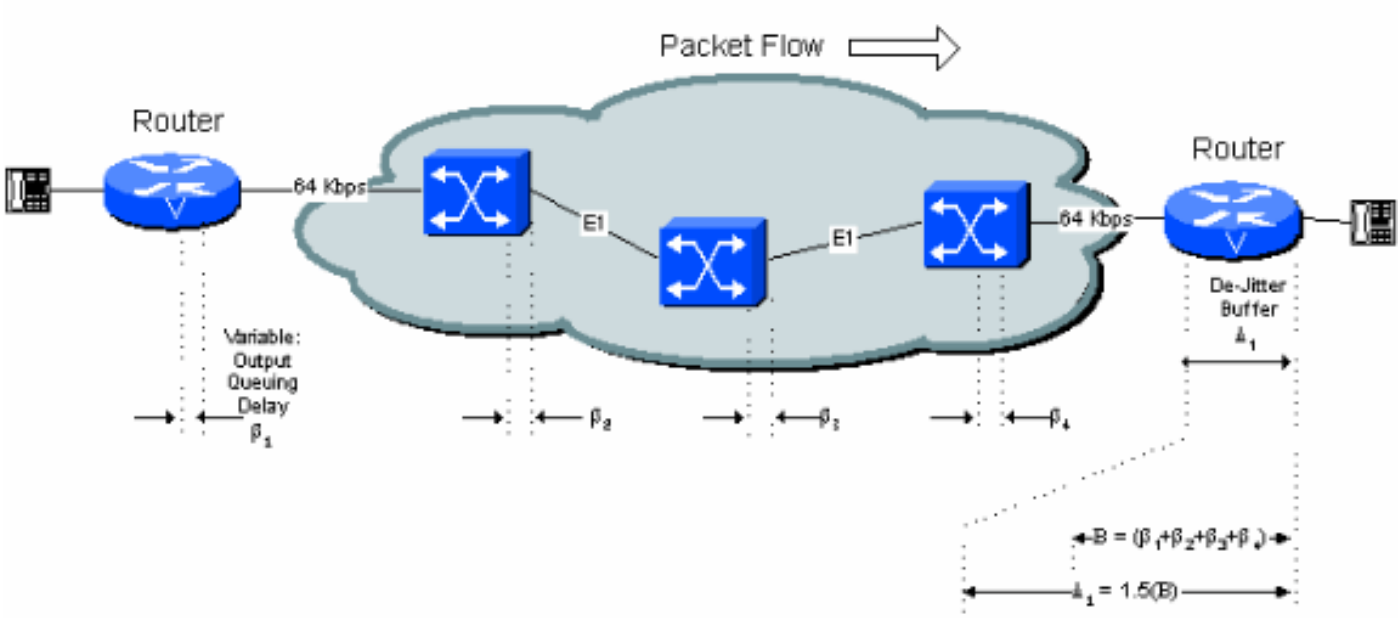
Es ist wichtig, den De-Jitter-Puffer richtig zu handhaben. Wenn Stichproben zu kurz gehalten werden, können Schwankungen der Verzögerung möglicherweise dazu führen, dass der Puffer nicht voll ausgeführt wird und Lücken in der Sprache entstehen. Wenn das Beispiel zu lange gehalten wird, kann der Puffer überlaufen, und die verworfenen Pakete verursachen wieder Lücken in der Sprache. Wenn Pakete zu lange gehalten werden, kann die Verzögerung der Verbindung insgesamt zu einem unannehmbaren Niveau führen.

Die optimale anfängliche Play-Out-Verzögerung für den De-Jitter-Puffer ist gleich der gesamten variablen Verzögerung entlang der Verbindung. Dies ist in Abbildung 5-4 dargestellt.

Hinweis: Die De-Jitter-Puffer können anpassbar sein, die maximale Verzögerung ist jedoch festgelegt. Wenn adaptive Puffer konfiguriert werden, wird die Verzögerung zu einer variablen Zahl. Die maximale Verzögerung kann jedoch im schlimmsten Fall für Designzwecke verwendet werden.

Weitere Informationen zu adaptiven Puffern finden Sie unter Verbesserungen der Playout-Verzögerung für Voice over IP.

Abbildung 5-4: Variable Verzögerung und De-Jitter-Puffer



Die anfängliche Wiedergabepause ist konfigurierbar. Die maximale Tiefe des Puffers vor dessen Überlauf wird normalerweise auf das 1,5- oder 2,0-fache dieses Werts festgelegt.

