ETSI TR 102 478 V1.1.1 (2006-03)

Technical Report

Corporate telecommunication Networks (CN);
Enterprise communication involving
Next Generation carrier Networks (NGN)

Reference

DTR/ECMA-00295

Keywords

network, service

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Copyright Notification

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Report (TR) has been produced by ECMA on behalf of its members and those of the European Telecommunications Standards Institute (ETSI).

Brief History

This Ecma Technical Report provides an overview of IP-based enterprise communication from/to Corporate telecommunication Networks (CNs) (also known as enterprise networks) including aspects of privately used home networks accessing public next generation networks (NGN).

The present document is based upon the practical experience of Ecma member companies and the results of their active and continuous participation in the work of ISO/IEC JTC1, ITU-T, ETSI, IETF and other international and national standardization bodies. It represents a pragmatic and widely based consensus.

This Ecma Technical Report has been adopted by the General Assembly of December 2005.

1 Scope

The present document identifies key use cases for communication with or between IP-based Next Generation Corporate Networks (NGCN) involving public next generation networks (NGN), analyses these use cases in terms of available or planned standardized technology and identifies requirements that will have to be met.

The present document investigates configurations involving NGCNs and NGNs and their interoperating requirements. Non-IP-based interoperation, i.e. using circuit-switched technology, between NGCNs and NGNs is outside the scope of the present document.

The present document does not discriminate between wireless and wired access technology.

All mobility aspects are outside the scope of the present document. They are covered by a companion Technical Report ECMA TR/92 [1].

Application considerations such as IP Centrex and CSTA (Computer Supported Telecommunications Applications) are outside the scope of the present document.

2 References

[1] ECMA TR/92: "Corporate Telecommunication Networks - Mobility for Enterprise Communication".

[2] ECMA-307: "Corporate Telecommunication Networks - Signalling Interworking between QSIG and H.323 - Generic Functional Protocol for the Support of Supplementary Services (June 2000)".

[3] ITU-T Recommendation H.323: "Packet-based multimedia communications systems".

[4] IETF RFC 3261: "(SIP) Session Initiation Protocol".

[5] IETF RFC 3489: "(STUN) Simple Traversal of User Datagram Protocol (UDP) Through Network Address Translators (NATs)".

[6] IETF RFC 3711: "The Secure Real-time Transport Protocol (SRTP)".

[7] IETF RFC 3761: "The E.164 to Uniform Resource Identifiers (URI) Dynamic Delegation Discovery System (DDDS) Application (ENUM)".

[8] IETF RFC 3966: "The tel URI for Telephone Numbers".

[9] IETF RFC 2401: "Security Architecture for the Internet Protocol (IPSEC)".

[10] ITU-T Recommendation E.164: "The international public telecommunication numbering plan".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document the following definitions apply:

Corporate telecommunication Network (CN): sets of equipment (Customer Premises Equipment and/or Customer Premises Networks) that are located at geographically dispersed locations and are interconnected to provide telecommunication services to a defined group of users

Next Generation CN (NGCN): self-contained corporate network designed to take advantage of emerging IP-based communications solutions and that can have its own applications and service provisioning

Next Generation Network (NGN): packet based public network able to provide telecommunication services, able to make use of multiple QoS enabled transport technologies and in which service related functions are independent of underlying transport related technologies

Virtual Private Network (VPN): virtual network that can deliver ubiquitous and secure connectivity over a shared network infrastructure (e.g. public carrier networks) using the same access policies as an enterprise network

Application Service Provider (ASP): entity that provides telecommunication applications

Session Service Provider (SSP): entity that intervenes in and adds value to signalling for the establishment and control of multi-media sessions and optionally intervenes in and adds value to the multi-media sessions themselves

Transport Service Provider (TSP): entity that provides IP connectivity

3.2 Abbreviations

ALG	Application Layer Gateway
API	Application Protocol Interface
ASP	Application Service Provider
B2BUA	Back-to-Back User Agent
CA	Certification Authority
CN	Corporate telecommunication Network
CSTA	Computer Supported Telecommunications Applications
ECC	Emergency Call Centres
ICE	Interactive Connectivity Establishment
IDSN	Integrated Services Digital Network
IP	Internet Protocol
IPSEC	IP SECurity
ISDN	Integrated Services Digital Network
MIKEY	Multimedia Internet KEYing
NAT	Network Address Translator
NGCN	Next Generation Corporate Network
NGN	public Next Generation Network
PKI	Public Key Infrastructure
PSTN	Packet Switched Public Data Network
QoS	Quality of Service
RTCP	RTP Control Protocol
RTP	Real-Time Protocol
SBC	Session Border Controller
SDP	Session Description Protocol
SIP	Session Initiation Protocol
SOHO	Single Office Home Office
SRTCP	Secure RTCP
SRTP	Secure Real-time Transport Protocol
SSP	Session Service Provider
STUN	Simple Traversal of UDP through NATS
TCP	Transmission Control Protocol
TE	Terminal Equipment
TLS	Transport Layer Security
TSP	Transport Service Provider
UA	User Agent
UAC	User Agent Client
UAS	User Agent Server
UDP	User Datagram Protocol
VoIP	Voice over IP
VPN	Virtual Private Network

4 Void

5 Background

There has been a major evolution in enterprise telecommunications during the last few years. Prior to that, enterprise telecommunication networks (or corporate telecommunication networks, CN) were based on 64 kbit/s circuit-switched technology, which had synergy with corresponding technology deployed in public Integrated Services Digital Networks (ISDN) and traditional analogue services. Those CNs primarily delivered a voice or telephony service to their users, although in principle they were capable of other services too, including video and various types of data service. For communication outside the enterprise, CNs were able to interwork with public ISDNs.

Many public networks also offered optional services to corporate customers, such as Centrex services and premise equipment leasing and maintenance.

With the advent of technologies for transmitting voice and other real-time media over the Internet Protocol (IP) (e.g. based on Real Time Protocol (RTP)) and corresponding new signalling protocols (e.g. H.323, SIP), there was potential for providing telephony and other real-time person-to-person services in the public Internet. Moreover, such services also became possible in the IP-based "intranets" already deployed in enterprises for data services such as corporate email, file transfer, corporate web services and access to the world wide web. Enterprises saw advantages such as savings on infrastructure costs (e.g. one wire to the desk) and the introduction of innovative services that exploited the convergence of real-time and data communication. The traditional PBX (Private Branch Exchange) was replaced by or evolved to an "IP-PBX" or soft switch that supported IP connectivity to the desktop and IP connectivity between nodes. Direct IP-based transmission of multimedia between endpoints meant that switching capabilities were no longer required, except gateways for interworking with "legacy" circuit-switched networks.

IP-based CNs are continuing to evolve, to support additional services, improved security, improved QoS, etc. A CN that fully embraces IP technology is referred to here as a Next Generation CN (NGCN).

At present, NGCNs generally fall back to legacy circuit-switched techniques for communication outside the enterprise, e.g. using public ISDN or circuit-switching over leased lines. Gateways provide the necessary interworking of signalling and media.

The next stage of evolution will be for NGCNs to extend IP-based communication outside the enterprise by using a public network that also supports IP-based communication. The public IP-based network may provide communication between the NGCN and another NGCN, between the NGCN and another IP-based user (e.g. residential user) or, via a gateway in the public network, between the NGCN and a legacy user or network. With this the NGCN no longer needs gateways to legacy networks (except where required by existing investment or economic considerations) and can enjoy the benefits of end-to-end IP-based communication with appropriately equipped communication partners.

The public Internet is one example of a public IP-based network that an NGCN can use for external communications. In addition, some service providers are planning to offer public IP-based networks that offer improvements compared with the public Internet, e.g. in terms of QoS, security, mobility, applications, etc. These value added public IP-based networks are collectively known as Next Generation Networks (NGN).

The present document aims to contribute to this next stage of evolution by looking at the issues involved in interoperating NGCNs and NGNs and to identify requirements on NGNs.

5.1 Provision of services by NGNs

5.1.1 Levels of service provision

An NGN may provide services to NGCNs and NGCN users at a number of different levels.

The most basic level of service provision is IP connectivity. Differentiation from the Internet can be in the form of improved or guaranteed quality of service or security. For the purposes of the present document an NGN that provides this level of service acts as a Transport Service Provider (TSP).

A second level of service provision is in session establishment and control of communication sessions, e.g. voice, multimedia, messaging. Here the NGN adds value by being involved in the signalling protocol used to establish and control media sessions. For the purposes of the present document the primary session control signalling protocol concerned is assumed to be the Session Initiation Protocol (SIP). Added value can include routing, provision of quality of service for media, provision of gateway services to legacy networks, assistance in NAT traversal, etc.. For the purposes of the present document an NGN that provides this level of service is known as a Session Service Provider (SSP).

A third level of service provision is at the application level. Applications can be many and varied, but for the purposes of the present document an application is assumed to be related to telecommunication in some way. An application may be able to monitor or control multi-media sessions (either directly or through a protocol or API such as CSTA) and may or may not be involved in media as well. Examples of applications that involve media include conferencing services, transcoding and translation services and call distribution centres. Examples of applications that monitor or control sessions but do not involve media include presence services, call logging services and UA configuration services. In addition, an application may be accessed through a session control protocol such as SIP. For the purposes of the present document an NGN that provides this level of service is known as an Application Service Provider (ASP).

An NGN may provide services at one or more of these levels. Not all services offered will be of interest to enterprise customers and of relevance for interworking with NGCNs. Enterprise customers may use different NGNs for different levels of service provision and may have different contractual relationships with each of these NGNs. In addition, for a given communication and depending on the number of parties to be interconnected and/or the number of services to be accessed, multiple providers may be involved.

Figure 1 shows an example of one or more NGNs providing services at the three levels.

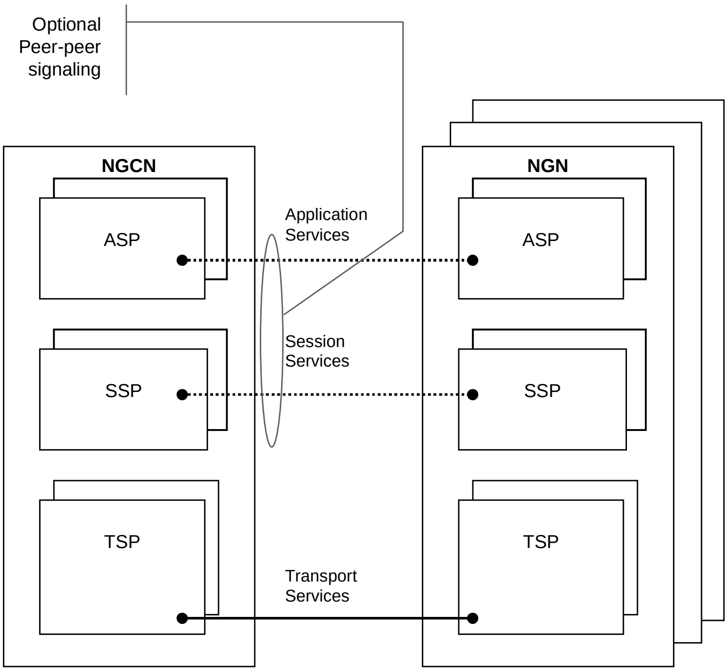

Figure 1: TSP, SSP, and ASP provided by NGN(s)

5.1.2 Use of NGN services by NGCNs

In providing communication services to its users, an NGCN can make use of services of one or more NGNs. For communication with another entity (e.g. another NGCN or a residential user), an NGCN may make use of just TSP services, TSP and SSP services, or TSP, SSP and ASP services from the same NGN or different NGNs. In addition, an NGCN can make use of ASP services in their own right, rather than as part of communication with another entity (e.g. a presence service).

