An Integration of Semantics in Multi Criteria Decision Making for Converged Multimedia Network Management

Sajjad Ali Mushtaq
Computer Engineering Dept.
Telecom Bretagne Brest France
Email: sajjad.mushtaq@telecom-bretagne.eu

Christophe Lohr
Computer Engineering Dept.
Telecom Bretagne Brest France
Email: christophe.lohr@telecom-bretagne.eu

Annie Gravey
Computer Engineering Dept.
Telecom Bretagne Brest France
Email: annie.gravey@telecom-bretagne.eu

Abstract—A framework for rule-based multimedia network management in multihoming environment is presented. Information diversity with multiple criteria and semantic variations has led us to integrate the Multi Criteria Decision Making (MCDM) theory and Ontology over an all IP platform. MCDM helps in achieving the multiple objectives on the basis of multi-disciplinary information domains. The ontology, on the other hand, overcomes the semantic matching and reasoning issues using conventional object-oriented techniques with the help of inference engine. A simple use-case involving the dynamic decision-based routing at private-public network border is presented to validate the proposed solution. The system gives higher throughput and lower call dropping probability with an add-on susceptible delay offering good QoS.

I. INTRODUCTION

Network Management and control has always been a challenging task especially within an all Internet Protocol (IP) converged environment. Traffic mix, heterogeneous wireline/wireless access technology, fixed-to-mobile and vice versa convergence difficulties, versatile data representation, diverse Quality of Service (QoS) requirements and variations/dynamics over the network are some of the main complexities involved for network administration and control. The underlying problem becomes more prominent when rule based network management is desired over a converged platform offering multimedia and data services all together with promising QoS and Quality of Experience (QoE). QoS-centered architecture shown in Fig. 1 integrates different services, binds versatile control interfaces, combines a number of protocols in addition to unification of transport and access technologies which are unified over a single platform. Voice, video, triple-play, quadruple services over the platform are being offered by different service providers requiring diverse resources with varying set of QoS parameters.

Application, control and network information are to be shared within those distinct domains and across these planes for network management. The attributes and parameters related to resources, profiles, services and dynamics taking place over those planes are disseminated and exchanged via different interfaces and protocols. The information sharing and exploitation process becomes more delicate when its control and management has to be performed under the dictation of certain set of pre-configured rules. There are some frameworks [1], [2], [3] for rule-based management and control but they are static and/or semi-dynamic and are limited to device level configuration and control with different granularity. Moreover, in heterogeneous unified communication systems, there are vendor-specific concepts and implementation dependencies with different means of information representation and processing. These systems take into account a few information among the set of available parameters over the platform, while computing the rules (service profile, reliability information, time of the day, business objectives of the company, latest state of the links, user profiles and Service Level Agreement (SLA)). Moreover, 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 information stated above for rule-based network management and control comes from different sources with different dimensions, hence formulating a multicriteria problem. The first challenge is to utilize the available information over the platform maximally so that the network management reflects dynamic control and effective resource utilization with good QoS and QoE. Moreover, the objective is to capture the semantics among these distinct domains (inter-domain and intra-domain) with all the dynamics and variations in order to avoid the conflicts and overlapping. Finally, the use of existing technologies (e.g. Network Address Translation (NAT)ing, Domain Name Service (DNS) Cycling, Hashing, Proxying etc.) without introducing overheads in the protocol stack for policy enforcement is another aim over the unified converged platform.

A framework for rule based network control and management has been presented. In order to avoid complexity, routing rules computation (taking into account all the knowledge from distinct information domains and the dynamics taking place over
the presented platform) at private-public network border is focused on while emphasizing the multimedia communication (voice, video). Multi Criteria Decision Making (MCDM) theory has been applied to handle the multidimensional problem with multiple objectives in addition to varying set of attributes and parameters. Ontologies [4] are integrated with MCDM paradigm to capture the inter and intra domain semantics and dynamics. A reasoner enables the hypothesis generation using the facts, meanings and feedback captured from context, the information base and environmental conditional parameters. The rest of the paper is organized as follows: In the following Section, platform’s architecture is described. Section III elaborates the system model, MCDM theory, ontology and their integration for the application of dynamic routing. Section IV presents TOPSIS application steps with simple scenario including an example. In Section V, the validity of the proposed solution is presented with comparisons. Section VI outlines the related work. Finally, in Section VII, concluding remarks are given while outlining the future work.

