A Solution for Secure SIP Conferencing over IMS and SAE

CRISTINA-ELENA VINTILA
Bucharest, ROMANIA
cristina.vintila@gmail.com

Abstract: Over the latest few years, most of the major telephony and services providers have got their attention on the LTE/SAE solution, in the attempt of getting the most bandwidth and features at the least implementation and operating price. One of the major challenges that 3GPP, the creator of LTE/SAE architecture, has faced is the IMS integration with SAE. The latest standard version available at this moment on IMS integration and its security challenges is TS 33.203, which is focused on 3G security aspects. When talking about IMS-SIP security, there are several studies that propose end-to-end security for a SIP conversation over EPS infrastructure.

This paper reviews the security issues that resides in the SAE-IMS interaction and, looking at the specificities of the SIP conferencing, proposes a security model that uses GDOI management to secure the SIP conference data over IMS and SAE. One important aspect of conferencing in the mobile world is to realize the user is never stationary. One chapter of this paper describes the most complicated type of mobility scenario and also introduces the role of the Diameter server into this architecture.

Keywords: SAE, LTE, EPS, EPC, IMS, security, SIP, conference, GDOI, GCKS, IPsec, key management

1 Introduction

SAE, or System Architecture Evolution, is the core network architecture of 3GPP, evolved from GPRS Core Network and, together with the highly performant radio interface called LTE (Long Term Evolution), is the newest answer to the increasing demand of high throughput and low latency issues of the mobile world. It comes with a simplified, all IP flat network and mobility between 3GPP and non-3GPP systems.

The picture below describes the main component of the SAE, EPC – Evolved Packet Core, which comprises three elements: MME – Mobility Management Entity, SGW – Serving Gateway and PGW – PDN (Packet Data Network) Gateway.

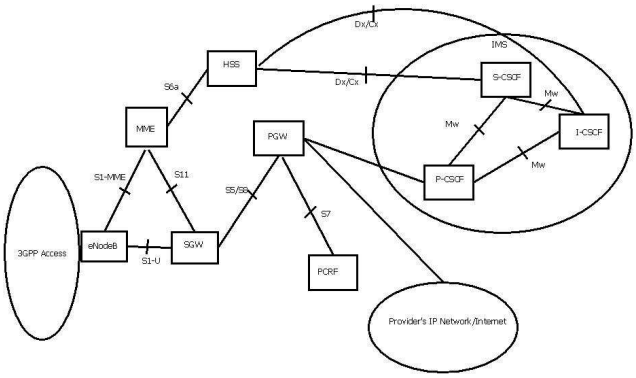

Figure 1. Simplified SAE – IMS architecture

SAE conceptually separates its signaling traffic from the users actual traffic, using the concepts of control-plane and user-plane.

MME is the key element of control-plane, being in charge of the UE's location and state tracking, as well as the negotiation of bearers between the UE, eNodeB (connected on the S1-MME interface) and SGW, based on the QoS rules from the PCRF – Policy and Charging Rules Function; the bearer negotiation is usually triggered from the PGW, the Create Bearer Request coming from the SGW via the S11 interface of GTP-C protocol. It is the authentication proxy for the UEs entering the network, through its connection to HSS via the S6a interface, the enforcer of roaming restrictions and provides the functions for mobility towards 2G and 3G networks. This entity supports also lawful interception of signaling/control plane traffic.

Unlike MME, which is esentially a control-plane entity, the SGW has both control plane and user plane functionality, being the anchor point of the user plane traffic in the handover scenarios; it is the manager of the UE bearers and stores the UE contexts. SGW is connected to MME via the S11 interface, a GTP-C interface, and to PGW via S5/S8 interface, which transports the bearer management messages. The S5 interface is for connection to a local PGW, while the S8 interface is created to a PGW from another network and it is used in the roaming scenarios. It is also connected to UMTS's SGSN entity via S4 interface. If configured so, SGW provides support for lawful interception of user plane traffic. As a user plane entity, the SGW is connected to eNodeB via S1-U interface, where the traffic encapsulation protocol is GTP-U. While the GTPv2/eGTP GTP-C protocol requires layer 3 connectivity between MME and SGW, as well as between MME and eNodeB, GTP-U requires layer 3 connectivity between eNodeB and SGW as well, in order to transport user plane traffic, as well as for the purpose of handover.

PGW is the exit point of the EPS, a gateway to the data network, more commonly the Internet. It is also the entry point in the EPS for the user plane traffic destined to a

certain UE. Connected to the PCRF via S7 interface, the PGW provides policy enforcement, packet filtering, charging support and lawful interception.

SAE is not a standalone technology, but rather modular and very flexible. It provides interfaces to a multitude of technologies, on the radio side 3GPP (UMTS, GPRS), as well as non-3GPP (WiMAX, HRPD, WLAN), and also on the services side. Being a flat, IP only infrastructure, the services are based on IP, being either the Internet or specific network operators, like IMS – IP Multimedia Subsystem. The current orientation of providers is mostly on Data Networks, but the Voice networks will be targetted in at most 2 years from now, IMS being the favorite approach, given its flexibility. The IMS on top of EPC is linked via a data interface (SGi) from the main gateway of the EPC (PGW), while the signaling/control interface is Rx coming from the P-CSCF (Proxy-CSCF) entity of IMS to the main EPC session controller in charge of authorization, admission control, resource reservation and QoS (PCRF).

