2009 IEEE International Conference on Networking, Architecture, and Storage

A Teletraffic Perspective on Relay-Node Selection Strategy in VoP2P System

Wen-Kang Jia
Department of Computer Science
National Chiao-Tung University
Hsinchu, Taiwan
wkchia@cs.nctu.edu.tw

Yaw-Chung Chen
Department of Computer Science
National Chiao-Tung University
Hsinchu, Taiwan
yccehn@cs.nctu.edu.tw

Abstract—**This paper gives a theoretical approach to modeling and simulating the Grade of Service (GoS) of Voice over Peer-to-Peer Session Initiation Protocol system which is created based on various teletraffic parameters. VoP2P features a particular service characteristic due to its different GoS models originated from circuit switched telephone networks. Especially, most voice session in VoP2P is indirectly connected because peer-nodes usually locate behind NATs. This situation forces peer-node to select one or more relay-node for establishing sessions. However, the relay-node is a constrained resource because every relay-node can only support very limited relay services in VoP2P network. There are two strategies, the S^{2}RS (Staged Single-Relay Strategy) and MRS (Multi-Relay Strategy), to solve the relay-node issue. We adopt three new parameters in VoP2P call-loss models, the first is callee-attendance rate, which is due to the characteristic of destination peer-nodes with a lower attendance probability; the second is call-blocking rate, which is a traditional definition just because the relay-node reaches the limitation of resource; the third is handoff-dropping rate (aka handoff-loss rate), which is caused by the relay-node being removed frequently. The simulation results provide useful reference in the study of teletraffic of VoP2P system.**

Keyword: VoP2P, GoS, Teletraffic, Relay-node, Blocking Rate, relay-node Handoff, S^{2}RS, MRS

I. INTRODUCTION

Voice over Peer-to-Peer Session Initiation Protocol system, abbreviated as VoP2P, is different from both traditional VoIP and classical PSTN systems in service provisioning. It may encounter degraded quality of service (QoS) due to low link reliability caused by packet loss, delay and jitter, as well as suffer degradation of grade of service (GoS) due to high service demand with insufficient resources.

In PSTN systems, both QoS and GoS are estimate of customer satisfaction with particular aspects of concerns such as noise, echo or blocking. In telecom networks, the principles of QoS and GoS are independent, GoS defines the percentage of call blocking or poor service (long delay), with one user being served by one circuit. A user is part of a source user group that is handled by one or more lines of service trunk. If a user has optional access to more than one line, the system has to allocate more lines for the user in order to achieve an optimal GoS. The goal of PSTN planning is to deploy a system such that GoS can be maximized, while the number of lines serving a certain number of users can be minimized. In fact, some studies to explore this type of GoS problem focus on the effect of coverage overlapping between lines and effect of decision algorithm for user allocation. When GoS is used without further explanation, it generally refers to blocking probability; and the studies to QoS problem concern the effect of electronic, radio and optical signal attenuation, interference and distortion etc. [21].

A classical GoS analysis deals with the probability of successful connection without call-blocking. This is based on Erlang-B model [18, 19], Extended Erlang-B model and Engset model [20, 21]. These models are special cases of the birth and death problem in statistical traffic theory, which is regarding the relation of service trunk, traffic load, blocking rate and number of nodes. However, these three models can only be justified in the environment where all users have access to all resources of the system (a situation of full availability).

The calculation of teletraffic attribute of VoP2P system is quite different from the traditional GoS. A VoP2P system lacks the concept of service trunk (circuit), and faces Call Admission Control (CAC) limitation because it is impossible to provide unlimited resource in peer-nodes. Actually, the number of total service trunks is equivalent to "concurrent-call based" CAC in VoIP system. Usually a user may be allowed to have unlimited access to system resources (making calls), however, he/she may experience a poor voice quality as a result. This concerns the scarcity of resource, so we cannot apply the classical approach to model VoP2P system. Especially, most voice traffics go through indirect connection because peer-nodes usually reside behind NATs. As a result, each peer-node may play the role of service trunk and serve as a media relay for other peer-nodes, and a peer-node must select one or more relay-node to establish sessions. But not every relay-node can support relay service satisfactorily. A relay-node may be shutdown or removed frequently, so it usually offers short and discontinuous service periods. There are two strategies to solve the problem, the staged single-relay-node strategy (S^{2}RS) and multi-relay-node strategy (MRS) in VoP2P system, their common goal is to share the relay loading. Therefore, it is necessary to make physical measurement and assessment of traffic in the provisioning network of VoP2P systems.

978-0-7695-3741-2/09 $25.00 © 2009 IEEE
DOI 10.1109/NAS.2009.27

II. VOICE OVER PEER TO PEER NETWORKS

A. *Hierarchy of VoP2P systems*

Nowadays Voice over IP (VoIP) system is highly centralized and lacking of standard. The ultimate goal of VoP2P is server-less call establishment and service provisioning. It is very useful in mobile ad-hoc network (MANET) and emergency situation. Unlike VoP2P, existing SIP-based VoIP system [1] is based on client-server architecture. SIP-based VoIP can be treated as a P2P system with a static set of super-nodes (SIP servers) where the lookup is based on DNS instead of hash keys. Using a pure P2P architecture improves the reliability and allows the system to dynamically adapt to node failures.

Since VoP2P has no central server to establish calls, it relies on the global index carried by the super-nodes and the peer host for directly connecting to the other peer. When a VoP2P user calls another VoP2P user, the easiest way to make the call is directly establishing a RTP connection between two peers. However, NAT problem usually causes the call unsuccessful, and the industry is working on possible solutions, such as STUN [25] and ICE [14], for NAT traversal between two peers. If it still doesn't work satisfactorily, the global index provides a small number of nodes called relay-nodes, which actually do route voice packets between two peers. Both caller and callee contact the relay-node that handles the call establishment between the two peers.

