An exemplary management method includes receiving at a first network a notice of an intended communication to a called party network, the notice including a first priority indicator, the intended communication requiring a resource for supporting a streaming data protocol in each network between a calling party network and the called party network; forwarding the notice to a second network and toward the called party network, the notice including the first priority indicator; in parallel with said forwarding, mapping the first priority indicator to a second priority indicator and initiating for the intended communication a determination of resource availability for the first network based on the second priority indicator; determining the determination of resource availability for the first network based on the second priority indicator, wherein the determination is for a first resource for the first network; and verifying resource availability for the intended communication.
RPH MAPPING AND DEFAULTING BEHAVIOR

RELATED APPLICATION

[0001] This application claims the benefit of U.S. Provisional Application, “Method And Apparatus For End-To-End Capacity And Priority Management Through Multiple Packet Network Segments”, 61/188,364 filed Aug. 8, 200 concurrently herewith. The entire teachings of the above application are incorporated herein by reference.

FIELD OF THE INVENTION

[0002] The invention relates to the field of communication systems, and, more specifically, to a resource control and policy-based management mechanisms adapted to implement services including priority services.

BACKGROUND INFORMATION

[0003] Often, a call may go through several service provider networks in order to connect from a calling party to a called party. Each service provider’s network may include several domains (or administrative boundaries). In Time Division Multiplexed (TDM) Networks, resource clearance using signaling messages is supported through serial resource clearance on a link-by-link basis. In the serial (or sequential) case, no node transmits a message until first obtaining resource clearance within its domain. Accordingly, the total time required to set up a call or session with a link-by-link signaling method is O(t^2), where t represents the time it takes for resource clearance in the i-th link.

[0004] Priority services is the name given to preferred treatment of certain kinds of traffic (e.g., media/data and signaling) over other kinds of traffic in, for example, the context of civilian or defense communication networks. Priority in a public switched telephone network (PSTN) is performed on a switch by switch basis. Calls for priority users, such as government entities pass through switches while non-priority traffic is blocked.

[0005] Three kinds of priority services exist as present: namely, wireless priority service (WPS), Government emergency telecommunications services (GETS) by national communications systems (NCS) and multi-level pre-emption (MLPP) by the United States Department of Defense.

[0006] GETS provides an ability to preempt calls at a lower priority than MLPP. GETS also provides for buffer-type queuing of a call such that the call will be the first call to pass when a necessary resource becomes available. Call control functionality includes the decision of choosing which call to allow and which call to preempt. Resource control and policy-based management functionality includes the decision of when to preempt a call or when to buffer-type queue a call.

[0007] Within the context of voice, video and other data communications (e.g., multimedia communications and the like) via packet based networks (e.g., Ethernet, etc.) resource clearance is handled on a sequential basis. Methods for pre-emption and priority handling of voice calls and multimedia services are also lacking in packet based networks.

SUMMARY OF THE INVENTION

[0008] A network element, apparatus and method for combining call admission control and priority services within different type of networks is provided. A logical functional entity denoted as a Priority Services Functional Element (PSFE) enables administration of priority services and call or session admission control policies through interactions with Call Session Control Function Element (CSCF) and Resource Access Control Facilities (RACF). Method steps in exemplary embodiments may include verifying the priority levels, receiving resource availability information, storing call admission and resource allocation information, identifying path and link information required for an originated session, checking available resource at links against required resources for the call or session establishment, and communicating resource availability indication(s) (or lack thereof) based on admission control policies after resource verification.

[0009] The PSFE allows for capacity and priority management in parallel over all the links in the bearer path within a particular network. PSFEs in different networks can also communicate to enable PSFEs in each network in the path to perform resource allocation independently in parallel (i.e., undertaking apparent or actual performance of more than one operation at a time). Proposed methods allow both kinds of parallelization in networks of different architectures, such as IMS and SIP, and also across varied networks. The PSFE introduces obtaining parallel resource allocation even with network boundaries, including heterogeneous networks. In this manner, calls or sessions may be established efficiently, and varied features provided in a multimedia network.

[0010] PSFE may use Session Initiation Protocol (SIP) messages to communicate with Signaling and Service Layer entities, such as Proxy (P-) or Interrogating (I-) CSCFs, Subscriber Information Databases (e.g., Home Subscriber Server or HSS), and Application Servers (AS). PSFE may also communicate with Transport Layer entities (such as, RACF in an all IMP Multimedia Subsystems (IMS) architecture) using, for example, Diameter messages. However, PSFE may integrate both IMS and non-IMS Next Generation Service Provider (NGSP) networks and enable efficient resource verification and allocation in the non-IMS NGSP network. Further, the PSFE may also determine resource availability using open loop control or other accounting methods in a network without a RACF.

[0011] Next generation networks are not only expected to provide various multi-media features as different media are packetized and transported over the same IP network infrastructure, but also to provide wide-ranging interconnectivity among different networks. In many cases, an end-to-end call/session may traverse multiple networks based on the location of the end-users. Parallel resource clearance provided by a PSFE is useful in establishing various multi-media services in all IP-based networks regardless of the types of users and may be utilized in Enterprise, Commercial Service Provider, Civilian Government, and Military network segments of IP networks.

[0012] Further, the provided PSFE enables use of the priority assignment from an originating network or end-user globally in a trusted manner with private mapping of such priority assignments at each separate network for specific internal usage within each network. The priority assignment associated with a call/session is communicated across multiple networks so that the priority of the session is intact end-to-end. At each provider network, the priority assignment is mapped for internal use according to pre-defined rules. In this fashion, the PSFE provides a trusted method of
communicating and abiding by Resource Priority Headers (RPH) among different service providers (SP).