The IMS, or IP Multimedia Subsystem, is a flat IP infrastructure as well, created initially to deliver the Internet services over GPRS. At the moment, IMS is a collection of functions that handle the management of voice-related services, like VoIP, messaging or multimedia over the network. The main components of the IMS core are the HSS (Home Subscriber Server), which keeps the UE identities, the Call/Session Control functions (taken care by three functions: P-CSCF, I-CSCF and S-CSCF) and the Application Servers. It may also provide Media Servers, Breakout and PSTN Gateways. On top of all the architecture, the charging function is employed.

2 Security aspects

While the eNodeB, which belongs to a certain operator, faces a series of vulnerability issues derived from its position in the SAE, at the boundary between the radio interface and the IP wired EPS, the PGW is a key element in an even more complex series of security aspects, related mostly to upper layer protocols (the security issues any Internet device would have to face), to the interoperability with other providers and also with other systems, like trusted and untrusted 3GPP and non-3GPP network cores and the interaction with the Internet.

One of the interfaces posing interoperability and security issues is the interface to P-CSCF. Proxy-Call/Session Control Function is a component of the IMS system, the entry point in the IMS architecture. This is usually an SBC device, handling requests from UAs wanting access to the IMS services; it can be located in the visited network, if that network is also IMS, or in the home network, in case the visited is a non-IMS one. Its purpose is to handle all the signaling requests from the UA located outside the home network and forward them to the I-CSCF, as well as to authenticate that UA and establish a secured session with it. It can also compress and decompress SigComp messages, may include a PDF (Policy Decision Function) to authorize QoS for media and can generate charging information. Because it is in the path of all signaling messages of the UE's, the P-CSCF is a major security asset and there are several ways to protect this device, both standalone and in the interaction with EPS, more specifically with its PGW component.

One of the biggest security challenges when talking about IMS and EPC is the secure SIP session establishment and secure voice and messaging delivery between UAs, with emphasis on the roaming and mobility scenarios. The SIP security mechanisms are defined in [2] – as generic SIP protocol security mechanisms. Extending the SIP establishment threats, but also security solutions and mechanisms to the SIP conferencing scenarios, this paper tries to create a security model for the IMS-SAE interaction, using the existing SIP extensions for 3GPP interaction and Privacy Headers. There will also be introduced a new type of P-Header extension, one that indicates the location of the GCKS server to the newly added/joining SIP party.

Regarding the P-CSCF and PGW interaction, there are 2 main scenarios where this can happen differently from the security point of view: the UE is located in the home network and the UE is in roaming. Both of these scenarios should cover the mobility of the user. Thus, the security measures architecture could be devided into measures applicable in the home network and measures applicable in the visited network.

2.1 Secure registration to the network

Leaving aside the scenario where the UE is located in its home network, this paper focuses on the roaming scenario where the UE-Alice is in a SAE visited network and tries to contact his friend UE-Bob at home, both of them being customers of the same SAE network in domainA. In order to emphasize the importance of the SIP Privacy Headers and the inter-domain security required here, let's assume that Bob has multiple identities on his SIP account, a business profile (bob-business) and personal profile (bob-personal), he keeps both of them active all the time, with different ring-tones and different redirect options: all the business calls are redirected to a voice mail server during week-ends and late hours, while the personal calls are forwarded directly to his cell-phone, no matter the hour nor day. Let's analyze the steps required for Alice to be able to talk to Bob during week-end.

The premises of this scenario are:

1. Alice is located in another SAE network, other than Bob's

2. Alice has to authenticate to her network in order to be able to call Bob

3. Alice calls Bob during week-end

4. The visited network where Alice finds herself has a roaming service agreement with her and Bob's operator (from domainA)

5. The visited network where Alice is located has multiple proxy servers, same as domainA network, so Alice's call will be forwarded between multiple proxies before it reaches Bob

Alice has to be connected to the visited network; once she turns on the cell-phone, the UE-Alice is detected by the closest eNode and the eGTP Initial Attach procedure takes place, so that Alice is registered to the EPS network. The details of the eGTP signaling for the Initial Attach are out of the scope of this paper, as well as the LTE and EPS specific security mechanisms.

After the Initial Attach, Alice wants to register to her home network. Located in domainB, she sends a SIP REGISTER message to her domainA registrar server.

The REGISTER message would look something like this:

```
REGISTER sip:registrar.domainA SIP/2.0

Via: SIP/2.0/UDP [5555::aaaa:bbbb:cccc:dddd]; comp=sigcomp;branch=z9hGjd446sh6rt

Max-Forwards: 20

P-Access-Network-Info: 3GPP-EUTRAN-TDD; eutran-cell-id-3gpp=12345667

From: <sip:alice@domainA>; tag=234h

To: <sip:alice@domainA>

Contact: <sip:[5555::aaaa:bbbb:cccc:dddd];comp=sigcomp>;expires=600000

Call-ID: wew8798k34jj3454389

Authorization: Digest username=»alice@domainA», realm=»registrar.domainA», nonce=»», uri=»sip:registrar.domainA», response=»»

Security-Client: ipsec-3gpp, alg=hmac-sha1-96, spi-c=12345678, spi-s=23456789;port-c=2545; port-s=1234

Require: sec-agree

Proxy-Require: sec-agree

CSeq: 1 REGISTER

Supported: path

Content-length: 0
```

Alice sends this message to the first P-CSCF proxy in her way in the visited network (domainB), as this is the only P she knows about. The Via header contains the IPv6 address the UE-Alice device received from the PGW during the Initial Attach procedure, as well as the Contact header. As this is a Registration procedure, both the From and To headers contain Alice's identity. And the following headers mark the security architecture that this client expects to have from the networks, both home and visited.