More recent work tried to combine SIP and P2P [6]. A set of extensions to SIP is proposed in SoSIMPLE (Self Organization SIP for Instant Messaging and Presence Leveraging Extensions) [7]. Basically, SIP can be combined with P2P in two ways: to replace the SIP user registration and do lookup by an existing P2P protocol, or to implement this P2P algorithm using SIP messaging. The former uses an existing P2P protocol [2, 3], whereas we focus on the latter approach that builds the P2P network among the peer-nodes using standard SIP messages without modifying message semantics [3, 4]. Its advantages include 1) use of existing SIP components such as voice gateway and IVR services, 2) independent of the existence of external VoP2P networks, and 3) built-in media relays for firewalls and NAT traversal [14], and its disadvantage is double sized transport message.

VoP2P is a fully decentralized, standard-based P2P communications system that utilizes existing clients. In this paper we discuss the challenges in developing a distributed signaling system that preserves the advantages of centralized systems. Our approach opens up new opportunities for decentralized communications systems that are readily available and extensible.

P2P systems inherently feature high scalability, robustness and fault tolerance because they do not rely on a centralized server, and the systems self-organize themselves. This is achieved with the cost of higher latency for locating the resources of interest in the P2P overlay network. Internet telephony can be viewed as an application of P2P system where the participants form a VoP2P overlay network to locate and communicate with each other.

B. *Relay-node in VoP2P systems*

The hierarchy of VoP2P is independent of SIP UA-Proxy hierarchy. Hierarchy of P2P overlay consists of both relay-nodes and peer-nodes. The former are also named intermediate-nodes in different applications, and their major purposes include the following:

- *NAT traversal:* Generally, a relay-node is reachable through predefined ports and protocols by any peer-node, and most likely owns a public IP address. If some peer-nodes behind symmetric NATs and/or port restricted cone NATs [14] could not make direct connection to each other, peer-node will relay their RTP stream [11] partially or totally. All calls through relay-nodes are same as that directly between peer-nodes. Today most mobile Internet access providers offer network services by WAP (Wireless Application Protocol), GPRS (General Packet Radio Service), EDGE (Enhanced Data Rates for GSM Evolution) and HSDPA (High Speed Downlink Packet Access) technologies; the mobile terminals are usually assigned a private IP addresses via dynamical allocation (behind NAT). Therefore unless working in an unordinary situation, developers of VoP2P application and middle-ware should always assume that their software will run on devices that are behind NAT. This causes VoP2P clients to highly rely on relay service for communication.
- *Audio/ Video Mixer:* Beside NAT traversal purposes, relay-nodes also play the role of media mixer. In a conference call scenario, a call from peer-node A was established between itself and B, and B decided to invite C to the conference, so B and C were sending their voice traffic in UDP to relay-node A. Assume that A is the most powerful one and acts as a media mixer, it mixed up its own packets with those of B, then sent them to C and vice versa. Even if user B or C started the conference, A is still the most powerful amongst the three, so it always gets elected by the particular algorithm. The above procedure also can be accomplished by full-mesh connection method, but each peer must maintain n-1 connections during the conferencing period (n is the number of participants). For saving the resources, the mixer method can reduce both bandwidth usage and CPU resource in each peer-node.
- *Audio/ Video Transcoder:* If the caller and callee can't come out an agreed codec type, especially in multi-party conference, the relay-node must be acting as a codec converter too.
- *Interceptor:* For security reason, the relay-node program includes an interceptor function if its voice session is monitored. The relay-node may replicate a copy during its relaying process. Of course, for encrypted form between peers, relay nodes cannot do eavesdropping.
- *QoS Proxy:* Occasionally, QoS through direct connection between peer A and peer B may be poorer than that through indirect connection via peer C. Ren et al. [26] addressed that about 1%~10% of connections with one-hop relay path experience better RTT quality than those connections using direct routing mode in voice sessions.
- *Mobile Proxy:* In mobile IPv4 architecture, when a Mobile Node (MN) moved to a foreign network (FN), the triangle routing method is necessary. The Home Agent (HA) must forward packets to MN's Care-of-Address (CoA) consistently, so HA actually plays the role as a mobile proxy.

Many VoP2P systems utilize a subset of peers with better capabilities such as sufficient network bandwidth, high computing power, low processing load, and large storage space, to enhance the quality and/or the functionalities of the service provisioning. Conventionally, high bandwidth Internet connection and transparent TCP/UDP protocol utility is necessary. These special peers are often referred to as relay-nodes. If the relay-node is far from the communicating peers, it may impair the voice quality. Therefore, it is important to design a method for searching a suitable relay-node that is close to the end-peers, so that unnecessary relaying delay can be reduced. Consequently, in this study, we adopt two strategies to reduce network traffic as well as to improve QoS through relay-node selection algorithm and evaluate the performance.

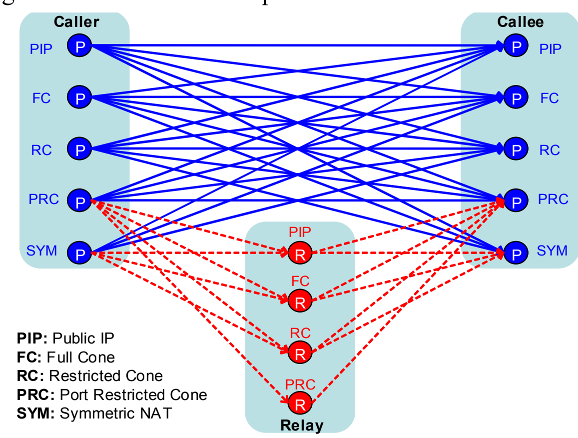

Figure 1. The connect relation between call peers behind various NATs types.

