

Decision Engine for SIP Based Dynamic Call Routing

Sajjad Ali Musthaq, Christophe Lohr, Annie Gravey

Department of Computer Science TELECOM Bretagne Brest, France,
Technopôle Brest-Iroise - CS 83818 - 29238 Brest Cedex 3 - France
<http://www.telecom-bretagne.eu>

Abstract. Enterprises nowadays are subscribing access to several Internet Service Providers (ISPs) for reliability, redundancy and better revenues underlying the service extension, while providing good Quality of Service (QoS). In this paper, a dynamic decision-making framework is presented for Session Initiation Protocol (SIP) based voice/video call routing in multihomed network. The decision engine takes multiple criteria into account while computing the routing decision (attributes from context of the request, platform's latest conditional parameters, business objectives of the company, etc.). Two Multi-Criteria Decision Making (MCDM) methods, namely Grey Relational Analysis (GRA) and an extended version of Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) are used for decision calculation in outsourcing and provisioning enforcement modes respectively. The proposed solution gives higher throughput and lower call dropping probability while fulfilling the desired goals, taking into account the multiple attributes for choosing the best alternative.

Key words: Decision Engine, Multi-Criteria Decision Making (MCDM), Grey Relation Analysis (GRA), Call Dropping Probability, Throughput

1 Introduction

Companies use Internet to deliver data, applications and services. Traditionally, multihoming to multiple Internet Service Providers (ISPs) has been employed to ensure performance, availability and reliability. However, over the past few years, multihoming has been increasingly leveraged for improving wide-area network performance, lowering bandwidth costs and optimizing the way in which upstream links are used. The correct link selection can optimize resource utilization by ensuring the required Quality of Service (QoS). With the advent of economical and high-bandwidth broadband connection technologies, multihoming is poised to emerge from a niche technique for large businesses and has become a dominant technology that underlies Internet connectivity solutions for small to mid-sized businesses. Although the main focus in multihoming is primarily on reliability, it is also being used for Load Balancing and latency reduction. A straightforward method, that is being used widely is to perform Round Trip Time (RTT) measurements and take the decision in favor of the destination with minimum RTT value. However, this method is not scalable and cannot take into account the platform-local preferences (business objectives, routing rules and Service Level Agreement (SLA), QoS for voice, video, etc). To address the reliability and scalability issues, businesses typically use Border Gateway Protocol

(BGP). However, BGP deployment is costly and requires lots of administration effort and hence does not suit small-to-medium business. Autonomous System Number, range of IP address prefixes, netmasks address assignment and allocation policies are required to configure the BGP. Stream Control Transport Protocol (SCTP) was launched to overcome the shortcomings of TCP and UDP. It provides multihoming support, offering failover and fault tolerance options only. There are two main factors involved in the design and development of multihoming load balancing systems: calculation of the decision for link selection among the available choices and the mechanism involving the enforcement of the calculated decision.

Multihoming Load Balancing systems being developed and deployed are examined from the functionality (algorithmic) and performance implications view points. Decision-making, i.e., the choice of link in these systems is usually static and/or semi-dynamic. Moreover, these systems take into account few attributes among the set of available parameters over the platform, while calculating the decision (service profile, Service Level Agreement (SLA) reliability information, time of the day, business objectives of the company, latest state of the links and user profiles). A system capable of taking into account Service Level Specifications (SLS), e.g. susceptible delay, jitter and packet loss may not accommodate the technology specific information. Systems considering user, service and QoS profiles do not compensate for dynamic context of the request. The first challenge is to utilize the available information over the platform maximally, which comes from different sources with different dimensions so that the final decision for link selection reflects dynamic control and effective resource utilization with good QoS. Another objective is to enforce the calculated decision using existing technologies (e.g NAT, DNS cycling hashing, etc.) so that we do not have to revamp the existing protocol stack. Both these issues (decision calculation and its enforcement) are addressed in this paper by taking into account the static and dynamic information available over the platform alongside the multihomed link data.

The rest of this paper is organized as follows: Section 2 describes the proposed architecture. Section 3 elaborates the MCDM theory and its application with two distinct methods. In section 4, the test bed for the validity of the proposed solution is presented. Section 5 outlines related work. Finally, in section 6, concluding remarks are made while outlining our future work.

