

以 SIP 為基礎之 VoIP 電話系統用戶代理者與伺服器的研究

A Research for User Agents and Servers of VoIP Phone System Based on SIP

詹凱富 應尚宏 洪永政 林強生 章博傑 葉生正*

銘傳大學資訊傳播工程學系

桃園縣 333 龜山鄉德明路 5 號

*E-mail: peteryeh@mcu.edu.tw

摘要

SIP (Session Initiation Protocol) 除了可建立 VoIP 網路電話之應用, 也可與 HTTP、SMTP、RTP... 等協定結合, 提供整合語音與其它多媒體的通訊服務, 如即時訊息 (Instant Messaging) 與個人現形 (Presence) 服務。另外, SIP 已經被公認為 IP 網路與傳統 PSTN 電信網路整合的關鍵技術, 而且亦被第三代行動通訊系統 (3G) 採用為未來無線多媒體通訊的技術標準。故本論文以 SIP 為標準通訊協定進行用戶代理者 UA (User Agent) 的研發, 完成與 Xten 的 X-lite 和工研院的 CCL Phone 之相容性測試。並對本研究所整合應用的伺服器 SER (SIP Express Router) 歸納分析其系統設定與操作管理程序, 且對穿越 NAT (Network Address Translation) 的技術提出初步的解決方法。

關鍵詞: VoIP、SIP、Instant message、User Agent、SER、3G、NAT

Abstract

SIP not only can make session of IP phone, but also can integrate the techniques including HTTP, SMTP, and RTP.... SIP provides integration voice and multimedia communication service, such as Instant Message and Presence, etc. Also, SIP is regarded as resolution to the problem of signal transmission between different types of network. The particular study intending to invite a real UA (User Agent) bases on SIP and the interoperation test of X-lite and CCL Phone. In addition, the study suggests the experiences of the deeper management of SER (SIP Express Router) and the reservation of NAT (Network Address Translation) issue.

Keywords: VoIP、SIP、Instant message、User Agent、SER、3G、NAT

1. 簡介

電信總局開放網路電話在即, 許多業者紛紛以網路電話產品進軍市場, 國內 VoIP (Voice over Internet Protocol) 的時代已即將啟動。目前已有許多相關產品正在網路上流通, 而最近 Google 發表 Google talk 以及 eBay 大手筆收購 Skype 等消息新聞, 說明了 VoIP 這塊領域不但商機無限, 也顯示眾多科技大廠爭奪客戶的決心。在 VoIP 如此熱門之際也衍生出許多問題, 其中如何將眾多不同廠商所研發的產品整合便是, 所以遵循 SIP (Session Initiation Protocol) 絕對是最好的解決方法。

SIP 是對應於 OSI 七層參考架構中的應用層 (Application Layer) 通訊協定, 用來建立、修改和終止使用者的連線 (Session)。而這些使用者的連線可以是點對點亦或是點對多點的關係, 且 SIP 可以支援網際網路多媒體、多方會談、電話、訊息傳遞亦或遊戲等服務。另外, SIP 被認為是一種輕量級協議, 因為該協議使用能被終端設備輕易產生與分析的簡訊命令, 只需使用 6 個指令進行管理呼叫控制資訊。同時在開發複雜度及成本上, SIP 都較 H.323 協定佔優勢。但其成熟度由於發展較晚, 不如 H.323 是理所當然的。若從定義範圍來看, H.323 從通訊到壓縮都有一定的標準, 這也代表著 H.323 協定在整合上的難度, 以及開發的元件必須耗費更大的系統資源。相對地, SIP 是以明文傳送, 可直接使用 HTTP 等格式直接溝通, 對於擴充性與相容性方面比起 H.323 佔有很大的優勢。最重要的是, SIP 已經被第三代行動通訊系統採用為多媒體通訊的技術標準, 顯見 SIP 足以適用於未來網際網路之環境[5][8]。

本論文即以 SIP 為標準通訊協定進行用戶代理者 UA (User Agent) 的研發, 完成與 Xten 的 X-lite 和工研院的 CCL Phone 之相容性測試。並對本研究所整合應用的伺服器 SER (SIP Express Router) 歸納分析其系統設定與操作管理程序, 且對穿越 NAT (Network Address Translation) 的技術提出初步的解決方法。本論文 Section 2 將詳述 SIP 的基本架構與原理; Section 3 則介紹本研究系統之 UA 開發和 SER 的應用。最後為扼要的結論與對未來相關