[0013] Although RPH may be provided between two networks currently, SPs may not support the RPH values received from other service providers for a host of reasons. For example, the receiving network simply may not be capable of providing the labeled priority; there may not be a business agreement concerning such support by the receiving network; the receiving network may not trust the priority indicated; and network policies of the receiving network may require temporary or long-term suspension of priority servicing as indicated in the received RPH, which effect is undesirable.

[0014] Apparatus and method are provided for capacity and priority management of a communications network. The provided apparatus and method may be used in a system having a plurality of networks. An exemplary method includes receiving at a first network a notice of an intended communication to a called party network, the notice of intended communication including a first priority indicator, with the intended communication requiring a resource for supporting a streaming data protocol in each network between a calling party network and the called party network. The notice of an intended communication is forwarded to a second network and toward the called party network. This forwarded notice of intended communication includes the first priority indicator. In parallel with said forwarding, the first priority indicator is mapped to a second priority indicator and a determination of resource availability for the first network is initiated based on the second priority indicator. Resource availability for the intended communication is then verified. For example, resource availability of resources necessary for the intended communication are not queried and reserved sequentially from calling to called party but query and reservation of necessary resources occur simultaneously or near simultaneously. Further, private mapping of the priority assignment of the originating network is utilized.

[0015] The mapping of the first priority indicator to the second priority indicator may be based on values negotiated via a key exchange protocol. The key exchange protocol may occur between the between the first network and the called party network. The mapping of the first priority indicator to the second priority indicator may be based on predetermined and default values. In one embodiment, the first priority indicator is included in a resource priority header.

[0016] In one embodiment, determining for the intended communication the determination of resource availability for the first network based on the second priority indicator includes reserving for the intended communication the first resource for the first network in the event the first resource is available for the intended communication. In another embodiment, determining the resource availability includes reserving for the intended communication the first resource for the first network in the event the first resource is available above a threshold level.

[0017] In one embodiment, the notice of intended communication includes a personal identification number, and an authentication of the intended communication is replayed based on the personal identification number. In a further embodiment, the mapping the first priority indicator to a second priority indicator maps the first priority indicator to a default priority indicator in the event a value for the personal identification number is unknown. Signaling for the determination of resource availability for the first network based on the second priority indicator may include both the first priority indicator and the second priority indicator.

[0018] Verifying resource availability for the intended communication may include receiving for the intended communication from a third network a second indication of resource availability; and the verifying may be based on the determination of resource availability for the first network and the second indication of resource availability. A re-invite for the intended communication may be forwarded in the event verifying resource availability for the intended communication indicates availability of resources from the calling party network to the called party network. The third network may be the second network in the event resource clearance is being fed back to the head end.

[0019] In one embodiment, an exemplary method includes receiving at a first network a notice of an intended communication to a called party network, the notice of intended communication including a first priority scheme, wherein the intended communication requires at least one resource for supporting a streaming data protocol between a calling party network and the called party network; mapping the first priority to a corresponding second priority for second priority scheme utilized in the first network; verifying resource availability of at least one resource in the first network based on the corresponding second priority; and in parallel with said verifying, forwarding the notice of an intended communication to a second network and toward the called party network prior to receiving an indication of resource availability for the at least one resource in the first network required for the intended communication.

[0020] The mapping of the first priority to the corresponding second priority may be based on a negotiation via a key exchange protocol between the first network and the calling party network. The mapping of the first priority to the corresponding second priority may be based on predetermined correspondences and on default correspondences. In one embodiment, the notice of intended communication includes a resource priority header.

[0021] One embodiment of the method includes receiving at the first network a second indication of resource availability of at least one resource at a third network required for the intended communication. The third network may be the second network when resource clearance is fed back to the head end. In another embodiment, a re-invite of the intended communication is forwarded to the second network and toward the called party network in the event the first indication and the second indication indicate resource availability for at least one resource in each network from the calling party network to the called party network.

[0022] In context of priority services, such a PSFE can provide higher priority to priority service users and queue such users’ requests for resource availability ahead of resource needs from other, routine users. Priority session requests and associated priority levels can be identified by the PSFE, for example, through Resource Priority Headers (RPH) in the SIP message. It is important to note that to queue priority sessions ahead of routine sessions, no routine session can be allowed in any of the links in the priority session path. Coordination of such activities across multiple network segments, which is quite complex among CSCFs, Back to Back user Agents (B2BUAs), and Session Border Controllers.
(SBCs) in separate network segments, can be achieved via communicating PSFEs deployed in these segments.

**BRIEF DESCRIPTION OF THE DRAWINGS**

[0023] The teachings of the present invention can be readily understood by considering the following detailed description in conjunction with the accompanying drawings, in which:

[0024] FIG. 1 depicts a network topology for facilitating calls through various networks according to an embodiment of the invention;

[0025] FIG. 2 depicts an overview of a network topology for facilitating calls through an exemplary network according to an embodiment of the invention; and

[0026] FIGS. 3a and 3b depict a flow diagram of an exemplary routine according to embodiments in accord with the invention.

[0027] To facilitate understanding, identical reference numerals have been used, where possible, to designate identical elements that are common to the figures.

**DETAILED DESCRIPTION**

[0028] The subject invention will be primarily described within the context of voice calls. However, since the subject invention is directed to the broader issues of multimedia services (including voice calls) references to calls and communications within this specification should be broadly construed to include voice, video, audio, email and other multimedia communications.

[0029] Specific structural and functional details disclosed herein are merely representative for purposes of describing example embodiments. Example embodiments may, however, be embodied in many alternate forms and should not be construed as limited to only the embodiments set forth herein.

[0030] The subject invention enables the establishment and management of VoIP traffic, including traffic of different priority levels, in a network (for example an Internet Protocol (IP) network) with parallel resource clearance and monitoring of criteria indicative of importance of a new voice call entering the network, network capability and instantaneous load. Accordingly, an exemplary telecommunications system is described as one potential environment in which a subject invention operates and exists.