2 Proposed Architecture

The architecture shown in Fig. 1 is proposed in the Company@ges [1] project, which provides a platform where companies are linked to the rest of the world via two or more network accesses offering multimedia services. QoS-centered architecture integrates devices and modules from different vendors over a single platform while offering multimedia services for public and private(local) networks. The global objective is the accommodation of dynamic modifications/variations into the decision-making criteria for request routing to different links by using enhanced general methods/techniques and protocols. Service, control and routing issues posing a multi-criteria problem are handled together without affecting

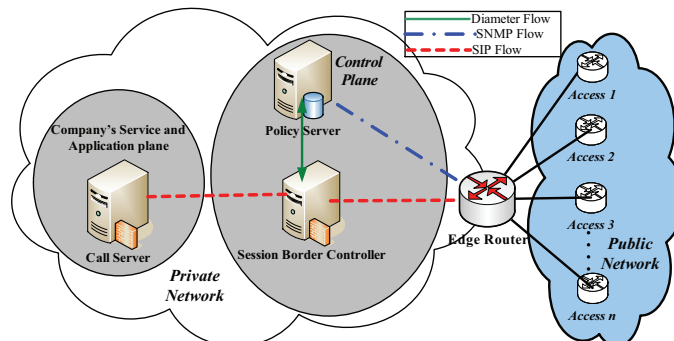


Fig. 1. Proposed Architecture.

the standard mechanisms and classical layered approach.

Policy Server (PS) is the main controller in the proposed architecture. It acts as a Policy Decision Point (PDP). It computes all the decisions by taking into account the static configurations and dynamics taking place over the platform, in addition to the policy enforcement supervision. The proposed dynamic decision engine partly constitutes the core of PS. Session Border Controller (SBC) in the offered framework is primarily dedicated to multimedia communication. It provides a number of vendor specific functionalities depending on the requirements and its deployment. More details are available in [2, 3]. In addition to SBC's standard functionalities, it is tweaked to act as a Policy Enforcement Point (PEP) in the proposed architecture. Call Server (CS) is an important component of IP-based PBX/Softswitch. It supports proxy, registrar, redirect and location services. CS here provide registration, user profile management and service control mechanism. It is modified to handle the user profile based Call Admission Control (CAC) functionality. It is worthwhile to mention here that we are targeting the Session Initiation Protocol (SIP) based multimedia communication over the platform while focussing on decision making and its enforcement. SIP is a Hypertext Transfer Protocol (HTTP) like request response signaling protocol used for creating, modifying and tearing down sessions [4]. Components of this platform (Fig. 1) are provided by partners: the platform's service and application plane is realized by modules from Alcatel-Lucent whereas SBC and PS are/will be developed and tweaked by two different teams at TELECOM Bretagne Brest. For detailed functionality, information sharing and communication between different devices over the presented architecture, the reader is referred to [2, 3].

The protocol chosen to communicate the information/decisions between PDP and PEP is Diameter with newly defined and developed Attribute Value Pairs (AVPs). Diameter is natively an Authentication Authorization Accounting (AAA) protocol. Due to its AAA characteristics, its enhancement orientations are becoming natural for decision-based network management. It has large AVP space and supports large number of pending requests. Common Open Policy Service (COPS) [6], a strong candidate for Policy Based Network Management (PBNM) [5] has not been chosen for decision(policy) provisioning and dissemination, as it is specifically designed for device-level configuration and management. However, dynamic session/call/data-connection management is required

while taking into account the variations and latest dynamics. SNMP has sometimes been proposed in the literature to be a candidate for PBNM [7]. SNMP-based information in our system is exploited to gauge the QoS parameters of access router interfaces. This paper addresses the private-public border traffic management issues for request routing decisions at the application layer (OSI) while taking multiple criteria into consideration. It supports dynamicity by using Multi-Criteria Decision Making (MCDM) theory. The calculated decisions are enforced during the signaling phase of SIP-based multimedia communication using existing mechanisms (NAT, Back-2-Back User Agent (B2BUA), Proxying etc.).

3 Multi-Criteria Decision Making Theory and its Application

The process of decision making involves choosing the best alternative, given a set of alternatives (available links in our case) and a set of criteria (context of the request and predefined configurations/settings over the proposed platform). These alternatives can also be ranked on the basis of multiple criteria using some specific MCDM method. MCDM methods have been used to help solve a wide variety of problems in many different applications such as telecommunications, manufacturing, transportation and software engineering [8],[9]. Experiences show that there is not a single MCDM technique to deal with all multi-criteria problems. Indeed each situation requires a specific MCDM technique. The choice of the technique and its impact on the decision making is not within the scope of this work and reader is referred to [10] for an overview of this particular domain. The targeted objectives in the multi-criteria decision making problems might sometimes be conflicting and/or overlapping. In the posed problem, SLA includes Delay, Jitter and Packet Loss (DJPL) which falls under the business objectives of the company when they sign the direct or reciprocal agreements with partners or companies. However, the same set of parameters (DJPL) are used to grade the QoS of the available links (a link has to be chosen). The triplet (DJPL) can be used to gauge the authorization and authentication of a particular user class (e.g., Gold user must have the best QoS profile, while Silver can be assigned either a good or a satisfactory QoS profile) while executing the context of the request. There are various approaches to deal with such sort of problems each having its pros and cons but we will not address this issue due to space limitations. Two MCDM methods have been chosen to address the problem of SIP-based multimedia traffic routing on the basis of multiple criteria, while making decision for the best link. Each MCDM problem is associated with multiple attributes. These attributes are linked to the goals and are referred as decision criteria. Since different criteria represent different dimensions of alternatives, they may conflict with each other (e.g., Cumulative Bandwidth may be confused with Total Bandwidth, traffic measurements, granularity (connection/session level) obsession, cost etc.). The criteria are assigned different weights according to context of the request and the rules defined over the platform. Conventional algorithms used for link selection in multihomed networks are either user-centric or motivated for efficient resource utilization over