研究議題的建言。

2. 相關協定之研究

2.1 SIP 協定

(1) 簡介

SIP 僅負責連線建立，之後則視需求去傳送需要的連線內容。且 SIP 必須搭配其他協定才有意義，例如與 RTP (Real-time Transport Protocol) 協定結合，能夠加強在 UDP 上的多媒體傳送品質，因為 SIP 只建立了基本連線，之後的多媒體傳送並不在其定義裡面。此外 SIP 是採用文字式 (Text-base) 的通訊協定，能在兩個或多個傳送點之間建立與控制多媒體傳送 (Multimedia session)。同時 SIP 亦具有 Client-Server 架構，可利用 HTTP 原有的封包格式資訊。再與 H.323 比較起來，從連線的管理到音訊編碼，以及 VoIP 系統服務的定義都全部包辦，而 SIP 卻只需要職司於連線的部分，故 SIP 簡單並容易廣為應用[5][8][10]。

(2) SIP 基本架構

SIP 的主要元件包括了使用者代理程式 (UA, User Agent)、代理伺服器 (Proxy Server)、重導伺服器 (Redirect Server) 和註冊伺服器 (Registrar)。SIP 的客戶端稱之為 User Agent Client (UAC)，伺服器端為 User Agent Server (UAS)；凡是送出 SIP request 的就是 SIP Client，而接收 SIP request 為 SIP Server，因此 UAC 和 UAS 可能為同一部電腦或 SIP Phone。代理伺服器的角色與 Web service 中的 proxy 相似，負責轉送連線控制訊息，也就是將 SIP message 傳送到正確的地方，如圖 1 所示。且因 UA 彼此之間只擁有對方的 SIP Address，但不知道實際的位置為何，所以訊息的傳送必須透過代理伺服器，代理訊息間的溝通。實際上，SIP message 傳送過程中必須經過許多代理伺服器轉送才能正確到達目的端；當然這些資訊都會紀錄在 message 裡面。所以被傳送的那一端回傳訊息時，便依照這些資訊轉傳回來。另外，代理伺服器亦可提供類似隨時傳送 (Time-of-day routing)、跟著我服務 (Follow-me service) 的功能[8][10]。

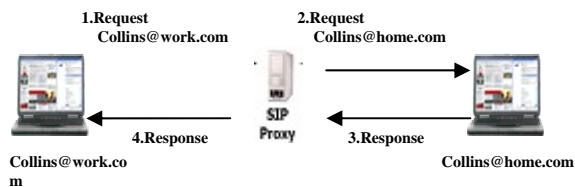


圖 1 代理伺服器工作原理

重導伺服器主要的功能是對使用者到不同的網路位址。當某使用者從一個網域移動到另一個網域時，若有 UAC 撥打給此用戶，便需要重導伺服器向註冊伺服器詢問此一用戶的目前位置並回傳

給 UAC，讓 UAC 的 message 能夠正確傳送，如圖 2 所示。重導伺服器本身並不參與該訊息任何處理的動作。

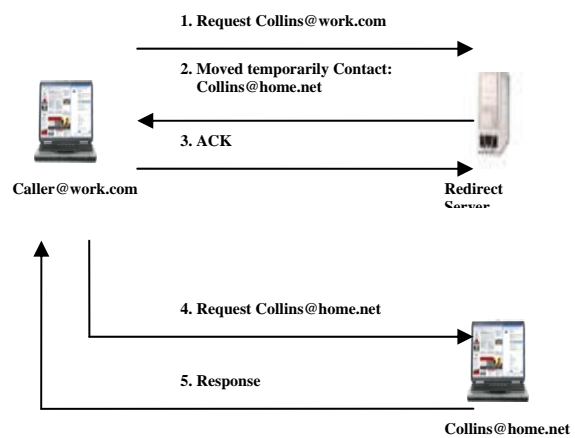


圖 2 重導伺服器工作原理

註冊伺服器是負責接收 SIP Register request，其表現方式近似於電子郵件位址，格式為 User@host。當使用者移動到別的地方時，可以透過 SIP REGISTER request 告知 registrar 新位置，通常註冊伺服器會搭配代理伺服器與重導伺服器，一起達成使用者行動性的功能[8][10]。

圖 3 是一個簡單的 SIP 建立通話流程。透過 Invite, OK, ACK 這樣的 Three Way Handshake 方式來建立連線。其中 Ringing response 為暫時回應 (provisional response)。所謂的暫時回應 是提供額外的資訊給 caller，在此處就是告知 UAC ring-back tone。而 OK response 為最終回應 (final response)，必須要送一個 ACK request 給 caller 來表示 final response 正確的傳到 callee[8][10]。

