Authenticated Mobile Groups To Secure The Backhaul: A New Paradigm And Challenges

Naïm Qachri
Université Libre de Bruxelles, Computer Science Dpt.
CP212, boulevard du Triomphe, 1050 Brussels. Belgium
Email: naim.qachri@ulb.ac.be

Jean-Michel Dricot
Université Libre de Bruxelles – Ecole Polytechnique
OPERA Dpt. – Wireless Communications Group
CP165/51, 50 Av. Roosevelt, 1050 Brussels. Belgium
Email: jdricot@ulb.ac.be

Abstract—In this paper, it is proposed to use group communication cryptographic protocols as a new security paradigm. This new paradigm aims at redesigning the security of small cell communications over an insecure mobile backhaul. The Heterogeneous Networks are positioned and a review of the current security mechanisms and their flaws is provided. A security and performance comparison between the current mechanisms and our proposition is developed, and this comparison leads to new development opportunities to increase the security of the overall backhauling.

keywords – network security, LTE, group communication cryptography, heterogeneous networks, mobile networks

I. INTRODUCTION

Small cell communications have evolved by integrating external entities such as WiFi access points and WLANs. Unfortunately, in order to provide secure links between the external entities, IPSec tunnels have to be deployed for each connection. As a consequence, the number of tunnels to maintain grows exponentially with respect to the external entities.

This evolution has led to an increase of the overhead, the number of context switching, and the amount of interfaces to secure. Moreover, the security assumptions are unclear (since security constraints on some channels are optional, most of the protocols are sometimes built over an undefined context). The trust given on some entities depends on the context and changes regarding the considered protocols. Therefore, this evolution has brought new security breaches in the chain of trust that open the door to mobile impersonations, network poisoning and theft of credentials.

In this paper, it is proposed to use group communication cryptographic protocols (such as group key agreement or group authentication protocols) as a new security paradigm in order to redesign the current security of small cell communications. In this paradigm, each mobile creates a securely authenticated group consisting of the neighboring entities of the user equipment (UE). This authenticated group will last from the admission to the disconnection of the user equipment, and, therefore, it will define and maintain the security of all the communicating entities. The new paradigm is a radical shift since it allows authentication and a secure communication over an insecure mobile backhaul, i.e., without the need of IPSec tunnels at all. Furthermore, it has been observed that those authenticated groups will increase the security and the overall performance while decreasing the assumptions made on the network entities and interfaces.

The remainder of the paper is structured in seven sections. The second section presents how the future core network and backhaul are currently designed, their mechanisms and limits. The third section defines the lifecycle of a mobile regarding the management of the the backhaul. The fourth section establishes the state of the art of the current security protocols and the proposed alternatives. In the fifth section, the current paradigm is compared, regarding the actual cryptographic protocols, to our proposition. The sixth section compares and discusses the global performances of both paradigms. The seventh section concludes the paper.

II. THE HETNET ARCHITECTURE

Heterogeneous Networks (HetNets) is a paradigm proposed in order to integrate different wireless technologies (such as WiFi, 3G, 4G-LTE, etc) in a converged network to improve the available bandwidth and to extend the coverage. HetNets are divided into three main steps: the access to the network, the transport of the communication (backhauling), and the routing of the packets within the core network (called also the evolved packet core – EPC). The complete architecture is build upon IP networking. Fig. 1 presents the different entities of the architecture.

The user equipment (UE) is the mobile device connecting to the network and it connects to a point of presence. A point of presence (PoP) is the point where a mobile node connects to the core network. Potential points of presence are WiFi access points, NodeBs (3G), or eNodeBs (4G). The Home Subscribed Server (HSS)
merges the role of the \textit{authentication center} (AuC) and the role of the Home Local Register (HLR) in previous architectures. A user equipment is authenticated to the core network thanks to the HSS. This server is often co-located with a Radius Server and a AAA Server to integrate WiFi PoPs. The points of presence will be connected to each other through dedicated interfaces depending on the technology. For instance, the X2 abstract interface will allow eNodeBs to interoperate and communicate autonomously in 4G-LTE and LTE-Advanced.

