Enhancing the QoS and Resource Management Aspects of the 3GPP IMS Emergency Service Architecture

May El Barachi
Concordia University, Montreal, Canada

Roch Glitho
Ericsson and Concordia University, Montreal, Canada

Rachida Dssouli
Concordia University, Montreal, Canada

***Abstract* --- Emergency sessions are the most fundamental and critical services offered by telecommunications networks. They require preferential treatment over regular sessions and this is achieved with QoS and resource management techniques. An IMS emergency service architecture has been recently proposed by 3GPP. However, this solution only provides preferential treatment to public-initiated emergency communications. It does not offer any special treatment to mission critical calls, PSAP callbacks, and urgent communications among citizens. Furthermore, the preferential treatment offered does not provide public-initiated emergency sessions with a prioritized access to the resources at the transport layer. This may lead to unacceptable emergency sessions' setup delays and even emergency sessions failures when there is a strong contention for transport resources. This paper tackles these issues by extending the existing 3GPP architecture. We introduce new QoS profiles for emergency sessions, and propose mechanisms for supporting them. We have also built a proof of concept prototype. Its main features are presented.**

***Keywords* — Emergency services, IMS-based mobile networks, QoS, resource management, prioritized call handling**

I. INTRODUCTION

Emergency services represent one of the fundamental and most valued services provided by communication networks. They enable the public to summon help in case of emergency, and the emergency service agencies to respond quickly in order to minimize life and property losses. There are four main categories of emergency communications: citizen to authority (used by the public to report problems and ask for help); authority to authority (used by authorities for coordinating efforts during emergency/disaster relief and mitigation operations); authority to citizen (used by government agencies to notify the public when disasters occur); and citizen to citizen (used by the public to learn the state of relatives and property in case of major events) [1]. All these categories rely on synchronous (or session-based) communications, except the third category that relies on broadcasting. This paper focuses on session-based emergency services.

Due to the importance of emergency calls, they need to be provided preferential treatment over regular calls (e.g. faster call setup times and higher probability of completion) and prioritized access to resources. This is especially important when there is a strong contention for scarce network resources.

The IP multimedia subsystem (IMS) is a key component of IP-based mobile networks [2]. It consists of an overlay control layer that enables the seamless delivery of IP multimedia services via the packet switched domain. An IMS emergency service architecture has been recently proposed by 3GPP [3]. However, it has several limitations in terms of prioritized emergency call handling. For instance, it only provides preferential treatment to the first category of emergency services. Furthermore, the preferential treatment offered does not provide emergency sessions with a prioritized access to the resources at the IP transport layer and this may lead to excessive session setup delays or even emergency sessions failure, in times of network congestion.

This paper enhances the 3GPP IMS emergency service architecture by proposing new QoS profiles for emergency sessions (reflecting their needs in terms of QoS parameters), and the mechanisms for realizing them. The implementation of a proof-of-concept prototype is also described. Our enhanced architecture, unlike the existing architecture, provides preferential treatment to all session-based emergency services. Furthermore, it prioritizes emergency calls when it comes to accessing the IP transport layer resources. The next section presents the 3GPP IMS emergency service architecture and identifies its limitations. The enhanced architecture is then discussed, followed by a presentation of the prototype and the conclusions.

II. THE 3GPP IMS EMERGENCY SERVICE ARCHITECTURE AND ITS LIMITATIONS.

In this section, we first introduce the requirements, then introduce the existing IMS emergency service architecture along with its shortcomings.

2.1 Requirements

The first and most important requirement is that the solution should provide preferential treatment to all categories of emergency communications, by prioritizing the access of emergency sessions to both signaling and transport level resources. Secondly, resources must be dynamically allocated to sessions according to their priorities, in order to achieve efficient resource utilization. Third, the system must be able to adapt its resource allocation strategy in order to continue to support emergency communications even during high load or crisis situations. Finally, the system should provide preferential treatment at the beginning and during sessions, in order to guarantee the consistency of the treatment offered.