the platform and/or they are centered towards application optimization for desired QoS. However, to cope with all these multi-criteria goals and objectives, the choice of MCDM is indispensable. Hence we picked GRA among the number of available MCDM methods to be applied to our problem of decision making in outsourcing decision enforcement mode. The reader is referred to [2] for detailed mechanism and information sharing (among different modules especially between PDP and PEP) about outsourcing decision enforcement mode. This mode takes latest platform conditions and network information into account and computes the decision on the fly, in accordance with the context of the request.

3.1 Grey Relational Analysis (GRA)

GRA is a decision-making technique that is based on grey system theory. Originally developed by Deng [11], Grey theory is widely applied in fields such as systems analysis, data processing, modeling and prediction, as well as control and decision-making. It is an effective mathematical means to deal with systems characterized by conflicting and partial information. Grey relation refers to the uncertain relations among things, among elements of systems, or among elements and behaviors. Due of its ability to use reference attribute vector, it is being applied in the proposed decision-making system in outsourcing mode. Moreover, the platform's latest conditions and the context of the request are taken into account while constructing the reference vector.

3.2 Problem Formulation and GRA Application

For brevity and to avoid the complexity of stringent mathematics, 6 attributes are chosen for the application of MCDM methods on 4 alternative links for routing the multimedia sessions. Fig. 2 illustrates the hierarchy of the desired goal, the criteria and the available alternative links. As mentioned before, we are focusing on SIP based multimedia communication, so let us have the QoS requirements for these services as follows:

Video Call: It requires a higher bandwidth than voice so the available bandwidth, transport cost and current utilization are important factors. Its ability to buffer a longer duration data before playback makes it less vulnerable to delay and jitter than voice.

Voice Call: It is very sensitive to delay and jitter, requiring low bandwidth but this service is susceptible to packet losses to some extent. Because of its low bandwidth usage, the transport cost factor is considered negligible. Total bandwidth and available bandwidth are not significant factors due to low bandwidth requirements. Since there is some correlation of utilization with jitter and delay, it is preferred to have a low utilization for the selected network.

There are four links L_1 , L_2 , L_3 and L_4 and for the sake of simplicity, we assume that the reference constituted contains 6 attributes denoted by UR , D , J , PL ,

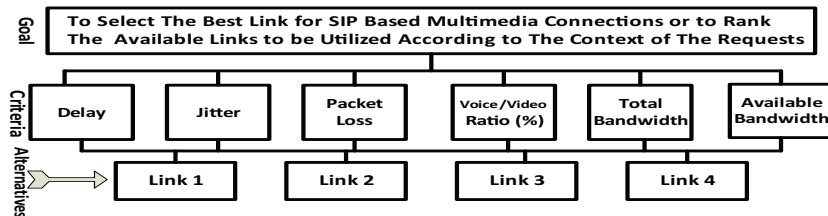


Fig. 2. Candidate Links, Attributes and Objectives Hierarchy.

TB and AB represents the voice/video Utilization Ratio of a link (UR), with Delay (D), Jitter (J), Packet Loss (PL), Total Bandwidth (TB) and Available Bandwidth (AB) respectively. Two user Alice (local) and Bob (remote) are supposed to communicate (outgoing SIP communication) with each other by using the available resources on the platform as shown in Fig.1. Alice initiates the communication and sends an initial INVITE to the SBC at the border of the platform to start a voice call with Bob. SBC extracts the information from the request and constructs the reference for decision making by taking into account the platform's pre-defined set of objectives for different services and users. At minimum, user and communication types (more details available in [3]) have to be known. This information is sent to the PS and the corresponding user, application and QoS profiles are loaded from the profile base. The information from the context of the request is bundled with the link latest information to construct the reference vector as follows:

$$Ref(L) = [UR\ D\ J\ PL\ TB\ AB] \quad (1)$$

The candidate link attributes constituting the Decision Matrix (DM) is given as follows:

$$DM = \begin{bmatrix} UR_1 & D_1 & J_1 & PL_1 & TB_1 & AB_1 \\ UR_2 & D_2 & J_2 & PL_2 & TB_2 & AB_2 \\ UR_3 & D_3 & J_3 & PL_3 & TB_3 & AB_3 \\ UR_4 & D_4 & J_4 & PL_4 & TB_4 & AB_4 \end{bmatrix} \begin{matrix} L_1 \\ L_2 \\ L_3 \\ L_4 \end{matrix} \quad (2)$$