[0031] Generally speaking, the invention assists in enabling the following system and network characteristics:

1. A profitable uniform architecture that addresses the requirements for both DoD and NCS;
2. A minimization in the overlap functionality in terms of development and maintenance of priority service features;
3. A minimization of the impact on existing architectures;
4. Addressing the requirements of IMS as well as non-IMS environments;
5. An ability to add on priority service at any time by a network provider;
6. An interoperability with network solutions of multiple vendors;
7. An ability to pass priority calls earlier than routine calls, when a congestion has been cleared and ability to maintain “current gets/wps voice functionality” which may require an ability to mimic priority signal queuing from a TDM environments;
8. An adaptation to IMS and non-IMS based Next Generation networks;
9. An ability to extend and support, at least architecturally, multimedia applications and mid-call features for priority services in the future; and
10. The providing of call admission control.

[0032] The present invention provides for a logical functional entity denoted as a priority services functional element (PSFE) that integrates with both IMS and non-IMS networks and environments. Within the context of network architectures adapted to utilize a PSFE, the PSFE is added to the architecture as an additional logical entity, such that the rest of the network architecture remains substantially unmodified. The PSFE interacts with the session initiation protocol (SIP) proxies, call session control functions (PCSCF), resource access control facilities (RACF), (NGSP) and other PSFE entities.

[0033] It will be understood that, although the terms first, second, etc. may be used herein to describe various elements, these elements should not be limited by these terms since such terms are only used to distinguish one element from another. For example, a first element could be termed a second element, and, similarly, a second element could be termed a first element, without departing from the scope of example embodiments. As used herein, the term “and” is utilized in the conjunctive and disjunctive senses and includes any and all combinations of one or more of the associated listed items, and the singular forms “a”, “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. Unless otherwise defined, all terms (including technical and scientific terms) used herein have the same meaning as commonly understood by one of ordinary skill in the art to which example embodiments belongs. It will be further understood that terms, such as those defined in commonly used dictionaries, should be interpreted as having a meaning that is consistent with their meaning in the context of the relevant art and should not be interpreted in an idealized or overly formal sense unless expressly so defined herein.

[0034] It is noted that the PSFE can also perform its intended function without the use of a RACF. In that instance, control of specific links is achieved via accounting based methods (whereas control of an entire network is performed with a RACF). In one embodiment, a standard or commercial network architecture is modified to use a PSFE by providers seeking NCS compliance, though mandatory use of PSFE is envisioned for DoD networks. Optionally, PSFE interaction with H323 systems is provided. The invention may be utilized within the context of SIP/H323, voice over IP (VoIP) or, more generally, X over IP network (XoIP), where X is a voice, video or multimedia communication.

[0035] The PSFE may provide priority signal message queuing (queuing tracked by links of a transport path) and preemption capabilities (i.e., it is able to identify the calls that need to be preempted to decongest certain links). Queuing in this context refers to the buffering-type queuing as is known in the prior art. Queuing in the inventive context is termed by the inventor as protocol level queuing and includes buffering-type queuing as well as call state queuing. Queuing a call or other service occurs when, for example, resource availability is verified for a link during parallel resource clearance. In this case the call is not dropped; rather, its state is stored by the PSFE until appropriate resources become available for the call in each network to be traversed by the call. Being held or stored means that the state information of the call is stored for subsequent use. The state of a call comprises the call source, the call destination, the resources needed to support the call, and so on.

[0036] Call admission control in a simple form is defined as determining which calls should be dropped and which calls should be allowed. Priority services in a simple form is defined as providing preferential treatment to some calls and not other calls. To effectively perform call admission control
and priority services, the network needs to have knowledge of the calls in the system, their actual call paths (control plane and media plane paths) as well as the transport resources available in the network. One embodiment of the invention includes within one entity (the PSFE), priority and call information (e.g., number of calls in system, available transport resources and the like) needed to combine and control call admission and priority service functions. Thus, the PSFE helps to simplify network control plane architecture by combining multiple priority-related functions (e.g., call admission control and priority services) in a single functional entity.

The PSFE optionally communicates via Diameter to a RACF. The PSFE may also be call-state aware such that it can identify calls over links in a transport network (transport resource aware). In some embodiments, the PSFE continuously polls a RACF to identify availability of resources such that resource clearance of multiple call paths may be handled efficiently. In some embodiments the PSFE operates without a RACF (e.g., for open loop or accounting based controls).

In some embodiments, additional support for the PSFE is beneficial. Specifically, at times it may be appropriate for a RACF to pass link information of a media path (e.g., via a PDEFE) to the PSFE. However, this will be needed only under certain cases. RACF link information is not needed when a call flows or when a priority call is cleared. For a priority call not cleared by the RACF, link information is needed for all subsequent RACF-cleared calls so that the availability of resources supporting the non-cleared priority calls can be determined.

Optionally, a P-SCSF (or a Proxy in non-IMS cases) will pass every SIP message it receives to the PSFE. Optionally, a new 5xxx message to convey network attempts (e.g., in the case of priority call queuing) is appropriate. Optionally, a protocol between the PSFE and NGSPs may be adapted to address fault tolerance issues such that fault tolerance functionality need not be distributed across a number of CSCFs.

FIG. 1 depicts a network topology facilitating calls through various networks according to an embodiment of the invention. Specifically, FIG. 1 depicts an exemplary communications system 100 according to an embodiment of the invention. In addition to architecture, FIG. 1 also depicts control and bearer signal paths, which signal paths are associated with one or more call elements indicative of a temporal order of utilization in establishing and tearing down calls. FIG. 1 will be discussed in conjunction with other FIGs to illustrate various call flow examples according to embodiments of the present invention.