2.2 Overview of the IMS emergency service architecture

Figure 1 illustrates the IMS emergency service architecture [3]. This architecture relies on four main functional entities: the UE, the P-CSCF, the E-CSCF, and the LRF. The P-CSCF (proxy-CSCF) and the E-CSCF (emergency-CSCF) are SIP

1-4244-1457-1/08/$25.00 © IEEE

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE CCNC 2008 proceedings.

servers, which handle different aspects of emergency sessions' establishment/termination. The P-CSCF is the first point of contact between the UE (user equipment) and the network. It acts as inbound/outbound proxy, and performs authentication/ authorization, emergency session prioritization, and application level routing to the appropriate E-CSCF. The E-CSCF is responsible for acquiring/validating the location of the UE (by interacting with the LRF) and routing the call to the appropriate (IP-enabled or legacy) PSAP (public safety answering point). The LRF (location retrieval function) interacts with location severs and/or the access network to retrieve the location of the UE that has initiated the session. It may also provide PSAP route determination services.

The UE is the equipment used to make the emergency call. It detects that an emergency session is being established (based on the number dialed), registers with the IMS using a special emergency public user ID, determines its own location if possible (either using an internal location measurement mechanism or by interacting with the access network), and sends an emergency session establishment request to the P-CSCF with the needed information (e.g. the emergency public user ID and the location information). The P-CSCF then performs authorization/ authentication of the session and the user, prioritizes the emergency session, and forwards the session establishment request to an E-CSCF in the same network. If the location information provided by the UE is insufficient (i.e. missing or not accurate), the E-CSCF interacts with the LRF to acquire/validate the information. After that, the E-CSCF determines the address of an appropriate PSAP (based on this information), and routes the call to this PSAP, thus completing the call establishment. It should be noted that the PSAP is the only entity that can terminate the call - if the caller hangs up or gets disconnected, the PSAP operator initiates a callback to re-establish the call.

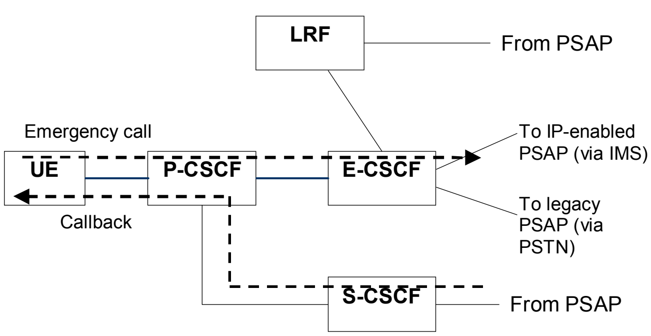

Figure 1. The IMS emergency service architecture

2.3 Limitations

In the previously described architecture, the P-CSCF detects the establishment of an emergency session and gives it priority over regular sessions, before passing it to a specialized signaling component (the E-CSCF) which routes it to the appropriate PSAP. This solution provides preferential treatment to the first category of emergency services (i.e. public to authority communications) only. As shown in figure 1, only public-initiated emergency calls (e.g. 911 calls) are recognized and prioritized by the P-CSCF (over regular calls), and have access to dedicated E-CSCF resources. This is not the case for PSAP callbacks, mission critical calls, and urgent calls between citizens, which are not recognized by the network as emergency communications, and are treated as regular calls, traversing the regular signaling path (i.e. going through the caller and the callee's P-CSCFs and S-CSCFs) without any preferential treatment. The second limitation is that the preferential treatment offered to public-initiated emergency calls guarantees prioritized access to SIP servers' resources (i.e. P-CSCFs and E-CSCFs resources), but not to transport level resources. In fact, all conversational traffic (including both regular and emergency calls) receives the same treatment in the IP-transport layer, since the 3GPP QoS solution [4] does not distinguish sessions within the same traffic class. Therefore, in the case of overload or network congestion, emergency calls related messages may experience too much delay or even get dropped. Finally, this solution offers preferential treatment at the beginning of sessions only (with no QoS guarantees during sessions).

III. An enhanced IMS emergency solution for mobile networks

The solution we propose consists of two new QoS profiles for emergency sessions, along with an architectural framework and resource management techniques and policies to support these profiles in the IMS. The QoS profiles represent the guarantees required by different types of emergency communication services on certain QoS parameters, while the architectural framework and the resource management techniques provide the means to offer those guarantees in the network. In the coming sub-sections, we start by discussing the QoS profiles, and then present the architecture.