5.1.3 Home NGN

An NGCN customer may have contractual relationships with one or more NGNs for the provision of certain services, each NGN being known as the Home NGN for the services that it provides. Some parts of the NGCN may not have direct connectivity to the Home NGN and will need to access services of the Home NGN via other NGNs. The customer may have contractual relationships with those other NGNs too (e.g. just for IP connectivity) or reliance may be placed on agreements between the Home NGN and those other NGNs for such use.

Figure 2 shows an example of two parts of an NGCN, each part accessing services of the same Home NGN, one directly and the other via a different NGN. The solid lines represent connectivity and the broken lines represent service access.

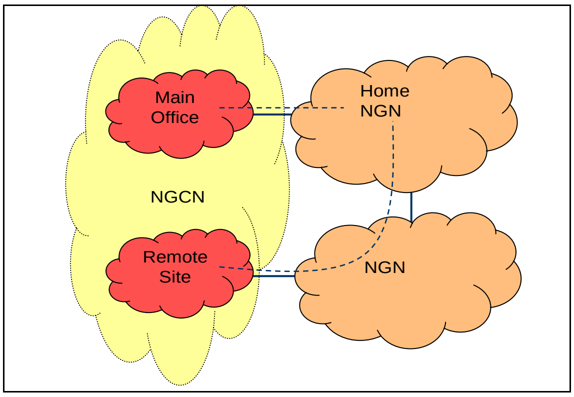

Figure 2: Home NGN and other NGN

5.2 Management considerations

Services provided by an NGN may need to be managed from an NGCN. In addition, an NGN may provide a service for managing aspects of an NGCN. For both these purposes industry standard management interfaces should be used between the NGN and the NGCN. However, consideration of these interfaces is outside the scope of this edition of the present document.

6 Basic Configurations and General Requirements

This clause lists important configurations which apply to an enterprise-specific next generation corporate network (NGCN) and its equipment (terminal equipment, servers, etc.) when connected to public next generation networks (NGN). In addition, general requirements on NGN are specified in clause 6.6.

The following interconnection **Scenarios** will have to be considered:

- Scenario 1 - Communication between NGCNs via an NGN using a VPN.

  Use cases: typically closely related partners.

  - Two peer offices of one enterprise.
  - Main office and branch office of an enterprise.
  - Two closely-related enterprises ("extranet").

- Scenario 2 - Communication between NGCNs via an NGN **not** using a VPN.

  Use cases: typically **not** closely related partners.

  - Two enterprises (not closely-related).

- Scenario 3 - Communication between NGCN and TE via an NGN using a VPN.

  Use cases: typically closely related partners.

  - Between enterprise and residential or SOHO user (e.g. home worker).

- Scenario 4 - Communication between NGCN and TE via an NGN **not** using a VPN.

  Use cases: typically **not** closely related partners.

  - Residential or SOHO user (e.g. member of public).

- Scenario 5 - Communication between NGCN and PSTN/ISDN via an NGN

Clause 6 discusses the Technical Issues and requirements for each of the main scenarios in detail.

6.1 Scenario 1 - Communication between NGCNs via an NGN using a VPN

Enterprise 1 and 2 are closely enough related to have a security association that allows use of a VPN, which provides a tunnel through which inter-NGCN signalling, media and other data can flow. The tunnel provides functions such as security, transparency for private IP addresses, NAT and firewall traversal, etc.

Typically this close relation is used between:

- Offices A and B of a single enterprise.
- Enterprise main office and branch office.
- An Enterprise 1 and a closely-related (e.g. partner) Enterprise 2.

Tunnel control may be directly between the NGCN(s) (i.e. tunnel control end-to-end) with no NGN involvement other than as a TSP, or tunnel control may be provided/owned by the NGN or a third party.

The NGN in this scenario acts only as TSP. Multiple TSPs within the NGN, e.g. one for each enterprise, may be involved.

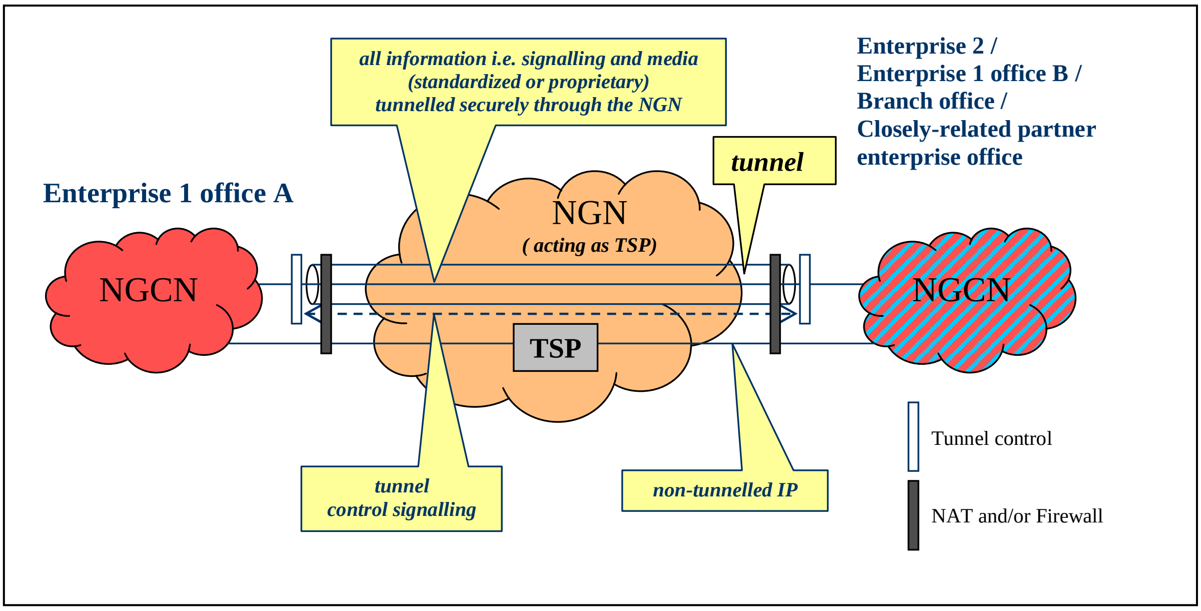

Figure 3: Scenario 1 - Communication between NGCNs using a VPN

NAT and firewall at each end of the tunnel are optional, depending on the relationship between the two enterprises.

6.2 Scenario 2 - Communication between NGCNs via an NGN not using a VPN

This scenario typically applies to not closely related enterprises (i.e. no common security, no secure tunnel). It may also apply to closely related enterprises (or even between two locations of the same enterprise) if, due to whatever reasons, they are not able to establish a tunnel through the NGN.

The NGN in this scenario as a minimum acts as a TSP. It may also act as a SSP (provision of session services) and/or as an ASP (provision of applications such as conference, presence, voice mail, unified mail, etc.). Depending on the number of services to be accessed also multiple TSPs, SSPs and/or ASPs may be involved.

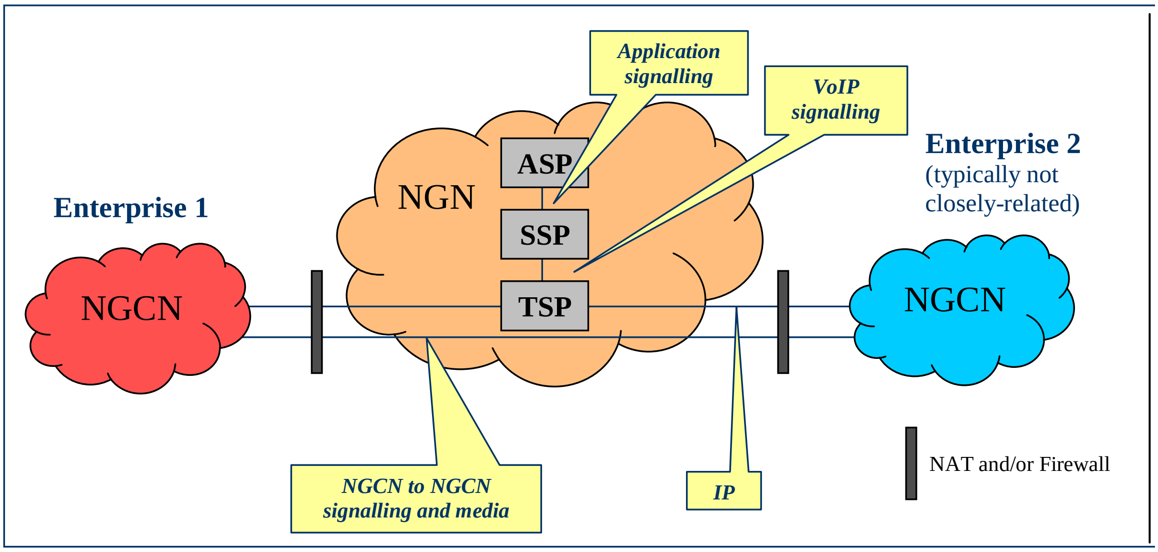

Figure 4: Scenario 2 - Communication between NGCNs not using a VPN

6.3 Scenario 3 - Communication between NGCN and TE via an NGN using a VPN

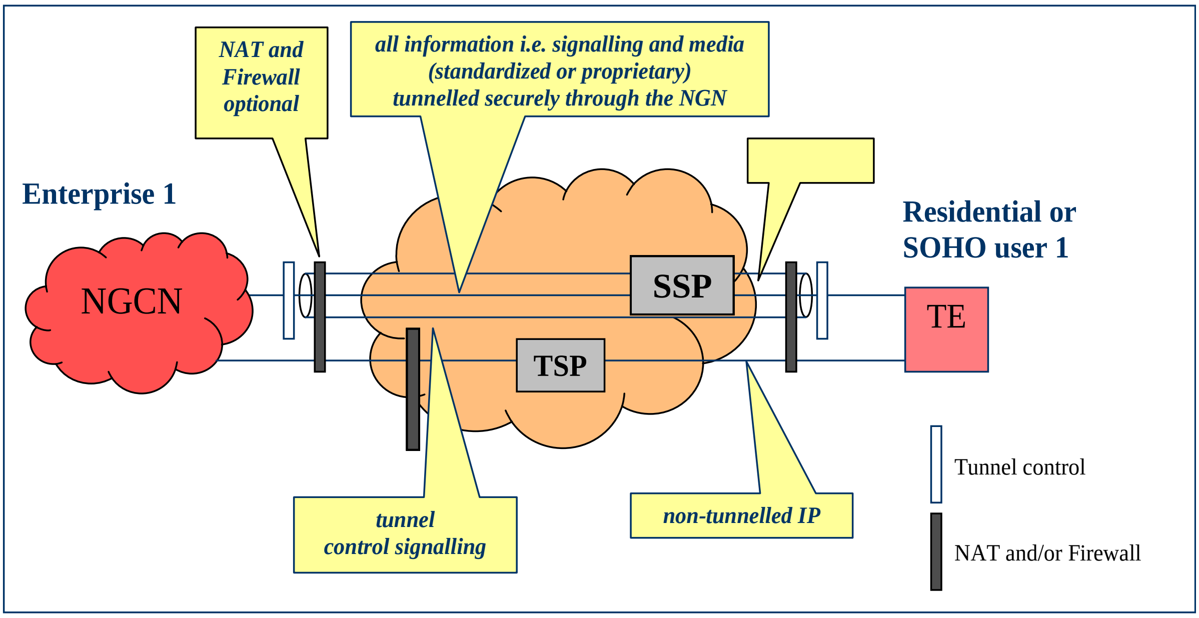
NAT and Firewall optional
all information i.e. signalling and media (standardized or proprietary) tunnelled securely through the NGN
Enterprise 1
Residential or SOHO user 1
NGCN
SSP
TSP
TE
Tunnel control
NAT and/or Firewall
tunnel control signalling
non-tunnelled IP

Figure 5 Scenario 3 - Communication between NGCN and TE using a VPN

NAT and firewall for the tunnel are optional, depending on the relationship between enterprise and TE.