The Serving Gateways (S-GW1 and S-GW2) route and forward user data packets. Serving GPRS Support Node (SGSN) is the interface between the 2G/3G Radio Access Network (GERAN/UTRAN) and the core network. The Mobility Management Entity (MME) acts as a geographic interface for mobility management between eNodeBs of the same operator. The Packet Data Network GateWay (P-GW) routes the data packets between the user equipment and external IP networks. Finally, the evolved Packet Data Gateway (ePDG) is the interface between the core network and non-3GPP trusted (or untrusted) access networks.

From Fig. 1, it is possible to group entities with respect to their role in the security of the communication. These groups represent levels of trust (trusted authority, trusted or untrusted entities). For instance, the entities of the EPC are on the same level of trust and can be considered as a group of trusted entities. The only exception of the EPC is the HSS, because this entity manages the accesses and the privileges of each entity in the network operator. Since the HSS is trusted authority of the network, it is a separated level of trust.

Another group and level of trust are the EnodeBs managed by a MME, which acts like an entry point to the EPC. None of the eNodeBs have a direct access to the core network and they can only communicate with the other eNodeBs managed by the MME (through the X2 interface), with the MME itself, and with the connected devices. The ePDG plays more or less the same role as the MME for the WiFi access points, except that the access points communicate exclusively with the ePDG and the connected devices in the scheme, but, in fine, this group has the same level of trust than the groups managed by the MME. Those two levels of trust and the three groups of entities are physically separated and clearly defined.

Finally, the mobile that communicates directly with the different points of presence is a level of trust by itself, and the device is supposed to be isolated. It can be noted that this is no longer true with the introduction of the Non-Access Stratum in 4G-LTE, the introduction of data offloading (increasing and optimizing the available bandwidth through multiple accesses) and the use of pre-authentication mechanisms (that authenticates the device to neighbor points of presence to fasten soft handovers). From those additions, a virtual group is implicitly present in the network (colored in red in Fig. 1). This group is initiated by the mobile, by choosing and introducing the different members (such as the eNodeBs and the WiFi access points) in its group of communication, but the group is managed by the network through the MMEs and ePDGs that accept and authenticate the sessions. The different points of presence provide a smart backhaul between the core network and the mobile.

Unfortunately, the mobile, its communications and its security have never been designed to be used in this way. It means that the current proposed mechanisms will not provide the needed flexibility and security to implement efficiently those new features. Indeed, most of the implemented security mechanisms have been designed to make point to point connections. In the remainder of the paper, these groups will be called the \textit{mobile groups}, and the term \textit{core entry point} will be used to group the following network entities on a unique denomination: S-GW1, MME and ePDG. For the eNodeB, NodeB and the WiFi access points, the term \textit{point of presences} will be used for the same purpose. Fig. 2 proposes a scheme representing the notion of mobile groups.

![Fig. 2. The representation of mobile group communications](image.png)

### III. The Lifecycle of a Mobile Group

The core network group and the EnodeBs groups are stable. Once the setup done, it is rare that new entities appears or disappears regularly. On the opposite, the
mobile group is highly dynamic and proposes its specific challenges due to its lifecycle.

The mobile device lifecycle can be defined as the following different procedures:

1) Initial Connection – the mobile connects himself to the network through one or more points of presence and core entry points;
2) Paging – it refers to the process when an operator wants to reach the mobile, or the mobile wants to reach another one;
3) Communication – this is the process where the mobile sends data to the network;
4) Handovers – this process occurs when a mobile moves from a set of points of presence to another;
5) Disconnection – this process closes explicitly (or implicitly in case of connectivity loss) the lifecycle and the connection of a mobile.

Each of this procedures happens an arbitrary number of times during the life of a mobile device. There is not a specific order for those procedures to appear, except for the first and the last listed. The procedures 2, 3 and 4 can appear at any time during the lifecycle (even sometimes together).

The number of instances of those procedures can be fixed: a connection and reconnection happens one time during a complete cycle. The paging happens a number of \( P \) times, handovers \( H \) times, and the device initiates communication a number of \( C \) times. This lifecycle will be the scenario that will be used to describe and analyze the behavior of a mobile group from a security point of view.