Wenn die nominale Verzögerungseinstellung von 40 ms verwendet wird, wird das erste Sprachmuster, das bei Leerung des De-Jitter-Puffers empfangen wird, 40 ms lang gehalten, bevor es ausgegeben wird. Dies impliziert, dass ein nachfolgendes Paket, das vom Netzwerk empfangen wird, ohne Verlust der Sprachkontinuität bis zu 40 ms verzögert werden kann (in Bezug auf das erste Paket). Wenn die Verzögerung mehr als 40 ms beträgt, wird der De-Jitter-Puffer geleert, und das nächste empfangene Paket wird 40 ms lang gehalten, bevor es abgespielt wird, um den Puffer zurückzusetzen. Dies führt zu einer Lücke in der Stimme gespielt für etwa 40 ms.

Der tatsächliche Beitrag des De-Jitter-Puffers zur Verzögerung ist die anfängliche Play-Out-Verzögerung des De-Jitter-Puffers zuzüglich des tatsächlichen Betrags, den das erste Paket im Netzwerk gepuffert wurde. Der schlimmste Fall ist die doppelte anfängliche Verzögerung des De-Jitter-Puffers (es wird davon ausgegangen, dass beim ersten Paket im Netzwerk nur eine minimale Pufferverzögerung auftrat). In der Praxis ist es bei einer Reihe von Netzwerk-Switch-Hops wahrscheinlich nicht erforderlich, den schlimmsten Fall anzunehmen. Die Berechnungen in den Beispielen im verbleibenden Teil dieses Dokuments erhöhen die Verzögerung bei der ersten Wiedergabe um den Faktor 1,5, um diesen Effekt zu ermöglichen.

Hinweis: Im empfangenden Router/Gateway erfolgt eine Verzögerung durch die Dekomprimierungsfunktion. Dies wird jedoch berücksichtigt, indem es zusammen mit der zuvor beschriebenen Verzögerung der Komprimierungsverarbeitung hochgeladen wird.

Erstellen des Verzögerungsbudgets

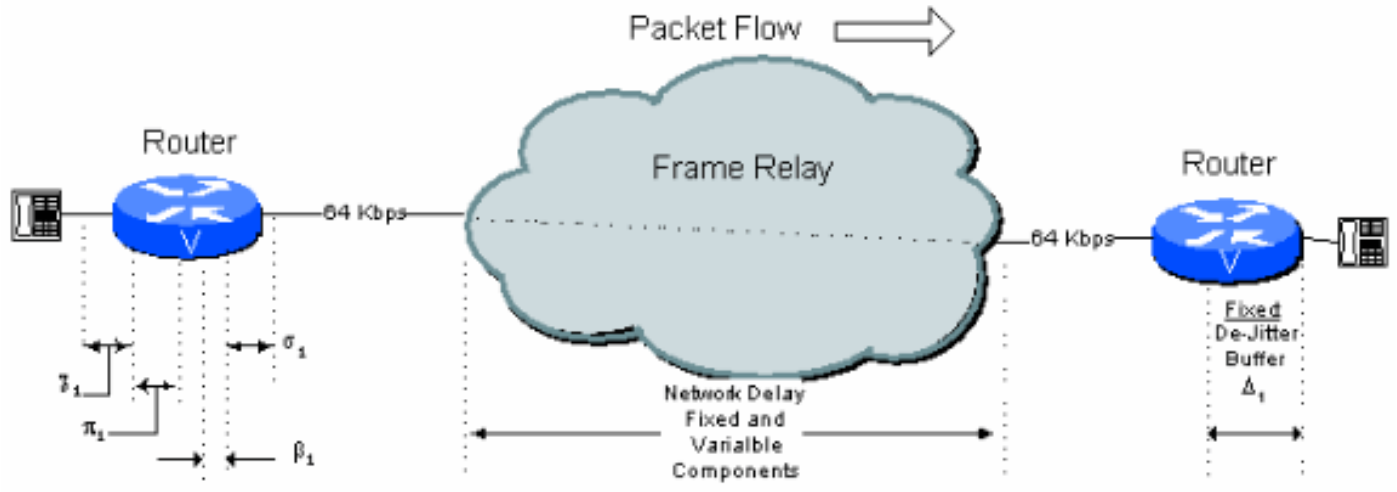
Die allgemein akzeptierte Grenze für eine qualitativ hochwertige Sprachverbindung beträgt 200 ms unidirektional (oder 250 ms als Grenzwert). Wenn sich die Verzögerungen über diese Zahl erhöhen, werden Redner und Zuhörer nicht synchronisiert, und oft sprechen sie gleichzeitig, oder beide warten, bis der andere spricht. Diese Bedingung wird häufig als Überlappung von Sprechern bezeichnet. Die allgemeine Sprachqualität ist akzeptabel, aber manchmal ist es für die Benutzer unannehmbar, dass das Gespräch auf dem Stich geblieben ist. Bei internationalen Telefongesprächen, die über Satellitenverbindungen erfolgen, kann eine Überlappung der Sprechanlage beobachtet werden (die Satellitenverzögerung liegt bei 500 ms, 250 ms hoch und