6.4 Scenario 4 - Communication between NGCN and TE via an NGN not using a VPN

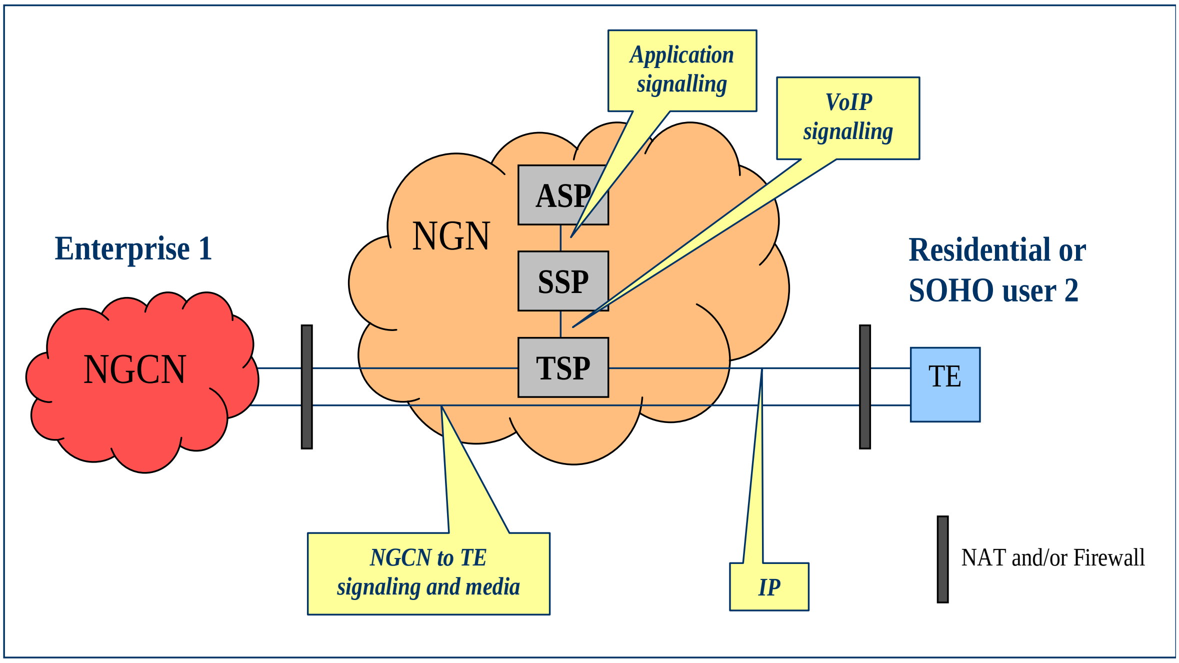

Figure 6: Scenario 4 - Communication between NGCN and TE not using a VPN

6.5 Scenario 5 - Communication between NGCN and PSTN/ISDN via an NGN

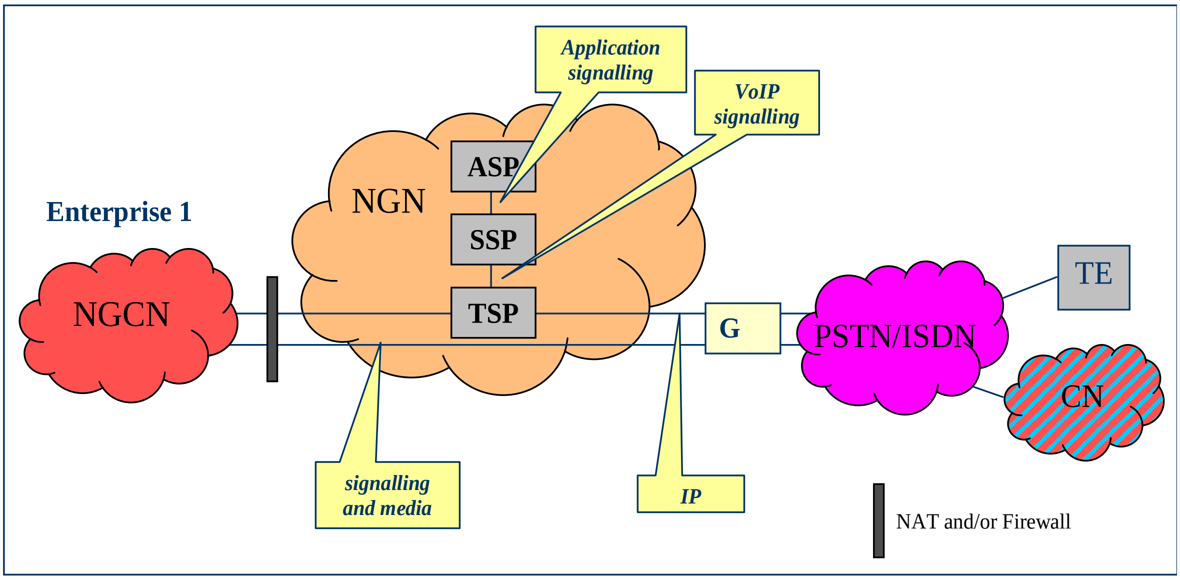

Figure 7: Scenario 5 - Communication between NGCN and PSTN/ISDN via an NGN

The GateWay (GW) may be provided by the NGN as part of its SSP or it may be a 3rd party gateway.

6.6 General requirements on NGNs

In order to support communications involving NGCNs, NGNs need to meet certain requirements. Some generic requirements are identified here. These requirements will apply in different ways to some or all of the scenarios identified in clause 6 and influence the technical considerations of clause 6. Where an NGN is mentioned, this can be either a single NGN or a number of cooperating NGNs.

REQUIREMENT ON NGN 1: An NGN shall support communication sessions established in either direction between an NGCN user and a user in another NGCN, a user in the NGN or a user in a PSTN/IDSN accessed via a gateway in the NGN.

REQUIREMENT ON NGN 2: An NGN shall allow the use of any IP-based media during a communication session subject to the availability of resources and contractual arrangements.

REQUIREMENT ON NGN 3: Except where explicitly agreed with the NGCN (by signalling or contract) or because of formal legal requirements, an NGN shall not intervene in media transport above the transport layer.

NOTE 1: Functions such as transcoding, translation and bridging can be provided where agreed.

NOTE 2: The applicability of formal legal requirements may depend on whether media is transported between two parts of the same NGCN (i.e. within an enterprise) or outside the enterprise.

REQUIREMENT ON NGN 4: An NGN shall offer service level agreements in which quality of service of media, signalling and other data is fit for enterprise use.

REQUIREMENT ON NGN 5: An NGN shall provide industry standard functionality for assisting with traversal of commonly encountered NAT arrangements at an NGCN boundary.

NOTE: In general the endpoint should be responsible for signalling IP addresses and ports that are suitable for use by the remote endpoint, but to achieve this the NGN may need to provide functionality such as a STUN server.

REQUIREMENT ON NGN 6: An NGN shall support identification of users and services by means or SIP or SIPS URIs and optionally by means of telephone URIs, when not in conflict with available end point user contractual or privacy settings.

REQUIREMENT ON NGN 7: An NGN shall provide authenticated caller identity to an NGCN when available, or an appropriate and meaningful datum when not.

REQUIREMENT ON NGN 8: An NGN shall be able to pass an authenticated caller identity from an NGCN towards a called user during call establishment and call re-arrangements.

REQUIREMENT ON NGN 9: An NGN shall offer a standard SIP signalling interface to an NGCN based on standards track RFCs from the IETF.

REQUIREMENT ON NGN 10: Except when providing a gateway to PSTN/ISDN or an application service, an NGN shall be transparent to signalling that is of an end-to-end nature and shall be tolerant of any that is not supported.

NOTE: This includes SIP bodies and certain SIP headers. In the case of SDP bodies the NGN is required to be tolerant of unsupported SDP extensions and next generation SDP.

REQUIREMENT ON NGN 11: When transporting signalling messages, an NGN shall not carry out actions that will invalidate cryptographic integrity checks on parts of messages intended for delivery unchanged, except when required by local regulation.

NOTE: The applicability of formal legal requirements may depend on whether signalling is transported between two parts of the same NGCN (i.e. within an enterprise) or outside the enterprise.

REQUIREMENT ON NGN 12: When transporting signalling messages, an NGN shall allow end-to-end information to be encrypted and shall not require access to the key.

NOTE: Examples are geographic location information and key information for media security. Where certain parts of encrypted end-to-end data need to be accessed, the NGCN should arrange for this to be delivered separately, either encrypted (for decryption by the NGN) or unencrypted.

REQUIREMENT ON NGN 13: An NGN shall be able to provide optional security (encryption, authentication and integrity checking) of signalling services at the NGCN-NGN interface.

REQUIREMENT ON NGN 14: An NGN shall be able to provide security (encryption, authentication and integrity checking) of signalling on each signalling hop between communication partners on request and each NGN shall be able to provide to the NGCN evidence that this is provided.

NOTE: This can be provided through use of the SIPS URI scheme.

REQUIREMENT ON NGN 15: An NGN shall permit media to be secured (encrypted, authenticated and integrity checked) end-to-end.

REQUIREMENT ON NGN 16: An NGN shall permit key management for the purpose of media encryption to take place within signalling between the end devices.

NOTE: Key management exchange between an NGCN user and an NGN intermediary will often be unacceptable to NGCN users and their partners. For example, a residential user calling a bank may require authentication based on a bank-provided certificate before talking. See REQ3 and REQ11.

REQUIREMENT ON NGN 17: An NGN shall be able to provide advice of charge information to an NGCN at the time charges are incurred.

6.7 General requirements on NGNs concerning measures for compliance with regulations

Potential issues for regulatory authorities and administrations are: emergency services, (including emergency telecommunication and disaster relief), lawful interception and privacy (to be considered in the context of presentation of business user's or terminal's identity, name, address and location information).

Several regulatory authorities, e.g. in the USA and in Europe, have started consultations on a regulatory framework for the provision of public VoIP-services.

However, at the time of preparation of the present document it is not clear, whether or to what extent enterprise networks and in particular their internal communications traffic will be affected by those regulations that already apply to public network operators (e.g. on lawful interception, data retention). Beyond that, it is a still an open issue how to define the boundaries of an enterprise network, which, especially in the case of remote access, could share the network infrastructure of a universal services provider to which regulations apply. The challenge is how to handle such virtual enterprise networks.

NOTE: In a circuit switched environment the T-reference point acted as demarcation between private and public network infrastructures and thus the boundary for regulation.

It is expected that enterprise networks, including virtual enterprise networks, will not be subject to more regulation than circuit switched enterprise networks.

REQUIREMENT ON NGN 18: An NGN shall be able to discriminate between normal traffic and intra-enterprise traffic when applying measures for complying with regulations, thus keeping enterprise network internal communications free from any unwanted intervention.

7 Technical issues and requirements on NGN related to session service provision

In order to realize the configurations identified in clause 6, certain technical issues will arise. This clause discusses these technical issues in general terms and also in relationship to specific configurations and derives requirements on NGN in addition to those identified in clause 6.6. This edition of the present document addresses only certain fundamental technical issues, and there are other technical issues that may need further consideration, e.g. charging, routing, transcoding, performance, performance monitoring, resilience, etc., and also applications such as presence.