UE-Alice signals via the Security-Client, Require and Proxy-Require headers the security capabilities it knows. The Security-Client header, along with Security-Server and Security-Verify headers are described by [2]. When this message arrives at the P-CSCF1 of domainB, it strips off the Security-Client and the «sec-agree» tag, adds itself in the Path and adds a P-Visited-Network-ID header to the message, having as content the identifier of the domainB network. It also adds a P-Charging-Vector header for charging tracking purposes. Then P-CSCF1 of domainB sends this message to I-CSCF server of Alice's home network, domainA. The DNS queries the P has to do in order to locate Alice's I representative is out of the scope of this paper, as well as the intrinsec IMS procedure that identify whether Alice is a valid subscriber of this operator. The I communicates Alice's REGISTER request to the S, this S server interogates the HSS database and obtains an authentication vector from the database. This authentication vector is essentially an HSS challenge for Alice to prove her identity, usually an IMS-AKA authentication scheme and it has the format defined by [10]:

```
AV = RANDn||AUTNn||XRESn||CKn||IKn, where

- RAND is a random number used to generate the XRES, CK, IK and part of AUTN; UE uses it to generate the RES

- AUTN is the authentication token and it includes MAC and SQN

- XRES is the correct/expected response from the UE

- CK is the cipher key

- IK is the integrity key
```

and it appears in the S response to I and from this (home network) I is propagated to the visited network's P in the WWW-Authenticate header, which looks like this:

```
WWW-Authenticate: Digest-realm="register.domainA", nonce=base64(RAND + AUTH + server data), algorithm=AKAv1-MD5, ik="kjhgasfdasjhfdskfgg",
```

ck="jhuguyipoipotytrtyryewr", where the CK is optional.

When forwarding this message back to UE-Alice, the P-CSCF from domainB adds the Security-Server header, which looks like this:

Security-Server: ipsec-3gpp, q=0.1, alg=hmac-sha1-96, spi-c=23442343, spi-s=12112345;port-c=1212; port-s=4534

Bob is located in his home network, but the Registration procedure is similar, except for the headers required when passing from one P-CSCF to another. The P-Access-Network-Info may still be used, because Bob wants to signal to his IMS core network that he is in a specific E-UTRAN cell, with specific delay, latency, and bandwidth.

2.2 Conference initiation

[5] (Conferencing Scenarios) defines a numerous types of conferencing, many of them with specific SIP signaling procedures. The SIP conferencing solution proposed in this paper tries to accommodate as many scenarios as possible, through the introduction of P-GCKS-Info header. Whether the type of conference is an Ad-Hoc one, an extension of a Point-to-Point Call to a Multi-point Call or a more advanced scenario, all the participants are required to authenticate to a group controller/key server (GCKS) in order to retrieve their group session keys they will use to secure the media they are sending each other.

2.3 Assumptions and limitations of the model

There are also a few assumptions and limitations to this model:

1. The model requires that all participants authenticate to the GCKS, so they must support the P-GCKS-Info header, as well as have IPsec support in their software. While the P-Header requires a decision logic of the SIP engine of each cell-phone, the IPsec client should be fairly easy supported on the smart-phone that are 4G compatible.

2. The Conference with Unaware Participants (loosely coupled conference) is not supported by this model. This type of conference defines, as per [5], that conference-unaware participants may be using a proxy function that proxies the advanced functionality between the different protocols and the Conferencing System, as an IVR or a web interface.

3. Many conferences use a media mixer for data streams. This mixer may be a separate entity or may be a separate function of one (or more) of the participants. When the mixer is centralized, it is either located on the focus (the name given to the UA that starts and usually manages the conference) or the focus indicates the location of the mixer it uses for that particular conference, but it can also be distributed among the participants, each device having its own mixer functionality. This is important information for the proposed model, because, in order to achieve media security, all of the parties accessing the RTP stream must be part of the GDOI- Group Domain of Interpretation. In order to simplify the SIP exchanges described in the model, this paper only describes the case where the mixer is on the focus device.

There are 2 aspects to consider in this type of scenarios: the signaling/control plane and the media/data plane. The signaling plane should be done as described above, using the hop-by-top security solution in order to be able to securely reach all the parties that are to be invited to the conference, no matter the domain they found themselves in, while the media plane should benefit of the end-to-end security solution, using an IPsec path created via GDOI protocol. The SIP signaling is done via SIP-IMS. Let's assume that each party, no matter its location (home network or visited network), has authenticated and securely registered to their home network. The next step is that the conference administrator invites the first callee. The signaling is done via the mechanisms described above, sending an INVITE message that is first encapsulated in GTP-U (GTPv1) while flowing from the UE to the PGW, then encapsulated in IPSec from the PGW to the P of the invited party. From there on, the home P of invited party forwards it the INVITE via 3GPP or non-3GPP network protocols. Once the party replies and the first SIP session is up, the media is going to be forwarded via the PGWs of the two networks (GGSN or the data network gateway specific to any other 3GPP or non-3GPP implementation), encapsulated in ESP by each UE, with the group key.