3.1 QoS profiles for emergency sessions

We have previously proposed a call differentiation scheme for 3G networks [5]. This scheme enables the definition of various categories of calls, with different QoS profiles. Three QoS profiles were defined for regular calls as examples (silver, gold, and platinum). We now define two new QoS profiles for emergency sessions. Before presenting these profiles, we review the differentiating factors used to distinguish between the different categories.

A. The differentiating factors

Five differentiating factors are used to distinguish between the different classes, namely: the call blocking probability (CBP); the forced call termination probability (FCTP); the multiparty session ability to grow; the media type guarantee; and the user perceived quality [5]. The *CBP* is the probability that a new call is blocked and not allowed admission to the core network, while the *FCTP* is the probability that an ongoing call is terminated by the network. The *Multiparty session ability to grow* represents the limit a multiparty session can reach in terms of number of participants. As for the *Media type guarantee*, it represents the ability to sustain a call with a certain media type, without downgrade to another type (e.g. dropping from video to voice). Finally, the last factor (*User perceived media quality)* is the quality of the media transmission as perceived by the user, which could vary with network conditions or be sustained by the system.

B. The QoS profiles

Our definition of the new QoS profiles is based on the needs of emergency communication services, in terms of QoS guarantees. For the first category of emergency communications (i.e. public to authority), the CBP and the FCTP should be as low as possible to guarantee a high probability of call completion and a low probability of call interruption. A guarantee on the media type used (either audio or video) is also needed in order to avoid affecting the communication quality. Furthermore, the user perceived quality should be sustained to guarantee intelligibility. As for the session size, it is rather limited in this case (the session potentially including the personnel involved in a limited rescue operation). The second category of emergency communications (i.e. authority to authority) has similar needs, except for the session size required. In this case, the size could be large depending on the size of the mission and the involved rescue teams (e.g. national authorities, international organizations, and non profit organizations). The last category of emergency calls (i.e. citizen to citizen) can be considered as "urgent" regular calls, and therefore could be supported using the highest profile defined for regular calls (i.e. the platinum profile).

Based on this analysis, we define two possible profiles for the first two categories of emergency communications (i.e. the emergency-public and the emergency-authority profiles). These profiles are presented in table 1, along with the three profiles previously defined for regular calls (i.e. silver, gold, and platinum).

TABLE I
THE DIFFERENT QOS PROFILES

Factor / Class	CBP				FCTP				Multiparty session ability to grow		Media type guarantee		User-perceived media quality	
	H	M	L	Nil	H	M	L	Nil	Ltd.	Ultd.	GA	GV	Var.	Sus.
Silver	•				•				• (L1)		•		•	
Gold		•				•			• (L2)		•			•
Platinum			•				•			•	•	•		•
Emergency-public				•				•	• (L3)		•	•		•
Emergency-authority				•				•		•	•	•		•

It should be noted that the first three classes (silver, gold, and platinum) are related to the user subscription. This last should indicate the highest subscription class (HSC). Any class up to and including the HSC can be chosen by the user on a per call basis, and the selected class can be changed by the user during the session. The fourth class (emergency-public) is special, as it is not related to the user subscription. In fact, it can be used even by non-subscribed users. Furthermore, calls made using an emergency dial-string (e.g. 911) should be automatically mapped to the emergency-public class identifier. As for the fifth class (emergency-authority), it is also related to the user subscription, but the subscription for this class should be reserved for persons in NS/EP (national security/emergency preparedness) leadership positions (e.g. emergency centers coordinators, senior command levels of law enforcement, fire and public safety functions…etc). This last class could also be used for PSAP callbacks.

3.2 An extended IMS emergency service architecture

We use the same architectural framework as in our previous work [5] but have extended it significantly to cater to emergency calls. Our previous architecture introduced two new functional entities to the standard IMS architecture, namely: the session prioritization function (SPF) and the context information base (CIB). The CIB is a support entity which is responsible of the management of the contextual information needed. The SPF makes resource allocation/re-allocation decisions, taking into consideration the contextual information it gets from the CIB, the sessions' QoS profiles, and the resource management policies. Two new interfaces were introduced in this architecture: the Pi and the Pa interfaces. The SIP event notification framework [6] was used on the Pi interface for the exchange of contextual information between different entities, while COPS [7] was used on the Pa interface for the exchange of policy-based resource allocation decisions.

