

第二章 VoIP 系統架構

近幾年來，VoIP 的產品慢慢廣為應用，但在應用上所必須面對的問題就是如何解決網際網路壅塞而造成語音封包時間延遲的問題，因網路壅塞而造成而造成語音封包播放延遲，往往會造成語音品質下滑。

在本章節中，會對 H.323 和所使用的語音壓縮標準（Codec），和 VoIP [6][7] 如何在網路上之使用方式提出說明和解釋，及 VoIP 在網路上會遭遇的壅塞情況作個簡單說明。

2.1 H.323 簡介

網際網路電話能讓通話雙方透過網際網路交換語音訊息，但是需要訂定一個通用的標準。因此藉由國際電信聯盟（ITU）訂定的 H.323 規範，正好可以來做為網際網路電話的標準。

H.323 (圖 2.1) 是一套包括視訊、語音、與資料傳遞的標準。然而在目前的網際網路上，頻寬不足的情況下，要實現即時多媒體的傳送是比較困難的。因此利用語音資料量較少的特性下，藉由網際網路來實現網際網路電話是比較可行的。所以利用 H.323 中的語音傳送部分來做為網際網路電話傳送是不錯的選擇。

H.323 系統包括了終端設備、閘道器、閘道網路管理系統、以及多

方會談控制單元。

H.323 所包含的資料標準有很多，其中在語音方面包括 G.711、G.722、G.723.1、G.728、G.729 等等多種壓縮標準，而其中以 G.723.1、G.729 為最廣泛使用在網際網路電話閘道器的語音壓縮標準。在影像方面包括 H.261、H.263，這兩個標準都是視訊會議（Video Conference）的標準。在資料傳送方面是以 T.120 為主。另外 H.323 的系統控制方面，則訂定了系統控制 H.245、播號設定 Q.931、使用管理 RAS（註冊、加入、狀態查詢）。圖 2.1 是 H.323 架構圖，用一個簡單的例子來說明雙方通訊建立的流程。首先是透過 Q.931 來播號設定，接著雙方建立 H.245 資料傳輸通道，交換容量能力資訊，決定主從關係，之後雙方各自開啟 RTP（Real-time Transport Protocol）通道以及 RTCP（Real Time Control Protocol）通道，並且互相開啟一條邏輯通道，以便做雙方通話。