250 ms unten).

Diese Beispiele veranschaulichen verschiedene Netzwerkkonfigurationen und die Verzögerungen, die der Netzwerkdesigner berücksichtigen muss.

Single-Hop-Verbindung

Abbildung 6 - 1: Single-Hop-Beispielverbindung



Aus dieser Abbildung kann bei einer typischen 1-Hop-Verbindung über eine öffentliche Frame-Relay-Verbindung das Verzögerungsbudget in Tabelle 6.1 verwendet werden.

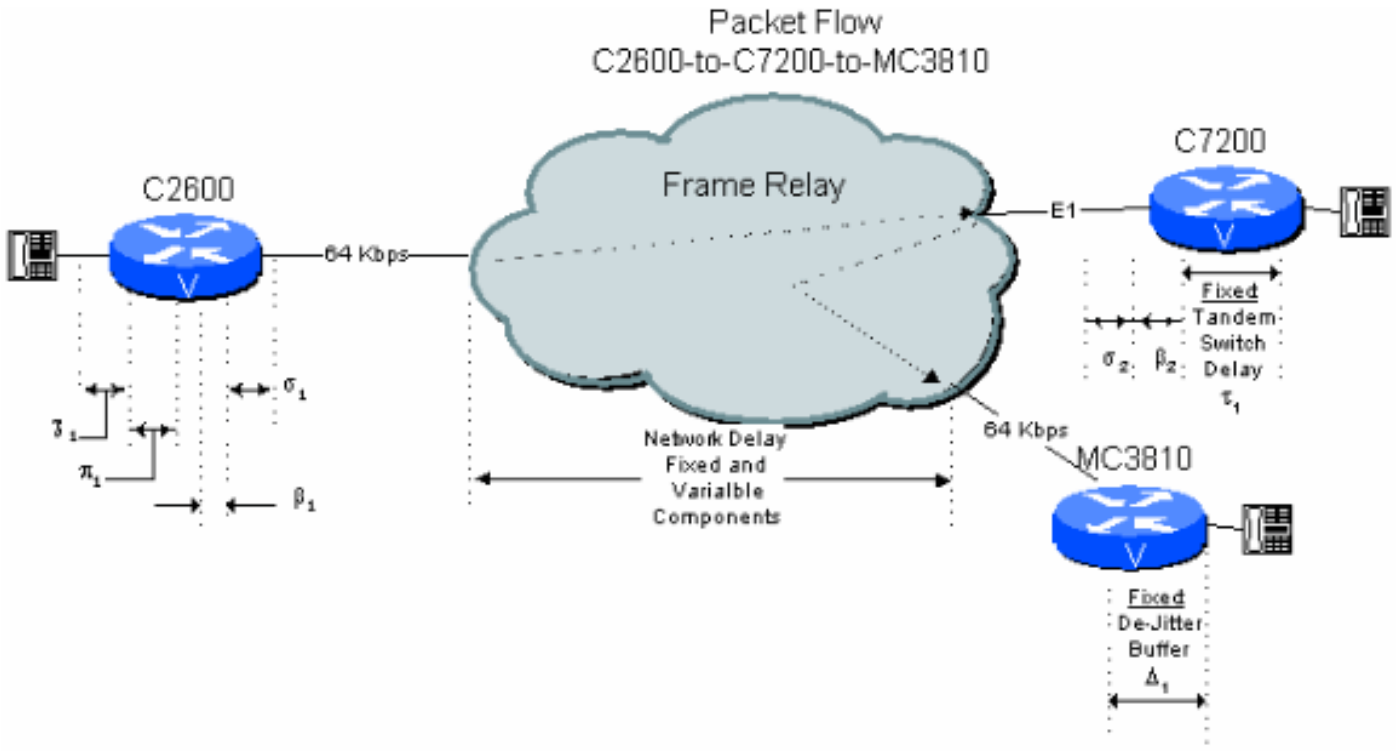
Tabelle 6.1: Berechnung der Single-Hop-Verzögerung