IV. STATE OF THE ART OF THE SECURITY MECHANISMS

A. Current cryptographic mechanisms and protocols

1) 3G, 4G-LTE and 4G-LTE Advanced: The security of 3G communication has been designed to share two keys between a NodeB and a mobile for authenticated encryption communications. In the case of 4G communications, the MME and the mobile share four keys for one link and a fifth is shared between the mobile and the eNodeB. The signaling between the MME and the mobile, and the data protection between the eNode and the mobile are separated and do not use the same security assets. The protocol to authenticate the mobile and to agree on keys is AKA’ (RFC 5448).

The links between the core network and the EnodeBs are secured with the use of IPSec (RFC 4301, 4303 and 4305). It means that each packets are decrypted and encrypted for the next hop before to be again decrypted and routed inside the core network.

2) 802.11: The proposed solution to integrate WLAN inside mobile operator networks is to use tunneled connections using IPSec and EAP-TLS. They are tunneled from the mobile to the access points (trusted access point) or to the ePDG (untrusted access point). S2a Mobility based on GPRS Tunneling Protocol (SaMOG)[1] details the integration of the WiFi base stations into the core network, and the last draft of the IEEE 802.11 standard [14] includes an integration of some 3GPP standards such as the EAP-SIM protocol (RFC 4186). WLAN integration does not currently support mobility and particularly handover procedures. Some works exist, such as [5], but no protocol has been currently ratified. Once a connection is established, a pair of keys is shared between the access point or the ePDG and the mobile, and in case of mobility, the connection is restarted.

3) IPSec: IPSec has been designed to connect securely computers. The protocols and the algorithms that have been integrated are varied. The core network uses point to point communications to secure the links between the core entry points and the points of presence. Once authenticated, each point shares a pair of keys with a core entry point. IPSec key agreement protocol (based on the IKEv2 protocol – RFC 4306 and 4555) has been subject to criticism and security flaws[3], [8]. Solutions exist such dynamic multipoint VPNs, but they are not suited for the backhaul since it cannot manage the mobility of a user equipment that needs to be connected to more than one entry point. Moreover, user equipments should not have a direct access to the core network for security reasons.

4) The threat model: This paradigm creates two distinct strata on the security data plane. The upside stratum covers the security of the communications between the core entry points and the points of presence, and the downside stratum covers the security of the communications between the points of presence and the mobile device. This notion presented in Fig. 2 shows the importance to consider the side effects on the routing of packets through the different channels.

This division creates multiple flows of communication and each of them are secured with different keys. Each flow is physically and virtually separated from the others, and the communication packets are decrypted and encrypted at each hop. The EPC must then trust the points of presence. That assumption opens the door to security breaches that the current paradigm cannot thwart.

a) Attacks from a corrupted point of presence: Since Self-Organizing Networks help to change the configuration of the network (with the possible help of the EPC). It is possible to transfer packets tagged with the IP of a mobile device in order to introduce errors (creating localized denial of services or corrupted communications). For instance, since the only authenticating information of decrypted packet is the IP address or the TIMSI (that could have been eavesdropped), a corrupted point of presence can transfer false packets to a safe
one, as if it was a trusted group member. There is no possible way to prove that they are not valid. With that configuration, the safe point of presence will transfer the corrupting packets to the core network. In other cases, corrupted points of presence could make a poisoning of the communications by transferring invalid packets directly to the SG-W1 entity of the group without being distinguished from the valid ones.

b) Attacks from a corrupted point of presence and a malicious node: Since the points of presence do not know which are the other group members, duplicating credentials is facilitated, and less noticeable by the operators. Before the mobile group paradigm, duplicating credentials meant having twice the user connected instead of one. At the group mobile era, there will not be difference between two or three valid sessions. Another threat is the use of context transfers to leak the current credentials to a malicious mobile device to maintain useless flows (creating a potential overbilling attack).

c) The point of failure: Those attacks are possible because the members of the group are not aware of the other members and cannot make the distinction between them and an attacker.