The values of these attributes are obtained from the SNMP traps and the Service Level Agreements (SLAs) of the corresponding links over the platform. As the parameters involved in the Decision Matrix come from different sources, the units representing the values are different. We need to normalize these parameters in order to make them unit-less. The attributes having bigger values (e.g., TB is in Mega) are divided by the largest value in the corresponding column vector while the smaller range attribute (e.g., D , which is in milliseconds) is divided by the smallest value in the corresponding column vector. The normalized Decision Matrix is given by

$$DM \sim = \begin{bmatrix} \widetilde{UR}_1 & \widetilde{D}_1 & \widetilde{J}_1 & \widetilde{PL}_1 & \widetilde{TB}_1 & \widetilde{AB}_1 \\ \widetilde{UR}_2 & \widetilde{D}_2 & \widetilde{J}_2 & \widetilde{PL}_2 & \widetilde{TB}_2 & \widetilde{AB}_2 \\ \widetilde{UR}_3 & \widetilde{D}_3 & \widetilde{J}_3 & \widetilde{PL}_3 & \widetilde{TB}_3 & \widetilde{AB}_3 \\ \widetilde{UR}_4 & \widetilde{D}_4 & \widetilde{J}_4 & \widetilde{PL}_4 & \widetilde{TB}_4 & \widetilde{AB}_4 \end{bmatrix} \begin{matrix} L_1 \\ L_2 \\ L_3 \\ L_4 \end{matrix} \quad (3)$$

The normalized reference vector is given by

$$Ref(\widetilde{L}) = [\widetilde{UR}\ \widetilde{D}\ \widetilde{J}\ \widetilde{PL}\ \widetilde{TB}\ \widetilde{AB}] \quad (4)$$

Now the distance between the corresponding normalized reference vector entities and the normalized Decision Matrix entities is calculated as follows:

$$\Delta_{UR_i} = |\widetilde{UR} - \widetilde{UR}_i|, i = 1, 2, 3, 4 \quad (5)$$

The Δ Decision Matrix is obtained by applying Eq. 5 to the corresponding entities in the normalized Decision Matrix and the normalized reference vector:

$$\Delta_{DM} = \begin{bmatrix} \Delta_{UR_1} & \Delta_{D_1} & \Delta_{J_1} & \Delta_{PL_1} & \Delta_{TB_1} & \Delta_{AB_1} \\ \Delta_{UR_2} & \Delta_{D_2} & \Delta_{J_2} & \Delta_{PL_2} & \Delta_{TB_2} & \Delta_{AB_2} \\ \Delta_{UR_3} & \Delta_{D_3} & \Delta_{J_3} & \Delta_{PL_3} & \Delta_{TB_3} & \Delta_{AB_3} \\ \Delta_{UR_4} & \Delta_{D_4} & \Delta_{J_4} & \Delta_{PL_4} & \Delta_{TB_4} & \Delta_{AB_4} \end{bmatrix} \begin{matrix} L_1 \\ L_2 \\ L_3 \\ L_4 \end{matrix} \quad (6)$$

Grey Relation Coefficients (GRCs) representing the measurement of similarity of an attribute to its reference are calculated (e.g. for voice/video Utilization Ratio of a link UR) as follows:

$$GRC_{UR_i} = \frac{\Delta_{\min} + \alpha\Delta_{\max}}{\Delta_{UR_i} + \alpha\Delta_{\max}}, i = 1, 2, 3, 4 \quad (7)$$

where $\alpha \in [0, 1]$ and Δ_{\min} and Δ_{\max} are calculated as follows:

$$\Delta_{\max} = \max_i (\Delta_{UR_i} + \Delta_{D_i} + \Delta_{J_i} + \Delta_{PL_i} + \Delta_{TB_i} + \Delta_{AB_i}) \quad (8)$$

$$\Delta_{\min} = \min_i (\Delta_{UR_i} + \Delta_{D_i} + \Delta_{J_i} + \Delta_{PL_i} + \Delta_{TB_i} + \Delta_{AB_i}) \quad (9)$$