Specifically, FIG. 1 depicts a plurality of networks including a calling party home network 110, a calling party visited network 130 and a called party visited network 140.

The calling party home network 110 comprises one or more application servers (AS) modules 112, a home subscriber server (HSS) 114 and a serving call session control function (S-CSCF) 116.

The called party home network 120 comprises one or more application servers (AS) 212, a home subscriber server (HSS) 214, a serving call session control function (S-CSCF) 216 and interrogating call session control function (I-CSCF) 128.

The calling party visited network 130 comprises a P-SCSF 131, a priority services functional element (PSFE) 132, a resource allocation control function (RACF) 133, a backbone packet network 136 and an access network 137.

The RACF 133 includes a policy decision functional element (PD-FE) 134 and a traffic resource control functional element (TRC-FE) 135.

[0045] The called party visited network 140 comprises a P-SCSF 141, a priority services functional element (PSFE) 142, a resource allocation control function (RACF) 143, a backbone packet network 146 and an access network 147. The RACF 143 includes a policy decision functional element (PD-FE) 144 and a traffic resource control functional element (TRC-FE) 145.

The network 100 FIG. 1 may be modified to represent a conventional network by removing each PSFE and RACF from the calling and called party visited networks. In this case, a call between A and B is initiated in accordance with the numbered signal paths 1-16; namely (1) a call invitation is communicated from Endpoint A via access network 137 to backbone packet networks 136, then (2) to P-SCSF 131 and then to (3) S-SCSF, which (4) queries HSS 114, (5) AS 112 and (6) DNS 150 to retrieve the appropriate information pertaining to the called party and its home network. S-SCSF 116 then (7) communicates with the I-CSCF 128 of the called party home network. I-CSCF 128 propagates the request (8) to S-SCSF 126, which communicates (9) with its local HSS 124 and (10) AS 122 together necessary information regarding Endpoint B. S-SCSF then (11) propagates a message to the P-SCSF 141 of the called party visited network 140 associated with Endpoint B.

The above steps (1-11) are fairly conventional. In a conventional network, the P-SCSF 141 communicates with a backbone packet network 146 to identify those routes and other output ports within the RTP stream capable of supporting the call. This information is propagated back to the P-SCSF 131. The P-SCSF 131 and P-SCSF 141 cause their respective backbone packet networks 136 and 137 to provide communication path supporting the call within the RTP stream.

The network 100 FIG. 1 is depicted as including a PSFE and RACF within each of the calling and called party visited networks. The interaction of these components with each other and existing infrastructure will be described in more detail below with respect to the remaining figures.

Within the context of the present invention, a SIP Proxy (or P-SCSF) continues to work as a transaction-stateful or call-stateless proxy as regards all other features. All calls, independent of priority, to or from a SIP Proxy (or P-SCSF) are locally looped to a PSFE. The PSFE optionally maintains call-state information (including link identification information for the traversed links associated with the call) for all calls. Communications between PSFE and RACF are via diameter for necessary resource clearance, and various demands and requests necessary transport path information.

Advantageously, the PSFE enables parallel resource clearance across network boundaries, queuing a call in itself until resources become available. In order to set up calls in parallel across multiple links, the following call control functions are provided: i) call-state awareness about existing and in-progress callsessions; ii) ability to gather information regarding resource usages at the links and network elements, and iii) ability to correlate call path within its domain from the addresses. In one embodiment, the PSFE obtains all necessary resource clearance for a particular call in one parallel operation. The PSFE performs necessary signal message queuing when it decides that resources are available with its domain for a call as part of resource verification. The PSFE
routinely polls the RACF to identify resource availability. The PSFE sends back appropriate (e.g., standard SIP) error messages to the SIP Proxy (or CSCF) to signal a call that is to be blocked, if need be (e.g., a higher priority call is queued). In the event of resource availability for the PSFE’s own network and receipt of an indication of resource availability for another network to be traversed by an intended call, the PSFE sends back appropriate (standard) SIP protocol messages to the SIP Proxy (or CSCF) to signal a call that is to be allowed. The PSFE initiates 3rd party calling for queued calls via the P-CSCF, and as when the resources for those calls become available. In the case of a 3rd party call, the network operates to communicate (i.e., call back) both the calling party and the called party at a later time.

[0051] While denoted as a real-time transport protocol (RTP) stream in FIG. 1, the backbone packet networks may communicate via any streaming protocol, such as transmission control protocol (TCP), user datagram protocol (UDP) and the like. Signaling messages are sent from the call control elements in the originating network up to the call control elements in the terminating network for setting up end-to-end voice, data and video sessions across multiple networks. Different mechanisms for carrying such messages can be put in place depending on the network topology, interconnectivity, and protocols supported by the underlying network.

[0052] FIG. 2 depicts an overview of a network topology for facilitating calls through an exemplary network according to an embodiment of the invention. A method according to the invention proposes to obtain resource clearance in parallel across all links within a network boundary, as well as obtain resource clearance across several network boundaries. A conventional system employs sequential resource clearance, in which no node transmits a message until that node first obtains resource clearance within its domain. (i.e., serial resource clearance is accomplished on a link-by-link basis). In contrast, in parallel resource clearance as provided by the PSFE, the apparent or actual performance of more than one resource clearance operation occurs at a time, by the same or different devices and modules. In the parallel case, the various time requirements for resource clearance (Ti) are triggered as and when a resource availability request is received in domain i, and the domain i does not wait for the completion of Ti before sending a resource availability request to domain i+1. In this manner, multiple domains may be obtaining resource clearance without having to wait for the completion of resource clearance in a previous one’s. When there are multiple links in each domain, resources clearance within a domain may also be performed in parallel. For example, by using PSFEs for obtaining resource clearance in parallel across all links within a network boundary, as well as obtaining resource clearance across several network boundaries, the time required for resource clearance will be reduced to O(Max(t)), where Max(t) represents the maximum of all ti, i.e., the maximum of all individual resource clearance times in individual links. Accordingly, the time required to process and transmit a message in a PSFE enabled network is much lower than the time required to wait for clearance on all links in a conventional network.