B. Group communication cryptography

Our paradigm states that the authenticated mobile group is managed through a unified secure design over an insecure backhaul. Only the mobile device and the core entry points are trusted. At the authentication, any trusted member should choose a common secure set of keys that would authenticate the authorized members. During the mobility of a device, the authenticated mobile group shall update the shared keys regarding the group changes to preserve the overall security. The authenticated group avoid the unauthenticated packet transfers, because the packets do not need to be decrypted during the transport over the backhaul. The authenticated group creates a trusted link between each members (simplifying, for instance, the security needs on the X2 interface). Similarly, a mobile device and its clone could not be easily set up parallel sessions since each group has a different set of keys (virtualizing the notion of one unique connected user).

A branch of the research in cryptography has been dedicated to develop group communication cryptographic protocols. They aims to secure communication of groups. If two-party protocols are dedicated to one-to-one communications, group communication cryptographic protocols are designed to many-to-many communications. In this model, there are two possibilities to model the group, every member are trusted for the protocol, or every member are untrusted and they use a trusted party.

The different members of the groups authenticate to the others. Once the authentication succeeds, the members can extract a pair of keys, in that case the protocol is called group key agreement. From those keys, anyone inside the group can securely communicate with the others. Depending on the propositions, the cryptographic protocol are developed with specific constraints to reach some security properties. Among them, if the designer wants the signature to be traceable to discover corrupted participants, traitor tracing properties will be added in the specification of the protocol.

Group communication cryptographic protocols cover a large set of protocols such as group key agreements[13], [4], [15], [19], [11], [6], threshold schemes[9], and traitor tracing[17], [10], [7], [2], [20], [19] use group communication cryptography in networking with a hierarchical context or through independent clusters.

Among the proposed protocols, most of them use elliptic curve Diffie-Hellman (ECDH)[16], RSA[17], [10], [18], or secret sharing[12] for group authenticated communication. The solutions are built essentially using asymmetric cryptography. They are not often contextualized, and they lacks of performance analysis. The solutions taken as references are therefore not perfect for the needed tasks, such as efficient dynamic rekeying. For instance, it does not exist, as far as we know, group key agreement purely based on symmetric key cryptography.

V. THE RELATED SECURITY OPERATIONS OF THE GROUP DURING A LIFECYCLE

The five presented procedures in Section III can be directly translated into security procedures build on tools, such as authentication and key agreement protocols, or encryption algorithms. Each procedure will be presented in both paradigms. The comparison will be based on three metrics: the number of contexts maintained to execute the processes(related delays), the number of keys (traceability of the system) and the number of messages sent over the network (and the related overhead). The two last criterion are straightforward, but the first criterion measures the number of context switch and the delays that they cost in the communication latencies for a mobile. The selected criterions are the most accurate to assess the performance quality of the solution. The number $n$ will refer to the number of members that are in a group. Since, the computational complexity is dependent of the choice of algorithms, It will not be accurate to use it as a metric.

The number of core entry points is fixed to $e$ and the number of point of presences to $p$ and the relation between the variables is $n = p + e + 1$. The model for group key agreement is fixed to the protocol of [16] to allow comparisons. The number of participants in the protocol are limited to the core entry points and the mobile, since the points of presence are untrusted.

1) Connection process:
TABLE I
PERFORMANCES DURING A CONNECTION

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Current Paradigm</th>
<th>Group Paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>keys</td>
<td>4 \times p keys, or</td>
<td>2 keys, and e + 1 cert.</td>
</tr>
<tr>
<td></td>
<td>2 \times p keys and n cert.</td>
<td></td>
</tr>
<tr>
<td>contexts</td>
<td>p</td>
<td>e + 1</td>
</tr>
<tr>
<td>messages</td>
<td>4 \times p</td>
<td>e \times 2 + 2p</td>
</tr>
</tbody>
</table>