As we are emphasizing on voice communication (outbound calls) and to meet the QoS requirements of voice, (Delay and Jitter are given more weight), we choose the weights corresponding to each attribute in the Decision Matrix. The available bandwidth is coupled with user profile loaded from the profile base (in case of gold profile, it is highly desirable to choose the link with good available bandwidth so AB and U will also be given suitable weight values). These assigned weights illustrate the relative importance of each attribute in Decision Matrix such that:

$$W = W_{UR} + W_D + W_J + W_{PL} + W_{TB} + W_{AB} = 1 \quad (10)$$

The weighted GRC coefficient representing an attribute column is given by:

$$GRC_{wUR_i} = W_{UR} * GRC_{UR_i}, i = 1, 2, 3, 4 \quad (11)$$

The resulting weighted GRC matrix is given by:

$$GRC_{wDM} = \begin{bmatrix} GRC_{wUR_1} & GRC_{wD_1} & GRC_{wJ_1} & GRC_{wPL_1} & GRC_{wTB_1} & GRC_{wAB_1} \\ GRC_{wUR_2} & GRC_{wD_2} & GRC_{wJ_2} & GRC_{wPL_2} & GRC_{wTB_2} & GRC_{wAB_2} \\ GRC_{wUR_3} & GRC_{wD_3} & GRC_{wJ_3} & GRC_{wPL_3} & GRC_{wTB_3} & GRC_{wAB_3} \\ GRC_{wUR_4} & GRC_{wD_4} & GRC_{wJ_4} & GRC_{wPL_4} & GRC_{wTB_4} & GRC_{wAB_4} \end{bmatrix} \begin{matrix} L_1 \\ L_2 \\ L_3 \\ L_4 \end{matrix} \quad (12)$$

The GRC value for individual link is calculated as follows:

$$\begin{aligned} Coef(GRC)_i &= GRC_{wUR_i} + GRC_{wD_i} + GRC_{wJ_i} + GRC_{wPL_i} + GRC_{wTB_i} \\ &+ GRC_{wAB_i}, i = 1, 2, 3, 4. \end{aligned} \quad (13)$$

The Candidate Link with the highest GRC coefficient value is the final decision, i.e., the best link for the request. There are two possibilities for calculating/declaring the reference attribute vector: the first is to compute the reference attribute vector before the request arrives and the second is to calculate it on the fly (discussed above in the GRA method). The susceptible QoS set of parameters are well defined and known for voice and video. The range of required bandwidth for different codecs (used by the end points during the multimedia communication) is also well documented. The attribute, available bandwidth is calculated by keeping the track of number of ongoing calls/requests on a particular link (i.e., Available Bandwidth=Total Bandwidth - Used Bandwidth). It is important to mention that the presented GRA includes the simplest possible case. Embedding the reference vector beforehand can be tedious and complex as the number of links and attributes increases. The business objectives of an enterprise might change (e.g., voice might be given priority over video, the silver profile might use gold profile service during night (free hours), etc.), the

user profile priorities/authentication/authorization parameter (QoS profile corresponding to a user profile) may go through modification, or the link resources might go through up-gradation/downgrading. Although this complexity can be handled but it requires extensive administrative efforts. The objective however, is to minimize these efforts at minimal while taking into account the dynamicity that the manual system is not able to accommodate.

3.3 Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) MCDM Method

TOPSIS was developed by Yoon and Hwang [12]. It is an alternative to ELECTRE [13] and is considered to be one of its variants. It is known as a double standard method that evaluates alternatives through two basic criteria. First, the chosen alternative should have the shortest distance from the positive ideal solution and secondly it must be farthest from the negative-ideal solution for a MCDM problem. The perceived positive and negative ideal solutions are based on the range of attribute values available for the alternatives. The distances are measured in Euclidean terms. The Euclidean distance approach is proposed to evaluate the relative closeness of the alternatives to the ideal solution. The reason for choosing TOPSIS in provisioning enforcement mode (i.e., pre-computed decisions are available at PEP in this mode, which is described in detail in [2]) is that it will rank/grade the available alternatives (links) whenever applied. Moreover, TOPSIS is extended to be applied on interval data (i.e. lower and upper values of an attribute) for the provisioning enforcement mode over the proposed architecture. In provisioning mode, the decision engine is not very much interactive with the platform's variations especially at the arrival of a new request so it provides half hearted dynamicity. Hence if the exact value of an attribute is not known, then these bounds (upper and lower) are used for the application of an extended TOPSIS. The best link among the available alternative links (ranked by the application of an extended TOPSIS) is assigned to request by following the predefined set of criteria. Due to space limitations and to avoid the complexity, only the Decision Matrix (equation 14) is expressed with lower and upper bounds while considering the same set of 6 attributes and 4 links as follows:

$$DM = \begin{bmatrix} (UR_1^L, UR_1^U) & (D_1^L, D_1^U) & (J_1^L, J_1^U) & (PL_1^L, PL_1^U) & (TB_1^L, TB_1^U) & (AB_1^L, AB_1^U) \\ (UR_2^L, UR_2^U) & (D_2^L, D_2^U) & (J_2^L, J_2^U) & (PL_2^L, PL_2^U) & (TB_2^L, TB_2^U) & (AB_2^L, AB_2^U) \\ (UR_3^L, UR_3^U) & (D_3^L, D_3^U) & (J_3^L, J_3^U) & (PL_3^L, PL_3^U) & (TB_3^L, TB_3^U) & (AB_3^L, AB_3^U) \\ (UR_4^L, UR_4^U) & (D_4^L, D_4^U) & (J_4^L, J_4^U) & (PL_4^L, PL_4^U) & (TB_4^L, TB_4^U) & (AB_4^L, AB_4^U) \end{bmatrix} \begin{matrix} L_1 \\ L_2 \\ L_3 \\ L_4 \end{matrix} \quad (14)$$

3.4 TOPSIS MCDM Method Application Steps

TOPSIS method is explained and applied here by using the standard approach to avoid rigorous mathematics in the following steps.

1. Normalize the Decision Matrix containing the link attributes: the process is to transform different scales and units among various criteria into common measurable units in order to allow comparisons across the criteria. The normalization procedure is the same as described in section 3.2.

2. Construct the weighted normalized Decision Matrix: it cannot be assumed that each evaluation criterion is of equal importance because the evaluation criteria have various meanings. The decision engine in provisioning mode calculates the weight of the corresponding column vector representing an attribute by using the environmental conditions and administrative rules/conditions at the time of TOPSIS execution. The context of the request is not taken into account, as opposed to GRA.
3. Determine positive and negative ideal solutions for each attribute: the positive ideal solution indicates the most preferable alternative, and the negative ideal solution indicates the least preferable alternative as follows (e.g. for voice/video link Utilization ratio, U)

$$UR^+ = (Max(UR_{w-norm})_i) \parallel (Min(UR_{w-norm})_i), i = 1, 2, 3, 4 \quad (15)$$

and

$$UR^- = (Min(UR_{w-norm})_i) \parallel (Max(UR_{w-norm})_i), i = 1, 2, 3, 4 \quad (16)$$

4. The Euclidean distance method is applied to measure the separation from the positive and negative ideal for each alternative

$$S_i^+ = \sqrt{\frac{((UR_{w-norm})_i - UR^+)^2 + ((D_{w-norm})_i - D^+)^2 + ((J_{w-norm})_i - J^+)^2 + ((PL_{w-norm})_i - PL^+)^2 + ((TB_{w-norm})_i - TB^+)^2 + ((AB_{w-norm})_i - AB^+)^2}{2}} \quad (17)$$

and

$$S_i^- = \sqrt{\frac{((UR_{w-norm})_i - UR^-)^2 + ((D_{w-norm})_i - D^-)^2 + ((J_{w-norm})_i - J^-)^2 + ((PL_{w-norm})_i - PL^-)^2 + ((TB_{w-norm})_i - TB^-)^2 + ((AB_{w-norm})_i - AB^-)^2}{2}} \quad (18)$$

5. Finally, the candidate links are ranked by measuring the relative closeness of an alternative (candidate links L_1 , L_2 , L_3 and L_4 under consideration represented by a row vector in the Decision Matrix) to the ideal solution S^+ as follows:

$$R_i = \frac{S_i^+}{S_i^+ + S_i^-} \quad (19)$$

	UR	D	J	PL	TB	AB
	%	Milliseconds(ms)	ms	%	Megabits per second(Mbps)	Mbps
L₁	66.65	300	50	100	100	65
L₂	53.84	200	25	25	100	71
L₃	81.81	100	15	10	100	81
L₄	25.00	150	30	30	100	46

Table 1. Links With Corresponding Parametric Values

The links L_1 , L_2 , L_3 and L_4 characterized by attributes voice/video Utilization ratio UR , Delay (D), Jitter (J) Packet Loss (PL), Total Bandwidth (TB) and Available Bandwidth (AB) respectively are represented by the values shown in Table 1. D , J and L are given higher weights (step 2) due to voice call (outgoing) while keeping in view the required bandwidth judged from the codec negotiated during the call setup. For the application of TOPSIS on the links represented by the corresponding row vectors in Table 1, all 5 steps are gone through in the order stated above in this section. The links are ranked with R values as mentioned in Table 2.