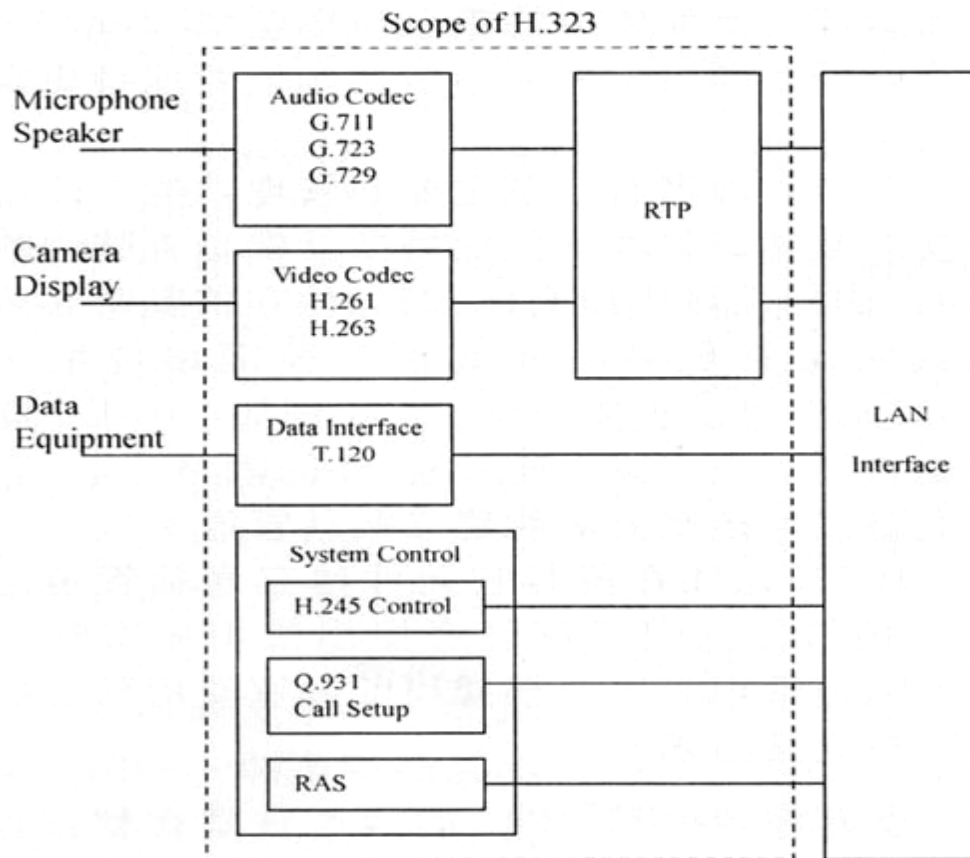


圖 2.1 H.323 架構圖

我所做的 VoIP 實驗架構就是架構在 H.323 下，並藉由兩端的電腦來作溝通。

2.2 G.729 基本介紹

G.729 為 ITU-T 於 1996 訂定的 8Kbits/sec 語音壓縮標準，其所接受的輸入語音資料為 8KHz 取樣頻率取樣，並以 16bits 量化而成的數位語音資料。它採用音框導向（frame-oriented）處理模式，將輸入的語音訊號，切成一連串固定長度為 10ms 的音框（即 80 個語音數位量化值），並等分成二個子音框（subframe），再加以分析每個音

框內的語音訊號予以壓縮編碼。

G.729 的編碼架構為共軛架構線性預測算術運算碼激發編碼 (Conjugate-Structure Algebraic-Code-Excited Linear-Prediction , 簡稱 CS-ACELP) , 亦屬於 CELP 編碼一類 , 它擁有傳統參數編碼法及波形編碼法的優點 , 以一 10 階線性預測濾波器 (10th order linear prediction filter) 來表示語音訊號的短時距 (short-term) 特性之數學模型 , 並以最小化誤差分析合成搜尋法來找出語音訊號的長時距 (long-term) 類周期激發源 , 最後再利用多重脈衝訊號來逼近殘餘訊號 , 其中長時距類周期激發源及殘餘訊號的增益編碼部份是使用共軛架構碼簿 (conjugate-structure VQ) 。

G.729 編碼架構區塊主要有「線性預測分析」、「感觀權重濾波器」、「開迴路基頻分析」、「適應性碼簿搜尋」 (adaptive codebook search) 、「固定性碼簿搜尋」 (fixed codebook search) 及「增益量化」六個主要部份[5]。

利用 G.729 語音壓縮標準處理過後的語音封包會有三種不同的封包形態 , 分別是有意義封包 (Active packet) 、不傳送封包 (Untransmitted packet) 、靜音封包 (SID) , 這些語音封包是根據語音能量大小來決定 , 以語音方面來說 , 有意義的語音封包所代表的意義是人的說話聲音 , 是不能任意丟棄的 , 所以有意義封包其重要性將

遠大於不傳送封包和靜音封包。當遇到網路壅塞時，丟棄不傳送封包和靜音封包將比丟棄有意義封包來的可行，也因為如此，在本論文中將不會有三種封包形態，而只有有意義封包(Active packet)和靜音封包(Silence packet)兩種封包形態，其中靜音封包包含原來的不傳送封包和靜音封包。

2.3 VoIP 基本架構

VoIP (Voice over IP) 就是把語音類比信號變成數位信號之後，包裝成封包然後透過網路傳送到遠端的接受端來完成通話，和一般透過電話和遠端做溝通方式相比下，傳統方式的電話溝通，除了費用過高之外並不能其他的服務，這對科技發達的現在是不合適的。相對地，VoIP 除了能傳送語音之外，還能傳送一般的文件資料，提供其他的加值服務，這在這個資訊爆炸的時代，產品多功能的應用是不可或缺的。