II. System Architecture

The architecture shown in Fig. 1 provides an all IP cost effective solution offering versatile access technology convergence while highlighting network traffic unification. It integrates devices and modules from different vendors over a single platform while offering diverse services for public and private (local) networks. The global objective is the accommodation of dynamic modifications/variations into the policy-making (decision-making) criteria for control and management of a multihomed platform by using enhanced general methods/techniques and protocols. Service, control, network/transport and routing issues posing a multi-criteria problem are handled together without affecting the standard mechanisms and classical layered approach. The underlying platform stems from competitiveness cluster for handling traffic management issues at the network border in either direction (inbound and outbound traffic).

Policy Server (PS) is the main controller in the proposed architecture. It acts as a Policy Decision Point (PDP). It computes all the rules (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. 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 provides registration, user profile management and service control mechanism. It is modified to handle the user profile based Call Admission Control (CAC) functionality. 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 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 [5], [6].

The protocol chosen to communicate the information/rules 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. SNMP-based information in our system is exploited to gauge the QoS parameters of access router interfaces. This paper addresses the policy based control and management in a highly dynamic, multidisciplinary and multivariate information domains (planes/layers) which are synchronized locally and asynchronous globally (Globally Asynchronous Locally Synchronous (GALS) systems).

III. Problem Formulation and System Model

The information sources representing different domains (e.g. service plane, control plane etc.) might be highly structured and synchronized locally (inter-domain/plane) but may have higher probability of asynchronous and un-structured information representation globally (intra-domains/planes). QoS profile of the links, user authentication/authorization profiles, service variants/profiles, business objectives of the company, fluent dynamics over the multihomed platform and traffic management issues at private-public network border constitutes a multi-disciplinary problem. The information coming from different sources with different dimensions reflects the complexity of the underlying problem when a single rule/decision has to be computed on the basis of multi-dimensional and multi-disciplinary information. Moreover, the situation becomes more mingled when multiple objectives have to be targeted in addition to the varying inter/intra-domain semantics that have to be captured for effective and efficient rule-based control and management.

Conventional solutions to handle such scenarios in multihomed converged environment 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, MCDM is used. The choice of the technique and its impact on the decision-making is not within the scope of this work and the reader is referred to [7] for an overview of this particular domain. Ontologies are used for representing the individual domain knowledge with semantics. These ontologies formally and explicitly specify the conceptualization via terms and relationships. The domains, presenting the knowledge base are complex and might have large number of sub-domains with frequent variations and dynamics. This complexity is handled by using the normalization technique (via inference engine). The ontologies with multiple inheritance are managed with automated reasoner using a set of axioms. Moreover, the ontologies are representation of relationships between classes and/or instances within a particular domain and as such, they cannot provide the knowledge based reasoning and feedback mechanism. The reasoner accommodates the feedback mechanism.

Block diagram of the system is shown in Fig. 2. All the units are connected via the adaptive control/trigger bus. Resource manager in the unified communication environment keeps all the business, system, network and service level resource information. Context manager detects any sort of changes in the application/service plane, network/transport plane and business objectives of the platform and triggers a new set of rules for control and adaptation. The inference engine drives the parser for ontology inheritance and overloading in addition to the reasoning and learning performed under the direction of embedded axioms. Knowledge base unit contains all the technology specific and technology independent service, profile and network data sets. Rule manager contains the predefined set of rules added by the administrator of the platform reflecting the targeted goals. Decision engine is analyzing, adapting and handling the multi-criteria and multidimensional information coming from different sources with multiple objectives. The underlying system precisely illustrates different components and their interactions. However, the aim here is to highlight the use of ontology in coordination with MCDM over the unified platform for capturing the semantic variations alongside multiple criteria and versatile objectives.