	L_1	L_2	L_3	L_4
R Value	0.0788	0.6852	0.7805	0.7186
Rank	4	3	1	2

Table 2. R Values and the Corresponding Grading of Links

4 Test Bed and Experimental Setup

OpenSIPS [14], an open source SIP server is tweaked to act as SBC and Load Balancer (LB). It is built around the core that is responsible for the basic processing and handling of SIP messages. The modules developed around its core are responsible for the majority of OpenSIPS functionalities. Its scalable and modular design provides a number of functionalities (registrar, router/proxy (LCR, dynamic routing, dialplan features), redirect server, B2BUA etc). It enforces the calculated decision and forwards the outgoing SIP request to different links for the experimental setup as shown in Fig. 3. SIPp [15] is used to generate extensive SIP requests (INVITE). It is a configurable traffic generator and is extensible via a simple XML configuration language. Call model with User Agent Client (UAC) sends an INVITE to OpenSIPS, and it is analyzed to judge its communication type. It is important to mention here that a random number is generated to send the codec information along with the SIP message. The bandwidth requirement of the call is judged from the codec information and the request is forwarded to an appropriate link (already ranked by using TOPSIS) by following the pre-defined criteria. Network Address Translation (NAT) is enabled on OpenSIPS and the decision is enforced during NAT implementation in provisioning mode. Details about the design and development of the parser for embedding the calculated decisions during NATing are avoided due to space limitations. The SIP server responds with 100 TRYING, 180 RINGING and 200 OK. UAC then sends an ACK and the call is established. The UAC closes the communication after variable pause by sending a BYE which is acknowledged by the SIP server with 200 OK. Wireshark is used to capture the traffic at different interfaces (links).

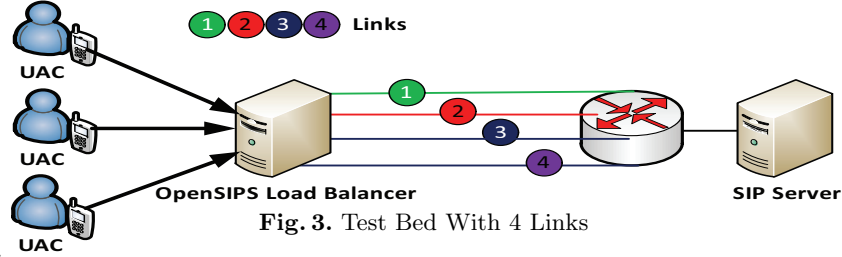


Fig. 3. Test Bed With 4 Links

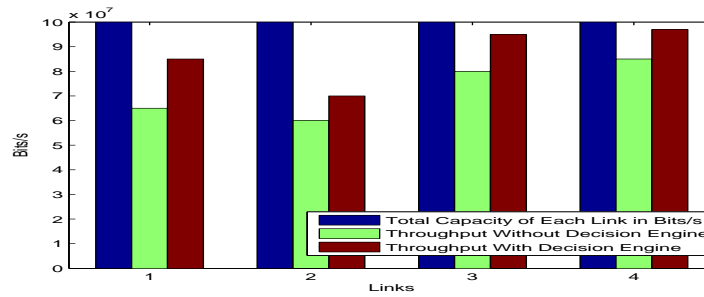


Fig. 4. Throughput of Each Link With and Without Decision Engine

OriginLab [16] is used for data analysis from the captured file. Throughput of each link is plotted with and without decision engine (i.e. using built-in LB in OpenSIPS) as shown in Fig. 4. It is observed that there is a significant improvement in the throughput for each link with decision engine while performing SIP call routing. The retransmission mechanism within SIPp is turned off when INVITE messages are sent in order to know that a call has been dropped. The aggregated call dropping probability (for the 4 links shown in Fig. 3) with the proposed decision engine has lower value than the ordinary Opensips's LB as shown in Fig. 5. The performance may improve by decision computation in outsourcing mode (on the fly) alongside attribute space enlargement as the present tests are performed in simplest possible scenario.

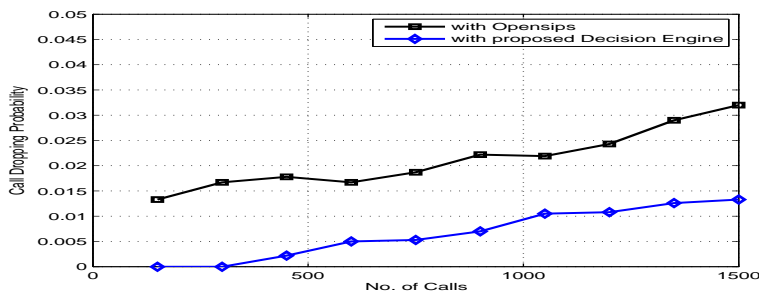


Fig. 5. Call Dropping Probability With and Without Decision Engine.