The issues are discussed with particular reference to SIP-based multi-media communication, but generally apply to any form of multi-media communication, e.g. communication based on H.323 [2] signalling.

7.1 Signalling architecture

Basically SIP is a protocol that operates between two endpoints known as User Agents (UAs), enabling those UAs to establish a communication session comprising one or more media connections (i.e. bearer streams) such as audio, video, fax, instant messaging or real-time text. A session is established by exchanging session descriptions in accordance with the Session Description Protocol (SDP) within SIP messages. Session descriptions indicate factors such as the media to be used, codecs, packet rates, security contexts and IP addresses and port numbers for media reception.

In principle, SIP can operate directly between two UAs. However, the SIP standards do define other SIP entities that can assist. In particular the main SIP standard [4] defines the concept of a location service at which UAs can register to assist in locating UAs representing a particular user. Related to this are the concepts of a registrar (a SIP entity that registers UAs at the location service) and a proxy (a SIP entity that queries the location service in order to assist in the forwarding of requests to appropriate UAs).

NOTE: Also there is a redirect, which is similar to a proxy but redirects rather than forwards requests. For the purposes of the present document a redirect is treated as a proxy unless otherwise stated.

The handling of a SIP request from one UA (the UA client, UAC) to another UA (the UA server, UAS) typically passes through a proxy, which queries the location service in order to locate UAs that might be able to handle the request. Also the UAC often sends requests to a local proxy, known as an outbound proxy) in order to reduce the routing burden on the UAC. This gives rise to the typical SIP trapezoid involving two UAs and two proxies (see figure 9).

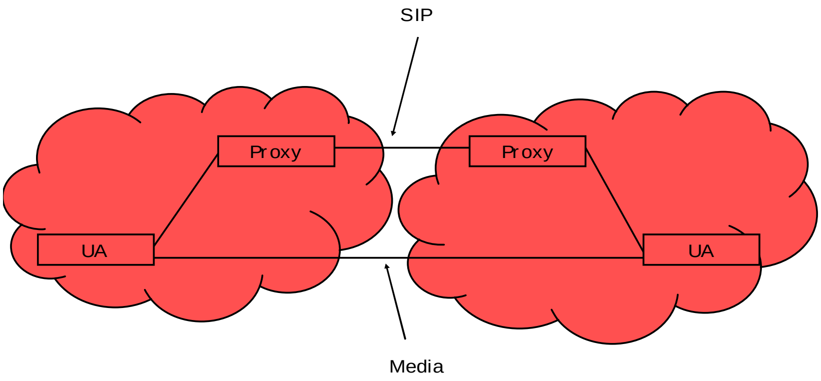

Figure 8: Typical SIP trapezoid

Sometimes a SIP request can result in the formation of an association between two UAs, in the context of which further SIP requests in either direction can be sent. This is known as a dialog. The most important dialog in SIP is that initiated by an INVITE request, because such a dialog has an associated session, comprising a collection of media between the UAs. An INVITE-initiated dialog and its associated session can be regarded as a call. Proxies involved in routing the dialog-initiating request may or may not remain involved in the dialog for routing subsequent requests.

Typically a proxy is collocated with a location service and a registrar, so the whole entity tends to be known as a proxy. The behaviour of a proxy is standardized in SIP [4] and other SIP RFCs and its behaviour is quite tightly constrained by the standards in order not to jeopardize the end-to-end character that is fundamental to much of SIP.

Although not defined in the SIP [4] the market has found the need for an entity known as a Back-to-Back User Agent (B2BUA). In some respects a B2BUA acts like a proxy, being an intermediate entity on a signalling path between two UAs, but in other respects a B2BUA acts as two UAs back-to-back. The behaviour of a B2BUA is not standardized, and because it is not constrained by standards a B2BUA can behave more flexibly than a proxy. This allows a B2BUA to fulfil certain roles that a proxy cannot (e.g. by offering transcoding capabilities for media). On the other hand by not behaving as a proxy a badly designed B2BUA can behave in ways that are inconsistent with the expectations of UAs and can cause problems.

Other forms of entity that can intervene either actively or passively in a SIP signalling path include Application Layer Gateways (ALGs) and Session Border Controllers (SBCs). These have knowledge of the SIP protocol and can monitor or modify it for certain purposes, e.g. control of NATs and firewalls, security, privacy, etc.

In the rest of this clause, the term SIP intermediary is used to refer to any intermediate entity involved either actively or passively in SIP signalling between two UAs, including but not limited to proxies, B2BUAs, ALGs and SBCs.

As stated above, SIP can in principle operate directly between UAs, and indeed there is a lot of interest at present in peer-to-peer (P2P) SIP where SIP intermediaries are eliminated. Whether and how soon this ideal will be realized is open to speculation, but the trend is expected to be away from the TDM concept of signalling passing through an arbitrary number of intermediaries (e.g. associated with switches) towards a lightweight solution with a minimal number of SIP intermediaries. There should generally be no need for more SIP intermediaries than in the typical SIP trapezoid mentioned above. An excessive number of SIP intermediaries can have the negative impact of adding to signalling delays (e.g. during call establishment). Also undue intervention by any SIP intermediary can have a harmful effect by not complying with the expectations of UAs.

These lead to general requirements on NGNs that act as SSPs as follows. More specific requirements apply in the context of specific scenarios and specific technical issues.

REQUIREMENT ON NGN 19: SIP intermediaries in a signalling path should be kept to a minimum.

REQUIREMENT ON NGN 20: A SIP intermediary in an NGN shall behave in a way that is not harmful to UAs that are trying to communicate through it. A SIP intermediary that obeys all the rules of a proxy would be compliant with this. This includes transparency to SIP signalling information that is not relevant to the intermediary, including any unsupported SIP extensions (standardized or proprietary).

The above requirements apply also where two or more NGNs in series act as SSPs for a given communication.

Although SIP signalling typically passes through one or more SIP intermediaries, media in general flow directly between the UAs concerned, with no involvement of intermediaries above the transport layer (intermediaries such as NATs and firewalls can be involved up to the transport layer). The exception to this is when an intermediary performs some function on the media such as transcoding, translating or bridging, in which case the intermediary really acts as two or more UAs in a back-to-back arrangement (similar to and perhaps coincident with a SIP B2BUA).

7.1.1 Scenario 1 - Communication between NGCNs using a VPN

Two SIP intermediaries, one in each part of the NGCN, should suffice. The NGN acts only as a TSP. In any case, encryption through the tunnel will mean that the NGN cannot provide any SIP intermediaries. Media as well as SIP signalling will flow through the tunnel.

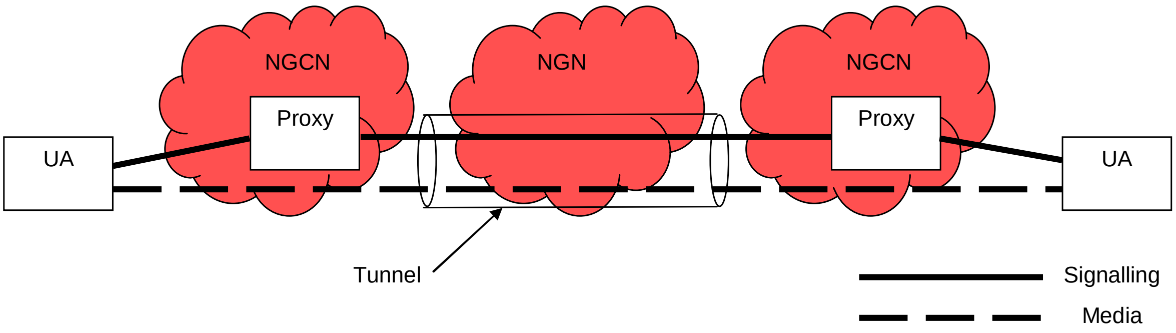

Figure 9: Typical signalling architecture for scenario 1

7.1.2 Scenario 2 - Communication between NGCNs not using a VPN

Two SIP intermediaries, one in each NGCN, should in principle suffice. SIP intermediaries in the NGN should be provided only if they allow the NGN to add value, e.g. by providing QoS guarantees for media. SIP intermediaries in the NGN should be transparent to SIP signalling information that is not relevant to those intermediaries, including any unsupported SIP extensions (standardized or proprietary). Whether or not the NGN provides any SIP intermediaries, media may or may not flow via the same NGN.

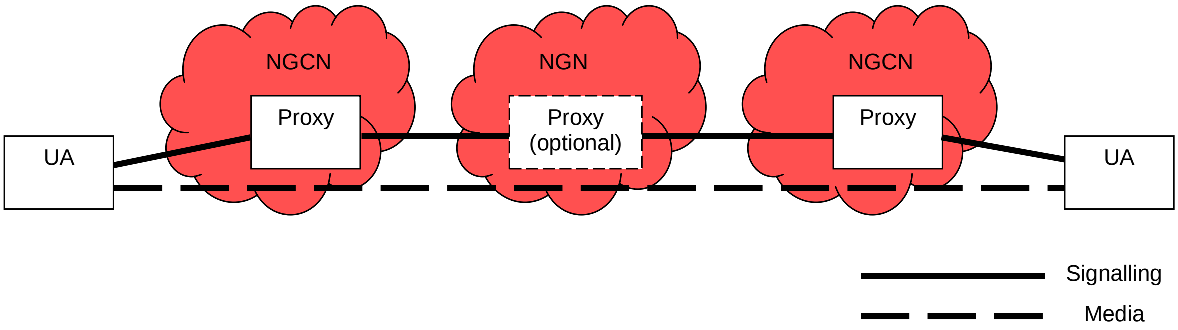

Figure 10: Typical signalling architecture for scenario 2

7.1.3 Scenario 3 - Communication between NGCN and TE using a VPN

SIP intermediaries should be confined to the NGCN. The NGN acts only as a TSP. The TE will incorporate a UA that communicates with an inbound / outbound proxy or other intermediary in the NGCN via the tunnel. Media as well as SIP signalling will flow through the tunnel.

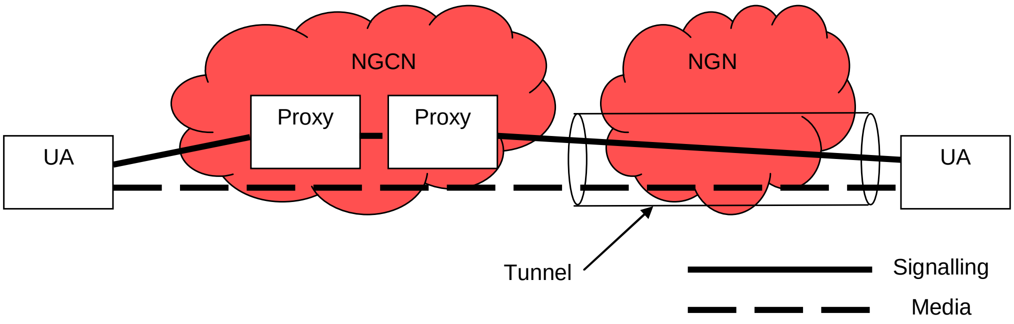

Figure 11: Typical signalling architecture for scenario 3

7.1.4 Scenario 4 - Communication between NGCN and TE not using a VPN

Two SIP intermediaries, one in the NGCN and one in the NGN, should suffice. Additional SIP intermediaries in the NGN should be provided only if they allow the NGN to add value, e.g. by providing QoS guarantees for media. SIP intermediaries in the NGN should be transparent to SIP signalling information that is not relevant to those intermediaries, including any unsupported SIP extensions (standardized or proprietary). Media may or may not flow through the same NGN.