The major reason to use relay-node is NAT traversal. Interactive Connectivity Establishment (ICE) is a novel method that tries to build NAT traversal intelligence into peer-nodes so that they can perform route discovery, relay lookup, path optimization, and even verify media flow before a call is deemed to be established. In P2PSIP, prior to sending an INVITE, the caller executes a sequence of steps to characterize the type of NAT with which it is associated. First, a caller obtains addresses of all available interfaces; then it checks the results of reachable peers from STUN server; sometimes a caller peer can't find a direct media path to another peer, then it needs to negotiate a usable port with the relay-node(s).

Afterwards, the caller attaches a list of available IP addresses/ports in the INVITE message with SDP to the super-node. As soon as the callee gets an INVITE message, it follows a similar diagnostics set of steps as did the caller. Next, it attempts to send STUN queries to the caller to check whether it is possible to directly send media to any IP addresses/ports presented in the INVITE message. Finally, the callee picks from INVITE message the address of highest preference to which it can confidently send media.

In the VoP2P system, ICE solution is similar to most known NAT types that support UDP. In other words, ICE supports RTP transport. For ICE to work properly, both caller and callee peers must support ICE client, and if a relay-node is behind NATs, the relay-node must support ICE client too. In P2P environment, any server facility should not be fixed, and there should be one more super-nodes acting as the role of STUN server. For this reason, it should be deployed only in a homogeneous and controlled environment. Furthermore, it is likely that the call setup will be delayed because of the process involved in media path discovery by the caller and callee.

TABLE I. TABLE TYPE STYLES

Symbol	***Description***	***Unit***
λ_1	Input rate for registered peer-node joining to system with Poisson distributed random arrival	1/s
λ_2	Input rate for active peer-node requesting a call with Poisson distributed random arrival	1/s
N	Number of registered peer-nodes in the system	#
ρ	Average probability of peer-nodes online	%
ρN	Number of concurrent peer-nodes in the system	#
σ	Percentage of relay-node in all peer-nodes	%
τ_1	Percentage of peer-node behind symmetric NATs in all peer-nodes	%
τ_2	Percentage of peer-node behind port restricted cone NATs in all peer-nodes	%
$P(dc)$	Percentage of direct connection between callers and a callees	%
$P(rc)$	Percentage of relayed connections between callers and callees	%
C	Codec bit rate (e.g. G.729/30ms=18.6k)	bps
R	Average offered bandwidth limitation per relay-node	bps
a_{sc}	Busy hour traffic with single call per node	Erlang
$a(R)$	Busy hour traffic with bandwidth limitation per relay-node as CAC	Erlang
$N_{handoff}$	Number of handoff	#
$W_{handoff}$	Handoff latency	ms
T_{AST}	Average survival time per peer-node	s
T_{AHT}	Average hold time per call	s
$P(b)$	Call-blocking rate	%
$P(d)$	Handoff-dropping rate	%
$P(p)$	Callee-attendance rate	%

III. VoP2P TRAFFIC MODEL

Actually, the GoS definition of VoP2P system is quite similar to cellular systems'. However, the original GoS function cannot show the impact of callee-attendance probability *P(p)* in VoP2P system. Traditional GoS of PSTN system is out of consideration for the effect of user absence. In POTS (Plain Old Telephone System) system, Customer Premises Equipment (CPE) is always assumed in presence by Central Office (CO). In PLMN (Public Land Mobile Network) system, mobile terminal maybe absent for some reasons, but is still out of consideration in GoS. In P2P system, whether a call attempt is successful or not depends on the scale of P2P population N (number of registered peer-nodes in the system) and its individual attendance rate ρ, which is defined as the average online probability of individual peer-node, or from an alterative view, the mean average survival time, so ρN is the average number of active peer-nodes in the system. Briefly, if *P(p)* is not considered in VoP2P system, the GoS value will be abnormal because the user behavior is quite unordinary in a VoP2P system. We derive a formula of callee-attendance probability based on binomial distribution:

$$P(p) = \sum_{x=2}^{N} \binom{N}{x} \frac{(x-1)\rho^{x}(1-\rho)^{N-x}}{N-1} \qquad (1)$$

According to the queuing theory, a user arrives at the system with an inter-arrival model represented by Poisson distribution, and a departure model represented by exponential distribution. There are two parameters for peer-nodes: the Average Survival Time (AST) T_{AST} and the Average Hold Time (AHT) T_{AHT}. The former is the duration for a peer-node from its joining till leaving a VoP2P system. During an AST, relay-nodes can perform relaying function and serve other peer-nodes, the distribution of AST ranges from few minutes to few days. The latter is the average duration a peer-node takes to hold a call. Each AHT is represented by exponential distribution. Figure 2 shows that peer-nodes joining and leaving a VoP2P system. Let VoP2P population N be the number of registered peer-nodes in the system, ρ is the probability of peer-nodes online, and ρN is the number of concurrent peer-nodes in the whole system, these parameters will be saved in the finger table and distributed to part of peer-nodes.

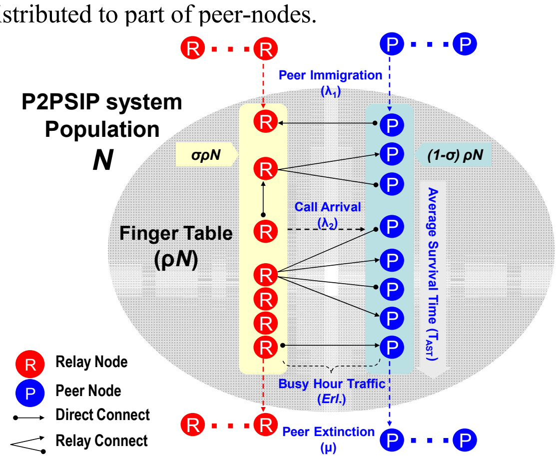

Figure 2. VoP2P peer-nodes joining and leaving the system

The challenges of VoP2P system include low attendance rate, high availability demands and high turnover rate. A peer-node may be shutdown or removed frequently, this causes relay-node to experience a short and discontinuous service period. Suppose each peer-node can only be called once or it accepts only one call, let $\lambda 2$ be the input rate with Poisson distributed random arrival, and T_{AHT} be the average holding time per call, then the maximum busy-hour traffic in VoP2P system will be:

$$BHT_{SINGLECALL} = \frac{\lambda_2 T_{AHT}}{2} \qquad (2)$$

Here the unit is in Erlang. There are two types of nodes, peer-node and relay-node, which cause 4 possibilities of call directions: relay-node to relay-node, peer-node to relay-node, relay-node to peer-node and peer-node to peer-node, their probabilities depends on the ratio σ. Among first three call directions, an RTP stream is directly connected without relay-node. However, an RTP stream from peer-node to peer-node must go through a relay-node, therefore the probabilities of direct connection *P(d)* and probability of relayed connection *P(r)* are:

$$P(dc) = P(R \to R) + P(P \to R) + P(R \to P) = 1 - \tau_1^2 - 2\tau_1\tau_2 \quad (3a)$$

$$P(rc) = P(P \to P) = \tau_1^2 + 2\tau_1\tau_2 \qquad (3b)$$

Suppose TAST is the average survival time, and ρN denotes the number of active peer-nodes over all registered peer-nodes in a VoP2P system, λ is the call arrival rate, σ is the coefficient of relay capacity (percentage of relay-node in all peer-nodes), and σN is the number of peer-nodes which can act as relay-node. Due to NAT problem, a peer-node behind the NAT can not act as a relay-node, which is required to relay traffic. From Little's formula, we have:

$$N(Active\_relay\_node) = \sigma\rho N = \sigma\lambda_1 T_{AST} \quad (4)$$

In single-call rule, let TAHT be the average holding time for calls, $\lambda 2$ be the input rate for an active peer-node to initiate a call with Poisson distributed random arrival. Then the traffic intensity in the VoP2P system will be:

$$C\lambda_2 T_{AHT}\left(1 + \tau_1^2 + 2\tau_1\tau\right) \qquad (5)$$

Assuming that T_{AST} is the average survival time of peer-nodes, $\lambda 1$ is the input rate for registered peer-nodes joining the system with Poisson distributed random arrival, then the maximum offered traffic is limited by total relay-node capacity. The capacity (in number of sessions) per single relay-node will be:

$$\sigma\left(R\lambda_1 T_{AST} - C\lambda_2 T_{AHT}\right)\Big/2C \qquad if \quad \frac{\lambda_1}{2} \geq \lambda_2 \qquad (6)$$

Since a call from the circuit-switched side may never reach a callee due to the blocking at PSTN trunk interfaces, it is important to estimate the grade of service (GoS) of calls and offered load at the PSTN network. Traditional Internet services are best effort basis, but VoIP service requires guaranteed service, so call admission control (CAC) is essential, because CAC determines how to allocate resource for new calls as well as to handover calls in order to fulfil the quality of service (QoS). But CAC may also affect the GoS because a call may be blocked due to insufficient resource.

If VoP2P system doesn't set restriction that only one call is allowed for each peer-node, multiple calls may be established in a peer-node which can act as an aggregation gateway, or it can form a sub P2P overlay. In general, a VoP2P system mostly has resource based CAC mechanism, which limits the maximum bandwidth loading for each relay-node. Once the measured traffic reaches the maximum bandwidth R of a relay-node, this relay-node will reject further call attempt. In this case, the maximum busy hour traffic (concurrent calls) of VoP2P system must meet the following condition:

$$\sigma\left(R\lambda_1 T_{AST} - C\lambda_2 T_{AHT}\right) \geq 2C\lambda_2 T_{AHT}\left(\tau_1^2 + 2\tau_1\tau_2\right) \quad (7)$$

Thus we revise the Erlang-B call-loss model [19, 20], in which traffic load a are the upper limit of total offered bandwidth for the relay-node, we consequently obtain the expected blocking rate *P(b)* as in:

$$P(b)_{CAC} = f(CAC, a = \lambda_2 T_{AHT}) = \frac{a^{CAC}}{CAC!}\Big/\sum_{k=0}^{CAC}\frac{a^k}{k!}$$

$$= \left(\left(\sum_{k=0}^{\left(\sigma\left(R\lambda_1 T_{AST} - C\lambda_2 T_{AHT}\right)/2C\right)} \frac{\left(\lambda_2 T_{AHT}\right)^k}{k!}\right) \times \left(\sigma\left(R\lambda_1 T_{AST} - C\lambda_2 T_{AHT}\right)\Big/2C\right)!\right) \quad (8)$$

$$\times \frac{1}{\lambda_2 T_{AHT} \wedge \sigma\left(R\lambda_1 T_{AST} - C\lambda_2 T_{AHT}\right)}$$

IV. THE STAGED SINGLE RELAY STRATEGY (S^2RS)

There are two strategies to choose from for relay-node. First, a peer-node behind NATs must randomly select a relay-node to relay its RTP traffic before making the call. This is called Staged Single-Relay Strategy (S^2RS).

The operation of S^2RS with SIP is defined in RFC 3312 [9], and updated by RFC 4032 [10], in which it describes two methods to implement the relay-node change during a call: one is UPDATE method and the other is re-INVITE method. The latter can be used during a session to change the characteristics of the session and restart it, and the former allows a client to update parameters of a session without impact on the state of the dialog. In that sense, it looks like a re-INVITE, but unlike re-INVITE, it can be sent before the initial INVITE has been completed. This makes it very useful for updating session parameters during relay-node handoff.