[0053] Such a methodology leads to one or more benefits. For example, setting up calls and sessions across an IP network will be more efficient and faster, which will reduce post-dial delay. Calls and sessions will have higher probability to complete set-up protocol exchanges before any particular timer expires, such as during set up when the networks are congested. For an end-to-end call which encounters multiple service provider (SP) networks, each SP network may be provided with information on resource needs and priority for the call/session, and the separate networks are able to assess and clear the calls. The time for set up of a call or a session that crosses several network domain boundaries (resulting from reasons, such as, administrative domains, geographical distribution, and population density, etc, even within a same SP network) will require only the maximum time required for setup in any of the domains.

[0054] A further overview of the mechanism and methodology for parallel resource clearance by the PSFE across multiple networks required for an intended call is described in U.S. patent application Ser. No. __________ entitled End-To-End Capacity And Priority Management Through Multiple Packet Network Segments, filed Dec. ________, 2008 which is herein incorporated by reference. Further, beneficial applications can be developed using the methodology provided. For example, many call scenarios require connection to media servers, at times prior to call establishment and in other cases in mid-call, and the access path to media server may differ from that carrying the bearer packets from the caller to the called. The proposed methodology enables reservations and resource management across such different paths in parallel. In addition, the methodology of the invention facilitates the establishment of multiple conference legs, where call legs for an intended communication may include legs for each end point (end point referring to users, media server, conferencing server, and the like) and for each media type (e.g., video, audio, text, data, and the like, and some combination thereof) and some combination thereof. For example, if capacity is constrained to support all media types in a multi-media call, reservations across media types in parallel can be utilized to enable early establishment of calls with one media requiring lesser capacity when required capacity to support all media types is unavailable at the start of the call. Later, other media types can be added when additional capacities become available as further described in U.S. patent application Ser. No. __________ entitled Incremental Addition And Scale-Back of Resources Adapting to Network Resource Availability, filed Dec. ________, 2008 which is herein incorporated by reference. Further conferences with multiple end points across various network may be initially established with less than all of the multiple end points, with endpoints called back as resources become available as described in U.S. patent application Ser. No. __________ entitled Network Call Back To Add Conference Participants And/or Media, filed Dec. ________, 2008, which is herein incorporated by reference.

[0055] An overview of the mechanism for the parallel resource clearance by PSFE with Resource Priority Header (RPH) mapping and defaulting across multiple networks is outlined in FIG. 2. For example, a priority call initiated from a caller in network A to a called end in network C will traverse network 200, which comprises networks A 210, B 220 and C 230. Alternatively stated, a call initiated from the caller to the called end will traverse networks N1, N2, N3, N4, . . . Nm, where N1 is the head end, and Nm is the tail end of the call. The head end refers to network components like proxies, CSCF and PSFEs closest to the caller, while tail end refers to those network components closest to the called user. Also note that only three networks, namely A, B and C are illustrated for simplicity.