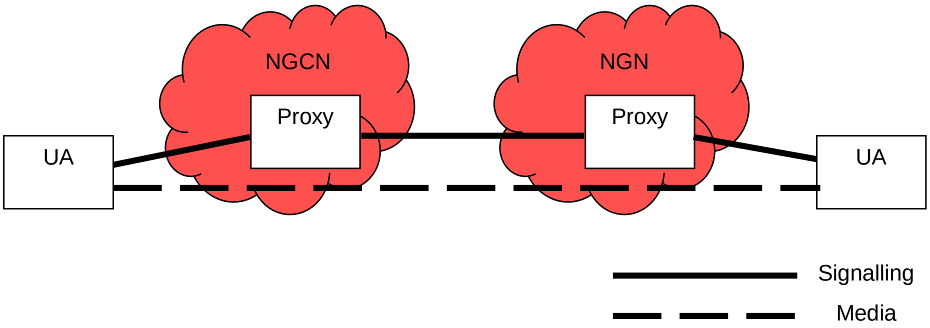

Figure 12: Typical signalling architecture for scenario 4

7.1.5 Scenario 5 - Communication between NGCN and PSTN/ISDN

Two SIP intermediaries, one in the NGCN and one in the NGN, should suffice. The SIP intermediary in the NGN will act as the inbound / outbound proxy for the gateway UA. Additional SIP intermediaries in the NGN should be provided only if they allow the NGN to add value. Media may or may not flow through the same NGN.

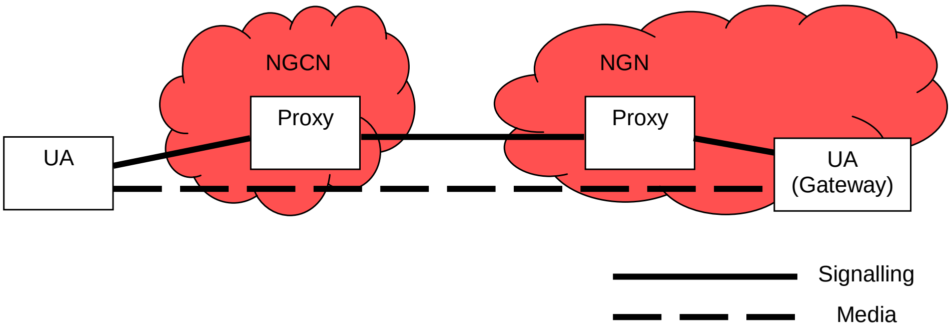

Figure 13: Typical signalling architecture for scenario 5

7.2 NAT traversal

Network Address Translation (NAT) devices are typically present at network borders to provide a binding, typically on a temporary basis, between IP addresses and ports belonging to the internal address space and IP addresses and ports belonging to the external address space. A NAT provides several benefits to the network concerned, including provision of a larger address space internally without using a corresponding amount of public address space and the firewall effect of not allowing packets to flow from the external network to the internal network (inbound) without first opening a binding by transmitting packets from the internal network to the external network (outbound).

NAT traversal can be a significant issue for SIP-based communication, firstly because SIP signalling may need to traverse a NAT and secondly because media streams may need to traverse a NAT.

7.2.1 NAT traversal for SIP signalling

The first NAT traversal problem concerns SIP signalling. In general this is not a problem if communication is initiated from inside the NAT, since the first outbound packet will open a binding, which will then apply to both outbound and inbound packets.

For the typical transport mechanism, Transmission Control Protocol (TCP) (including Transport Layer Security (TLS) over TCP), this means the connection must be established in the outbound direction. It also means the connection must be retained even after completion of any outbound transactions (outbound request and inbound response) in case inbound requests arrive subsequently. By refreshing a connection periodically, NAT bindings can be kept alive. Where there are multiple SIP entities on one side or both sides of the NAT. Multiple TCP connections might need to be established and maintained for load sharing purposes.

For User Datagram Protocol (UDP), an initial packet must be sent outbound to open the binding. Because a response is not associated at the transport level with the request, the response to an outbound request needs to be sent explicitly to the port from which it was received rather than to the port indicated in the Via header of the SIP request. Keep-alive packets need to be sent periodically to keep the binding open for inbound requests. Because of the unreliability of UDP and its inability to be secured by TLS and also the limited maximum payload size of UDP, it is considered an unlikely transport protocol for NGN access and therefore is not considered further in the present document.

7.2.1.1 Scenario 1 - Communication between NGCNs using a VPN

NAT traversal should not be an issue if the VPN tunnel links two parts of an NGCN sharing the same IP address space, even though the VPN may extend across any number of NATs.

If the VPN tunnel links two parts of an NGCN having different IP address spaces (e.g. two closely related enterprises), this is equivalent to the use of NATs between different parts of an NGCN when there is no intervening NGN. Since it is not an issue arising from the involvement of an NGN, it is considered outside the scope of the present document.

7.2.1.2 Scenario 2 - Communication between NGCNs not using a VPN

In this scenario there is a need for SIP requests to be sent from the SIP intermediary in one NGCN to the SIP intermediary in the other NGCN. In this scenario it is likely that a NAT will exist at the boundary between an NGCN and the NGN and likewise between the peer NGCN and the NGN. Assuming the NGCN SIP intermediaries are located on the internal side of their respective NATs, SIP requests (and responses) may have to traverse two NATs. Each NGCN SIP intermediary must therefore have a globally routable IP address (or an IP address from the address space in use within the NGN) in order to receive SIP requests from SIP intermediaries in other NGCNs and each NAT must be configured to allow unsolicited inbound packets from any source to be mapped through to the appropriate internal IP address and port of the SIP intermediary. Otherwise unsolicited SIP requests would be unable to reach the SIP intermediary. A SIP intermediary cannot know where unsolicited requests will come from and therefore cannot open NAT bindings in advance by issuing outbound packets.

Alternatively the SIP intermediary could be located on the external side of the NAT, i.e. between the NAT and the NGN. In this case NAT traversal is an intra-NGCN matter.

Additional considerations will apply where requests pass through multiple NGNs having different IP address spaces or when SIP intermediaries exist within the NGN(s).

7.2.1.3 Scenario 3 - Communication between NGCN and TE using a VPN

The considerations of clause 7.2.1.1 apply, where the TE acts as a part of the NGCN.

7.2.1.4 Scenario 4 - Communication between NGCN and TE not using a VPN

In this scenario there is a need for SIP requests to be sent from the SIP intermediary in the NGCN to the SIP intermediary in the NGN and vice versa. It is likely that a NAT will exist at the boundary between the NGCN and the NGN. Assuming the NGCN SIP intermediary is located on the internal side of its NAT, SIP requests (and responses) have to traverse the NAT. The NGCN SIP intermediary must therefore have a globally routable IP address (or an IP address from the address space in use within the NGN) in order to receive SIP requests from SIP intermediaries in the NGN and the NAT must be configured to allow unsolicited inbound packets from any source to be mapped through to the internal IP address and port of the NGCN SIP intermediary. Otherwise unsolicited SIP requests would be unable to reach the NGCN SIP intermediary. It is assumed that the NGCN SIP intermediary cannot know where unsolicited requests will come from and therefore cannot open NAT bindings in advance by issuing outbound packets.

Additional considerations will apply where requests pass through multiple NGNs having different IP address spaces or when SIP intermediaries exist within the NGN(s).

7.2.1.5 Scenario 5 - Communication between NGCN and PSTN/ISDN

Considerations are similar to scenario 4.

7.2.2 NAT traversal for media streams

The second and more serious issue with NAT traversal concerns media streams. IP addresses and ports for reception of media are signalled in SIP messages using the Session Description Protocol (SDP). If an internal UA signals its reception IP address and port to an external UA, the internal IP address and port will in general be useless to the external UA. The address and port may not exist in the external address space or they may be assigned to a different device. Therefore the external IP address and port need to be signalled in SDP. To do this the internal UA needs to discover its external IP address and port.

For most situations in which the medium is transported over UDP, STUN (Simple Traversal of UDP through NATs) [5] provides an adequate means of discovering the IP address and port. The internal UA sends a request to a STUN server in the external network and the STUN server responds with the IP address and port from which the request appeared to have come, i.e. the external IP address and port. This requires a STUN server in the external network. The act of sending a STUN request will open a NAT binding if one did not already exist, and this same NAT binding is used for the medium itself. There are a few NAT types for which this does not work, but these are believed to be quite uncommon.

NOTE: In particular it does not work for the NAT type commonly known as "symmetric", where a binding is to a particular address and port on the public side, and therefore a binding opened to the STUN server cannot be used for media from a different address and port.

Although STUN provides a means of discovering the external IP address and port, there is a further issue of determining whether NAT traversal is applicable, since if the peer UA is internal the internal IP address and port should be used, and not the external IP address and port. Furthermore, if the internal network has interfaces to more than one external network, different NAT bindings will apply to each and it is important to know which external network applies and obtain the external IP address and port applicable to that external network. Further complications arise if more than one NAT is traversed between the two UAs. Most of these problems can be solved if the UAs concerned implement

ICE (Interactive Connectivity Establishment) (ongoing work in the IETF), which provides a means for candidate IP addresses and ports to be tested before use.

Other considerations apply to media transported over TCP, e.g. text messages within the context of a session in accordance with the Message Session Relay Protocol (MSRP) (ongoing work in the IETF). These are not addressed in this edition of the present document.

Another approach sometimes taken to solve the problem of NAT traversal for media streams is the use of an Application Layer Gateway (ALG) that is on the SIP signalling path (it may be combined with a SIP intermediary, for example) and opens NAT bindings and adjusts SDP addresses and ports accordingly. This is contrary to SIP architectural principles, since it requires examination and modification of a SIP body (SDP) by a SIP intermediary. In particular, examination will fail if the body is encrypted and modification will fail if the body is integrity protected. Therefore this approach cannot be recommended.

Yet another approach involves the use of a policy server accessed by UAs to establish NAT bindings (see clause 7.7).

7.2.2.1 Scenario 1 - Communication between NGCNs using a VPN

The considerations in clause 7.2.1.1 apply.

7.2.2.2 Scenario 2 - Communication between NGCNs not using a VPN

For a medium transported over UDP, each UA must determine the external IP address and port corresponding to the internal IP address and port on which that medium will be received and include this in SDP offer or answer to the peer UA. The use of ICE may be required to determine whether this is required. Each UA will need to use a STUN server in the NGN to discover its own external IP address and port.

REQUIREMENT ON NGN 21: The NGN is required to provide a STUN server.

Alternatively the UAs may make use of policy servers (see clause 7.7).

7.2.2.3 Scenario 3 - Communication between NGCN and TE using a VPN

The considerations in clause 7.2.1.3 apply.

7.2.2.4 Scenario 4 - Communication between NGCN and TE not using a VPN

For a medium transported over UDP, a UA in the NGCN must determine the external IP address and port corresponding to the internal IP address and port on which that medium will be received and include this in SDP offer or answer to the peer UA. The use of ICE may be required to determine whether this is required. The NGCN UA will need to use a STUN server in the NGN to discover its own external IP address and port. The use of ICE requires corresponding support at UAs in the NGN.

Alternatively the UAs may make use of policy servers (see clause 7.7).

7.2.2.5 Scenario 5 - Communication between NGCN and PSTN/ISDN

For a medium transported over UDP, a UA in the NGCN must determine the external IP address and port corresponding to the internal IP address and port on which that medium will be received and include this in SDP offer or answer to the gateway. The use of ICE may be required to determine whether this is required. The NGCN UA will need to use a STUN server in the NGN to discover its own external IP address and port. The use of ICE requires corresponding support at the gateway.