In order to be able to accomplish this scenario using GDOI, one of the networks must have a GCKS, group controller/key server. This could be a separate device, or a function embedded into one of the IMS or EPS entities. As the GCKS should have closest proximity to all the other UEs participating in the conference, no matter their location, this should be better positioned close to the border of the EPS network, rather than in the middle of it; this is why the PGW looks like the best candidate. In this scenario, I have considered that the PGW also takes the roles of a GCKS. Once the UE that initiates the conference calls the first party, it also triggers the creation of 2 session keys inside the GCKS. The key distribution is done via Phase 1 component of GDOI, as described in [4], the actual way to deliver the keys is not of concern at this point. Once both of the parties are in possession of their corresponding key, the media transmission can start.

2.4 GDOI support signaling

The actual type of conference does not matter, so let's take the case of a three way conference, when the third party is added

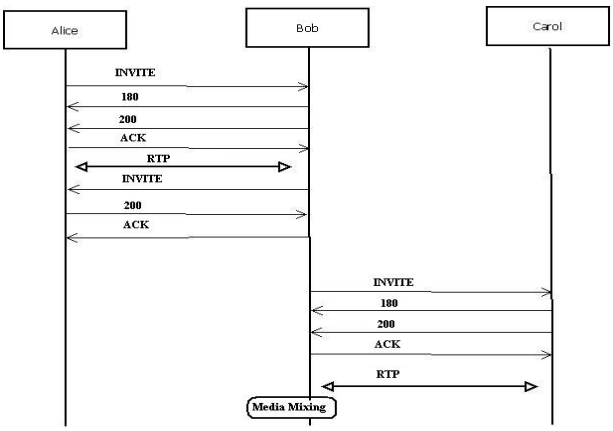

Figure 2. SIP conference flow – Third party added

The UA-Alice opens the SIP session, sending a usual INVITE message to Bob, via P-CSCF from Alice's visited network and through the P-CSCF server on Bob's network. This INVITE should look like this:

```
INVITE sip: bob-personal@domainA SIP/2.0

Via:                              SIP/2.0/UDP
[5555::aaaa:bbbb:cccc:dddd]:1234;
comp=sigcomp;branch=z9hGjd446sh6rt

Max-Forwards: 20

Route:
<sip:pcscf1.domainB;lr;comp=sigcomp>,
<sip:icscf1.domainA;lr>

P-Preferred-Identity:              "Alice"
<sip:alice@domainA>

P-Access-Network-Info:  3GPP-EUTRAN-TDD;
eutran-cell-id-3gpp=12345667

P-GCKS-Info: pgw1.domainB

From: <sip:alice@domainA>; tag=234h

To: <sip:bob@domainA>

Call-ID: wew8798k34jj3454389

CSeq: 123 INVITE

Require: precondition, sec-agree

Proxy-Require: sec-agree

Supported: 100rel

Security-Verify:  ipsec-3gpp,  alg=hmac-
sha1-96,      spi-c=12345678,      spi-
s=23456789;port-c=2545; port-s=1234

Contact:                                <sip:
[5555::aaaa:bbbb:cccc:dddd]:1234;
comp=sigcomp>

Content-Type: application/sdp

Content-length: (...)

<<< SDP headers – omitted >>>
```

Once Bob decides to also invite Carol into the conference, it send another INVITE to Alice, similar to the one above, but having a Contact header like this:

```
Contact:                                <sip:
[5555::aaaa:bbbb:cccc:ddee]:1254;
comp=sigcomp>;isfocus
```

Then Bob will invite Carol, sending her an INVITE similar to the one sent to Alice, stating that his UE is focus for the SIP conference, and indicating Alice's PGW as GCKS.

2.5 Leaving the conference

When one of the conference participants decide to leave the conference, they have to signal this intention, by sending a BYE message to the focus of the conference. This one will further inform the GCKS of the leaving party, in order for it to update the session keys for the remaining participants. This scenario may still be of not too much security value, as one of the party leaving means its key is not used anymore. What would be a concern in the real IPsec networks (one former group member being able to use its old key to read the new messages after it has left the group), in the conference scenario may not necessarily need to be enforced, unless the remaining parties explicitly want to ban the former group member from the new topics on the agenda.

2.6 Mobility aspects

When talking about security and persistent end-to-end flows in the perspective of Mobile Network architecture is very hard not to take into account the mobility aspect of the entities involved.

There are several cases when mobility appears. One of them is when the UE moves from one cell to another, triggering an update in the eNB signaling. Should the cells all be served by the same eNB, this signaling is not necessarily propagated in the core network. When the UE moves from one eNB (called source eNB) coverage to another eNB (called target eNB), the signaling required to keep track of this UE is no longer a matter of LTE signaling and management. Moving from one eNB to another raises two more cases:

a) when the eNB entities have an X2 link between them

b) when the eNB entities do not have an X2 link between them

In case a), the signaling required to update the UE's position is done via EPC. The air medium security is covered by the LTE sessions keys and this is not the scope of this paper. What happens to the signaling and, more importantly, what happens to the encrypted voice payload? First of all, the signaling is nevertheless hop-by-hop based, involving the SGW, PGW, P-CSCF, I-CSCF and S-CSCF. The new, target eNB, will have to update the EPC that the UE is now under its supervision. This is done via an eGTP/GTPv2 message exchange, called generically "handover". In case a), where there is an X2 link between source and target eNB and both eNBs are served by the same MME entity, the handover case is called "X2 based handover". The first phase of this mechanism is the preparation, involving mostly LTE computations. During the handover execution at the eNB level, the source eNB already forwards the downlink data it has for the UE in question to the target eNB. This process itself can be secured by an eNB to eNB authentication and mutual trust. The implications over the EPC are also very important, because you can attack an entire EPC network only by using a forged eNB. This aspect wasn't taken into consideration before, as few could believe a forged UMTS Node-B station could actually be put in place. Nowadays, this threat is overcome by mutual authentication procedures, enforced by the UE firmware, as well as the eNB firmware.