[0056] Each network includes a CSCF (212, 222, 232), a PSFE (212, 224, 234), a RACF (216, 226, 236), and a back-
bone packet network (218, 228, 238). The backbone packet network in each network may include a plurality of servers (219, 229, 239). The RACF may include a policy decision functional element (PD-FE) and a traffic resource control functional element (TRC-FE) (not shown). [0057] All calls, independent of priority, to or from a CSCF are locally looped to a PSFE. The PSFE maintains call-state information (including link identification information for the traversed links associated with the call) for all calls. Communications between PSFE and RACF are via Diameter for necessary resource clearance, and various demands and requests necessary transport path information. [0058] When a call passes through networks A, B and C, then the PSFE element in each of those network clouds is responsible for obtaining necessary resource clearances in its own network respectively. In parallel with the verification of resource availability within its network, the PSFE forwards the notice of an intended communication to the next network and toward the called party network. The priority of the intended communication may be indicated in a Resource Priority Header (RPH) of the notice. [0059] As resources become available in each of the networks, the PSFE in that network awaits the clearance to arrive from the tail end of the call. In other words, network B waits for the clearance from network C, and network A waits for the clearance from network B. Thus, network A will not receive clearance until both networks C and B have necessary resources for the call. At this point the PSFE in network A, communicates to the CSCF in network A (or the SIP proxy in network A) and triggers a re-invite from the caller to the called user. In an alternative embodiment, PSFEs in each of the networks await for the resource clearance indication to arrive from the head end direction of the call. [0060] The relationship between two networks (e.g., network A, network B) or service providers may vary in one of two ways. In a first instance, SPs may trust the priority indicator received in a notice of intended communication and forward the notice without verification of the priority therein. For example, as the originating network A sends the session INVITE through the RPH, the PSFE in receiving network B trusts the priority indicated in the RPH and needs no extra verification. Trust may be achieved through either implicit business agreements, or through a technological component. The technological component may be associated with the PSFE. For example, the PSFE in the receiving network may exchange keys with PSFEs of known service providers, via a protocol (e.g., the peer interface of the PSFE). In this case, the receiving network B uses the same priority indicated in the RPH received from network A, within its own network (here, network B). If network B is not the call terminating network and needs to send the INVITE to follow-on networks (network C), the received RPH from network A will be sent without any modification. [0061] On the other hand, when there is no trust mechanism established between SPs, the receiving network (e.g., network B or C) needs to verify the priority of the call as indicated in the notice of intended communication. For example, priority calls may be first authenticated by a personal Identification Number (PIN) before the notice of intended communication is signaled for session start. After verification of the PIN at the originating network A, the PSFE in the originating network A inserts the PIN in a header extension and sends it toward the destination network. The PSFEs in receiving networks (networks B and C) in the call/session path are able to identify the PIN in the header and replay the authentication of the call through an authentication server. For security, the PIN number may be sent in an encrypted manner, and not in clear text. After authentication, the PSFEs in networks B and C are able to map and assign an RPH for the session solely for use in their own networks according to a previously set policy. [0062] Should a PSFE receive a priority call without an accompanied PIN, the priority in the receiving network (networks B and C) may be assigned a default value for internal use, which may be set by a predetermined policy. [0063] PSFEs in receiving networks (networks B and C) may store or add the internal priority information as additional header extensions (to be used only in their own networks) over the original header, and remove these additional header extensions if the notice of intended communication needs to be sent to other follow-on networks (e.g., from network B to network C). In effect, if the receiving network sends the notice of intended communication to follow-on networks, the original RPH is sent as received, without any modification. [0064] In each of these network segments one or multiple PSFE communicates with the corresponding CSCF and RACFs. For example, if resource clearance is propagated toward the tail end, on receiving a new session or call request through SIP INVITE and from the identification that the call or session needs to be forwarded to the CSCF in Network B, the PSFE in Network segment A forwards a resource assessment request to PSFE in Network segment B, while looking into the resource requirement versus availability in Network segment A. The resource assessment request carries a RPH for network A and a PIN. The PSFE in Network segment A also follows up with a message to the CSCF and PSFE in network segment B regarding resource availability in Network segment A after its determination. [0065] The PSFE in Network segment B performs the same function as the PSFE in Network segment A, which is sending a resource assessment request to PSFE in Network segment C by recognizing that the call or session needs to be sent to Network segment C while it gets engaged in verifying resource availability in Network segment B. The PSFE in Network segment B verifies the PIN received from the PSFE in Network segment B, if necessary. The PSFE in Network segment B maintains a map of RPHs between Network segment A and B, so as to allow internal use of corresponding RPHs for Network segment B while replying with RPHs for Network segment A. PSFE in Network segment B also follows up with a message to the CSCF and PSFE in Network segment C regarding resource availability in Network segment B after its determination. PSFE in Network segment B also forwards the resource availability determination that it received from PSFE in Network segment A. [0066] The CSCF in Network segment C determines that the called end is in Network segment C. PSFE in Network segment C determines the resource availability for the requested call or session for the Network segment C using a mapped RPH and together with messages received from PSFEs in Network segments A and B, determines the end-to-end resource availability before informing its decision on call or session admission to the CSCF in Network segment C. CSCF in Network segment C then admits or disallows the call based on PSFEs’ decision (based on resource availability information for PSFEs in A, B and C). [0067] In other embodiments, a transit network may be interposed two of the Network Segments illustrated (e.g.,
between Network segments A and B, Network segments B and C). A transit network is used for bulk transport between networks providing user services. Such transit networks provide bulk transport services only and do not have call or session managers associated therewith but may honor Service Level Agreements (SLA), which are primarily based on statistically averaged call or session performance over time. SLAs are generally based on connectivity metrics such as, downtime, reliability and availability measures, and bytes sent/received, and Quality of Service metrics such as, delay, delay variation, packet loss and packet throughput.

[0068] Transit networks do not provide individual call or session control and management. Thus, the transit network and its partners will typically have an overall bulk transport carrying agreement but the transit network does not support specifically any individual call or session related characteristics. For example, a transit network may be outfitted with a PSFE, RACF and a Backbone network but not have a CSCF.

[0069] For calls passing through such core transport networks or transit networks, obviously call or session management is not performed by the core transport service provider. Thus, the originating and/or terminating network service provider on either side of the transit network is responsible to provide end-to-end call or session admission, and management. Accordingly, the PSFE also is enabled to negotiate and determine resource availability from the transport-only networks. In addition, as described above, the PSFE performs resource assessment in its own network and forwards a resource assessment and clearance request through SDP offers to PSFEs in the call path as determined from the destination URI. Based on received the availability information for its own network and other networks in the call path including any interposed transit networks, a PSFE decides on call admission and instructs the Call Session Control Function Elements (CSCF) for implementing call or session admission decisions.

[0070] FIGS. 3a and 3b depict a flow diagram of an exemplary routine according to embodiments in accord with the present invention. The call blocking call flow, call queuing and call creation for the parallel resource clearance of 300 of FIGS. 3a and 3b will be described within the context of the communications system 200 of FIG. 2. Specifically, control signal paths between calling Endpoint-A 302 in network A (i.e., calling party network) and called Endpoint-C of network C (i.e., called party network) as can be gathered from FIG. 2 comprise, in the order named, CSCF-A 212, PSFE-A 214, CSCF-B 222, PSFE-B 224, CSCF-C 232, PSFE-C 234.

[0071] Referring now to FIG. 3a, at 301 an invite message is propagated from calling Endpoint-A to CSCF-A 212, then at 302 to CSCF-B 222, and finally at 303 to CSCF-C 232. In general, each network sends an invite message toward the called network until the called network is reached by the notice of intended communication. The invite message includes the Resource Priority Header (RPH) for Network A (RPH-A). The invite message may also include a Personal Identification Number (PIN) associated with the authentication of the call. The PIN may be encrypted.

[0072] In parallel with the propagation of the invite message, each PSFE (214, 224, 234) verifies resource availability for a resource required for intended communication in response to a reserve message (311, 312, 313).