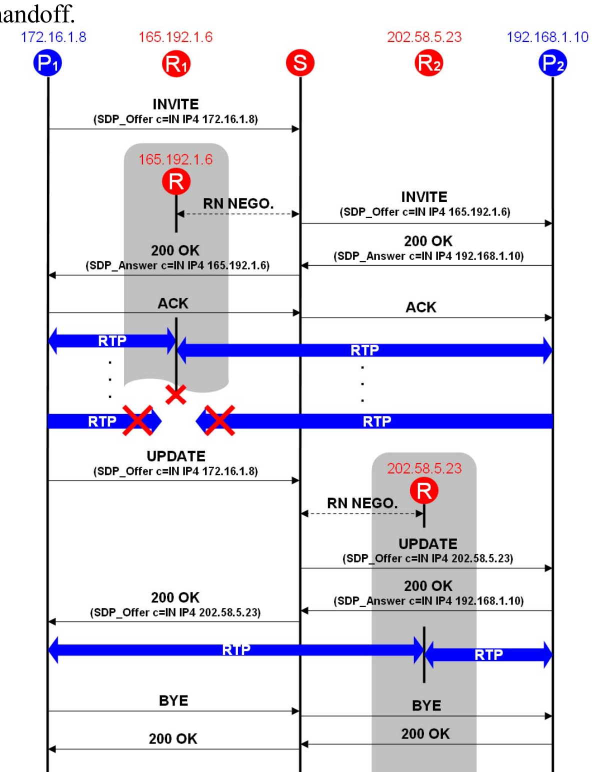

Figure 3. Relay-node handoff through SIP re-INVITE in S^2RS

We present a call procedure in Figure 3. In call attempt phase, where a super-node S acts as a proxy server in traditional SIP, which will randomly assign R1(165.192.1.6) to relay the RTP traffic between peer-nodes P1(172.16.1.8) and P2(192.168.1.10). S informs P1 and P2 through SDP [5] to establish a session between them, and the conversation can start. Unfortunately, R1 leaves the VoP2P system and results in discontinued conversation between P1 and P2 until one of the peer-nodes detects this situation. As a consequence, P1 sends an UPDATE message to inform super node S, and S reassigns R2 (202.58.5.23) as the relay-node for P1 and P2, then it informs both P1 and P2 by UPDATE message with SDP. Finally, P1 and P2 establish an RTP session through the new relay-node R2 and recover the conversation. This is an instance for relay-node-handoff during a call session.

We denote T_{AST} as the average survival time of relay-node, and T_{AHT} as the average holding time. During the conversation, the relay-node might leave the system, and the peer-node must perform relay-node-handoff procedure. The peer-node then randomly selects a new standby relay-node within the shortest time and re-establishes or recovers the connection. Under the S^2RS, each connection encountering a relay-node leaving must perform the relay-node-handoff procedure. When the first handoff occurred, the remaining holding time is decreased, and it affects the probability of the second handoff. During the whole conversation period, peer-node may perform $N_{HANDOFF}$ times of relay-node-handoff procedure under S^2RS; and the more handoff occurs, the much voice quality will be degraded. Suppose each relay-node-handoff blanks out a portion of the conversation time, and $W_{HANDOFF}$ denotes handoff latency, then the probability of relay-node-handoff $P(d)$ during a call will be:

S^2RS is a simple strategy for deploy VoP2P services. For instance, Skype is based on S^2RS with selection algorithm based on Round-Trip-Time (RTT) measurements. Little work has been done on peer node selections to relay voice packets. This is a critical issue for the quality, scalability and cost in a VoP2P system.

V. THE MULTI-RELAY STRATEGY (MRS)

The second strategy is Multi-Relay Strategy (MRS). When a peer-node behind an NAT initiates a call, it needs to randomly choose several relay-nodes from the list to perform the relaying function, so that the call can be established. The RTP traffic is probably shared equally between multiple relay-nodes based on the Round-Robin (RR), Least Recently Used First (LRUF), Shortest Path First (SPF) or Minimal Delay First (MDF) policies, or other High-Level Load Balance algorithm and Equal-Cost Multi-Path algorithm. The details of these algorithms are beyond our discussion in this article, and we only address the difference in GoS between S^2RS and MRS.

RFC 3388 defines two SDP attributes, "group" and "mid", which allow SIP to group together media streams (m) and connections (c) information in SDP massage for two purposes: lip synchronization and media receiving from a single flow that are encoded in various formats in a particular session, on different ports and host interfaces [16].

The most popular multi-relay strategy (MRS) is implemented by using MRTP (Multi-flow Real-time Transport Protocol) [12], Figure 4 shows MRTP operations in MRS. Multi-path transport has high potential in ad-hoc networks in which link bandwidth may fluctuate and paths are unreliable. Using multiple paths can provide higher aggregate bandwidth, better error resilience, and load balancing for a multimedia session like audio or video streaming based on multiple servers or peer-nodes in wired networks. MRTP is modified and extended from the RTP [11]. Multi-flow Real-time Transport Control Protocol (MRTCP) provides essential support for MRTP, including session and flow management, data partitioning, traffic dispersion, time-stamping, sequence numbering, and QoS feedback. In VoP2P system, a relay-node linked path may be created and maintained by P2P protocol, any report from a remote peer-node may show whether the QoS of relay-node, such as delay, loss, and jitter is good or not. A peer-node can learn newer relay-node, remove or remark the un-qualified relay-nodes dynamically, and look for the highest QoS.