Only when the UE has successfully authenticated the target eNB, its source eNB forwards the downlink packets destined to this UE to the target eNB. After this step finishes, the UE's uplink data (encrypted voice) is forwarded to the SGW via the new, target eNB. Case a) assumes the existance of the X2 link between eNB, a pre-existing link and, in most of the cases, the existance of a security clearance between these entities. From the target eNB, the uplink encrypted RTP packets could go to the same SGW as the before, or to a different SGW (there is no rule that two separate eNBs, even though server by the same MME, should also be server by the same SGW). This is why, when talking about an X2 handover, there are two sub-cases involved: X2 handover with SGW relocation and X2 handover without SGW relocation. In both cases, the target eNB sends a Path Switch Request message to the MME. Now, if the SGW is relocated, the MME has to create a new session/bearer with the new SGW, by exchanging a Create Session Request/Create Session Response with it, in order to move all this UE's bearers on the new SGW/PGW. This operation itself can be successful (Cause Accepted in the Response message), rejected, or partially accepted by the EPC, this decision being based on the capabilities of the new eNB and the new SGW, because these two entities are involved in the user-plane GTPv1 traffic path. If the SGW is not relocated, the MME only signals to the (existing) SGW the modification of the status of this UE's bearer, via a Modify Bearer Request/Modify Bearer Response exchange, propagated also to the PGW.

These steps are important for the end-to-end security of the system and, before the handover process is completed, any voice packets heading to the UE are sent to the former, secured, eNB, which in turn forwards them to the target eNB, which delivers them to the UE.

But what happens if there is no X2 link between the source and target eNB? Case b) has many other sub-cases. First of all, Case b) is more realistic from this point of view. Maybe the eNBs don't trust each other. Maybe be also that the two eNBs are served by different MME entities and/or by different SGW entities. The new path of the data should be signaled and secured again, hop-by-hop. For the moment, let's described the cases involved in this mobility process.

When the UE first attached to the network, via a process called Initial Attach, it authenticated the network (meaning the eNB, as far as the UE is concerned) and the network also authenticated the user by verifying its credentials stored in the HSS database (IMS network also authenticated the user, located in its home network or in roaming, via the same/or a different HSS database). This Initial Attach procedure gave the UE something called "mobility anchor", which is in fact the address of the PGW. The PGW is the user's virtual connection to the Internet or to his IMS network, and also the entity having a DHCP pool or some other similar mechanism via which gave the user an IP address. The entire purpose of the mobility is to make sure that the UE has the same IP address given to it in the moment of the Initial Attach, no matter via which cells this UE might travel during his staying in this network. This persistent IP address also ensures that the end-to-end exchange of information (as encrypted conferencing voice data is) is not interrupted in any way. No matter if the UE moves to a different eNB, MME and/or SGW and even to a different PGW device, it is still "attached" to the same PDN – there may be two or more PGW devices, for failover of load balancing purposes, but we are still talking about the same PDN.

Assuming there is no X2 link between the two eNBs, the UE and the EPC could find themselves in one of the following situations:

1. the MME is relocated and the SGW is not relocated
2. the MME is not relocated and the SGW is relocated
3. neither MME, nor SGW are relocated
4. both MME and SGW are relocated

Even if the eNBs are connected via an X2 link, there can also be the case where they are served by different MME entities. 3GPP has named this case, where the eNBs are served by different MME entities, S1 handover, after the name of the interface between eNB and MME, called S1-MME. Taking into account the entities that change and the presence or absence of the X2 link between the eNB, the handover cases can further be divided into Direct Tunneling and Indirect Tunneling. By definition, all the X2 scenarios (two) are implicitly Direct Tunneling cases. Among the four

scenarios enumerated above, only scenario 3., in the presence of the X2 link, is the same as the X2-based handover, with neither MME, nor SGW relocation. The other three remaining scenarios (1., 2. and 4.) can be either Direct Tunneling and Indirect Tunneling scenarios. Scenario 3. can only be Indirect Tunneling (as if there is an X2 link we wouldn't be talking about an S1, but rather about an X2 based handover). Summing them up, there are nine mobility scenarios possible, each having its specific signaling and security aspects. Let's consider the most complex of them, the S1-based handover with MME and SGW relocation and Indirect Tunneling, and describe the signaling necessary to seamlessly move the UE from one part of the network to another, without interfering with the encrypted end-to-end RTP packets of the voice conference.

This complex S1-based scenario is described in the figure below.