In order to handle emergency calls, we now introduce an additional Pa interface between the SPF and the E-CSCF, and an additional Pi interface between the CIB and the LRF. Furthermore, we make enhancements to the UE, the E-CSCF, and the LRF. The UE is enhanced with the ability to map public-initiated emergency calls to the appropriate category. If the emergency session is not detected by the UE, this mapping should be performed by the P-CSCF. As for the E-CSCF, it is enhanced with the ability to communicate with the SPF for resource allocation decisions and the ability to receive triggers (concerning the re-negotiation of emergency session parameters) and take the necessary actions. The LRF is enhanced with the ability to interact with the CIB, in order to obtain location information. The extended architecture is depicted in figure 2.

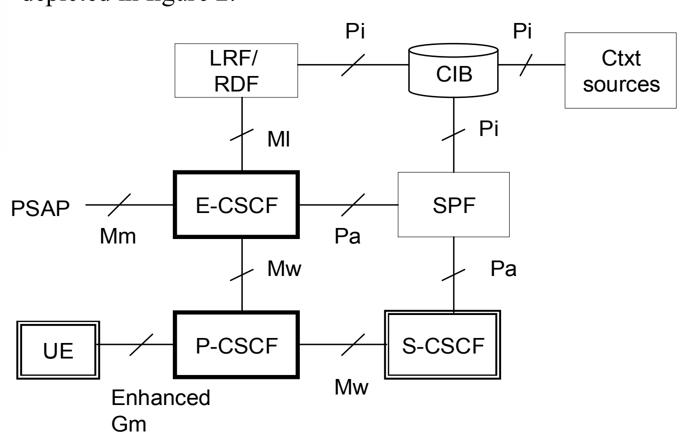

Figure 2. The extended IMS emergency service architecture

For public initiated emergency calls (e.g. 911 calls), session prioritization is achieved as follows: when the user initiates

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE CCNC 2008 proceedings.

the emergency call, the UE detects this, maps the call to the emergency-public service class, and forwards the session initiation request (including the service class) to the P-CSCF. The P-CSCF sends the request to the E-CSCF, which communicates with the SPF in order to allocate resources to the call. If resources are available, the SPF renders a positive decision and the call is established normally. If no resources are available, the SPF triggers one or several S-CSCFs to downgrade and/or preempt one or more ongoing (regular) calls, in order to free resources for the emergency call, which is admitted afterwards. After session establishment, the SPF triggers the E-CSCF to re-negotiate the session parameters in order to sustain its user perceived quality. Furthermore, a limit is imposed on the session size in case of lack of resources. Meanwhile, the LRF may consider the CIB as location server and interact with it to obtain more accurate location information, or location information obtained from alternate sources (e.g. wireless sensor networks).

For mission critical calls (i.e. calls initiated by authorities), emergency callbacks (initiated by PSAP operators), and regular calls, a similar procedure is followed, except that the call goes through a S-CSCF that communicates with the SPF for resource allocation decisions. In these cases, the call category is explicitly chosen by the user when the session is established. Furthermore, in the case of regular calls, depending on the call category and its CBP, the call may be rejected if there are not enough resources and no additional resources can be freed. This is not the case for the two categories of emergency calls, which have a CBP of zero.

As shown, whether the call is a regular call or an emergency call initiated by a regular user; a mission critical user; or a PSAP operator, it receives the appropriate treatment due to the interaction of E-CSCFs and S-CSCFs with the SPF, which dynamically allocates resources to sessions, based on their QoS profiles. It is important to note that achieving prioritization at the control level implicitly leads to prioritized access to transport level resources. In fact, blocking, downgrading, or terminating low priority sessions at the signaling phase directly impacts the amount of traffic that they generate (or may have generated), thus freeing/reserving those resources for more important sessions. This signaling level scheme achieves a finer level of granularity in terms of prioritization, than what is achieved by the existing network level scheme [4]. Furthermore, it enables more control over the different communication aspects that can be used as differentiating factors.