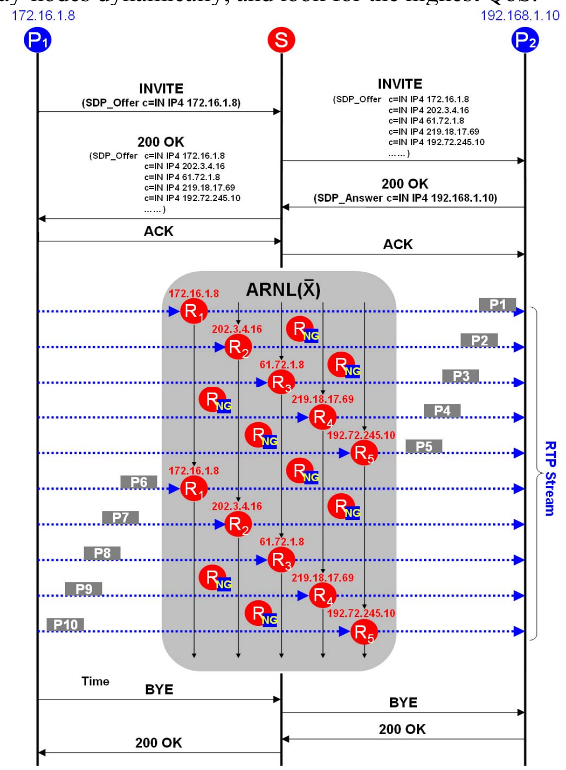

Figure 4. MRTP round-robin based packet striping in MRS

In MRS, no matter which load balancing algorithm is used, QoS feedback should not be included. In other words, whenever a relay-node leaves, it should not be eliminated from the relay-node list. If we use N relay-nodes at the same time, the probability of handoff-dropping rate *P(d)* will be:

$$P(d)=\prod_{1}^{N}\left(\int_{0}^{(T_{AHT}-N*W_{Handoff})}\frac{2}{N*T_{AST}}e^{-1/N*T_{AST}}dt\right) \quad (10)$$

This is an extremely low figure, *P(d)* will approach to 0 when N is increased to a certain value. Of course, the voice quality is degraded before the handoff-dropping rate approach to 0. We show the packet-loss rate as in Equation (11), however, we don't discuss much of this subject in this article.

$$\sum_{k=1}^{\infty}\left(\left(\int_{0}^{T_{AHT}}\frac{2}{T_{AST}}e^{-1/T_{AST}}dt\right)^{k}*k*W_{Handoff}\Big/T_{AST}\times T_{AHT}\right) \quad (11)$$

VI. Simulation and Results

The grade of service in VoP2P system is evaluated through simulation. We use the simulation tool --MATLAB to model several scenarios and emulate their behaviours. The results are presented in Figures 5, 6, 7 and 8.

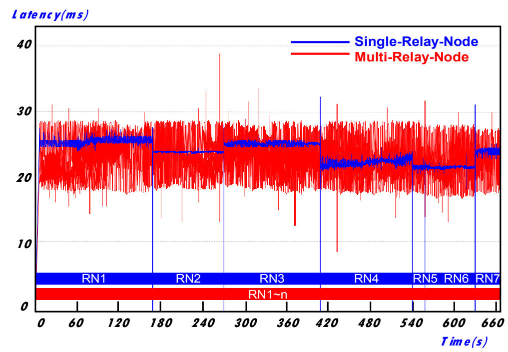

Figure 5. The latency and jitter in S^2RS and MRS

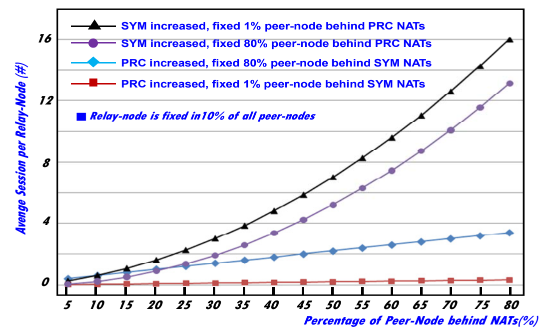

Figure 6. Growth coefficient of traffic intensity with Single Relay-node under Single-Call in S^2RS

Figure 5 shows the simulation results of packet delay during 660 seconds of the call session. We obviously observed that the MRS' latency variation is smaller than that in S^2RS. However there is a short voice conversation recess after a certain period of time when the relay-node leaves the system. The recess time depends on the RTP report frequency and the time complexity of handoff algorithm. Sometimes it may reach as long as several seconds. A new relay-node may experience a performance gap in the latency or jitter comparing with previous relay-node. The jitter (variation of latency) in MRS is bigger but much stable, the out-of-sequence phenomenon may occur frequently. During a conversation period, a relay-node may affect the voice quality if it leaves the system. Nevertheless it is unlikely to cause the short conversation recess, which often happens in S^2RS during a call.

Figure 6 shows the simulation results of traffic intensity of each relay-node using single-call model in S^2RS with full load (concurrent number of calls = number of peer-nodes/2). The percentage of relay-node in all peer-nodes is fixed at 10%. The X-axis is the percentage of peer-nodes which are behind symmetric or port restricted cone NATs, and the Y-axis is the number of sessions undertaken for each relay-node in average (one call consists of two sessions). We can observe that only 10% of relay-nodes in the system are under full loading, and each relay-node must support 16 sessions (8 calls) in average. If we take G.729 that uses 24Kbps per session as an example, it means that each relay-node undertakes the relay traffic with an average of 384Kbps approximately.