Alternatively the UAs may make use of policy servers (see clause 7.7).

REQUIREMENT ON NGN 22: Gateways in an NGN may need to support ICE.

7.3 Firewall traversal

Firewalls are generally present at network borders to prevent unwanted traffic across the border, particularly incoming traffic. NATs also provide a partial firewall capability by virtue of not allowing bindings to be opened by requests from the external side to the internal side, and this is dealt with in clause 7.1. Firewalls can impose additional restrictions, some of which can be harmful to SIP-based multi-media communication (e.g. the denial of all UDP traffic).

In general, the opening of firewall holes by sending initial packets from inside the firewall to outside the firewall will help. This needs to be done anyway for NAT traversal.

Firewalls can normally be expected to be configured to allow SIP signalling to pass through, since SIP signalling normally involves specific SIP intermediary IP addresses and specific port numbers. Media streams are likely to be a far bigger problem because port numbers are chosen dynamically and it is unlikely that firewalls will keep holes open for the full range of possibilities.

If firewalls can be opened in each direction by sending an initial packet in the outbound direction, then this might be an acceptable solution. Sending and receiving on the same port is generally a pre-requisite for this to work.

An approach sometimes taken to solve the problem of firewall traversal for media streams is the use of an Application Layer Gateway (ALG) that is on the SIP signalling path (it may be combined with a SIP intermediary, for example) and opens firewall holes based on IP addresses and ports in SDP bodies of SIP messages. This is contrary to SIP architectural principles, since it requires examination of a SIP body (SDP) by a SIP intermediary. In particular, examination will fail if the body is encrypted. It may also fail if there is information in the SDP that needs to be taken into account but is not understood by the ALG. End-to-middle security studies in the IETF may provide a solution to the encryption problem.

Another approach involves the use of a policy server accessed by UAs to open firewall holes (see clause 7.7).

7.3.1 Scenario 1 - Communication between NGCNs using a VPN

Firewall traversal should not be an issue if the VPN tunnel passes across any intervening firewalls. VPN tunnel establishment is outside the scope of the present document.

7.3.2 Scenario 2 - Communication between NGCNs not using a VPN

Firewalls will generally exist at the edge of each NGCN, protecting the NGCN from the NGN. For cases where firewalls are not opened simply by sending outbound packets, other solutions will be required. Holes could be opened by the SIP intermediary in the NGCN concerned, but this is subject to the difficulties described in clause 7.3. A preferable solution would be for each UA to use a policy server to open holes, based on the send and receive IP addresses and ports exchanged via SDP.

Firewalls might also exist at the edge of the NGN, protecting the NGN from each NGCN. In this case, the use of outbound packets to open holes is not possible. The opening of holes by a SIP intermediary in the NGN would be subject to the difficulties described in clause 7.3. The use of an NGN policy server by each UA (under instruction from a SIP intermediary in the NGN) would be preferable.

REQUIREMENT ON NGN 23: If an NGN provides firewalls, it shall provide a means to open holes through the firewall for media between two NGCN UAs without relying on the ability to examine SDP passing between those UAs.

7.3.3 Scenario 3 - Communication between NGCN and TE using a VPN

The considerations in clause 7.3.1 apply.

7.3.4 Scenario 4 - Communication between NGCN and TE not using a VPN

The considerations in clause 7.3.2 apply.

7.3.5 Scenario 5 - Communication between NGCN and PSTN/ISDN

The considerations in clause 7.3.2 apply, except that here it may be possible for media packets transmitted by the gateway to open a firewall in the NGN at the boundary to the NGCN.

7.4 Identification

With the SIP as the means of signalling, identification of users and other entities is assumed to be by means of sip (or sips) Uniform Resource Identifiers (URIs) [4] or tel URIs [8].

A SIP URI is of the form SIP:user@domain. Each domain (as identified by the domain part) is responsible for allocating user parts to its users (human users, services, etc.). An NGCN will comprise one or more domains. A request from an NGN towards an entity in an NGCN will normally bear the SIP URI of that entity as its desired destination. The NGN only needs to understand the domain part of the SIP URI in order to route such a request to the NGCN domain, which is then responsible for routing to the entity concerned. Thus there is no need for an NGN to have prior knowledge of user parts within an NGCN domain, and an NGCN should be at liberty to assign or revoke SIP URIs within its domain(s) without reference to the NGN.

However, it is necessary that any NGCN entity that needs to be reachable on calls from PSTN/ISDN has an identity that can be mapped to/from an E.164 [10] telephone number that is one of a block of telephone numbers allocated to the NGCN. This could be achieved by use of tel URIs, whereby an identifiable entity in the NGCN has a tel URI as its sole or alternative identity. If the entity also has a SIP identity, this could be simply derivable from the tel URI, e.g. by having the telephone number as the user part of the SIP URI. Either the NGCN or the NGN could be responsible for mapping telephone numbers to/from SIP URIs.

For a call from PSTN/ISDN to an entity in an NGCN, the NGN shall first identify the NGCN domain concerned (e.g. by ENUM [7] look-up). There are then various possibilities for proceeding:

- the NGN forms a tel URI and uses this as the desired destination of the request to the NGCN, which then routes using the tel URI to the destination UA;
- the NGN forms a tel URI and uses this as the desired destination of the request to the NGCN, which then converts to a SIP URI for onward routing;
- The NGN converts the telephone number to a SIP URI for the NGCN domain concerned and uses this as the desired destination of the request to the NGCN.

REQUIREMENT ON NGN 24: An NGN shall be able to accept any user part in a SIP URI for an NGCN domain as valid.

REQUIREMENT ON NGN 25: An NGN shall be able to map telephone numbers assigned to an NGCN to/from URIs that are valid within that NGCN.

7.5 Provision of identification information

The recipient of a SIP request (the UAS) often requires reliable identification of the source of the request. Reliable identification is crucial to the filtering of unwanted requests, e.g. to support the use of black lists and white lists.

There are two basic methods of achieving this: hop-by-hop assertion of identity or cryptographically signed identity.

With hop-by-hop assertion, each SIP entity in the signalling path receives an assertion of the identity of the source of the request from the previous SIP entity, which has been authenticated by some means. The outbound SIP intermediary for the UAC will in general be able to authenticate the UAC and therefore can assert the UAC's identity to the next SIP entity in the signalling path, which in turn can assert that identity to the next entity. This only works if there is an unbroken chain of trust and authenticated signalling hops. If an entity does not trust the asserted identity from the preceding entity, it should not use that information and should not pass it on to other entities.

A cryptographically signed identity can be inserted by the UAC itself (if the UAC has a private key and certificate suitable for this purpose) or by an identity server within the same domain (e.g. collocated with the outbound SIP intermediary) that has authenticated the UAC by other means. A recipient of a cryptographically signed identity can assess whether to trust the identity based on the certificate of the signer.

There is a similar requirement for a UAC to be able to obtain reliable identification of the source of a response. Because a request can be retargeted, a response does not necessarily come from an entity with the same identity as that in the original request. Because of technical difficulties, no solution has been found yet to this problem.

Provision of identification information will depend on privacy requirements of the user concerned. If an NGCN does not wish his identity to be disclosed to other parties, the NGCN will either not supply the identity to the NGN or, if there is a trust relationship in place, pass the identity with a privacy marking. In the latter case the NGN is responsible for ensuring that it is not disclosed.

Similar privacy considerations apply in the opposite direction.

7.5.1 Scenario 1 - Communication between NGCNs using a VPN

This is an intra-NGCN or inter-NGCN problem with no impact on NGN.

7.5.2 Scenario 2 - Communication between NGCNs not using a VPN

If an NGCN provides a cryptographically signed identity, this is an inter-NGCN problem with no impact on NGN (other than transport).

If there is no SIP intermediary in the NGN, an NGCN could provide an asserted identity to the other NGCN, although this might not be accepted unless there is a trust relationship between the two NGCNs.

If there is a SIP intermediary in the NGN, an asserted identity would depend on transitive trust: the NGN trusting the first NGCN and the second NGCN trusting the NGN.

7.5.3 Scenario 3 - Communication between NGCN and TE using a VPN

The considerations in clause 7.5.1 apply.

7.5.4 Scenario 4 - Communication between NGCN and TE not using a VPN

For identification information sent from the NGN to the NGCN, the NGCN should be able to accept either an asserted identity or a signed identity. The former requires trust in the NGN and the latter requires trust in the signer of the identification information as indicated by a certificate.

REQUIREMENT ON NGN 26: An NGN shall be able to supply either an asserted identity or a cryptographically signed identity to an NGCN.

In the opposite direction the NGCN could supply either an asserted identity or a signed identity. If the NGN regards the NGCN as a peer it may accept an asserted identity and in turn assert this identity when forwarding the request towards the UAS.

REQUIREMENT ON NGN 27: An NGN should be able to accept an asserted identity as the source of a request from an NGCN if that NGCN is trusted.

Otherwise the NGN might only accept a cryptographically signed identity and even then may require the signer's certificate to be signed by a known Certification Authority (CA) in order to trust it. Whether or not the signed identity is trusted by the NGN, the NGN should still be able to forward the signed identity with the request towards the UAS. The UAS or the end user can then decide whether to trust the identity.

REQUIREMENT ON NGN 28: An NGN shall be able to accept a cryptographically signed identity as the source of a request from an NGCN if the NGCN does not provide an asserted identity or if the NGN is unable to trust an asserted identity.

7.5.5 Scenario 5 - Communication between NGCN and PSTN/ISDN

For identification information sent from the NGN to the NGCN the NGCN should be able to accept either an asserted identity or a signed identity. The former requires trust in the NGN and the latter requires trust in the signer of the identification information as indicated by a certificate.

REQUIREMENT ON NGN 29: An NGN shall be able to supply either an asserted identity or a cryptographically signed identity to an NGCN.

In the opposite direction the NGCN could supply either an asserted identity or a signed identity. If the NGN regards the NGCN as a peer it may accept an asserted identity and in turn use this asserted identity to derive an identity to be sent to the PSTN/ISDN.

Otherwise the NGN might only accept a cryptographically signed identity and even then may require the signer's certificate to be signed by a known certification authority (CA) in order to trust it and derive an identity to be sent to the PSTN/ISDN.

In either case the NGCN should supply an identity from which the NGN can derive a telephone number.

REQUIREMENT ON NGN 30: An NGN shall be able to accept a cryptographically signed identity as the source of a request from an NGCN if the NGCN does not provide an asserted identity or if the NGN is unable to trust an asserted identity.

7.6 Security

Security is a major consideration for the deployment of any application on an IP network, because of the ability for an attacker with access to the network to eavesdrop, impersonate, modify data in transit, etc. Whilst obviously true for the public Internet, it is also true for "closed" IP networks (e.g. within an enterprise). In fact, security is a major consideration for NGNs, for NGCNs and for interworking of NGNs and NGCNs.

One aspect of security is privacy. Privacy considerations in the context of NGCNs are expected to differ compared with considerations for NGNs and are likely to be governed by the requirements of the enterprise concerned rather than regulatory requirements. Whereas in public networks the regulator might have a public interest in collecting information about certain users and their communications, this would not be acceptable for intra-NGCN communications between business users of a single NGCN.

REQUIREMENT ON NGN 31: An NGN shall be able to discriminate intra- from extra-NGCN communications, and thus to be able to decide where and when regulatory requirements (e.g. lawful interception, provision of location information, etc.) may apply and where such requests shall be refused.

For SIP-based multi-media communications, security has to be considered both for signalling (SIP) and for media.