Verzögerungstyp	Fest (ms)	Variable (ms)
Reichweite$_1$	18	
Packetisierungsverzögerung, $_1$. Quartal	30	
Warteschlangenverwaltung/Pufferung, ß$_1$		8
Serialisierungsverzögerung (64 Kbit/s), $_1$	5	
Netzwerkverzögerung (öffentlicher Frame), Lokalisierung 1	40	25
De-Jitter-Pufferverzögerung, richtig$_1$	45	
Gesamtsumme	138	33

Hinweis: Da Warteschlangenverzögerungen und die variable Komponente der Netzwerkverzögerung bereits in den De-Jitter-Pufferberechnungen berücksichtigt werden, entspricht die Gesamtverzögerung im Prinzip nur der Summe aller Festverzögerungen. In diesem Fall beträgt die Gesamtverzögerung 138 ms.

Zwei Hops in einem öffentlichen Netzwerk mit einem C7200, der als Tandem-Switch fungiert

Abbildung 6 - 2: Beispiel für ein öffentliches Hops-Netzwerk mit Router/Gateway-Tandem



Betrachten Sie jetzt eine Verbindung zwischen Zweigstellen in einem Sterntopologienetzwerk, in dem der C7200 im Hauptsitz den Anruf an die Zielverzweigung weiterleitet. In diesem Fall bleibt das Signal durch den zentralen C7200 im komprimierten Format. Dies führt zu erheblichen Einsparungen beim Verzögerungsbudget im Vergleich zum nächsten Beispiel, der Zwei-Hop-Verbindung über ein öffentliches Netzwerk mit einem Tandem-PBX-Switch.

Tabelle 6.2: Berechnung der öffentlichen 2-Hop-Netzwerkverzögerung mit Router/Gateway-Tandem

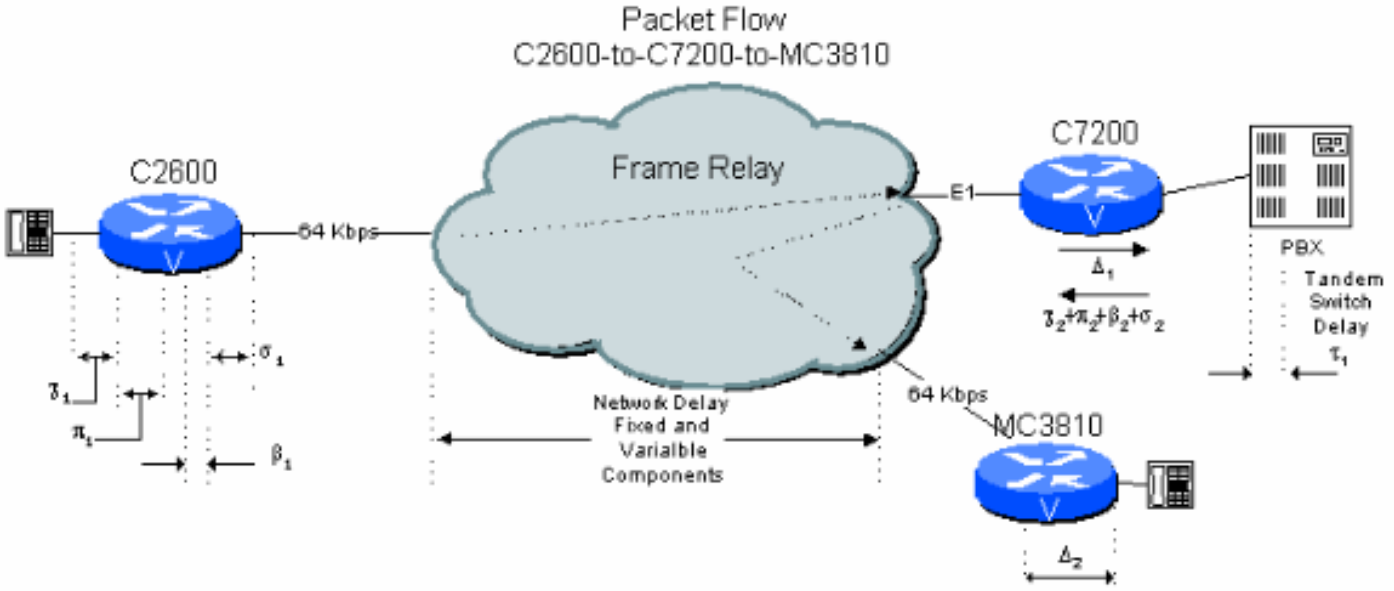
Verzögerungstyp	Fest (ms)	Variable (ms)
Reichweite$_1$	18	
Packetisierungsverzögerung, $_1$. Quartal	30	
Warteschlangenverwaltung/Pufferung, ß$_1$		8
Serialisierungsverzögerung (64 Kbit/s), $_1$	5	
Netzwerkverzögerung (öffentlicher Frame), Lokalisierung 1	40	25
Tandem-Verzögerung in MC3810, Überlagerung$_1$	1	
Warteschlangenverwaltung/Pufferung, ß$_2$		0.2
Serialisierungsverzögerung (2 Mbit/s), $_2$	0.1	
Netzwerkverzögerung (öffentlicher Frame), Lokalisierung$_2$	40	25
De-Jitter-Pufferverzögerung, richtig$_1$	75	