A. Multi Criteria Decision Theory and Ontology Integration for Decision Engine

We focus on multimedia services over the multihomed unified platform due to the requirements of rule based QoS demands and control. Moreover, it is not possible to present an ontology on a piece of paper totally representing a particular domain. So a simple use-case is presented to understand the ontology and MCDM integration due to space limitation. It is worthwhile to mention here that Session Initiation Protocol (SIP)-based voice communication routing over the platform for rule (decision) computation and its enforcement is emphasized. SIP is a request response signaling protocol. Its infrastructure is highly open and flexible facilitating the services [8]. It allows the creation, modification and termination of service sessions independently of the underlying data-link layer technologies and transport protocols.

Voice Call: It is very sensitive to delay and jitter, requiring low bandwidth however it is susceptible to packet losses to some extent. Owing to its low bandwidth usage, the transport cost factor is considered negligible. Total bandwidth and available bandwidth are not significant factors due to the 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.

To avoid the stringent mathematics, 6 attributes are chosen for the application of MCDM method on 4 alternative links over the multihomed platform shown in Fig. 1. SIP-based voice call will be routed to one of the 4 alternative links by using the proposed decision engine. $L_1$, $L_2$, $L_3$ and $L_4$ are the four links and $UR$, $D$, $J$, $PL$, $TB$ and $AB$ are 6 attributes representing the voice/video Utilization Ratio, Delay, Jitter, Packet Loss, Total Bandwidth and Available Bandwidth 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. Fig. 3 illustrates the hierarchy of the desired goal, the criteria, sub-criteria and the available alternatives. Bob initiates the communication and sends an initial INVITE to the CS. Resource priority tag indicating the user profile is added and the request is forwarded to the SBC. Communication type is evaluated at SBC and this information (resource priority tag, communication type) is sent to the PS. Decision engine within the PS maps the corresponding profiles from the central profile base and then generates a trigger and takes the snapshot of the platform’s information for further processing. The ontology space overloads the corresponding

Fig. 2. System’s Block Diagram.

Fig. 3. Goal, Criteria, Sub-Criteria and Candidate Links Hierarchy.
concepts, classes relations and instances in accordance with the axioms of the inference engine. A glimpse of service, QoS and SLA ontologies is shown in Fig. 4. Relevant goal, criteria and sub-criteria are computed keeping in view the context and the business rules in consent with the reasoner. The attributes (corresponding to the candidate links) constitutes the Decision Matrix (DM) as follows:

$$DM = \begin{bmatrix} UR_1 & D_1 & J_1 & PL_1 & TB_1 & AB_1 \\ UR_2 & D_2 & J_2 & PL_2 & TB_2 & AB_2 \\ UR_3 & D_3 & J_3 & PL_3 & TB_3 & AB_3 \\ UR_4 & D_4 & J_4 & PL_4 & TB_4 & AB_4 \end{bmatrix} < -L_1$$

(1)

The values of these attributes are obtained from the SNMP traps and the SLAs of the corresponding links over the platform.

IV. TECHNIQUE FOR ORDER PREFERENCE BY SIMILARITY TO IDEAL SOLUTION (TOPSIS) MCDM METHOD

The system supports two decision computation and enforcement modes namely, outsourcing (on-the-fly) and provisioning (off-line) modes respectively. Outsourcing mode is presented here due to more dynamics and interactions involved over the platform during this mode. TOPSIS, developed by Yoon and Hwang [9] is used for decision computation in this mode. It is an alternative to ELECTRE [10] 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 is that it will rank/grade the available alternatives (links) whenever applied by taking into account all the variations/dynamics and static configurations of the platform. TOPSIS application steps are as follows: As the parameters involved in the DM come from different sources, the units representing the values are different. Normalization of these parameters is required 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

$$\tilde{DM} = \begin{bmatrix} \tilde{UR}_1 & \tilde{D}_1 & \tilde{J}_1 & \tilde{PL}_1 & \tilde{TB}_1 & \tilde{AB}_1 \\ \tilde{UR}_2 & \tilde{D}_2 & \tilde{J}_2 & \tilde{PL}_2 & \tilde{TB}_2 & \tilde{AB}_2 \\ \tilde{UR}_3 & \tilde{D}_3 & \tilde{J}_3 & \tilde{PL}_3 & \tilde{TB}_3 & \tilde{AB}_3 \\ \tilde{UR}_4 & \tilde{D}_4 & \tilde{J}_4 & \tilde{PL}_4 & \tilde{TB}_4 & \tilde{AB}_4 \end{bmatrix} < -L_1$$