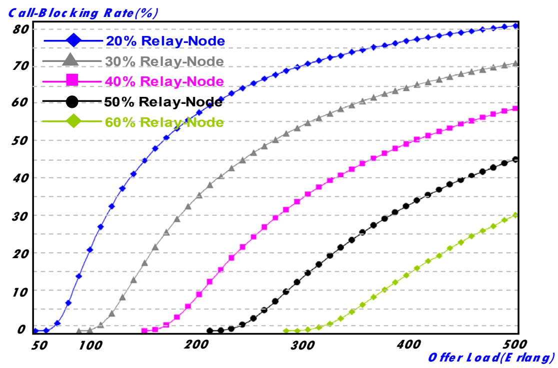

Figure 7. Call-blocking rate in various rate-based CAC under multi-call in S^2RS

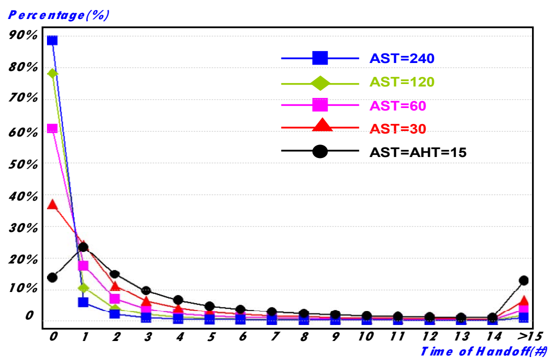

Figure 8. The relay-node handoff probability under various average survival time in S^2RS

Figure 7 shows the simulation results of call-blocking rate with 50~350 Erlangs traffic load under Multi-Call in S^2RS. Here we have 400 active peer-nodes including relay-nodes in VoP2P system, and the percentage of relay-nodes grows from 20% to 60%. We assign relay-nodes with 64K rate-limited call admission control, and observe the variation of call-blocking rate. We obviously observe that it experiences 60% of call-blocking with 20% peer-nodes in 200 Erlangs traffic intensity under 64K CAC, and the blocking rate is near 0 with 60% peer-nodes in 200 Erlangs under the same condition.

Figure 8 shows the simulation results of handoff frequency for relay-node under conditions of AST=30, 60, 120, 240 minutes, and each call has AHT=15 minutes. The X-axis represents the handoff frequency and the zero means no handoff occurred during a call. We can see that relay-node-handoff will occur at least once with 63.21% probabilities under AST=60 and AHT=15. By increasing AST, the probability of relay-node-handoff will be decreasing.

Sometimes users may use a particular system like Skype^{TM}. For circumvent the responsibility of playing relay service role for anybody else, some users may make a call immediately after join VoP2P networks, and exit from the system after the call. Now we have a situation of AST≈AHT, and we can see the service quality which will be critically aggravated by quickly increased relay-node-handoff probability.

VII. CONCLUSIONS

Given the complexity of today's NAT and firewall-protected networks in residential, campus and enterprise environments, VoP2P system can only be successfully provided with certain help from NAT traversal mechanism, which must therefore be able to solve peers' NAT and firewall problems remotely in VoP2P system. Many solutions such as STUN, TURN, and ICE have been proposed, but each can reasonably be deployed only in certain situations, in other words, those are only partial solutions. Practically, for VoP2P service, it is necessary to locate a relay-node for relaying voice traffic between peers which are behind NATs. Moreover, the relay capacity also decides how much traffic intensity can be offered in the VoP2P system.

There are two relay-node selection strategies to accomplish the operation of VoP2P under limited capacity of peer-nodes, the staged single-relay strategy and multi-relay strategy. We have presented a novel approach to investigate the grade of service for a VoP2P call attempts under certain system scale, relay-node ratio, traffic load and their relations under single-call session on S^2RS. Mostly, we verify the traffic load of relay-nodes. When the traditional signaling system is changed to VoP2P, the relay link can still handle the task of relaying media session for RTP stream. Under single-call session in S^2RS, we find that if the number of relay-nodes or percentage of relay-nodes is decreased, the loading per relay-node will grow exponentially.

We think that the GoS should include two implications, one is user call-blocking for call attempts, and the other is user call-dropping during a call session. The former is caused by callee's absence and/or insufficient resource, while the latter is due to the departure of relay-node and/or inappropriate relay-node selection. We discussed two particular call-loss models in VoP2P system, the first one is regarding call-blocking rate, which is a traditional definition and it happens because the relay-node reaches its limit of service resource. The other is handoff-loss rate, sometimes relay-nodes may disappear frequently under either S^2RS or MRS. A call session may encounter leave or removal of its relay-node in VoP2P system, in that case the call session will be blocked suddenly, and call peers must perform relay-node handoff procedure. In S^2RS, the probability of handoff depends on the percentage of average call holding time of a call and average survival time of relay-nodes. We addressed the handoff-dropping ratio during a call session. The handoff-dropping ratio of MRS will be comparatively less than that in S^2RS. However MRS will pay for QoS with high jitter rate and even high out-of-sequence rate.

Internet Connectivity Establishment (ICE) is becoming increasingly important for P2P systems on the open Internet, as it enables NAT-bound peers to provide accessible services. ICE provides best-effort direct connection between peers, and it can help peers discover qualified candidates for relay-nodes; that is, the STUN enabled super-node and relay-node which provides hole-punching and relaying services, respectively. An ICE (STUN) service deployed in super-peers is suggested (i.e. Skype). A super-node may instruct peer-nodes to find one or more relay-peers and randomly select one with acceptable forward latency. An efficient relay selection algorithm is helpful to reduce latencies of the call setup, voice transport and relay-handoff process in VoP2P system. [27]

Since frequent relay-node-handoff causes the handoff-dropping ratio to increase, we suggest avoiding relay node whenever possible because of the unpredictable behavior of the single relay node under S^2RS strategy. In other words, MRS is better than S^2RS. If peers must select S^2RS for some reasons, those peers should find more relay nodes to deal with the relay-node failure. This can minimize the relay-node handoff delay, and a good relay-node selection algorithm will improve smoothness and seamlessness of the handoff.

Although MRTP has been developed in the context of mobile ad hoc networks, we believe that MRTP can be applied on the Internet and P2P system when an institutional network has multiple access routers; on infrastructure wireless networks when multiple base stations can be accessed in parallel; and on multimedia data sharing under P2P overlay networks for improving QoS and GoS.

