ANALYSIS AND OPTIMIZATION OF RESOURCE CONTROL SCHEMES IN NEXT GENERATION NETWORKS

*Mostafa Safavi Hemami†, Mahmoud Pirhadi‡, Abbas Iravani Tabrizipoor**

†Amirkabir University of Technology, Tehran, Iran, msafavi@aut.ac.ir
‡ Islamic Azad University, Science and Research Branch, Tehran, Iran, pirhadi@ties.itu.int
* Iran Telecom Research Center (ITRC), iravani@itrc.ac.ir

ABSTRACT

Resource and Admission Control Function (RACF) is the main part of resource management architecture in Next Generation Networks (NGN). In this paper we analyze the effect of different resource control schemes in RACF architecture on some network performance parameters and show that a dynamic policy for scheme selection will outperform the static one. The dynamic policy results in better resource utilization for Best Effort traffic while providing the required resource for QoS-guaranteed traffic. Likewise, we present some suggestions about Service Control Functions (SCF) and RACF responsibilities so that NGN layers independence will be maintained more than before. This led us to a bit alteration in DIAMETER protocol used between RACF and SCF.

Keywords— Next Generation Networks, Quality of Service, Resource Control, RACF

1. INTRODUCTION

One of the critical issues in Next Generation Networks (NGN) development is resource management. This issue has forced clear-sighted organizations such as ITU and ETSI to propose some models and architectures for provision of resource management in NGN networks. An architecture which has been introduced by ITU-T for the sake of resource management is called RACF (Resource and Admission Control Function), whose simplified structure is indicated in Fig. 1 [1]. ETSI has also recommended a model for resource and admission control in NGN that is envisaged as an instance of ITU RACF for fixed access networks [2].

The Service Control Functions (SCF) represents an abstract notion of the functional entities in the service stratum of NGN, such as call servers, SIP proxies, etc. that requests the Quality of Service (QoS) resource and admission control for media flows of a given service via its interface to RACF.

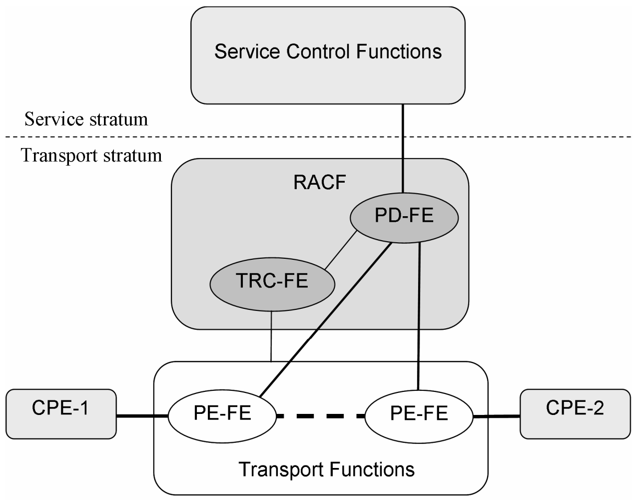

Fig. 1 Generic resource and admission control functional architecture in NGN

RACF acts as the mediator between SCF and transport functions for QoS-related transport resource control within the access and core networks. One of the basic functionalities of RACF is policy decisions based on transport resource status and utilization information, Service Level Agreements (SLA), network policy rules, and service priorities. The RACF applies control policies to transport resources, e.g., routers, upon SCF requests, determines whether transport resource is available, and makes admission decisions. RACF interacts with the transport functions to control the following tasks in the transport stratum from the QoS point of view: bandwidth reservation and allocation, traffic classification, traffic marking, traffic policing, and priority handling [1].

RACF consists of two resource and admission control functional entities, i.e. PD-FE (Policy Decision Functional Entity) and TRC-FE (Transport Resource Control Functional Entity). The main functionality of PD-FE and TRC-FE is to make policy decision and to determine network resources availability, respectively. Dividing RACF into two distinct functions enables it to support variant networks within a general resource control framework.

The PE-FE (Policy Enforcement Functional Entity) in the transport stratum is a packet-to-packet gateway at the boundary of different packet networks and/or between the CPE (Customer Premises Equipment) and access network. It is the key injection node to enforce dynamic QoS and resource control policies [1].