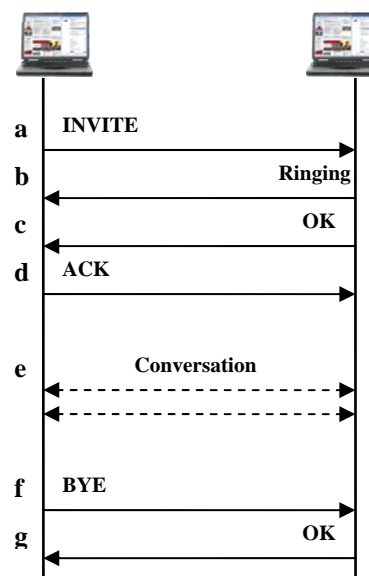


圖 3 SIP 建立連線簡易圖

(3) SIP命令與控制訊息格式

SIP的message結構是以文字為主，而且類似Http (Hypertext Transmission Protocol)，相較於Library-Base而言，SIP在語法分析上簡單，但是會浪費更多的頻寬。SIP message的內容中，第一行用來表示SIP Request或Response的命令列，之後連續數行是message header用換行 (CRLF)，最後才是message的主要內容 (message-body)，而內容的陳述通常使用SDP[7]。

如果要利用SIP來建立兩個端點的連線，呼叫的那一端必須找一個代理伺服器或重導伺服器，然後送出SIP Request。Request目前有六種方法 (Methods)，分別是邀請 (INVITE)、回應 (ACK)、結束 (BYE)、取消 (CANCEL)、註冊 (REGISTER) 和選項 (OPTIONS)。每個方法的用途列舉如表1所示[10]。

表1 Request Method

INVITE	邀請使用者參與session
ACK	確認接收到INVITE的最終回應
BYE	結束session
CANCEL	取消session傳輸過程
REGISTER	向伺服器確認使用者目前所在的位置
OPTIONS	透過伺服器查詢SIP Request的能力

當代理伺服器或重導伺服器收到呼叫端的SIP Request，Server會回傳SIP Response來通知呼叫Server如何處理這個Response。SIP Response目前定義成六個類別，其中，除了100-199這個類別為暫時回應，其餘都為最終回應。而這兩大類別之間的差異在於最終回應送出後，接收到Response的那一端必須回傳ACK來做確認。

3. 網路電話系統開發

本研究的工作流程為相關資料蒐集、熟悉Winsock元件與使用方法、語音傳輸的研究與開發、利用Borland C++與繪圖軟體設計User Agent的介面、視訊傳輸方法研究、即時訊息傳輸、SER管理操作研究、User Agent與Server端的整合、簡易的穿越NAT方法等，以下將分別敘述其重點：

3.1 Winsock 元件

一個socket就是一個通訊端點，而每個網路應用程式都會一個通訊端點，分別為伺服器 (Server) 與客戶端 (Client)。用戶端會先送出第一個封包，由一個伺服器接收，在接收到封包後，用戶端和伺服器便能夠開始收送資料，再依據Socket所提供的服務來將它分類傳遞。由於session的建立是利用SIP，所使用的封包格式已經固定，只需要加上適當的內容，因此這部分的研究著重於文字與影音的

傳輸。文字的部分是以SIP/SIMPLE (SIP for Instant Messaging and Presence Leveraging Extensions) 為基礎，透過Proxy Server傳送；語音透過RTP協定傳輸，而視訊部份則是使用UDP協定。語音和視訊的傳輸皆使用Windows API，這些API將Winsock元件包裝在裡面，因此很容易設定使用。

3.2 語音傳輸的研究與開發

基於可靠性和即時傳輸資料等性質，本研究的語音傳輸採用 RTP (A Transport Protocol for Real-Time Applications) 標準。實做上使用Vovida所提供的開放原始碼 WinRTP 元件。WinRTP 是Cisco IP SoftPhone 產品的一部分，以C++所撰寫並包裝成易於呼叫的元件，可以方便的在高階語言中使用。傳送過程為UA的語音透過WinRTP元件處理，依G.711的標準壓縮，封裝成為RTP封包，經由指定的通訊埠傳送至另一端的UA；同時間收到語音的一方將存放在緩衝區內的資料播放出來。基本設定如下：