7.6.1 Signalling security

Signalling security is important for several reasons. First, it is important to authenticate the source of a message, in order to authorize any action to be taken on the message. Otherwise an attacker could gain unauthorized access to services or information or carry out a denial of service attack. Secondly, even if the source of a message is authenticated, it is important to be sure that data integrity has not been compromised en route. Thirdly, signalling can contain sensitive information (e.g. the identities of parties in communication, IP addresses and ports to be used for media reception, etc.). This information needs to be protected against eavesdropping.

Because SIP uses SIP intermediaries, signalling between two UAs in general comprises one or more hops (e.g. from UA to SIP intermediary, SIP intermediary to next SIP intermediary, etc.). Signalling can be secured separately on each hop using general purpose security protocols at the transport layer (TLS) or the network layer (IPSEC [9]). These can provide authentication, integrity protection and secrecy. Authentication is achieved using public key cryptography, whereby an entity has a private key for signing messages and a certificate that it can publish to allow other entities to verify its signature. Between two SIP intermediaries mutual certificate-based authentication is the norm. However, between a SIP UA and its local SIP intermediary, a shared secret is normally used to allow the SIP intermediary to authenticate the UA (a certificate still being used to allow the UA to authenticate the SIP intermediary). An established secure connection should normally be retained on a semi-permanent basis for use by multiple SIP transactions in either direction, because of the overhead involved in establishing a new secure connection for each transaction.

Security should be considered essential for any SIP signalling hop, whether it be wholly within an NGCN, wholly within an NGN or spanning more than one network. For any given signalling transaction security will be expected on every hop linking the peer UAs (hop-by-hop security). The SIPS URI scheme is aimed at assuring hop-by-hop security, although it falls short of mandating the method of achieving security on a request's last hop, i.e. from the intermediary responsible for the domain in the request URI to the UAS (TLS is mandated on all other hops).

However, from the perspective of the UA, SIPS and hop-by-hop security might be insufficient. It relies on the UAs trusting the SIP intermediaries to conform to the SIP specification and not to disclose or modify information that is intended for end-to-end use. Whilst there is some information that SIP intermediaries need to inspect and perhaps also modify for routing or other purposes, there is other information that should not be relevant for SIP intermediaries. If the integrity or secrecy of that other information is important to the UAs, they may require that information to be secured end-to-end.

For example, the session description carried in SDP might be relevant only to the UAs and not to SIP intermediaries. Thus SDP is a candidate for being secured end-to-end. In particular, with end-to-end encryption, IP addresses and ports need not be revealed to SIP intermediaries. More importantly, encryption keys for media security need not be revealed.

Another example is the transport of geographic location information between UAs. In general this is not relevant to SIP intermediaries (emergency calls excepted).

SIP allows bodies of SIP messages to be encrypted, authenticated and integrity protected using S/MIME. To use this capability, UAs need private keys and certificates, implying the need for a public key infrastructure (PKI) to provide this information. Similar considerations apply to key management for media security (see clause 7.6.2), and indeed S/MIME is one mechanism that can be used for securing key management. The implications of this are discussed in the context of key management for media security.

On the other hand, SIP intermediaries may require access to some aspects of SDP, (e.g. for ensuring firewall traversal or ensuring bandwidth availability). This might lead to the need for "end-to-middle" security, whereby a UA can secure information between itself and a known SIP intermediary.

7.6.1.1 Scenario 1 - Communication between NGCNs using a VPN

The VPN tunnel will provide encryption and integrity protection of SIP signalling as it traverses the NGCN. However hop-by-hop and end-to-end security of SIP signalling may still be required, but this is an internal matter for the NGCN and outside the scope of the present document.

7.6.1.2 Scenario 2 - Communication between NGCNs not using a VPN

SIP signalling will normally be routed via a SIP intermediary in each NGCN. The hop of interest is that between these two SIP intermediaries. It is essential that this hop be secured by TLS or IPSEC since it traverses an NGN (TLS according to SIPS). The two SIP intermediaries will need to accept each other's certificate, which implies possession of the Certification Authority (CA) certificate concerned.

If the NGN does provide a SIP intermediary, the considerations in clause 7.6.1.4 apply.

End-to-end security may be applied between the UAs for certain bodies of SIP messages. If either of the NGCN SIP intermediaries requires access to encrypted data, then end-to-middle security techniques will need to be applied. This is an NGCN matter with no impact on the NGN.

7.6.1.3 Scenario 3 - Communication between NGCN and TE using a VPN

The considerations in clause 7.6.1.1 apply.

7.6.1.4 Scenario 4 - Communication between NGCN and TE not using a VPN

SIP signalling will normally be routed via a SIP intermediary in the NGCN and a SIP intermediary in the NGN. The hop of interest is that between these two SIP intermediaries. It is essential that this hop be secured by TLS or IPSEC (TLS according to SIPS), since part of it traverses the NGN. The two SIP intermediaries will need to accept each other's certificate, which implies possession of the CA certificate concerned.

REQUIREMENT ON NGN 32: An NGN shall be able to support establishment of secure connections from an NGN SIP intermediary to an NGCN SIP intermediary and vice versa and retain connections on a semi-permanent basis for use by multiple SIP transactions in either direction.

REQUIREMENT ON NGN 33: An NGN shall be able to hold the CA certificate for a given NGCN and be able to authenticate an NGCN SIP intermediary's certificate during establishment of a secure connection to/from that SIP intermediary.

REQUIREMENT ON NGN 34: An NGN SIP intermediary shall have a private key and associated certificate and use them to allow authentication by an NGCN SIP intermediary.

Furthermore, it is essential that the NGN conform to the rules of SIPS to give users a guarantee of hop-by-hop security (subject to trust).

REQUIREMENT ON NGN 35: An NGN shall support SIPS by not allowing a request to a SIPS URI to be forwarded on an insecure link.

End-to-end security may be applied between the UAs for certain bodies of SIP messages. If either the NGCN SIP intermediary or the NGN SIP intermediary requires access to encrypted data, then end-to-middle security techniques will need to be applied.

REQUIREMENT ON NGN 36: An NGN SIP intermediary shall not require to modify end-to-end SIP signalling.

REQUIREMENT ON NGN 37: Ideally an NGN SIP intermediary should not require read access to end-to-end secured signalling. If it does it will need to employ end-to-middle security techniques, but this will require the cooperation of the UAs concerned and might not be acceptable to users or their enterprises.

7.6.1.5 Scenario 5 - Communication between NGCN and PSTN/ISDN

The considerations in clause 7.6.1.4 apply, except that here the gateway in the NGN is one of the UAs and therefore will have access to secured end-to-end signalling, if applicable. For this purpose the gateway will need a private key and certificate, which the UA in the NGCN will need to be able to verify in order to authenticate the gateway.

REQUIREMENT ON NGN 38: An NGN shall support SIPS on calls to/from a PSTN/ISDN gateway.

7.6.2 Media security

Media security is discussed here in terms of security of real-time media (voice and video) transported over RTP. Media transmitted over RTP can be secured by means of Secure RTP (SRTP) [6], which extends RTP by allowing media to be encrypted and by adding information for authentication and integrity checking. Secure RTCP (SRTCP) provides similar capabilities for RTP Control Protocol (RTCP).

Because RTP normally flows end-to-end between UAs, SRTP likewise is end-to-end. Intermediate entities may, however, be involved to provide specific services (e.g. transcoding, conferencing), in which case SRTP will operate between each UA and the intermediate entity (which itself appears as a UA to each of the other UAs).

To enable the use of SRTP (and SRTCP), a key management capability is required for providing the two UAs with a secret master key from which session keys can be derived. The key management capability must ensure that:

- the secret key (or information from which to derive it) is delivered to or accepted from only a known, authenticated entity; and
- the secret key (or information from which to derive it) cannot be eavesdropped en route.

For the first of these, basically during any key exchange a UA must ensure that the peer UA to which it sends the key or from which it receives the key is a UA with credentials to prove that it belongs to the user with whom the local user wishes to communicate. The use of a shared secret for this purpose is not normally scalable beyond a relatively small community, and therefore the use of public key cryptography is normally the only feasible solution. The sender of key information signs it using the private key associated with a certificate that the peer user can associate with a user with which he expects to or is willing to communicate. This may be a direct association or an indirect association (e.g. the user can associate the certificate with the identity delivered in signalling (see clause 7.4), which the user can then associate with a user with which he expects to or is willing to communicate.

Therefore to support media security in a scalable way, UAs need to have private keys and associated certificates. Furthermore they need to be in possession of a means of verifying signatures from peers, either by having an independent secure means of obtaining peer certificates or by having appropriate CA certificates for verifying certificates received from peers. This imposes a significant Public Key Infrastructure (PKI) requirement on environments in which media security is required. In particular:

- a private key and associated certificate have to be produced for each user;
- this information has to be installed securely on any UA that the user uses, e.g. by secure download from a credentials server;
- peer UAs need to be able to obtain certificates from a reliable source for checking signatures (either peer certificates for checking peer signatures or CA certificates for checking the signatures of peer certificates).

Given that such a PKI is in place, it can also be used as the basis for encrypting key information for transmission end-to-end between UAs, thereby preventing eavesdropping. A UA can generate a key, sign it with its own private key and encrypt it using the peer UA's public key (from the peer UA's certificate). The peer UA can verify the signature using the sending UA's public key (from the sending UA's certificate) and decrypt the key using its own private key. This can be achieved using SIP's general purpose end-to-end security mechanism (i.e. S/MIME) to secure SDP, including any key management information. Alternatively, security can be achieved by embedding a key management protocol such as Multimedia Internet KEYing (MIKEY) within SDP. MIKEY provides its own means for authenticating and encrypting key information, again based on PKI. An advantage of MIKEY is that it is not necessary to send a key in each direction: a single key sent in one direction can be used to derive two different SRTP master keys, one for each direction. Another advantage is that it does not require the entire SDP to be secured.

A disadvantage of sending an encrypted key is that the sending UA needs to know in advance the public key of the receiving UA. Even if a certificate can be obtained for the user represented by the request URI in a SIP INVITE request, retargeting by SIP intermediaries can result in the request reaching a UAS that does not have the key to decrypt. This can result in additional rounds of SIP signalling. An alternative being considered for MIKEY is for the UAC to send its certificate to the UAS and then, in the response, the UAS can encrypt key information using that certificate.

An alternative is to use Diffie-Hellman key agreement, whereby a Diffie-Hellman component can be sent signed but not encrypted, thereby avoiding the need to acquire the certificate of the destination UA. MIKEY provides support for Diffie-Hellman key agreement protected by digital signatures based on PKI.

7.6.2.1 Scenario 1 - Communication between NGCNs using a VPN

Although the VPN tunnel will provide security for all traffic across the NGN, including media, SRTP will normally still be required end-to-end between the UAs to provide security within the NGCN as well.

Key management will be between the UAs with no impact on the NGN. The VPN tunnel provides secures the key exchange across the NGN, but in practice security will be needed end-to-end between the UAs to cover also transmission within the NGCN.

Within the NGCN it is likely that all certificates are signed by a single CA, and therefore verification of certificates is relatively easy.

Because in this scenario the NGN is transparent to media security and key management, media security in this scenario is not discussed further in the present document.

7.6.2.2 Scenario 2 - Communication between NGCNs not using a VPN

SRTP will normally need to be provided end-to-end between the UAs in the different NGCNs, thereby securing the media through each NGCN and the NGN. The exception is if the NGN intervenes in media in some way, e.g. monitoring, transcoding and conferencing.

Key management will be between the UAs with no impact on the NGN. It will need to be secured end-to-end.