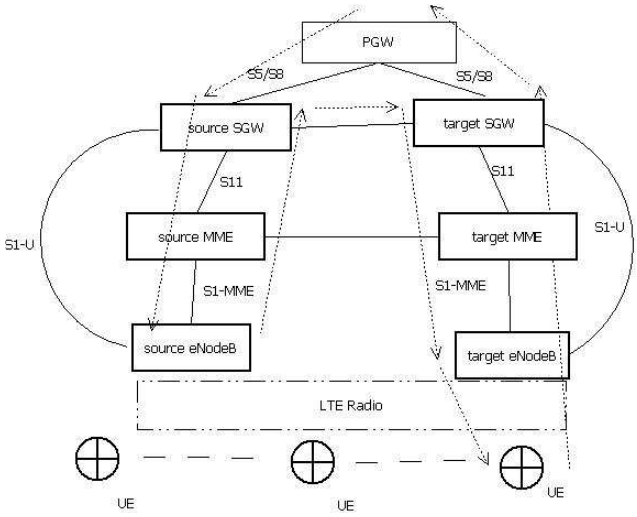

Figure 3. S1 handover, MME and SGW relocation, Indirect Tunneling

As per [8] and [17], when the UE moves from source eNB to target eNB, without having a (direct) X2 link between them, the source eNB signals to its serving source MME that a handover is required (the message is called Handover Required). The source MME is responsible for selecting the target MME, taking into account the new location of the UE. It sends a Forward Relocation Request message to the target MME. This MME exchanges a Create Session Request/Create Session Response with the target SGW, trying to move all this UE's bearers on the new MME/SGW/eNB entities. Then the target MME signals back to its eNB that the Handover has begun, exchanging therefore Handover Request/Handover Request Acknowledge messages.

The most important challenge in this type of scenarios is not to lose the downlink packets that have already got passed the source SGW to the source eNB. In this case, the source eNB must "reflect" the packets back to its SGW and this SGW must be instructed to forward these packets to the target SGW, which, in turn, will forward them to the target eNB. In order to accomplish this behavior, both the target MME, then the source MME, do a Create Indirect Data Forwarding Tunnel Request / Create Indirect Data Forwarding Tunnel Response exchange with their respective SGW entities. Once this EPC signaling ends with no failures, the source eNB reflects the packets it got in downlink during the handover to the SGW. The source SGW sends them to the target SGW, then this one forwards them to the target eNB and then they reach the UE. On the UE, mechanisms should be in place in order to deal with the possible out or order RTP packets. Nevertheless, the end-to-end security requirement is maintained.

The purpose of the Indirect Tunnel negotiation is to create Tunnel Identifiers that correctly signal the path of the data for the forwarding entities. This means that, the Create Indirect Data Forwarding Tunnel Request/Create Indirect Data Forwarding Tunnel Response message contain in their Bearer Context Grouped IE > TEID (Tunnel Identifier) header numerical values. The latest value negotiated is the first value used, while the first value negotiated is the last used. The first value negotiated, let's call it TEID1, appears in the Create Indirect Data Forwarding Tunnel Request sent from the target MME to the target SGW, and it represents the Tunnel Identifier that the downlink data from the target SGW to the target eNB will use. The standard defines this interface as being of type 19 - eNodeB GTP-U interface for DL data forwarding. The reply coming from the target SGW to the target MME, Create Indirect Data Forwarding Tunnel Response, contains the TEID (let's call this TEID2) for the SGW – user plane, interface type 23 - SGW GTP-U interface for data forwarding between the source SGW and target SGW.

The second set of Create Indirect Data Forwarding Tunnel Request/Response is exchanged between source MME and source SGW. The request contains the TEID of the source eNB, while the Response contains the TEID (let's call this TEID 3) of the source SGW interface for data forwarding (interface type 23). At this moment in time, when the source SGW already managed to send downlink data packets to the source eNB, packets destined for UE which has just started the handover procedure, the UE is no longer able to receive these packets. The Indirect Tunneling procedure is already triggered at this point, and the source eNB, knowing the TEID of its (source) SGW for this Indirect Tunnel, encapsulates these packets with this TEID (TEID 3) and sends them back to the source SGW, through a mechanism called "reflection". The source SGW, having the TEID for Indirect Tunnel of the target SGW, encapsulated these packets with that TEID and forwards them to the target SGW. This SGW, realizing the packets came on an Indirect Tunnel ID (TEID 2), encapsulates them in a different TEID (1) and forwards them to the target eNB

– which, at this moment in time, is managed by the target eNB.

During this entire process, the UE is located behind the second, target eNB, which raises the following question: what path does the uplink data of this UE take? As per the standard [22], the uplink of the UE during the handover process should be originated from behind target eNB – section 5.5.1.2.2 – S1-based handover, normal. Nevertheless, taking into account that, in order to send uplink (same as for downlink) traffic, an entity has to have defined a TEID for that traffic – previously negotiated, at this moment, the UE does not yet have a TEID for uplink. It's TEID for uplink is negotiated in the Modify Bearer Request/Modify Bearer Response exchange between target MME and target SGW. This is a situation that needs to further be investigated.

The figure below represents the path taken by the user-plane (secured RTP) packets on their way, from the PGW to the destined UE, on this indirect path negotiated.