5 Related Work

There are commercial and proprietary solutions available for SIP-based call routing at application layer. Publicly available information does not reveal the decision making mechanism and the LB algorithms. The core design and lower-level functionality are hidden because of commercial implications. However some vendors provide Software Development Kit (SDK) for customization of the specific solution with limited interaction and access to the core [17, 18]. Some products offer partial dynamicity with limited controls, while others are enforcing static decisions/rules. F5 networks [19] uses NAT for Load Balancing the SIP traffic to multiple links with static configurations. The proposed solution in this work accommodates the dynamic behavior of the platform and the context of the request with the provision of off-line (provisioning mode) and on the fly (outsourcing mode) decision making using MCDM theory. This theory is used for access network technology (UMTS, GSM, WLAN, etc.) selection during the handoff based on user preferences [20]. A user priority scheme for admission control using Analytic Hierarchy Process (AHP) is proposed in [21]. The proposed solution here uses GRA and an extended TOPSIS for online (on the fly) and off-line decision making respectively.

6 Conclusions and Future Work

Context of the request, state of the links and variations over the platform constitutes a multidisciplinary problem. This multi-criteria issue outlines more complexity when a single decision is required for routing a SIP request in a multihomed network. Traditional algorithms used for link selection in multihomed setups are either user-centric or service-oriented. Nevertheless, these solutions

either focus on performance optimization or they are technology specific. To cope with all these multi-disciplinary objectives, Multi Criteria Decision Making (MCDM) theory is chosen. A dynamic decision engine for SIP-based call routing has been presented. It is capable of handling dynamicity and fluctuations over the platform by taking into account a large number of attributes with corresponding weights. Two MCDM methods, namely GRA and TOPSIS are used in outsourcing (on the fly) and provisioning (off-line) enforcement modes respectively. A test bed is developed to validate the proposed solution. Few calls are dropped with the proposed decision engine giving lower aggregated call dropping probability than the ordinary Load Balancer. The throughput of the individual multihomed links is improved significantly. Decision and enforcement modules can be integrated in a single box but the solutions are developed by partners/teams independently so it might be an emerging step. Future work includes development of an automated linguistics to specify goals, criteria and alternatives along with their translation. This landmark will connect MCDM to conventional Policy Based Network Management (PBNM), but with dynamicity. Introduction of parallelism while handling the user and network-based information constitutes another dimension of our future work.

References

1. <http://www.images-et-reseaux.com/en> (12/01/2011, Last Visited (LV))
2. S. A. Mushtaq, O. Salem, C. Lohr, A. Gravey: Policy-based QoS Management for Multimedia Communication 14th EUNICE Open European Conference (2008)
3. S. A. Mushtaq, O. Salem, C. Lohr, A. Gravey: Distributed Call Admission Control in SIP Based Multimedia Communication; NEM Summit 2008
4. J. Rosenberg et al: Session Initiation Protocol; RFC 3261; June 2002.
5. M. Brenner: Diameter Policy Processing Application; RFC 5224 March, 2008
6. D. Durham et al: The COPS (Common Open Policy Service) Protocol; RFC 2748.
7. S. Boros: Policy-based network management with SNMP, October 2000
8. Semih, S. S. Kara, Elif: Long term supplier selection using a combined fuzzy MCDM approach: A case study for a telecommunication company, 2008.
9. P.P. Bonissone, R. Subbu, J. Lizzi: MCDM: A framework for research and applications; Computational Intelligence Magazine, IEEE, Aug. 2009.
10. E. Kornysheva, C. Salinesi: MCDM Techniques Selection Approaches; State of the Art; IEEE Symposium on Computational Intelligence in MCDM, 2007
11. J.L. Deng: Introduction to Grey System; J. Gerry Syst. 1(1), 1-24, 1989.
12. C.L. Hwang, K. Yoon: Multiple Attribute Decision Making Methods and Applications; Springer Verlag, New York, 1981.
13. R. Benayoun, B. Roy, N. Sussmann: Manual de reference du programme electre, Note de Sythese et Formation; Direction Scietifique SEMA, N. 25, 1966.
14. F. E. Goncalves: Building Telephony Systems with OpenSIPS 1.6; 2010.
15. SIPp: Test Tool Traffic Generator for the SIP Protocol; <http://sipp.sourceforge.net>
16. OriginLab: <http://www.originlab.com> (26/12/2010, LV)
17. <http://www.metaswitch.com/sbc-session-border-controller> (05/01/2011, LV)
18. <http://www.acmepacket.com> (05/01/2011, LV)
19. <http://f5.com/products/big-ip/local-traffic-manager.html> (05/01/2011, LV)
20. A. Seghal, R. Agrawal: QoS Based Network Selection Scheme for 4G Systems; IEEE Transactions on Consumer Electronics, 2010.
21. H. Pervaiz: A Multi-Criteria Decision Making (MCDM) Network Selection Model providing enhanced QoS differentiation to customers; MCIT 2010.