TABLE II
PERFORMANCES PAGING PROCESS

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Current Paradigm</th>
<th>Group Paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>keys</td>
<td>4 \times p keys</td>
<td>2 keys</td>
</tr>
<tr>
<td>contexts</td>
<td>p</td>
<td>1</td>
</tr>
<tr>
<td>messages</td>
<td>e + p broad. req.</td>
<td>e + p multi. req.</td>
</tr>
<tr>
<td></td>
<td>p resp.</td>
<td>1 + p multi. resp.</td>
</tr>
</tbody>
</table>

a) Current: The mobile establish with each point of presence a secure connection. During the establishment a key, a set of keys is shared between the mobile and a the point of presence. This step (see Fig.3) is repeated for each point of presence that the mobile want to add to its group. The average number of message to succeed an authentication is 4 (for AKA', IKEv2 or EAP-TLS). An instance of AKA' protocol generates 5 secret keys, and instance of EAP-TLS and IKEv2 generate 2 keys and require the use of 1 certificate for the mobile and 1 per access point or ePDG.

b) Mobile Auth. Group: The model of protocols uses n certificates. At the end 2 secret keys are computed for the group. Each participant uses multicasting or broadcasting for the requests and responses of the protocol. Their will be an exchange of e^2 messages and each comparisons are summarized in a table.

![Fig. 3. Connection in the current paradigm](image)

![Fig. 4. Connection in the authenticated group paradigm](image)

2) Paging process:

a) Current: The different entities of the operating network could want to reach a mobile (in case of incoming call or data for instance). Core entry points will send a paging request authenticated with the mobile shared keys. For each paging request, a response will be authenticated to each needed member. The number of requests and responses are at worst p using for each sequence 1 or 2 keys.

b) Mobile Auth. Group: The paradigm could optimize the process by multicasting the requests and the responses to the different members of the group. Furthermore, the paging could easily be forwarded to the different members without changing the encryption and the authentication of the requests. The p paging requests and responses messages could be broadcasted directly (e + p + 1 for the requests).

3) Handover process:

a) Current: The disconnections are explicit during a handover for the points of presence that will not remain in the group. If the new points of presence are not on the same geographical area, the core entry points transfer the context to the new ones. If the point of presence shares the same geographical area, they will use dedicated interfaces (such as the X2 interface) to transfer directly the keys.

From the security point of view, using interfaces to just transfer the keys between the points of presence is not secure. Another mechanism is to pre-authenticate the neighbors of the connected points of presence. After a handover, it will remain to pre-authenticate the new neighbors and to release the credentials of the points of presence that are not more in the neighborhood.

l members are leaving and r members are entering in the group. The counting is then q disconnection messages and r pre-authentications protocols with an average of 4 messages per authentication protocol. l is considered equal to r, in the remainder of the publication, to simplify the assessment of the process.

b) Mobile Auth. Group: The handover mechanisms introduce new points of presence (and the related core entry points) and releases others from the group. Group key agreement protocols are not often designed to allow flexibility and high mobility. Most of the time, the procedure is to reset completely the group, or to replay the group authentication phase (pessimistic case). Improvements can be done in the process management to cover this issue by sending messages to the core entry points and receiving the resulting responses that will integrate new members in the group with a fresh key (optimistic case).

4) Communicating process: During a communication, the device is sending encrypted and authenticated packets directly to the different entities following a predefined scenario. In the current paradigm, the packets are decrypted and encrypted following the upside and downside strata that increase by 2 the number of contexts. The main difference between the current and the mobile authenticated group paradigm is the possibility to optimize the decryption of the packets and their ordering thanks to the unique encryption of packets.

5) Disconnection process: In case of an explicit complete disconnection, the mobile sends authenticated
disconnection requests to each of the concerned entities. In the current case, the requests must be authenticated and sent to each point of presence. In the other, just 1 multicasted message is enough.

VI. PERFORMANCE COMPARISONS

The comparisons will be done on the basis of two scenarios: a high mobility scenario and a high consumption scenario. The high mobility scenario makes the assumptions that a mobile device moves and triggers many handovers, but does not consume much data. The high consuming scenario makes the assumptions that a mobile device does not move often but consumes a large amount of data. Once both scenarios presented, the analyses will assess the scenarios from an asymptotical point of view.