- (1) Initialize(): 起始化 WinRTP 物件
- (2) SetAudioCodecRX(): 設定受話端語音壓縮格式以及每毫秒封包容量。本研究設定為 G.711 Ulaw Codec, 30ms packet size。
- (3) SetAudioReceivePort(): 指定受話端連接埠，依照 RTP 設定為 9000
- (4) SetAudioCodecTX(): 設定發話端語音壓縮格式、優先權、Silence suppression 和每毫秒封包容量。本研究設定為 G.711 Ulaw Codec, 沒有 Silence suppression, 30ms packet size。
- (5) SetAudioDestination(): 設定收話端的 IP 位址和連接埠。IP 位址由 SIP 封包取得。
- (6) StartRX(): 開始接收封包並收音。
- (7) StartTX(): 開始錄音並傳送封包。
- (8) SetSpeakerVolume(): 設定揚聲器音量。
- (9) SetMicrophoneVolume(): 設定麥克風音量。
- (10) StopRX(): 停止接收封包與收音。
- (11) StopTX(): 停止錄音及傳送封包。
- (12) UnInitialize(): 結束使用 WinRTP 物件。

在使用者登入 SER 後，撥出有效的號碼，並在對方接通之後，則語音通訊即建立完成。語音通話時的軟體介面顯示如圖 4。



圖 4 語音連線介面顯示圖

為了驗證 SIP SoftPhone 之間互通性，經過 SIP signaling 程序後，本研究實做的 UA 能夠順利的和 X-lite 以及 CCL Phone 建立、開始和結束語音連線。

3.3 視訊傳輸方法研究

影像擷取的部分，主要是利用 Video For Window 函式庫 (vfw.h) 來偵測視訊硬體，當偵測到硬體 (如 WebCam) 時，把顯示在螢幕上的即時影像擷取後存成 Bitmap 格式，再經過壓縮分割形成三張 JPEG 圖片，透過 Borland C++ NMUDP 元件來傳送與接收，每秒約需要 5Kbytes 的頻寬。其基本的設定如下：

- (1) 以 VideoSourceDlg() 設定影像來源。
- (2) 以 VideoFormatDlg() 函數設定壓縮格式；VideoCompressDlg() 設定壓縮方式；SetCompressionQ() 函數設定 JPG 壓縮品質 (壓縮率越低，但圖片品質越好)。
- (3) 以 SetCameraMode() 函數設定模式 (設定為 PreviewMode)。
- (4) 以 CameraDisplay() 函數偵測並連結視訊裝置；CameraClose() 函數停止視訊裝置。
- (5) 以 CapBmpPic() 擷取 BMP 圖檔；CapJpgPic() 擷取 JPG 圖檔 (為本程式所使用，需配合 SetCompressionQ() 函數)；CapAVI() 擷取 AVI 格式影像。

視訊傳輸之程式開發方法為利用 Video For Window 函式庫連結到視訊裝置後，計算出顯示在電腦桌面上的影像座標，並擷取該影像且依設定存成特定格式的圖檔。儲存圖檔時，先檢查預設的資料夾內是否有相同檔名的檔案，如果沒有則會建立一新圖檔，若有則覆蓋過去。其即時影像傳送與接收的流程如下：

- (1) 透過 Video For Window 函式庫擷取視訊畫面，並在 UA 顯示本端影像。
- (2) 將本端影像擷取並轉換成分割成 JPEG 格式，透過 NMUDP 元件傳送到對方。
- (3) 透過 NMUDP 元件接收影像資料。
- (4) 將收到的影像資料顯示在程式畫面上。

當使用者順利啟動語音以後，可選擇性的開啟

視訊畫面，視訊開啟時的軟體介面顯示如圖 5。



圖 5 視訊傳輸顯示介面圖

3.4 SER 的研究與應用

(1) 簡介

SER (SIP Express Router) 是由 iptel 開發的 SIP Router。SER 依循著 SIP 標準開發，可以作為 SIP Proxy、Registrar 與 Redirect Server。SER 亦可結合資料庫來執行完整的 AAA 程序。SER 除了能夠依照 RFC 3261 在 TCP 及 UDP 的環境下支援 SIP，也能夠與 ENUM 結合運用、搭配 rtpproxy 或 mediaproxy 達成穿越 NAT 的需求等，功能非常齊全，作業系統需求為 BSD、Linux、Solaris 等，目前尚未出現適用於 Windows 的版本。