而 VoIP 是為了減少電話花費成本及得到更多的加值服務而發展出來的，再加上現在網路已經很普遍，不論在學校或者是企業都常常利用個人電腦 (Personal Computer) 在網路上瀏覽資料，就連一般的家庭也經常藉由數據機撥接來一窺這個廣大的網路世界，透過網路我們不必再負擔昂貴的電話費用，只需負擔部分的撥接費和上網費便

可，傳統電話頻寬是 64kbps 也就是有自己的專線，只要連上了對方電話直到掛上電話之中雙方都有 64kbps 的頻寬可做溝通，直到雙方都掛上電話為止這條連線才斷掉，使用者如果沒說話時，所佔之頻寬也不會釋放給別人使用，造成頻寬浪費，除此外，在長途電話或者是國際電話因距離的增加所造成的頻寬浪費更多，那就所費不貲了，因這 64kbps 頻寬的效率，便是 VoIP 努力克服的地方，除此之外，對於 VoIP 還能提供傳統電話無法做到的一些加值服務。

VoIP 是透過封包在網際網路傳送，所以傳送的是數位信號，也因傳送的是數位信號，所以各種的壓縮技術便因應而生，在 VOIP 的架構中一般都是採用 G.723.1 和 G.729 兩種，這兩種主要的差別是在於 G.723.1 有 5.3kbps 和 6.3kbps 兩種壓縮率，而 G.729 有 8kbps 一種壓縮率。

在整個網際網路電話的架構上，包括了終端設備（Terminal - 輸入及輸出設備）、閘道器（Gateway）、閘道網路管理系統（GateKeeper）、網際網路（Internet）、網際網路程式（Internet Programming）等等[6]。其中的閘道器是語音壓縮及解壓縮編碼的心臟引擎，是佔據最多運算資源的部分。在閘道器為提供多通道同時通話的特性下，處理速度是一個相當嚴重的問題。以一條 T1 線為例，可提供 24 的語音通道，如接上閘道器，則閘道器一定要能同時處理

完 24 個語音通道，以目前市面上一般處理器的能力而這，是很困難的，因此利用數位訊號處理器（Digital Signal Processor - DSP）來實現閘道器是比較可行的。

一般的 VOIP 都是先把原始原始透過不同的語音壓縮技術，把原始原始成不同的 bit rate，再透過 Gatekeeper 的控制，把壓縮的語音通過 Gateway 傳送到 Internet，並且由 Internet 傳送到目的地的 Gateway，再透過目的地 Gatekeeper 的控制，把傳送過來的語音經過相同的語音壓縮技術來解壓縮，再播放出來給對方聽，這就是一般的 VOIP 傳送原理，圖 2.2 是 VOIP 的架構。