Gesamtsumme	209.1	58.2

Hinweis: Da Warteschlangenverzögerungen und die variable Komponente der Netzwerkverzögerung bereits in den De-Jitter-Pufferberechnungen berücksichtigt werden, entspricht die Gesamtverzögerung im Prinzip nur der Summe aller Festverzögerungen. In diesem Fall beträgt die Gesamtverzögerung 209,1 ms.

Zwei-Hop-Verbindung über ein öffentliches Netzwerk mit einem Tandem-PBX-Switch

Abbildung 6-3: Beispiel für ein öffentliches Netzwerk mit zwei Hop und PBX Tandem



In einem Zweigstellennetzwerk, in dem der C7200 im Hauptsitz die Verbindung zum PBX-System des Hauptsitzes durchläuft, beispielsweise eine Verbindung zwischen Zweigstellen und Zweigstellen. Hier muss das Sprachsignal dekomprimiert und entfernt werden. Ein zweites Mal muss es wieder komprimiert und entfernt werden. Dies führt zu zusätzlichen Verzögerungen im Vergleich zum vorherigen Beispiel. Darüber hinaus reduzieren die beiden CS-ACELP-Komprimierungszyklen die Sprachqualität (siehe Effekte mehrerer Komprimierungszyklen).

Tabelle 6.3: Berechnung der Verzögerung öffentlicher Netzwerke mit zwei Hop und PBX Tandem

Verzögerungstyp	Fest (ms)	Variable (ms)
Reichweite$_1$	18	
Packetisierungsverzögerung, $_1$. Quartal	30	
Warteschlangenverwaltung/Pufferung, ß$_1$		8
Serialisierungsverzögerung (64 Kbit/s), $_1$	5	
Netzwerkverzögerung (öffentlicher Frame), Lokalisierung 1	40	25
De-Jitter-Pufferverzögerung, richtig$_1$		40
Coder Delay, Reichweite$_2$	15	

Packetisierungsverzögerung, $_2$. Tag	30	
Warteschlangenverwaltung/Pufferung, ß$_2$		0.1
Serialisierungsverzögerung (2 Mbit/s), $_2$	0.1	
Netzwerkverzögerung (öffentlicher Frame), Lokalisierung$_2$	40	25
De-Jitter-Pufferverzögerung, richtig$_2$	40	
Gesamtsumme	258.1	58.1