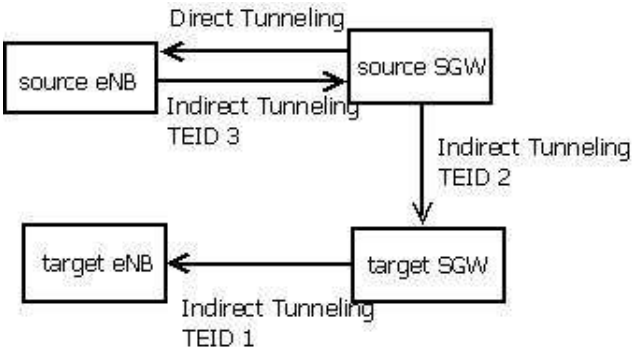

Figure 4. Indirect Tunnels created in the Indirect Data Forwarding Tunnel Request/Response exchange procedure

The remaining scenarios are simplified cases of this one, the most complex of all, when both EPC entities (MME and SGW) are relocated and, even more, the two eNB station do not have a connection between them.

2.7 Diameter

One cannot imagine an IMS network without the Diameter server. The older Radius server, even though flexible and secure, still had reliability and security issues and, most of all, flexibility and scalability issues. The original DIAMETER request for comments (RFC) states: "The basic concept behind DIAMETER is to provide a base protocol that can be extended in order to provide AAA services to new access technologies. Currently, the protocol only concerns itself with Internet access, both in the traditional PPP sense as well as taking into account the ROAMOPS [Roaming Operations] model, and Mobile-IP." This way, Diameter peer may now exchange many more messages, and it is also compatible with Radius, for the ease of the upgrade to this newer protocol. Initially planned for IMS and also heavily used in the IMS architecture, Diameter was developed by IETF.

In the IMS architecture, Diameter protocol is found on the HSS and SLF databases. HSS – Home Subscriber Server is the main database with the user location information, while the SLF – Subscriber Location Function has the purpose of interrogating multiple HSS databases (or a single one, if only one available), and provide an interface for the IMS servers, when these entities need to have fast access to user information. SLF is a lot faster than HSS and provides an additional layer between IMS servers and the database. Both HSS and SLF implement the Diameter server functionality, providing Cx, Dx and Sh interfaces to the Diameter clients.

At any moment of the call conference establishment, as well as during the conference, when the UA simple uses the network for participating in the conference, as well as when the UE performs a mobility maneuver or when a UA disconnect, the S-CSCF server communicates with the HSS via the Cx interface. This happens when the UE registers to the IMS network (as described above) and when it is necessary to retrieve UA – related information. The same S-CSCF server interrogates the HSS via interface Dx this time, when it needs to locate a particular HSS database, serving the UA via SLF. The Sh interface is used by the application servers to retrieve and also to update the UAs' profiles. These profiles my contain call directories, presence information or screening lists.

Another very important function of the Diameter is to take care of the charging for every IMS and SAE resource consumed in the conference process. The charging function, although not directly involved in the security of this conference, is a very important aspect. In order to implement the charging functions into Diameter, 3GPP created the Ro and Rf interfaces of this protocol, used for on-line charging and off-line retrieval of CDRs – Call Detail Records.

The exact procedure by which the charging takes place is out of the scope of this article, but, when talking about IPsec and GDOI and high computational requirements, the charging is an important part of the model, as it provides the operator with the means of Return On Investment.

3 Conclusions and Future Work

The proposed model takes advantage of the existing SAE and IMS infrastructures. At most, there should be a GCKS solution implemented, which may also be already implemented. It assumes the UE has support for P-Headers, which usually is of no concern at the UE level, but rather at the proxy-level. Still, there could be other ways of transmitting this information from one UE to another in case the P-Headers approach is not preferred. The UE CPU would have to deal with the cryptographic operations necessary to derive the session keys for the GDOI.

The security of the model inherits the security strengths and weaknesses of the IMS authentication and authorization procedures, as well as the LTE physical layer security aspects and those of the SAE core, EPC. Taking this aspect into consideration, the model indirectly provides another layer of security enforcement, because it is linking the SAE and IMS security procedures to those of the GDOI model. An attacker would not only have to overcome the LTE powerful authentication procedures, but also those of the EPC, the IMS registration, authentication and authorization to the home-network, and also the GDOI registration mechanisms. The GDOI implementation provides confidentiality, integrity and authentication, as well as protection against the man-in-the-middle and replay attacks. Because of the complexity of IMS and SAE interaction, there a lot of aspects not taken into consideration when designing this model. Still, this model can be extended and adapted according to each scenarios specificities. One aspect is the way of the SIP UA functionality and the IPsec group member functionality. On the UE, the SIP client software should be able to trigger the IPsec negotiation with GCKS.

This model describes a most common ad-hoc conference scenario, where one of the SIP peers also has the focus and mixer capabilities. A more general scenario, where the focus is a conference bridge, can be extended from the model proposed. Either the first attendee, or the focus itself, creates the trigger on the GCKS to derive a new group of keys. Also, this focus would have to be able to participate in multiple GDOI associations, one for each of the SIP GDOI-secured conferences it handles. Should the mixer be a standalone entity, it should also be notified of the newly existing conference and of the GDOI IPsec requirements it has. Also, the peers would have to know which of the existing mixers have GDOI capabilities. The GDOI capabilities of the focus and of the mixer would have to be published or advertised through some means. Either they are hard-coded on the cell-phone's firmware/software, they can be configured when purchasing the SIM from a particular vendor. These settings can also be downloaded automatically on the cell-phone in the moment of the SAE Initial Attach procedure (in case the model preferred is visited-network based, using a GCKS from the visited network) or when the UA registers to its home network (in case the model preferred is a home-network based GCKS or the visited network is not SAE). The GCKS should be a powerful machine, capable of handling at least hundreds of independent GDOI sessions.