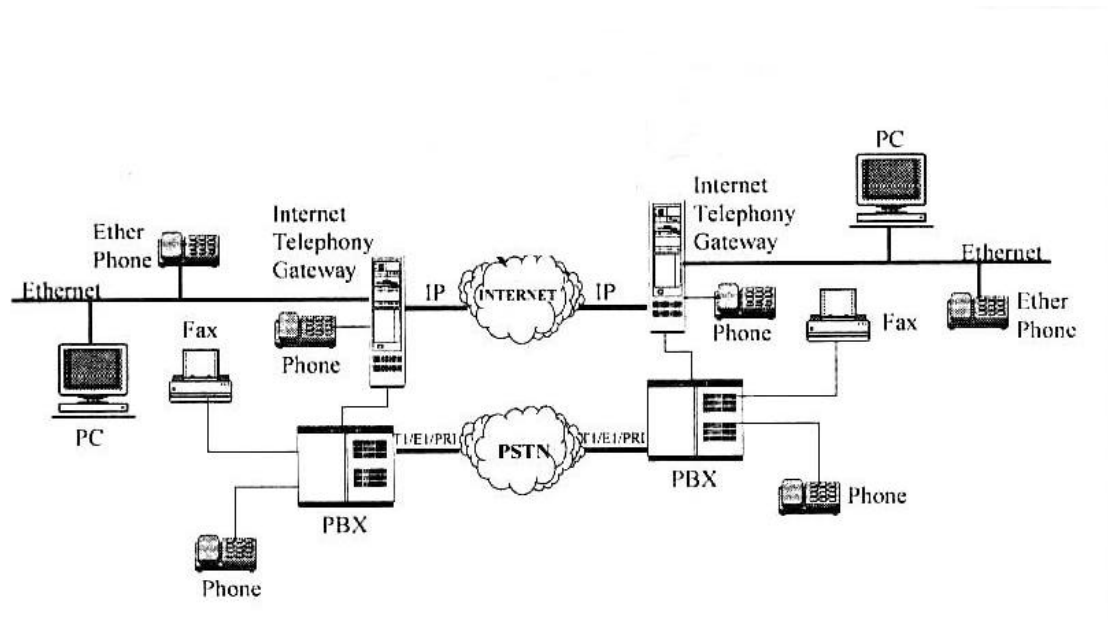


圖 2.2 VOIP 的架構

2.4 網路延遲

雖然透過網路已經可以達到無遠弗屆、世界一家的境界，但是在網際網路上頻寬有限，上網人數無限的情形下而造成網路壅塞卻是不爭的事實。也因如此，網路壅塞正是網路使用者共同面對的難題，這對即時的語音傳送更是急需解決的難題，而對於網路壅塞所造成對語音品質傷害降低最低的方法也就是本篇論文所研究的主題。

語音具有即時性（Real-time）和連續性（Continuous），所以語音封包延遲到達時，會對下一個封包產生連續性的影響，如果網路延遲變化大，對目的地的接收者而言，會聽到斷斷續續的語音而不是連續的一段語音，尤其當連續的語音封包被延遲播放或丟棄時，對語音品質傷害最大，不僅造成了聽覺上的不舒服外，甚至會無法清楚明白了解語音播放出的內容。所以保持語音封包的完整性和及時性正是本篇論文所探討的重點。

而本論文所採用的方法就是情願讓第一個語音封包在我們所建立的緩衝區先等待一定的封包數之後然後在一起播放出去，讓撥出的語音能夠更平順，不會有斷斷續續的情形發生。

圖 2.3 說明語音延遲時會發生的情形，在圖 2.3 中，可以知道要經過網路傳送的語音，第一步一定要先封包化（Packetized），然後再經由網路傳送到接受端，因為網路壅塞情形的發生，到達接受端時語

音封包並不會像原始語音一樣，是連續的到達，語音封包和語音封包之間會有間斷的情形產生，如果以這種情況下播放語音封包的話，會讓聽話端無法明白對方所要傳達的意思，所以如何有效的避免網路延遲對語音品質的傷害，便是一項重要的研究課題。

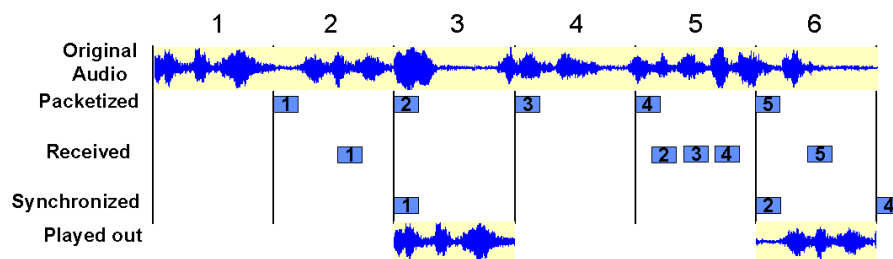


圖 2.3 網路延遲

在圖 2.3 中，利用語音壓縮標準（G.723.1、G.729）把原始語音封包化，封包化成六個封包，從圖中可以知道其封包是有順序的，藉由這個順序號碼可以讓接收端把原始語音還原，除此之外，在每個封包之間也都間隔一定時間，但經過網路傳送時卻造成每個封包之間的間隔時間已經不是固定的，因為之間已經受到網路延遲的影響，這種狀況對接受端是很困擾的，並會造成語音播放時，語音之間會有不連續的情形發生，像圖 2.3 中語音波形之間會有間斷的情形產生。