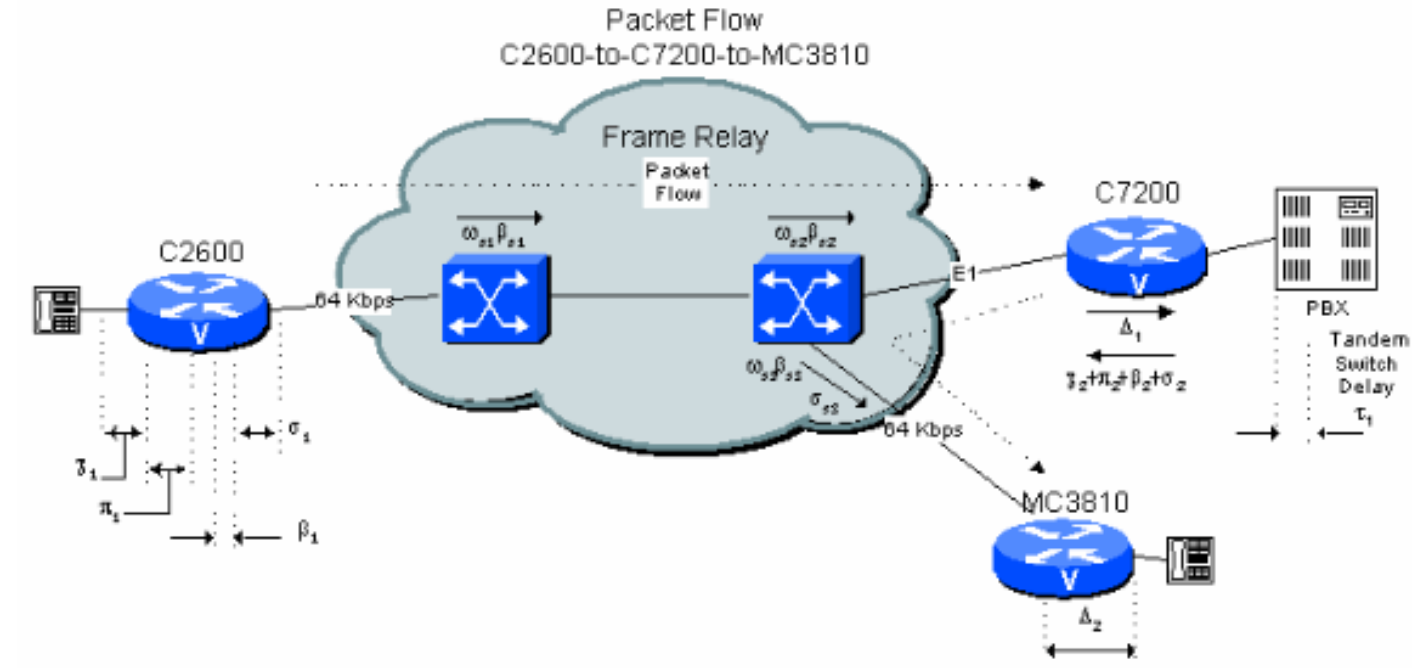
Hinweis: Da die Warteschlangenverzögerung und die variable Komponente der Netzwerkverzögerung bereits in den De-Jitter-Puffer-Berechnungen berücksichtigt wurden, entspricht die Gesamtverzögerung im Prinzip nur der Summe aller Festverzögerungen plus der De-Jitter-Pufferverzögerung. In diesem Fall beträgt die Gesamtverzögerung 258,1 ms.

Wenn Sie das PBX-System an einem zentralen Standort als Switch verwenden, erhöht sich die Verzögerung der unidirektionalen Verbindung von 206 ms auf 255 ms. Dies entspricht in etwa den ITU-Grenzwerten für eine unidirektionale Verzögerung. Bei dieser Art der Netzwerkkonfiguration muss der Techniker das Design sorgfältig verfolgen, um eine möglichst geringe Verzögerung zu vermeiden.

Der schlimmste Fall wird bei variabler Verzögerung angenommen (obwohl in beiden Abschnitten des öffentlichen Netzes keine maximalen Verzögerungen gleichzeitig auftreten). Wenn Sie optimistischere Annahmen für die variablen Verzögerungen machen, verbessert dies die Situation nur minimal. Mit besseren Informationen über feste und variable Verzögerungen im Frame Relay-Netzwerk des Carriers kann die berechnete Verzögerung jedoch reduziert werden. Lokale Verbindungen (z. B. innerstaatlich) haben vermutlich wesentlich bessere Verzögerungstoleranzen, aber die Betreiber zögern häufig, Verzögerungstoleranzen einzuführen.

Zwei-Hop-Verbindung über ein privates Netzwerk mit einem Tandem-PBX-Switch

Abbildung 6-4: Beispiel für ein privates Netzwerk mit zwei Hosts und PBX Tandem



Beispiel 4.3 zeigt, dass es unter der Annahme von Verzögerungen im schlimmsten Fall sehr schwierig ist, die berechnete Verzögerung unter 200 ms zu halten, wenn eine Verbindung zwischen Zweigstellen und Zweigstellen einen Tandem-Hop am zentralen Standort mit öffentlichen Frame-Relay-Netzwerkverbindungen auf beiden Seiten umfasst. Wenn die Netzwerktopologie und der Datenverkehr jedoch bekannt sind, kann die berechnete Anzahl erheblich reduziert werden. Dies liegt daran, dass die von den Betreibern im Allgemeinen angegebenen Zahlen durch die schlimmste Übertragung und Warteschlangenverzögerung über einen großen Bereich begrenzt sind. Es ist viel einfacher, in einem privaten Netzwerk vernünftigere Grenzen zu setzen.