(2) 系統平台與設定

本研究所採用的作業系統為 Linux，版本為 Fedora Core 4。將 SER 安裝以後，便能夠做最基本的登入動作，但對於使用者和伺服器本身並不安全，管理者在管理上也相當不方便，因此必須做一些設定來達到更完善的功能，如使用者認證等。

SER 最重要的設定檔為 ser.cfg，分成四個部分：通用設定 (Global Configuration)、模組載入 (Module Loading)、模組參數 (Module-specific Parameter)、通話路徑規則 (Routing Logic)。主要修改的部份為模組載入和模組參數；要讓 SER 能夠有使用者認證的功能，必須將 SER 資料庫模組載入，並做進一步的設定，如加入 domain name 的參數等。完成以後管理者能夠新增、修改及刪除使用者 SIP address 的內容，而使用者在 SIP address 建立以後，登入 SER 時必須透過密碼認證機制才能夠正確登入。

認證機制所採用的是 MD5 (Message-Digest algorithm 5) [6]：在登入的過程中，UA 第一次送出的 SIP Register 封包並沒有包含認證的訊息，因為有些 SIP Server 並不需要認證，若附加多餘的訊息只會浪費頻寬；如果該 Server 要求使用者驗證身份，便會回傳 401 Unauthorized 要求驗證，當 UA 收到此 SIP Response 後才會加入驗證訊息；UA 根據 401 Response 所附帶的驗證參數和 SIP address 的密碼，經過 one-way hash function 產生 128 位元的 16 進位亂數向 SER 作身份認證 [9]，UA 認證過程如圖 6 所示。



圖 6 使用者登入 SER 介面顯示圖

經過測試，除了本研究所開發的 UA，X-lite 和 CCL Phone 也能夠順利註冊認證，且能夠與本研究所開發的 UA 互相通話，說明了 SIP 的互通性。

(3) SERweb

SERweb 是一套圖形化的 SER 網頁管理頁面，提供管理者與使用者諸多服務，如註冊新使用者、管理使用者帳戶、紀錄未接來電、紀錄來電訊息、傳送即時訊息、通訊錄、伺服器監視、伺服器流量統計等，不同於文字介面下，所有動作都必須由告知並且由管理者操作，安裝 SERweb 後，使用者也能夠自行註冊及管理個人的資料。安裝 SERweb 前必須先架設好 Web Server，並且能夠支援 mysql 4.1 以上的版本和 PHP，其他更詳細的需求可參閱安裝說明。安裝後的 SERweb 網頁，分為管理者和使用者兩種形式，分別如圖 7 和圖 8 所示。

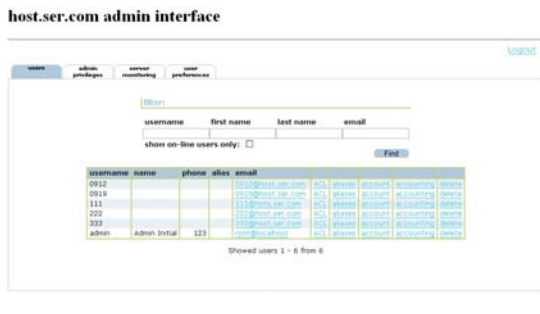


圖 7 管理者操作頁面



圖 8 使用者操作頁面

3.5 即時訊息傳輸

即時訊息以 SIP/SIMPLE 為基礎，使用的是 SIP extension 的 MESSAGE header，透過 UDP 傳送[4]。

當 UA 成功建立語音後，使用者可以選擇性開啟即時訊息的介面。與語音及視訊不同的是前兩者為 Peer to Peer 傳送，也就是在建立 session 以後，所有資料不經過 SER 轉傳。即時訊息必須過 SER 傳送；傳送後，如果接收端收到訊息，則立即回傳 200 OK 給傳送端，讓彼此知道訊息傳送成功；若傳送端沒有收到 200 OK，則會提示使用者訊息沒有傳送成功。因此即時訊息雖然是透過 UDP 傳送但仍維持其可靠性 (Reliability)。其 UA 即時訊息傳送的介面如圖 9 所示。



圖 9 即時訊息傳送介面圖

3.6 SIP Event Notification

此部分主要運用於 UA 通訊錄，透過 SIP 的 SUBSCRIBE 與 NOTIFY，可以讓使用者知道目前通訊錄裡的聯絡人是否在線上或離線之資訊。送出 SUBSCRIBE 端為 Subscriber，回覆 NOTIFY 則為 Notifier；透過固定時間回傳 NOTIFY，則 Subscriber 可以知道對方是否還在使用 SIP 服務[11]。其 UA 通訊錄介面顯示如圖 10 所示。