Because two NGCNs are involved it is likely that certificates will be signed by different CAs. Therefore each UA has to recognize the other UA's CA. If there has been no previous contact the CAs may not be readily acceptable and the users may need to be consulted.

If the NGCN does provide any SIP intermediaries in this scenario they will need to be transparent to SRTP and to key management.

REQUIREMENT ON NGN 39: Except where an NGN intervenes in media in some way, in order to add value and with the knowledge and permission of the users concerned, the NGN shall be transparent to SRTP and to key management.

If the NGN intervenes in media in any way, it will need to terminate SRTP and key management. In other words, it will need to appear as a UA to the UAs in the NGCNs. This has significant impact, since it means the UAs in the NGCN will need to perform key management not with each other but with the intervening entity. This implies that a UA must accept a certificate from the intervening entity rather than from the peer UA in the remote NGCN. A UA must have the means to verify such a certificate, which means having the appropriate CA certificate available and also being able to associate the subject of the certificate with something meaningful. It is likely that users will need to be involved in sanctioning such an arrangement. A user who is unwilling to sanction such an arrangement would need to find some other means of achieving the required communication, i.e. without the NGN being involved in media.

REQUIREMENT ON NGN 40: An NGN shall be able to avoid intervening in media if media is required to be secured end-to-end.

REQUIREMENT ON NGN 41: An NGN that intervenes in media will need to provide media security and associated key management between the intervening entity and each UA.

REQUIREMENT ON NGN 42: In order to provide key management between an intervening entity and a UA in an NGCN, the NGN will need to provide a certificate for that entity that can be verified by the NGCN UA.

7.6.2.3 Scenario 3 - Communication between NGCN and TE using a VPN

The considerations of clause 7.6.2.1 above apply.

7.6.2.4 Scenario 4 - Communication between NGCN and TE not using a VPN

SRTP will normally need to be provided end-to-end between the UA in the NGCN and the UA in the NGN. Considerations are very similar to scenario 2.

7.6.2.5 Scenario 5 - Communication between NGCN and PSTN/ISDN

SRTP will normally need to be provided end-to-end between the UA in the NGCN and the gateway in the NGN, thereby securing the media through the NGCN and the NGN. The exception is if the NGN intervenes in media in some way other than at the gateway, e.g. monitoring, transcoding and conferencing.

Key management will be between the UA and the gateway, with no other impact on the NGN. It will need to be secured end-to-end.

The UA and the gateway will each require a private key and a certificate and will need to accept each other's certificate. The UA should normally be provisioned to accept certificates from any anticipated gateways. Similarly the gateway should be provisioned to accept certificates from legitimate users.

REQUIREMENT ON NGN 43: A PSTN/ISDN gateway in an NGN shall be able to supply a certificate that NGCN users are able to verify as being genuine.

REQUIREMENT ON NGN 44: A PSTN/ISDN gateway in an NGN shall be able to verify certificates from legitimate NGCN users.

7.7 Session policy

A network may wish to impose certain policies on use of its communication infrastructure. In particular it may wish to insert a media intermediary or impose limitations on the media that make up a session. Examples of reasons for inserting a media intermediary include traffic monitoring, enforcing service level agreements and access control, load balancing and traffic shaping, QoS marking, NAT/firewall traversal, media-level topology hiding and IPv4/IPv6 interworking. Most of these can be achieved through intervention at the transport layer or below. Intervention at a higher layer (e.g. RTP) may not be feasible if media are secured. Examples of limitations on media include media types allowed, codecs allowed and maximum bandwidth allowed.

Limitations for the local network (or segment thereof) will typically be made available to UAs during UA configuration, but the insertion of media intermediaries will depend on the session being established. Some sessions will extend beyond the local network segment and involve other segments or networks, which may also wish to insert a media intermediary or impose limitations on sessions.

SIP intermediaries are not necessarily well placed to impose policy by inserting media intermediaries and limiting session capabilities. This is because of the end-to-end nature of SDP in SIP and in particular the fact that it can be encrypted, thereby preventing viewing by SIP intermediaries, or integrity checked, thereby preventing modification. The purpose of SIP intermediaries is to establish signalling communication between SIP UAs by routing SIP requests and responses. The separate function of imposing session policy is better left to other entities, known as policy servers. SIP UAs can contact policy servers with details of the session to be established (based on sent and received SDP) and the policy servers can propose adjustments in order to insert intermediaries or limit session capabilities. Different domains will typically have different policy servers, and a SIP intermediary for a given domain can signal the identity of the policy server for that domain to the UAs. Thus the UAs can contact any relevant policy servers. Details of this mechanism are still being studied in the IETF.

7.7.1 Scenario 1 - Communication between NGCNs using a VPN

NGN policy may impact VPN tunnel establishment, but this is outside the scope of the present document. NGCN policy is an internal mater for the NGCN.

7.7.2 Scenario 2 - Communication between NGCNs not using a VPN

The NGN and the two NGCNs can each potentially impose session policy.

7.7.3 Scenario 3 - Communication between NGCN and TE using a VPN

The considerations of clause 7.7.1 apply.

7.7.4 Scenario 4 - Communication between NGCN and TE not using a VPN

The NGN and the NGCN can each potentially impose session policy.

7.7.5 Scenario 5 - Communication between NGCN and PSTN/ISDN

The NGN and the NGCN can each potentially impose session policy.

7.8 Emergency calls

Networks need to allow users to make emergency calls to Emergency Call Centres (ECC), which may be PSTN/ISDN-based or IP-based. Some NGCNs may operate their own enterprise ECCs. There are several challenges involved in this, including:

- identification of an emergency call;
- routing an emergency call to the most appropriate ECC;
- allocating appropriate resources to an emergency call, if necessary at the expense of less important calls;
- providing correct identification information to the ECC to allow a return call to be made;
- providing caller location information to the ECC.

Identification of an emergency call in the SIP INVITE request from the UAC can be explicit (by means of the SOS SIP URI) or implicit. In the latter case a local SIP intermediary must determine it is an emergency call be analysing the request URI.

In order to route an emergency call to the most appropriate ECC, a SIP intermediary (generally the first, i.e. NGCN- or NGN-based) must attempt to obtain geographic location information (geographic coordinates and/or civil location) for the UAC. It may be able to obtain this from the network infrastructure (e.g. from LAN switches or by wireless base station triangulation), or information may be supplied by the UAC itself. Location information may be available from more than source, in which case all information should be forwarded to the ECC.

Identification information should be supplied in accordance with considerations in clause 7.5. However, privacy requirements must be overridden. Any other information that could possibly be of relevance should also be forwarded, subject to the capabilities of the signalling protocol.

Some NGCNs may require identification and location information sent to an ECC via an NGN to be encrypted. Likewise sensitive information may be divulged in the media, and therefore media encryption may also be required.

REQUIREMENT ON NGN 45: An NGN shall provide a means for encrypting sensitive information supplied by an NGCN for delivery to an ECC.

REQUIREMENT ON NGN 46: An NGN shall provide a means for encrypting media between an NGCN and an ECC.

7.8.1 Scenario 1 - Communication between NGCNs using a VPN

Emergency calls are not applicable to this scenario unless the NGCN provides its own ECC for emergencies inside the enterprise. This would not involve the NGN.

7.8.2 Scenario 2 - Communication between NGCNs not using a VPN

Emergency calls are not generally applicable to this scenario, except when used to route an emergency call to an ECC in a different part of same NGCN (intra-enterprise).

REQUIREMENT ON NGN 47: An NGN shall not interfere with the routing of a call to an enterprise ECC and shall not modify information for delivery to the enterprise ECC.

7.8.3 Scenario 3 - Communication between NGCN and TE using a VPN

The considerations of clause 7.8.1 apply. However, the NGCN proxy may be unable to determine the location of the TE and therefore it would be the responsibility of the TE to supply its location.

7.8.4 Scenario 4 - Communication between NGCN and TE not using a VPN

Emergency calls could be applicable to this scenario, the ECC being the TE in the NGN. In this case the SIP intermediary in the NGCN should determine that the call is an emergency call and forward the necessary information (caller identity and location). The SIP intermediary in the NGCN may not be able to determine the most appropriate ECC, in which case the SIP intermediary in the NGN should perform this function. All SIP intermediaries should assign appropriate resources.

NOTE: The calling TE in this scenario could be an enterprise ECC, which makes an emergency call to a public ECC.

7.8.5 Scenario 5 - Communication between NGCN and PSTN/ISDN

Emergency calls could be applicable to this scenario, the ECC being in the PSTN/ISDN. In this case the SIP intermediary in the NGCN should determine that the call is an emergency call and forward the necessary information (caller identity and location). The SIP intermediary in the NGCN may not be able to determine the most appropriate ECC, in which case the SIP intermediary in the NGN should perform this function. All SIP intermediaries should assign appropriate resources.

7.9 Geographic location

A UA may determine its geographic location (geographic coordinates and/or civil location) and make this available to authorized entities. These might be only the peer UA or they might include one or more SIP intermediaries. However, in general this information is sensitive and will need to be encrypted at least on a hop-by-hop basis. If it is not to be made available to SIP intermediaries then end-to-end encryption (using S/MIME) would be appropriate. If it is to be made available to designated SIP intermediaries, then end-to-middle security would be appropriate. There is ongoing work in the IETF on conveying location information in SIP.

Alternatively a SIP intermediary local to a UA may have the means to determine the UA's geographic information from the network infrastructure (e.g. from LAN switches or by wireless base station triangulation). Again it would need to be signalled securely to those SIP entities that are authorized to receive it. One solution being considered in the IETF is to signal the location information back to the UAC (using TLS on that hop) and ask the UA to signal it securely end-to-end.

In the case of an NGCN-owned TE connected to an NGN, the NGCN may wish to obtain location information for that TE from the NGN.

REQUIREMENT ON NGN 48: An NGN shall provide a means for an NGCN to obtain available location information of an NGCN-owned TE.

The provision of geographic location to an ECC is an important part of emergency calls (see clause 7.8).

7.9.1 Scenario 1 - Communication between NGCNs using a VPN

Geographic location is an internal matter for the NGCN.

7.9.2 Scenario 2 - Communication between NGCNs not using a VPN

Geographic location is a matter for the two NGCNs. There should be no reason for a SIP intermediary in the NGN to generate or have visibility of geographic location information. This information may be encrypted across the NGN.

7.9.3 Scenario 3 - Communication between NGCN and TE using a VPN

Geographic location is an internal matter for the NGCN. However, the NGCN proxy may be unable to determine the location of the TE and therefore it would be the responsibility of the TE to supply its location.

7.9.4 Scenario 4 - Communication between NGCN and TE not using a VPN

Geographic location information concerning the TE in the NGCN could be delivered to the TE in the NGN. There should be no reason for a SIP intermediary in the NGN to generate or have visibility of this information, which may be encrypted.

Geographic location information concerning the TE in the NGN could be generated by the TE or by a SIP intermediary in the NGN. If delivered to the TE in the NGCN, there should be no reason for a SIP intermediary in the NGCN to have visibility of this information, which may be encrypted.

7.9.5 Scenario 5 - Communication between NGCN and PSTN/ISDN

Geographic location information concerning the TE in the NGCN could be delivered to the PSTN/ISDN. There should be no reason for a SIP intermediary in the NGN to generate or have visibility of this information, which may be encrypted as far as the gateway.

Geographic location information concerning the TE in the PSTN/ISDN could be generated by the PSTN/ISDN, the gateway, or a SIP intermediary in the NGN. If delivered to the TE in the NGCN, there should be no reason for a SIP intermediary in the NGCN to have visibility of this information, which may be encrypted.

History

Document history		
V1.1.1	March 2006	Publication