Another aspect of this model is the way GCKS server authenticates the UE members. This clearly cannot be done statically, due to the enormous number of home subscribers, not to mention the ones in roaming. Most probably the GCKS would take advantage of the SAE and IMS architectures and, by means of either SAE or IMS capabilities and procedures, should be able to interrogate the HSS in order to get information about the UEs.

REFERENCES

[1] RFC 3455 Private Header (P-Header) Extensions to the Session Initiation
Protocol (SIP) for the 3rd-Generation Partnership Project (3GPP) - http://www.ietf.org/rfc/rfc3455.txt
[2] RFC 3329 Security Mechanism Agreement for the Session Initiation Protocol (SIP) - http://www.faqs.org/rfcs/rfc3329.html
[3] RFC 4457 The Session Initiation Protocol (SIP)P-User-Database Private-Header (P-Header)- http://www.ietf.org/rfc/rfc4457.txt
[4] RFC 3547 The Group Domain of Interpretation - http://www.ietf.org/rfc/rfc3547.txt
[5] RFC 4597 Conferencing Scenarios - http://tools.ietf.org/html/rfc4597
[6] RFC 4579 – SIP Call Control – Conferencing for User Agents - http://tools.ietf.org/html/rfc4579
[7] RFC 4353 - A Framework for Conferencing with the Session Initiation Protocol (SIP) - http://www.ietf.org/rfc/rfc4353.txt
[8] TS 33.401 - 3GPP SAE - Security architecture - http://www.3gpp.org/ftp/Specs/archive/33_series/33.401/
[9] TS 33.402 - 3GPP SAE - Security aspects of non-3GPP accesses - http://www.3gpp.org/ftp/Specs/archive/33_series/33.402/
[10] TS 33.203 - Access security for IP-based services - http://www.3gpp.org/ftp/Specs/archive/33_series/33.203/
[11] TS 33.210 – Network Domain Security; IP network layer security - http://www.3gpp.org/ftp/Specs/archive/33_series/33.210/
[12] TS 33.310 – Network Domain Security; Authentication Framework - http://www.3gpp.org/ftp/Specs/archive/33_series/33.310/

[13] "IP Multimedia Subsystem (IMS) signaling core security" - Ivan Tirado Kennesaw State University, Kennesaw, GA, 2008
[14] "Security issues with the IP multimedia subsystem (IMS)" - Michael T. Hunter, Russell J. Clark, Frank S. Park - Georgia Institute of Technology, Atlanta, GA, 2007
[15] "Seamless mobility and standards" - David Binet France Telecom, Rennes – 2009
[16] Tech-Invite: http://tech-invite.com/
[17] TS 29.294 – Tunneling Protocol for Control plane (GTPv2-C) http://www.3gpp.org/ftp/Specs/archive/29_series/29.274/
[18] "Media Independent Handover in Broadband Mobile Networks" - Jong-Moon Chung, Jae-Han Seol, Sang-Hyouk Choi Communication & Networking Laboratory (CNL) School of Electrical & Electronic Engineering Yonsei University Shinchon-Dong 134, Seodaemun-Gu Seoul 120-749, Republic of Korea. http://www.yonsei.ac.kr/jmc, Proceedings of the 6th WSEAS Int. Conf. on Electronics, Hardware, Wireless and Optical Communications, Corfu Island, Greece, February 16-19, 2007
[19] "Quality of Service and Security as Frameworks toward Next-Generation Wireless Networks" - ZORAN BOJKOVIĆ, BOJAN BAKMAZ Faculty of transport and traffic engineering University of Belgrade Vojvode Stepe 305, Belgrade SERBIA AND MONTENEGRO, WSEAS publication
[20] "Dual Identity Return Routability for the Security of Mobile Ipv6 Binding Updates within the Distributed Authentication Protocol" - ANDREW GEORGIADES, DR YUAN LUO, DR ABOUBAKER LASEBAE, PROF. RICHARD COMLEY Department of Computing Science Middlesex University Hendon campus, The Burroughs, London, NW4 4BT UNITED KINGDOM, www.cs.mdx.ac.uk, Proceedings of the 6th WSEAS International Conference on Applied Informatics and Communications, Elounda, Greece, August 18-20, 2006 (pp406-411)
[21] "A Framework to Mobility and Interactivity for Convergent Technologies" - Rodrigo F. Maia, Denis Gabos, Eduardo Bertassi, Ian Korolkovas, Edison Spina, Moacyr Martucci Jr. Department of Computing Engineering and Digital Systems Polytechnic School – University of Sao Paulo Av. Prof. Luciano Gualberto, trav. 3, no. 158. São Paulo, SP – Brazil, http://www.pcs.usp.br, Proceedings of the 5th WSEAS Int. Conf. on MULTIMEDIA, INTERNET AND VIDEO TECHNOLOGIES, Corfu, Greece, August 17-19, 2005 (pp95-100)
[22] "TS 23.401 - General Packet Radio Service (GPRS) enhancements for Evolved Universal Radio Access Network (E-UTRAN) access Terrestrial - http://www.3gpp.org/ftp/Specs/2009-06/Rel-8/23_series/