(2)

The next step is to construct the weighted normalized DM: it cannot be assumed that each evaluation criterion is of equal importance because the evaluation criteria have various meanings. As we are emphasizing over voice communication (an outbound call in the presented use-case), the Delay and Jitter are given more importance to meet QoS requirements and hence, the weights corresponding to each attribute in the DM are chosen according to the context. The available bandwidth is coupled with the user profile loaded from the profile base (in the case of a gold profile, it is highly desirable to choose the link with good available bandwidth so $AB$ and $U$ will also be given appropriate weight values). Business objectives of the platform and the preconfigured configurations over the architecture are pivotal indicators for suitable weight values. These assigned weights illustrate the relative importance of each attribute in the DM such that:

$$W = W_{UR} + W_D + W_J + W_{PL} + W_{TB} + W_{AB} = 1$$

(3)

The corresponding weighted normalized entities in the DM are represented by subscript $wn$ (e.g. for $UR$ will be $UR_{wn}$). Now positive and negative ideal solutions for each attribute are computed: the positive ideal solution indicates the most preferable alternative and the negative ideal solution indicates the least preferable alternative as follows (e.g. voice/video link Utilization Ratio, $UR$):

$$UR^+ = \left(\frac{Max\ (UR_{wn})}{\| Min\ (UR_{wn})\} \right), i = 1, 2, 3, 4$$

(4)

and

$$UR^- = \left(\frac{Min\ (UR_{wn})}{\| Max\ (UR_{wn})\} \right), i = 1, 2, 3, 4$$

(5)

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

$$s_i^+ = \sqrt{\left(\frac{(UR_{wn})_1 - UR^+}{(PL_{wn})_1 - PL^+} + \frac{(UR_{wn})_2 - UR^+}{(PL_{wn})_2 - PL^+} + \frac{(UR_{wn})_3 - UR^+}{(PL_{wn})_3 - PL^+} + \frac{(UR_{wn})_4 - UR^+}{(PL_{wn})_4 - PL^+} \right)^2 + \frac{(UR_{wn})_1 - UR^+}{(AB_{wn})_1 - AB^+} + \frac{(UR_{wn})_2 - UR^+}{(AB_{wn})_2 - AB^+} + \frac{(UR_{wn})_3 - UR^+}{(AB_{wn})_3 - AB^+} + \frac{(UR_{wn})_4 - UR^+}{(AB_{wn})_4 - AB^+} \right)}$$

(6)

and

$$s_i^- = \sqrt{\left(\frac{(UR_{wn})_1 - UR^-}{(PL_{wn})_1 - PL^-} + \frac{(UR_{wn})_2 - UR^-}{(PL_{wn})_2 - PL^-} + \frac{(UR_{wn})_3 - UR^-}{(PL_{wn})_3 - PL^-} + \frac{(UR_{wn})_4 - UR^-}{(PL_{wn})_4 - PL^-} \right)^2 + \frac{(UR_{wn})_1 - UR^-}{(AB_{wn})_1 - AB^-} + \frac{(UR_{wn})_2 - UR^-}{(AB_{wn})_2 - AB^-} + \frac{(UR_{wn})_3 - UR^-}{(AB_{wn})_3 - AB^-} + \frac{(UR_{wn})_4 - UR^-}{(AB_{wn})_4 - AB^-} \right)}$$

(7)

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 DM) to the ideal solution $S^+$ by using the utility function as follows:

$$R_i = \frac{S_i^+}{S_i^+ + S_i^-}$$

(8)

The links $L_1$, $L_2$, $L_3$ and $L_4$ characterized by attributes voice/video Utilization Ratio $UR$, Delay ($D$), Jitter ($J$) Packet
V. Solution Validation