Since most of services to be delivered in NGN require a guaranteed QoS, there should be a method to provide QoS guarantee before service delivery to the customers. The Call Admission Control (CAC) mechanism is one of the most important mechanisms of RACF affecting the resource management efficiency and QoS guarantees provided to users.

In this paper, we introduce a dynamic policy that, in conjunction with CAC, can selectively use one of the different resource control schemes for optimum resource utilization regarding the status of the network traffic.

Section 2 introduces different resource control schemes with their usage in a generic simple call flow. The effect of each scheme on different parameters, i.e. call setup delay, resource availability, and blocking probability is investigated in section 3. Some suggestions are proposed for more efficient resource utilization in section 4; and the last section will conclude the paper.

2. RESOURCE CONTROL SCHEMES

The QoS resource control process consists of three logical states. These states can occur in one or more steps as described below:

Authorization: The QoS resource is authorized based on policy rules. The authorized QoS bounds maximum amount of resources that can be allocated to a specified user.

Reservation: The QoS resource is reserved based on the authorized resource and resource availability. The reserved resource can be used by best effort media flows when the resource has not yet committed in the transport functions.

Commitment: The QoS resource is committed for the requested media flows when the gate is opened and other admission decisions (e.g., bandwidth allocation) are enforced in the transport functions.

According to the diversity of application characteristics and performance requirements, the RACF supports three different schemes of resource control:

Single-phase scheme: Authorization, reservation and commitment are performed in a single step. The requested resource is immediately committed upon successful authorization and reservation.

Two-phase scheme: Authorization and reservation are performed in one step, followed by commitment in another step. Alternatively authorization is performed in one step, followed by reservation and commitment in another step.

Three-phase scheme: Authorization, reservation and commitment are performed in three steps sequentially [1].

The paper focuses on the first two schemes and their influences on network performance along with some suggestions for improvement.

Fig. 2 illustrates two simplified signaling call flows for session establishment between two end users. These call flows are indicated in two different schemes of resource control, i.e. single-phase and two-phase based on the indicated architecture in Fig. 1. In this architecture SCF is assumed to be a SIP proxy server and hence it uses SIP [3] to communicate with the CPEs. Q.3301 [4] (DIAMETER [5]) is used between SCF and RACF and COPS-PR [6] is used between RACF and the transport layer elements according to ITU-T recommendations.

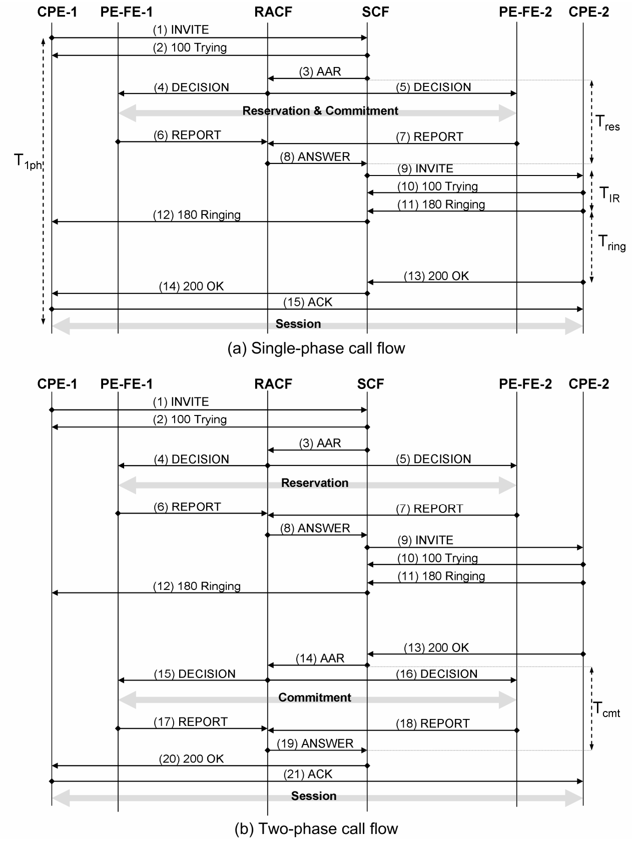

Fig. 2 Signaling call flow for (a) single-phase scheme and (b) two-phase scheme

Fig. 2-a depicts single-phase scheme signaling call flow from the instant of service request to that of session establishment. The stages of this call flow are as follows:

- Stages 1 & 2: Service is requested by CPE-1 from SCF. When the request is received, reservation and commitment steps are initiated. SCF does not forward the *INVITE* message to CPE-2 before the end of reservation and commitment steps.
- Stage 3: Resource reservation and commitment are requested by SCF from RACF. The request is based on the DIAMETER protocol and is sent through the *AAR* command. According to Q.3301 this request should have the *Resource-Reservation-Mode=2* option that means authorization, reservation, and commitment steps should be performed in a single step [4].
- Stages 4 & 5: *DECISION* command is issued from RACF to PE-FEs for resource reservation and commitment in transport layer in a bidirectional path and is based on COPS-PR protocol [6].

- For the sake of simplicity extra COPS-PR and RSVP messages are not indicated.
- Stages 6 & 7: *REPORT* messages are reported from PE-FEs to RACF which means resource reservation and commitment have been performed successfully [6].
- Stage 8: Receiving *REPORT* messages from both PE-FEs, RACF answers to SCF which means resource reservation and commitment have been performed successfully [4], [5].
- Stage 9: *INVITE* request is forwarded to CPE-2.
- Stages 10-15: These stages relates to the session establishment with respect to the RFC 3261 [3].

Fig. 2-b depicts the call flow of two-phase scheme from the instant of service request to that of session establishment.

The differences between two call flows are as follows:

- Stage 3: Resource reservation is requested by SCF from RACF. The request is based on the DIAMETER protocol and is sent through the *AAR* command. According to Q.3301 this request should have the *Resource-Reservation-Mode=1* option that means only authorization and reservation steps should be performed in a single step [4].
- Stage 14: Resource commitment is requested by SCF from RACF. The request is based on the DIAMETER protocol and is sent through the *AAR* command. According to Q.3301 this request should have the *Resource-Reservation-Mode=3* option that means only commitment step should be performed [4]. This request will be sent when the message *200 OK* is received from CPE-2 to SCF (Off-hook state).

The next section will investigate the effect of each resource control scheme on network performance according to the exampled call flows.

2. EFFECT OF RESOURCE CONTROL SCHEMES ON NETWORK PERFORMANCE

According to ITU-T recommendations, Call Setup Delay (CSD) is one of the important characteristics of QoS. CSD is the total call establishment time regardless of the delay associated with the called party answer to the incoming call [7].

CSD for single-phase scheme call flow can be obtained by means of the following equation according to Fig. 2-a.

$$CSD_{1ph} = T_{1ph} - (T_{IR} + T_{ring}) \qquad (1)$$

In which T_{1ph} is the duration between sending the *INVITE* request and receiving the *200 OK* message by CPE-1, T_{IR} is the duration between receiving the *INVITE* request and sending *180 Ringing* message by CPE-2, and T_{ring} is the average time required for answering to the incoming call.

CSD for two-phase scheme call flow can be obtained by means of the following equation according to Fig. 2-b:

$$CSD_{2ph} = CSD_{1ph} + T_{cmt} \qquad (2)$$

Where T_{cmt} is the required signaling time for committing the reserved resources.

Obviously, call setup delay of two-phase scheme is higher than that of single-phase, since reservation and commitment are handled in two phases. However, CSD is only one parameter that influences the quality of service. The major issue that should be considered in scheme selection is, in fact, optimum resource utilization in different network statuses.

Network resources can be utilized for both best effort traffic and QoS-guaranteed traffic. The resources that have not been committed yet, can be used for best effort traffic even if they had been reserved before; but committed resources cannot be used for other traffic flows. Accordingly, commitment would better be postponed as much as possible in order to utilize the resources optimally. This means that using the two-phase scheme can result in better resource utilization for best effort traffic.