[0073] Every SIP invite message sent out by N contains the RPH value sent in the original INVITE from the calling party network. However, internally N, the corresponding PSFE, might use an RPH value that is understood and utilized by network N, only. Thus, a mapping between the RPH values (viz., RPH-A and RPH-I is maintained by P). Messages going out toward the head end or tail end from P will contain RPH-A, and messages for internal consumption will contain RPH-I. These RPH values are negotiated via a key exchange protocol between various PSFE's. If no protocol is exchanged between various P's to exchange RPH information, a default RPH value may be utilized. Thus at 310, PSFE-B verifies the PIN for the call and maps RPH-A of the notice of intended communication to RPH-B for internal use. At 312, PSFE-C verifies the PIN for the call and maps RPH-A of the notice of intended communication to RPH-C for internal use.

[0074] It should be noted that RPH is a specific SIP header extension. In general, ones of the networks may use standard or proprietary signaling other than SIP. Thus in alternate embodiments, mapping between RPH and proprietary protocol fields indicating priority of the call/session will be undertaken.

[0075] When an invite message enters a first network (e.g., network C), the corresponding PSFE (e.g., PSFE-C 234) checks the RPH to provide treatment appropriate to its priority level, such as attempting to reserve necessary bandwidth for the call and queues the call in its stack. If P has a trusted relationship with P, then P uses the RPH-A for priority treatment within N. The key exchange between various P's is not addressed here as any appropriate key exchange mechanism between servers should suffice. If P does not have a trusted relationship with P, then P uses the PIN in the INVITE from P, authenticates the call and its priority level through an authentication server. After authentication, P sets RPH-I for use in N, solely, mapping RPH-I from RPH-A according to predetermined policy. RPH-A is not discarded from the signaling; rather RPH-I is added on top. In case P receives an INVITE without an accompanied PIN, P sets a default RPH-I value according to the predetermined policy of N. If P needs to send messages upstream to P, or downstream to P, P removes the RPH-I header extension from the message leaving the originally received RPH-A intact.

[0076] The corresponding PSFE (e.g., PSFE-C 234) determines from corresponding RACF (e.g., RACF-C 236) if resources are available for all the links within the first network (e.g., network C) for this particular call. In one embodiment, if requested resources are not available, the call is queued, and no lower or equal priority call is allowed until this particular call gets a chance to obtain the requested resource capacity.

[0077] For example, in case of dynamic real-time bandwidth management, corresponding RACF (e.g., RACF-C 236) listens to the SNMP messages from multiple switch/router in that network in parallel and determines if all the necessary link/s that constitute the call have enough resource/s available. In cases where dynamic bandwidth management is not required and mere call counting or bandwidth counting will suffice, the corresponding PSFE or the corresponding RAC may be directly involved in bandwidth or call counting and determine the availability of the necessary resource/s.

[0078] The act of queuing the call and communicating with the corresponding RACH ensures that resources are obtained in parallel in first network (e.g., network C). In case of dynamic bandwidth management, all corresponding routers in the network send updates in parallel to the corresponding
RACH of the first network. The corresponding PSFE then makes decisions based on the query to the corresponding RACH of the network as described above.

[0079] In another embodiment when a transport network is interposed networks A and B, the originating network’s (A) PSFE learns about the network where the call/session needs to be forwarded to (i.e., network B) and the intermediate transport network (A’) from the called URI. The PSFE then verifies the SLA agreed with the transport network (A’) and determines if the SLA assures the requirements for the call/session in progress. The originating network’s PSFE forwards the call/session request along with resource needs to the PSFE in the next network (B) in the call/session path (via the CSCF’s).

[0080] A PSFE sends a Held message toward the next PSFE closer the head-end PSFE only when the resources in all subsequent networks are available for the intended call. If the PSFE is in the called party network (e.g., network C), at 321 the held message may be forwarded based on resource availability within the called network alone. The held message is forwarded between PSFE via corresponding CSCF at 322, 323. Referring to FIG. 3b, for PSFE in a first network other than called party network (i.e., intermediate networks (e.g., network B) and the calling party network (e.g., network A)), at 324 a Held message is forwarded only if each network from first network in question to the called party network have resource availability for the intended communication. Again, the held message is delivered to the head end PSFE via the corresponding CSCF of the first network at 325 and the head end network at 326. In other words, the held message is forwarded between PSFE via corresponding CSCF (325 and 326). A Held message is essentially a command that instructs the CSCF that resources are held for a user, but are not yet consumed. The Held message may be implemented as a 1xx message.

[0081] As resources become available in each of the networks, the PSFE in that network await for the clearance to arrive from the tail end of the call. In other words, when a call passes through networks A, B, and C, B waits for the clearance from C, and A waits for the clearance from B. Thus A will not receive clearance until both B and C have necessary resources. All the intermediate networks’ PSFEs and the final network PSFE sends a 1xx message indicating that resources are held when the resources become available. Note that in this embodiment with a transport network A’ interposed networks A and B, PSFE in A also has verified that the transport network A’ can assure the resources required in A’ through communication with the transport network.

[0082] In an alternate embodiment, the PSFE sends an indication of resource clearance to the head-end as soon as resources are determined to be available in its own network and the decision about final call admission may be based on resources availability across the various PSFE as determined by the PSFE at the head end network (i.e., calling party network). In another embodiment call admission may be based on minimum possible resource is available across the various PSFE.

[0083] When PSFE at the calling party network determines that there is of resource availability for end-to-end communication contemplated, at 327 that PSFE forwards a “held message” that triggers corresponding CSCF to initiate a SIP Re-Invite from the calling network CSCF to the called network CSCF (328, 329, 330). Alternatively, an SDPReoffer may be triggered.

[0084] In response to the re-invite message, a ringing message is propagated from the called endpoint to the calling endpoint (331, 332, 333). When a PSFE obtains the Re-Invite message, the necessary resource is locked in using the queued information associated with the call via lock message (335, 336, 337). A Lock message is essentially a command that instructs the PSFE that resources are now being used. The Lock message may be implemented as another 1xx message or could be piggybacked with ‘ringing’ message. Prior to the PSFE receipt of the Lock message, the call is merely queued and the resources are not being consumed. Messages need not be in SIP, but can be implemented in DIAMETER or any other protocol.