The overhead is assessed on the number of messages exchanged during a lifecycle that is not linked to a communication process: messages = \( P \times (\text{paging}) + H \times (\text{handover}) + \text{connection} + \text{disconnection} \)

Once replaced by their value from the Section V, the following equations are obtained for each paradigm (pessimistic and optimistic case separated):

- \( Curr = P \times (e + 2p) + H \times (6r) + 4p + 2p \) (1)
- \( AuthPess = P \times (e + p + 1) + (H + 1) \times (e^2 + 2p) + 1 \) (2)
- \( AuthOpt = P \times (e + p + 1) + (H + 0.5) \times (4p + e^2 + 1) \) (3)

The equation (2) shows that if the number of handovers grows, the overhead grows with a quadratic factor.

B. The delays of context switching

The delays of mobile context switching communication introduced by a paradigm is the number of switch from a context to another multiply by the average time for one switch. In a high mobile scenario, the communications are short and not so many and there is no significant differences between the paradigms. In a high consumption scenario, each communication process could have a fixed data trunk a size. The following equation resume the statement: delay = \( \frac{numMsg}{numContexts} \times \text{switchTime} \).

The latest equation is derived and inserted in the lifecycle equation for the paradigms:

- \( Curr = (P + 2) \times \frac{p + H \times 2r + C \times \text{switchTime}}{\text{numMsg}} \) (4)
- \( AuthPess = (H + 1) \times \frac{e + 1}{\text{switchTime}} \) (5)
- \( AuthOpt = (H + e + 1) \times \text{switchTime} \) (6)

The network delays are bounded for just one mobile group. It means that there will be more groups and much more contexts to manage over the backhaul for each points of presence and core entry points.

C. Asymptotic behavior analysis

The equations (1) to (3) differs since the new paradigm adds a quadratic complexity in terms of message numbers due to the actual design of group communication cryptographic protocols. This quadratic addition is negligible in a high consumption scenario, but not during a high mobility scenario. In both cases, if the number of core entry points does not exceed 3 or 4 entities and that \( r \) number remains small, the differences between the two scenarios will not be so large (particularly if we consider the optimistic case).

In the case of context switching (equations (4) to (6)), the asymptotic behavior of both paradigms shows that the delays will increase linearly with the number of communications in the current paradigm and with much more terms in the addition. In the group communication
paradigm, the only factors influencing the delays is the number of handovers, and these delays the lowest in the optimistic case.

A more pragmatic approach can confirm the current analysis by fixing some parameters: \( P = 5, p = 4, e = 2, r = 2 \) for the high mobile scenarios, and \( H = 15, \text{switchTime} = 10^{-2} \text{seconds}, \text{numMsg} = 69905 \) based on 100MB communications with 1500 bytes Maximum Transmission Unit for IP packets for the high consumption scenario. It can be extracted from those parameters, that the overhead of the authenticated group communication paradigm exceeds the overhead of the current paradigm only after reaching a certain number of handovers. Similarly, if large data is sent over and over, the delays will grow dramatically.

Something that appears is that both paradigms offer interesting advantages. The current paradigm cost a bit less in terms of overhead, but cost much in terms of delays and the paradigm is much less insecure. Our paradigm relaxes many constraints, is more secure, avoids a maximum of delays, but does not yet fit perfectly in a highly mobile context. The differences come from the fact that the current paradigm offers dedicated channels (one-to-one communications) and that authenticated mobile group paradigm offers the opportunity to communicate securely with any entity of a group (many-to-many communications). The compromise would be to develop a secure mobile authenticated group based on one-to-many communications. It means that designing hybrid cryptographic protocols would be a target to capture the best parts of both paradigms.

VII. CONCLUSION AND OPEN CHALLENGES

Hybrid mechanisms are an open challenge, since few cryptographic protocols are designed in this way and with that specific context. Some aspects are difficult to update in the current paradigm (for instance the SIM card within user equipments), and hybrid mechanisms could relax some of the evolution contraints by designing new cryptographic protocols.

Finally, a new secure playground has been created and presented to build future secure protocols for the future of mobile networking. The mobile authenticated group paradigm offers many advantages over the current paradigm, even though the mobility performances can be largely improved with further studies.

REFERENCES

[1] 3GPP. “Study on s2a mobility based on gms tunnelling protocol (gtp) and wireless local area network (wlan) access to the enhanced packet core (epc) network (samog),” 3GPP, Tech. Rep. 3GPP TR 23.852, June 2012.