In order to compare best effort resource availability in single-phase and two-phase schemes the steps below can be followed:

Assuming that T_{call} is the average call holding time, T_{res} is the signaling time for resource reservation, and $T_{use-1ph}$ is the time during which the resources are in use in the single-phase scheme, the following relationship holds.

$$T_{use-1ph} = \frac{T_{res}}{2} + T_{IR} + T_{ring} + T_{call} \qquad (3)$$

By the same way the time during which the resources are in use in the two-phase scheme ($T_{use-2ph}$) equals to the following value:

$$T_{use-2ph} = \frac{T_{cmt}}{2} + T_{call} \qquad (4)$$

The ratio of in-use time of two-phase to single-phase shown by α, is a measure of resource availability for best effort traffic.

$$\alpha = \frac{T_{cmt}/2 + T_{call}}{T_{res}/2 + T_{IR} + T_{ring} + T_{call}} \qquad (5)$$

T_{ring} is constituted of two parts, one part is the mean time to answer (T_{MTA}) and the other one is the ring time limit (T_{RTL}). T_{MTA} relates to a situation in which the called party answers to the incoming call while T_{RTL} relates to when the called party does not answer. Assuming that p represents the probability of call answering by the called party, the following equation holds:

$$T_{ring} = T_{MTA}.p + T_{RTL}.(1-p) \qquad (6)$$

Replacing (6) in (5) the following equation is achieved:

$$\alpha = \frac{T_{cmt}/2 + T_{call}}{T_{res}/2 + T_{IR} + T_{MTA}.p + T_{RTL}.(1-p) + T_{call}} \qquad (7)$$

Fig. 3 demonstrates α versus average call holding time assuming that $T_{MTA} = 6$ *sec*, $T_{RTL} = 60$ *sec*, $T_{cmt} = T_{res} = 1$ *sec*, and $T_{IR} = 0.1$ *sec*, considering different values for p.

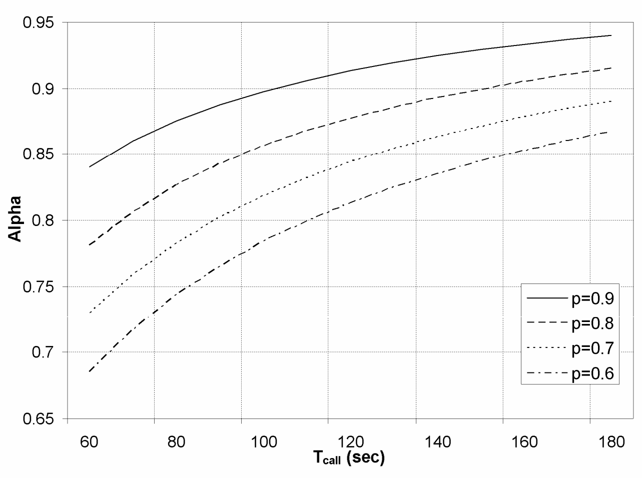

Fig. 3 The ratio of in-use time of two-phase to single-phase schemes (α) versus mean call holding time

As the figure shows, α is always less than one, which means that network resources are more available for best effort traffic in two-phase scheme. Also it can be seen that increasing call holding time will increase α and the difference between two schemes in terms of their effect on resource availability for best effort traffic will decrease.

Another important QoS parameter is blocking probability that should be considered in both single-phase and two-phase scheme. In order to investigate blocking probability the following equation can be used:

$$P_B = \frac{\rho^N / N!}{\sum_{n=0}^{N} \rho^n / n!} \tag{8}$$

In this equation N is network capacity, i.e. the maximum number of sessions that can be established in the network before any blockage and $\rho=\lambda/\mu$ where λ deals with request arrival rate, and μ implies the service rate, i.e. $1/\mu$ states the total time during which resources are reserved. This value equals to $T_{use\text{-}1ph}$ for the single-phase scheme and $T_{use\text{-}1ph} + T_{cmt}$ for the two-phase. Therefore blocking probability in the two-phase is more than that of the single-phase.

Fig. 4 indicates blocking probability ratio of two-phase to single-phase versus arrival rate with $N=100$, and $T_{cmt}=1\ sec$, considering different values for $T_{use\text{-}1ph}$.

As the figure shows blocking probability ratio is close to unity and increasing the arrival rate the ratio gets closer to one, i.e. the behavior of both schemes will be approximately the same. This means that T_{cmt} does not have a sensible effect on blocking probability and can be neglected.

The discussion concludes that although CSD in the two-phase is higher than that of the single-phase, two-phase will result in better resource availability for best effort traffic. On the other hand two-phase and single-phase behave nearly the same in terms of blocking probability. These all lead us to a dynamic approach for scheme selection to provide the best possible situation, i.e. the best trade-off between all important parameters.