A. Session management scenario

Figure 3 illustrates the case where an ongoing regular session (a session between UE2 and UE3) must be terminated in order to free resources for an emergency session (initiated by UE1) to be established. In this scenario, we assume that UE1 has already registered with the IMS and that the destination PSAP is IP-enabled. The scenario begins when the user operating UE1 attempts to establish an emergency call. UE1 then sends a SIP INVITE message with a resource-priority header set according to the session category (Q735.0 in this case) to the P-CSCF assigned to the user (i.e. P-CSCF1). P-CSCF1 ignores the resource-priority header and forwards the request to an E-CSCF within the same network. The E-CSCF sends the SPF a call admission request using a COPS REQ message (including the session info). In this case, the SPF determines that an ongoing session must be terminated in order to free resources, and sends a “trigger termination” decision to S-CSCF1, in order to modify the decision made about the (previously admitted) session between UE2 and UE3. After receiving this decision, S-CSCF1 sends a SIP REFER message instructing UE2 to contact UE3 in order to terminate the session established between them. UE2 then sends a SIP BYE message to UE3, containing a reason header that indicates “hard preemption” as reason for the termination. After the session is terminated, UE2 sends a notification to its S-CSCF, using a SIP NOTIFY message. This S-CSCF informs the SPF (using a COPS RPT message), which sends an “install” decision (using a COPS DEC message) to the E-CSCF to authorize the admission of the emergency call. The E-CSCF then carries the rest of the emergency session establishment procedure normally, and sends a COPS RPT message indicating that it has complied with the SPF decision.

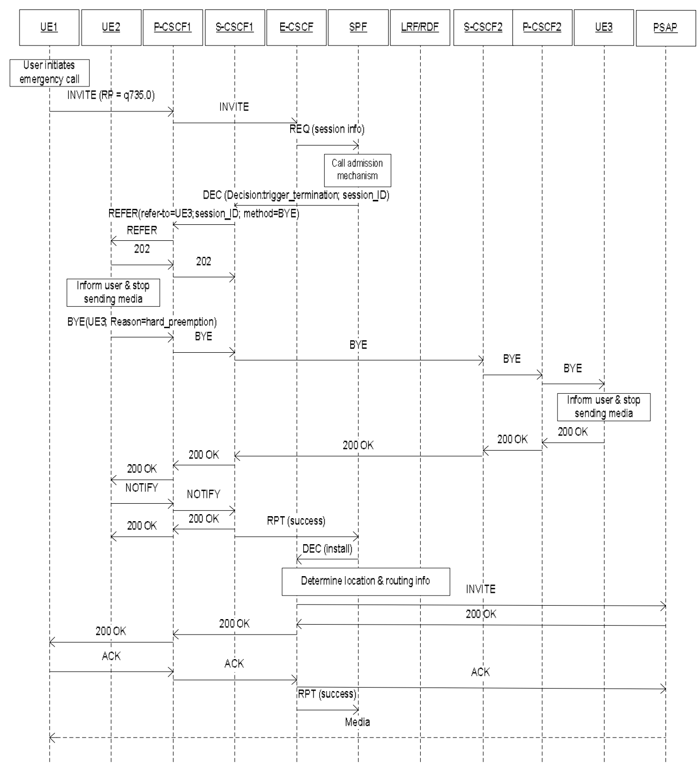

Figure 3. Successful emergency session establishment, after the termination on an ongoing session

B. Resource management techniques

Two resource management techniques are used to enable preferential treatment at the beginning and during sessions – call admission control and media parameter control. We re-use the call admission control mechanism presented in [5], but take into consideration the case of emergency sessions. This mechanism implements a hybrid upper limit and preemptive policy, as follows: In light to regular loading conditions, upper limits are imposed on regular call classes (i.e. silver, gold, and platinum) so that their overloads do not affect emergency classes. As for the two emergency classes, no thresholds are imposed. In fact, if these classes exceed their engineered loads, they may use any additional capacity available. However, in high loading conditions or crisis situations, a more aggressive approach (i.e preemption) is used to adapt to this situation by transferring resources between classes (e.g. from regular classes to emergency ones) when needed.

As for media parameter control, it consists of modifying the characteristics of ongoing sessions (in terms of media format used) in response to important variations in the network capacity by triggering the re-negotiation of session parameters between the involved parties. As shown in table 1, the media parameter control mechanism is applied to all classes of calls, except the silver class. This mechanism operates as follows: When an important drop in the network capacity is detected, a format downgrade procedure is invoked to trigger the re-negotiation of the session parameters to a media format that suits the network situation (i.e., a format that consumes less resources). The goal of this procedure is to prevent degradation of the session performance in terms of user-perceived quality, reducing by the same token the session's load on the network. Upon a re-increase in the network capacity, a format upgrade procedure is invoked to restore the initial session characteristics. It should be noted that several degrees of downgrade/upgrade may be used, depending on fluctuations in the network situation and the media codecs supported by the terminals.