圖 10 通訊錄介面圖

3.7 穿越 NAT 的方法研究

目前網路環境仍以 IPv4 為主，由於 IP 數量有限，所以大多採用 NAT (網路位址轉換技術)，以便讓私有網路下之電腦能夠存取公眾網路的資源。但是在 NAT 下的 SIP UA 在私有 IP 和公眾 IP 轉換時會造成 SIP 封包無法順利的傳送，使得語

音，視訊等多媒體應用無法正常運作[1]。

SER 提供了穿越 NAT 的模組，另外搭配的軟體為 rtpproxy 或 mediaproxy，另外必須載入 nathelper 或 mediaproxy 模組；本研究採用 SER + rtpproxy 作為測試環境。主要的概念為 SER 的 nathelper 藉由改寫 SDP 的內容[7]，並在 SIP Proxy Server 上開啟 socket，將 RTP 的封包導引至 SIP Proxy 上，再轉至 NAT 下的 UA[3]。

必須注意的是此方法僅支援 Symmetric RTP 的 UA，即傳送與接收語音接使用同一個連接埠；另外此方法會造成 SIP Proxy Server 額外的流量負擔，因此在使用者過多的環境下，若 SIP Proxy Server 無法負擔，不建議使用此方法。

4. 結論

本論文以 SIP 為標準通訊協定進行用戶代理者 UA 的研發，研究內容為語音傳輸的研發、User Agent 的介面設計、視訊傳輸方法研究、即時訊息傳輸、SER 管理操作研究、User Agent 與 Server 端的整合、穿越 NAT 的方法研究等。本研究除了完成以 SIP 為基礎之 VoIP 電話系統用戶代理者與伺服器之研究，遵循標準通訊協定加強軟體實用性，並完成與 Xten 的 X-lite 和工研院的 CCL Phone 之相容性測試。但仍有許多地方值得更進一步探討和加強的地方：語音部份可以改為 G.729 甚至是 G.723 來壓縮，以節省更多頻寬；在視訊傳輸方面，應利用更適合的壓縮方式以及串流傳輸等方法，使其效果更好；SER 仍有許多可以延伸的功能，例如與 ENUM 結合、Accounting 功能等；針對不同的 NAT 分別討論並提出更詳細的穿越方法。另外，以 SIP 為標準所開發的 UA，除了在有線網路環境，在無線網路環境下所達到之效能、可能遇到的困難與解決方法，也是未來值得深究的議題。

參考文獻

- [1] 吳順良、黃博俊，“以 SIP 為信令系統的多媒體傳輸服務品質之研究，” 2005 主動式網路研討會。
- [2] Archan Misra, Subir Das, Anthony J MacAuley, “Hierarchical Mobility Management for VoIP service,” IEEE, vol. 1, pp. 372- 377, October 2001.
- [3] <http://www.csie.nctu.edu.tw/~yhsung/voip/ser.html>
- [4] <http://www.ietf.org/html.charters/simple-charter.html>
- [5] M. Handley ACIRI, H. Schulzrinne, Schooler Cal Tech, J. Rosenberg, “SIP: Session Initiation Protocol,” Request for Comments: 2543, Category: Standards Track, March 1999.
- [6] [RFC 1321] R. Rivest, “The MD5 Message-Digest Algorithm,” IETF. April 1992
- [7] [RFC 2327] M. Handley, V. Jacobson, “ SDP: Session Description Protocol,” IETF. April 1998
- [8] [RFC 2543] M. Handley, H. Schulzrinne, E. Schooler, J. Rosenberg, “ SIP: Session Initiation Protocol,” IETF. March 1999
- [9] [RFC 2617] J. Franks, P. Hallam-Baker, J. Hostetler, S. Lawrence, P. Leach, A. Luotonen, L. Stewart, “HTTP Authentication: Basic and Digest Access Authentication,” IETF. June 1999
- [10] [RFC 3261] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, “ SIP: Session Initiation Protocol,” IETF. June 2002
- [11] [RFC 3265] A. B. Roach, “Session Initiation Protocol (SIP)-Specific Event Notification,” IETF. June 2002
- [12] Sheng-Cheng Yeh, Wen-Chang Lee and Cyung-Yu Sung, “Plan and Application of Wireless Campus Networking System,” TANET 2002, pp. 870-877, Taiwan, October 2002.