Die allgemein akzeptierte Zahl für die Übertragungsverzögerung zwischen Switches beträgt 10 Mikrosekunden pro Meile. Je nach Gerät muss die Verzögerung des Trans-Switches in einem Frame-Relay-Netzwerk zwischen 1 ms fest und 5 ms variabel sein, um in die Warteschlange aufgenommen zu werden. Diese Zahlen beziehen sich auf Geräte und Datenverkehr. Die Verzögerungswerte für die Cisco MGX-WAN-Switches betragen weniger als 1 ms pro Switch insgesamt, wenn E1/T1-Trunks verwendet werden. Bei Annahme einer Entfernung von 500 Meilen, 1 ms fest und 5 ms variabel für jeden Hop, ergibt sich die Verzögerungsberechnung wie folgt:

Tabelle 6.4: Berechnung der privaten 2-Hop-Netzwerkverzögerung mit PBX Tandem

Verzögerungstyp	Fest (ms)	Variable (ms)
Reichweite$_1$	18	
Packetisierungsverzögerung, $_1$. Quartal	30	
Warteschlangenverwaltung/Pufferung, ß$_1$		8
Serialisierungsverzögerung (64 Kbit/s), $_1$	5	
Netzwerkverzögerung (Privater Frame), $_{S1}$ + ß$_{S1}$ + Audioformate $_{S2}$ + ß$_{S2}$	2	10
De-Jitter-Pufferverzögerung, richtig$_1$	40	
Coder Delay, Reichweite$_2$	15	
Packetisierungsverzögerung, $_2$. Tag	30	
Warteschlangenverwaltung/Pufferung, ß$_2$		0.1
Serialisierungsverzögerung (2 Mbit/s), $_2$	0.1	
Netzwerkverzögerung (Privater Frame), $_{S3}$ + ß$_{S3}$	1	8
Serialisierungsverzögerung (64 Kbit/s), $_{S3}$	5	
De-Jitter-Pufferverzögerung, richtig$_2$	40	
Übertragungs-/Entfernungsverzögerung (nicht aufgeschlüsselt)	5	
Gesamtsumme	191.1	26.1

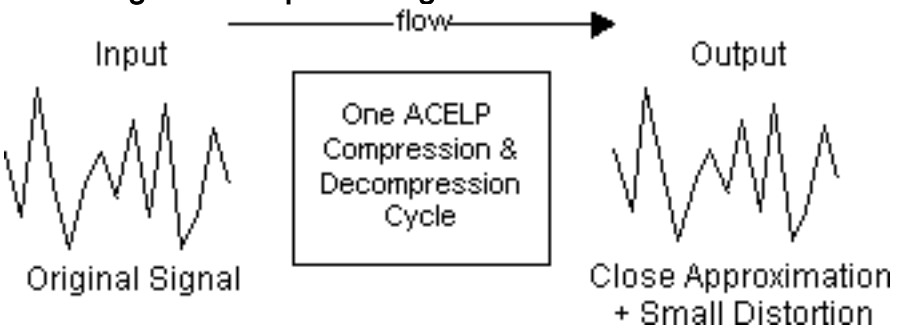
Hinweis: Da Warteschlangenverzögerungen und die variable Komponente der Netzwerkverzögerung bereits in den De-Jitter-Pufferberechnungen berücksichtigt wurden, entspricht die Gesamtverzögerung nur der Summe aller Festverzögerungen. In diesem Fall beträgt die Gesamtverzögerung 191,1 ms.

Wenn Sie über ein privates Frame-Relay-Netzwerk laufen, ist es möglich, eine Spoke-to-Spoke-Verbindung über das PBX-System am Hub-Standort herzustellen und innerhalb der 200-ms-Zahl zu bleiben.

Auswirkungen mehrerer Komprimierungszyklen

Die CS-ACELP-Komprimierungsalgorithmen sind nicht deterministisch. Das bedeutet, dass der Datenstrom der Eingabedaten nicht genau mit dem Datenstrom der Ausgabedaten identisch ist. Bei jedem Komprimierungszyklus wird eine geringe Verzerrung eingeführt, wie in Abbildung 7-1 dargestellt.