[0085] The PSFE of the present invention advantageously isolated Parallel Call Admission and Priority Services development so that such services are easier to develop and maintain. Moreover, such services may now be provided without service and software replication with respect to vendor product offerings. Parallel Call Admission and Priority Services can be an optional service required by a customer and can fit IMS and non-IMS networks, including H323. With the novel PSFE, newer DISA NGN and NGPS features are easier to develop. In this manner, new features like “priority conference call reservation” of priority conferences may be implemented. In addition, the PSFE relieves the burden of having a CSCF or any other fully operational SIP Proxy maintain resource awareness (i.e., call states with respect to links). The PSFE provides resource control and policy-based management mechanism adapted to implement priority services in a packet network environment, such as described herein.

[0086] The above-described embodiments of the invention may be implemented within the context of methods, computer readable media and computer program processes. Generally speaking, methods according to the invention may be implemented using computing devices having a processor as well as memory for storing various control programs, other programs and data. The memory may also store an operating system supporting the programs. The processor cooperates with conventional support circuitry such as power supplies, clock circuits, cache memory and the like as well as circuits that assist in executing the software routines stored in the memory. As such, it is contemplated that some of the steps discussed herein as software processes may be implemented within hardware, for example as circuitry that cooperates with the processor to perform various steps. Input/output (I/O) circuitry forms an interface between the various functional elements communicating with the device.

[0087] A computing device is contemplated as, illustratively, a general purpose computer that is programmed to perform various control functions in accordance with the present invention. The invention also can be implemented in hardware as, for example, an application specific integrated circuit (ASIC) or field programmable gate array (FPGA). As such, the process steps described herein are intended to be broadly interpreted as being equivalently performed by software, hardware and a combination thereof in various alternative embodiments.

[0088] The invention may also be implemented as a computer program product wherein computer instructions, when processed by a computer, adapt the operation of the computer such that the methods and/or techniques of the present invention are invoked or otherwise provided. Instructions for invoking the inventive methods may be stored in fixed or removable media, transmitted via a data stream in a signal
What is claimed is:

1. A method comprising:
   receiving at a first network a notice of an intended communication to a called party network, the notice of intended communication including a first priority indicator, wherein the intended communication requires a resource for supporting a streaming data protocol in each network between a calling party network and the called party network;
   forwarding the notice of an intended communication to a second network and toward the called party network, the notice of intended communication including the first priority indicator;
   in parallel with said forwarding, mapping the first priority indicator to a second priority indicator and initiating for the intended communication a determination of resource availability for the first network based on the second priority indicator;
   determining for the intended communication the determination of resource availability for the first network based on the second priority indicator, wherein the determination is for a first resource for the first network; and
   verifying resource availability for the intended communication.

2. The method of claim 1 wherein the mapping of the first priority indicator to the second priority indicator is based on values negotiated via a key exchange protocol.

3. The method of claim 1 wherein the mapping of the first priority indicator to the second priority indicator is based on values negotiated via a key exchange protocol between the first network and the calling party network.

4. The method of claim 1 wherein the mapping of the first priority indicator to the second priority indicator is based on predetermined values.

5. The method of claim 1 wherein the mapping of the first priority indicator to the second priority indicator is based on at least one default.

6. The method of claim 1 wherein the first priority indicator is included in a resource priority header.

7. The method of claim 1 wherein determining for the intended communication the determination of resource availability for the first network based on the second priority indicator comprises:
   reserving for the intended communication the first resource for the first network in the event the first resource is available for the intended communication.

8. The method of claim 1 wherein determining for the intended communication the determination of resource availability for the first network based on the second priority indicator comprises:
   reserving for the intended communication the first resource for the first network in the event the first resource is available above a threshold level.

9. The method of claim 1 wherein signaling for the determination of resource availability for the first network based
on the second priority indicator includes both the first priority indicator and the second priority indicator.

12. The method of claim 1 wherein verifying resource availability for the intended communication comprises: receiving for the intended communication from a second network a second indication of resource availability; and wherein the verifying is based on the determination of resource availability for the first network and the second indication of resource availability.

13. The method of claim 1 further comprising: forwarding a re-invite for the intended communication in the event verifying resource availability for the intended communication indicates availability of resources from the calling party network to the called party network.

14. A method comprising:
   receiving at a first network a notice of an intended communication to a called party network, the notice of intended communication including a first priority according to a first priority scheme, wherein the intended communication requires at least one resource for supporting a streaming data protocol between a calling party network and the called party network;
   mapping the first priority to a corresponding second priority for a second priority scheme utilized in the first network;
   verifying resource availability of at least one resource in the first network based on the corresponding second priority; and
   in parallel with said verifying, forwarding the notice of an intended communication to a second network and toward the called party network prior to receiving an first indication of resource availability of the at least one resource in the first network required for the intended communication.

15. The method of claim 14 wherein the mapping of the first priority to the corresponding second priority is based on negotiation via a key exchange protocol between the first network and the calling party network.

16. The method of claim 14 wherein the mapping of the first priority to the corresponding second priority is based on predetermined correspondences.

17. The method of claim 14 wherein the mapping of the first priority to the second priority is based on a default correspondence.

18. The method of claim 14 wherein the notice of intended communication includes a resource priority header.

19. The method of claim 14 further comprising:
   receiving at the first network a second indication of resource availability of at least one resource at a third network required for the intended communication.

20. The method of claim 19 further comprising:
   forwarding a re-invite of the intended communication to the second network and toward the called party network in the event the first indication and the second indication indicate resource availability for at least one resource in each network from the calling party network to the called party network.

* * * * *