Dynamic SIP-based call routing decision-making at private-public network border under the control of Decision Engine is investigated. The validation and proof of the concept is accomplished by using the architecture shown in Fig. 1 with 4 links of 100Mbps each. The system offers the provision of on-the-fly and off-line decision computation depending on the chosen enforcement mode. SIPp [11] is used to generate extensive SIP requests (INVITE messages). It is a configurable traffic generator and is extensible via a simple XML configuration language. Call model with User Agent Client (UAC) at CS sends an INVITE to softswitch at CS that adds the resource priority tag and forwards the request to SBC. SBC analyzes the request to judge its communication type and pairs it with the resource priority tag sent by the CS. Resource priority type and communication type tuple is mapped to an appropriate user profile from central profile base at PS. 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 (decision can be computed on-the-fly by following the criteria in sections III-A and IV. Network Address Translation (NAT) is enabled at SBC and the decision is enforced during NAT implementation. Details about the design and development of the parser for embedding the calculated decisions during NATing are avoided due to space limitations. The remote UAC 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 a variable timespan by sending a BYE which is acknowledged by the SIP server with 200 OK. Wireshark [12] is used to capture the traffic at different interfaces (links). OriginLab [13] 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 SBC) as shown in Fig. 5. It is observed that there is a significant improvement in the throughput for each link with decision engine while performing SIP-based 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. 1 with the proposed decision engine has lower value than the ordinary SBC’s LB as shown in Fig. 6. The delay introduced by the system with and without Decision Engine is also calculated over the same test bed. The graph shown in Fig. 7 indicates that addition of Decision Engine in the system introduces a minor overhead (delay). This calculation is performed in outsourcing enforcement mode due to more dynamics involved in that particular mode (for detailed functionality, information sharing and communication between different devices over the presented architecture in outsourcing and provisioning modes respectively, the reader is referred to [5], [6]). The delay increases almost linearly as the number of calls/requests increases and is small enough having very little impact on services. It is due to the fact that the decisions are being executed and enforced during call/connection setup time.
VI. RELATED WORK

MCDM [14], [15], [16] and ontology [17], [18], [19] has been used for rule-based network management and control independently. To the best of our knowledge, there is no reported work that uses MCDM in coordination with ontology for dynamic network management. There are commercial and proprietary solutions available for multimedia traffic management at higher layers (above OSI transport 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 [20], [21]. Some products offer partial dynamicity with limited controls, while others are enforcing static decisions/rules. F5 networks [22] 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 with the provision of off-line (provisioning mode) and on the fly (outourcing mode) decision-making by integrating MCDM theory and ontology.

VII. CONCLUSIONS

QoS profile of the links, user authentication/authorization profiles, business objectives of the company, reciprocal SLAs with providers, technology specific and technology independent information over converged platform, fluent dynamics over the multihomed platform and traffic management issues at private-public network border constitutes a multi-disciplinary problem. The scalability of the service, control and network/transport planes in converged network environment cannot be guaranteed without semantic reasoning and introducing object oriented techniques while modifying/updating and/or adding new datasets. The information coming from different sources with different dimensions reflects the complexity of the underlying problem when a single decision has to be taken on the basis of multi-dimensional and multi-disciplinary information. A dynamic framework for converged multimedia network management is presented to overcome these issues.

MCDM theory is used to address the multi-criteria and multi-dimensional facet of the problem while the semantic variations and the reasoning are captured by applying conventional object-oriented concepts to ontologies. The proposed solution is tested for dynamic routing decision-making at the private-public network border (SIP-based multimedia traffic). The system supports two decision enforcement modes. Decisions are computed on-the-fly in outsourcing mode in order to validate the dynamism. Existing standards and mechanisms are used for decision enforcement without introducing overheads in the protocol stack. Throughput of the individual links improved significantly where the resources are being used efficiently and effectively at the cost of susceptible delay. Aggregated call dropping probability with the proposed Decision Engine has lower values than the SBC’s built-in load balancer for call routing.

Future work includes the interconnection of MCDM and conventional Policy Based Network Management through the development of an automated lingua franca in order to specify goals, criteria, alternatives and ontologies.

REFERENCES