The QoS of MRS and modeling the conference call between peers is an important issue for future works, which would be more complicated. On the other hand, the alternative is the multi-homing solution, which has not been standardized in both network layer and transport layer protocol such as IPv6, SCTP (Stream Control Transmission Protocol) and DCCP (Datagram Congestion Control Protocol). Issues about various possible approaches to multi-homing are still unresolved.

We investigated design and implementation of a useful GoS assessment model for VoP2P system, including quantity of relay-node/ peer-node, CAC and relay-node handoff probability with various AST and AHT. Our study provides practical reference in the research of teletraffic of VoP2P system, and simulation results may be useful in teletraffic planning and maintenance for Internet service providers. In the future, we intend to design a mechanism that reduces the jitter of the MRTP flow for facilitating nowadays VoP2P systems.

ACKNOWLEDGMENT

This work was sponsored by the National Science Council of R.O.C. under Grant No. 96-2752-E-009-006-PAE.

REFERENCES

[1] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. R. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, SIP: session initiation protocol, RFC 3261, Internet Engineering Task Force, 2002.

[2] A. Johnston, SIP, P2P, and Internet Communications, draft-johnston-sipping-p2p-ipcom-02, Internet Engineering Task Force Internet Draft, 2006. (work in progress.)

[3] J. Hautakorpi and G. Camarillo, The Peer Protocol for VoP2P Networks, draft-hautakorpi-VoP2P-peer-protocol-00, Internet Engineering Task Force Internet Draft, 2007 (work in progress.)

[4] K. Singh and H. Schulzrinne, Peer-to-peer Internet Telephony using SIP, Columbia University Technical Report CUCS-044-04, New York, NY, 2004.

[5] H. Handley and V. Jacobson, SDP: session description protocol, RFC 2327, Internet Engineering Task Force, 1998.

[6] B. Campbell, J. Rosenberg, H. Schulzrinne, C. Huitema, and D. Gurle, session initiation protocol (SIP) extension for instant messaging, RFC 3428, Internet Engineering Task Force, 2002.

[7] D. A. Bryan, B. B. Lowekemp, and C. Jennings, SOSIMPLE: A Serverless, Standards-based P2P SIP Communication System, Proceedings of Interelay-nodeational Workshop on Advanced Architectures and Algorithms For Internet Delivery and Applications, Orlando, FL, 2005.

[8] S. Baset and H. Schulzrinne, Peer-to-Peer Protocol, draft-baset-VoP2P-p2pcommon-01, Internet Engineering Task Force Internet Draft, 2007 (work in progress.)

[9] W. Marshall and J. Rosenberg, Integration of Resource Management and Session Initiation Protocol (SIP), RFC 3312, Internet Engineering Task Force, 2002

[10] G. Camarillo and P. Kyzivat, Update to the Session Initiation Protocol (SIP) Preconditions Framework, RFC 4032, Internet Engineering Task Force, 2005

[11] H. Shulzrinne, S. Casner, R. Frederick, and V. Jacobson, RTP: a transport protocol for real-time applications, RFC 1889, Internet Engineering Task Force, 1996.

[12] S. Mao, D. Bushmitch, S. Narayanan and S. S. Panwar, MRTP: a multiflow real-time transport protocol for ad hoc networks, IEEE Transactions on Multimedia, Apr. 2006, pp.356–369

[13] S. Narayanan. D. Bushmitch, S . Mao, and S.S. Panwar. The multi-flow real-time transport protocol, draft-narayanan-mrtp-00, Internet Engineering Task Force Internet Draft, 2004 (work in progress.)

[14] J. Rosenberg. Interactive connectivity establishment (ICE): a methodology for network address translator (NAT) traversal for the session initiation protocol (SIP), draft-rosenberg-sipping-ice-02, Internet Engineering Task Force Internet Draft, 2003. (work in progress.)

[15] R. Stewart and X.Qiaobing, Stream Control Transmissions Protocol (SCTP): A Reference Guide, Addison-Wesley, 2001.

[16] G. Camarillo, G. Eriksson, J. Holler, and H. Schulzrinne, Grouping of Media Lines in the Session Description Protocol (SDP), RFC 3388, Internet Engineering Task Force, 2002.

[17] ETSI/RE53, A Guide to DECT feature,RES3(92)21, 1992.

[18] A. K. Erlang, The Theory of Probabilities and Telephone Conversations, Nyt Tidsskrift for Matematik B.20, 1909

[19] A. K. Erlang, Solution of some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges, Elektrotkeknikeren.13, 1917

[20] A. Myskja, An introduction to Teletraffic , Telektronikk 2/3(1995)3-40, 1995

[21] R. Parkinson, Traffic Engineering Techniques in Telecommunications, Infotel Systems Corp., 2002

[22] SKYPE TECHNOLOGIES, S.A. http://www.skype.org/.

[23] J. Rosenberg, R. Mahy and C. Huitema, Traversal Using Relay NAT (TURN), draft-rosenberg-midcom-turn-08, Internet Engineering Task Force Internet Draft, 2005. (work in progress.)

[24] P. Srisuresh, J. Kuthan, J. Rosenberg, A. Molitor, and A. Rayhan, “relay-node communication architecture and framework", Internet Engineering Task Force, RFC 3303, 2002.

[25] J. Rosenberg, J. Weinberger, C. Huitema and R. Mahy, STUN - Simple Traversal of User Datagram Protocol (UDP) Through Network Address Translators (NATs), RFC 3489, Internet Engineering Task Force, 2003

[26] S. Ren, L. Guo, and X. Zhang, ASAP: an AS-Aware Peer-relay protocol for high quality VoIP, Proceedings of the 26th International Conference on Distributed Computing Systems (ICDCS'06), Lisbon, Portugal, Jul. 2006.

[27] Jim Dowling, Jan Sacha and Seif Haridi, Improving ICE Service Selection in a P2P System using the Gradient Topology, International Conference on Self-Adaptive and Self-Organizing Systems (SASO), 2007.