IV. PROOF-OF-CONCEPT PROTOTYPE

We extended the prototype presented in [5] to demonstrate the feasibility of the solution proposed. The previous prototype leveraged an existing IMS simulated environment (simulating a P-CSCF, a S-CSCF, an HSS, and an application server) and extended it with an SPF, a CIB, and a dummy context source. In this work, two new components (related to emergency sessions) were added to this prototype: a simplified E-CSCF (mainly implementing PSAP routing functionality) and a LRF. As shown in figure 4, the 3G environment was simulated using seven laptops and two PCs, forming a WLAN. The clients were running on laptops, while one of the PCs represented the IMS simulated environment and the other PC represented a PSAP. For simplicity, all clients were stationary and their location information was pre-configured in the dummy context source. Three pairs of clients were used to establish three regular calls with different categories (i.e. silver, gold, and platinum), and the context source was configured to simulate different loading conditions. Then, the remaining client and PC (representing the PSAP) were used to test the different scenarios related to a 911 session establishment. Three scenarios were successfully tested, namely: session initiation without attempt to control ongoing sessions, session initiation after downgrade on an ongoing session, and session initiation after termination of an ongoing session. The same tests were successfully repeated with the client and PC representing mission critical users, therefore demonstrating the applicability of the solution proposed to different categories of emergency communications.

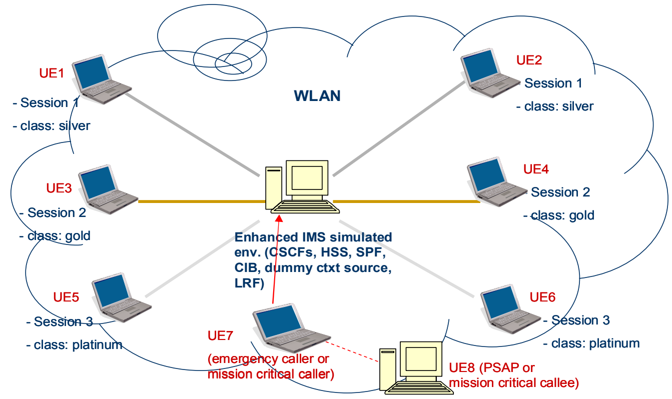

Figure 4. Prototype Setting

V. CONCLUSIONS

In this paper, we have proposed an enhanced emergency solution for IMS-based mobile networks. The enhancements focused on the provision of preferential treatment to all categories of emergency communications, by prioritizing their access to transport resources, in an efficient and adaptive manner. Two new QoS profiles were defined for emergency sessions, and an extension of the standard IMS emergency service architecture was proposed to support these profiles. The resource management techniques used were also discussed and the proof-of-concept prototype presented. The system proposed leverages contextual information to allocate resources efficiently and adapt to different network situations. Furthermore, the solution presented is applicable to all session-based categories of emergency calls (including public-initiated emergency calls, PSAP callbacks, mission critical calls, and calls among citizens during emergencies), and provides preferential treatment at the beginning and during sessions. Next steps consist of a detailed performance evaluation of the system and the enhancement of other aspects of emergency communications.

REFERENCES

[1] ETSI EMTEL – Emergency Telecommunications official website, http://www.emtel.etsi.org

[2] 3GPP TS 23.228, "IP multimedia subsystem (Release 7 – v.7.3.0)", March 2006

[3] 3GPP TS 23.167, "IMS emergency sessions (Release 7 – v.7.3.0)", December 2006

[4] 3GPP TS 123.107, "QoS concept and architecture", March 2006

[5] M. El Barachi, R. Glitho, and R. Dssouli, "Context-aware signaling for call differentiation in IMS-based 3G networks," IEEE ISCC 2007, Aveiro, Portugal, July 2007

[6] A. Roach, "Session Initiation Protocol (SIP) – Specific Event Notification", RFC 3265, June 2002

[7] D. Durham et al., "The COPS (Common Open Policy Service) Protocol", RFC 2748, January 2000

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE CCNC 2008 proceedings.