Abbildung 7-1: Komprimierungseffekte



Infolgedessen führen mehrere CS-ACELP-Komprimierungszyklen schnell zu erheblichen Verzerrungen. Dieser additive Verzerrungseffekt ist bei ADPCM-Algorithmen (Adaptive Differenzial Pulse Code Modulation) nicht so ausgeprägt.

Diese Eigenschaft hat den Vorteil, dass der Netzwerkdesigner zusätzlich zu den Auswirkungen der Verzögerung die Anzahl der CS-ACELP-Komprimierungszyklen im Pfad berücksichtigen muss.

Die Sprachqualität ist subjektiv. Die meisten Benutzer stellen fest, dass zwei Komprimierungszyklen immer noch eine angemessene Sprachqualität bieten. Ein dritter Komprimierungszyklus führt in der Regel zu einer deutlichen Beeinträchtigung, die für einige Benutzer inakzeptabel sein kann. In der Regel muss der Netzwerk-Designer die Anzahl der CS-ACELP-Komprimierungszyklen in einem Pfad auf zwei beschränken. Wenn mehr Zyklen verwendet werden müssen, lassen Sie es dem Kunden zuerst zuhören.

In den vorherigen Beispielen wird gezeigt, dass bei einem Tandem-Switching einer Zweigstellenverbindung über das PBX-System (in PCM-Form) im Hauptsitz eine erheblich höhere Verzögerung auftritt, als wenn es im Hauptsitz C7200 als bei einem Tandem-Switching betrieben würde. Wenn das PBX-System zum Umschalten verwendet wird, sind im Pfad zwei CS-ACELP-Komprimierungszyklen vorhanden, und nicht ein Zyklus, wenn die gerahmte Sprache vom zentralen C7200 geschaltet wird. Die Sprachqualität ist mit dem C7200-Switch-Beispiel (4.2) besser, auch wenn es andere Gründe geben kann, z. B. die Verwaltung des Anrufplans, die die Einbindung des PBX-Systems in den Pfad erfordern.

Wenn eine Verbindung zwischen Zweigstellen und Zweigstellen über ein zentrales PBX-System hergestellt wird und der Anruf von der zweiten Zweigstelle über das öffentliche Sprachnetzwerk

weitergeleitet und dann in einem Mobilfunknetz terminiert wird, befinden sich drei CS-ACELP-Komprimierungszyklen im Pfad sowie eine erheblich höhere Verzögerung. In diesem Szenario wird die Qualität spürbar beeinträchtigt. Auch hier muss der Netzwerk-Designer den im schlimmsten Fall möglichen Anrufpfad berücksichtigen und entscheiden, ob dieser unter Berücksichtigung des Netzwerks, der Erwartungen und der geschäftlichen Anforderungen der Benutzer akzeptabel ist.

Überlegungen für Verbindungen mit hoher Verzögerung

Es ist relativ einfach, Packet-Voice-Netzwerke zu entwerfen, die die ITU überschreiten, allgemein akzeptierte 150-ms-Verzögerungsgrenze.

Beim Entwurf von Sprachnetzwerken für Pakete muss der Techniker berücksichtigen, wie oft eine solche Verbindung verwendet wird, welche Anforderungen der Benutzer stellt und welche geschäftlichen Aktivitäten damit verbunden sind. Es ist nicht ungewöhnlich, dass solche Verbindungen unter bestimmten Umständen akzeptabel sind.

Wenn die Frame-Relay-Verbindungen keine große Distanz überschreiten, ist es sehr wahrscheinlich, dass die Verzögerungsleistung des Netzwerks besser ist als die in den Beispielen.

Wenn die Gesamtverzögerung für Tandem-Router-/Gateway-Verbindungen zu groß wird, besteht eine Alternative häufig darin, zusätzliche permanente virtuelle Schaltungen (PVCs) direkt zwischen den terminierenden MC3810s zu konfigurieren. Dadurch entstehen wiederkehrende Kosten für das Netzwerk, da die Carrier in der Regel pro PVC berechnen. In einigen Fällen kann dies jedoch erforderlich sein.

Zugehörige Informationen

- Internationale Fernmeldeunion
- Unterstützung von Sprachtechnologie
- Produkt-Support für Sprach- und Unified Communications
- Fehlerbehebung bei Cisco IP-Telefonie
- Technischer Support – Cisco Systems