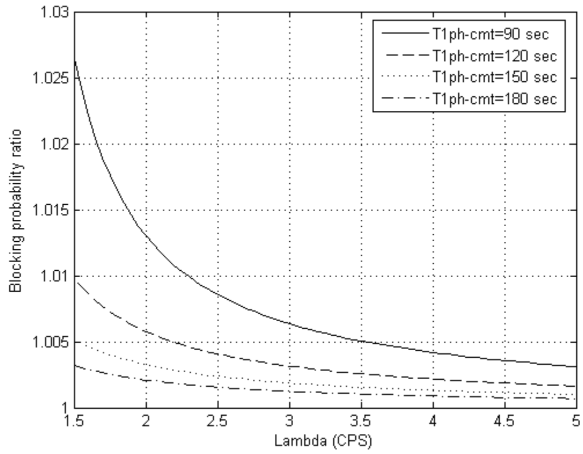

Fig. 4 Blocking probability ratio of two-phase to single-phase versus arrival rate (λ)

4. APPLYING DYNAMIC POLICY FOR SCHEME SELECTION

PD-FE is the major functional entity for QoS policing and decision-making in RACF architecture. Resource control schemes have been considered statically in the existing RACF architecture. However, according to our discussion it was proved that a dynamic policy would outperform a static one. Dynamic scheme selection requires information exchange between TRC-FE and PD-FE so that PD-FE can make its decisions regarding network traffic status.

The dynamic policy can be handled in different ways. For example, two threshold values, i.e. high and low, can be considered for the amount of network traffic. In high-load traffic the scheme is two-phase. If the load falls below the low-threshold the scheme will convert to single-phase. Likewise in normal-load traffic the scheme is single-phase. If the load exceeds the high-threshold the scheme will convert to two-phase.

The second suggestion relates to RACF and SCF responsibilities. In the ITU recommended architecture resource control scheme is determined by SCF considering some predefined policies. According to Q.3301 reservation and/or commitment request is sent from SCF to RACF via the *AAR (Resource-Reservation-Mode=0-4)* message in which *Resource-Reservation-Mode* value determines the scheme [4]. In fact SCF is responsible for scheme selection policy and RACF can only accept or reject the received request. This contradicts the main purpose of NGN layered architecture, i.e. layers independency and can cause some interoperability problems. Hence it is proposed that resource control scheme selection should be the RACF's responsibility.

Effecting the proposal requires some modifications in protocols between different network elements. These modifications mainly relate to the protocol used between SCF and RACF, i.e. DIAMETER. In this case the Request message sent from SCF to RACF only indicates SCF's request for resource access and has nothing to do with the scheme determination. Receiving the request RACF selects the proper scheme according to network status information and applies it to the transport layer. Thus the Answer request that has been sent back from RACF to SCF should contain *Resource-Reservation-Mode* value to determine the scheme. Receiving the message SCF recognizes if it is required to send another Request for resource commitment. Consequently, SCF follows RACF in scheme selection.

5. CONCLUSIONS

In this paper we investigated two resource control schemes, i.e. single-phase and two-phase, and their effect on different network parameters. Our investigation showed that the two-phase scheme has a higher call setup delay than single-phase. However, two-phase has better resource availability for best effort traffic, although it has a higher blocking probability for normal-load traffic. According to the results we found out that a dynamic approach for resource control scheme selection will have a better network performance than a static one. We also suggested a modification for RACF structure and consequently, a little revision in DIAMETER protocol implementation.

REFERENCES

[1] ITU Rec. Y.2111, "Resource and admission control functions in Next Generation Networks (Version 0.2)," Jan. 2007.

[2] ETSI ES 282 003 v1.1.1, "Resource and Admission Control Sub-system (RACS); Functional Architecture," June 2006.

[3] J. Rosenberg et al. "SIP: Session Initiation Protocol," RFC 3261, June 2002.

[4] ITU Draft Q.3301.1, "Resource control protocol no. 1 (rcp1) Protocol at the interface between service control entities and the Policy Decision Physical Entity (PD-PE)," Nov. 2006.

[5] P. Calhoun et al., "Diameter Base Protocol," RFC 3588, Sept. 2003.

[6] K. Chan et al., "COPS Usage for Policy Provisioning (COPS-PR)," RFC 3084, Mar. 2001.

[7] ITU Rec. Y.1530, "Call processing performance for voice service in hybrid IP networks," May 2004.