**3GPP TSG-SA3 Meeting #107-e-Ad Hoc *draft\_S3-221329-r2***

**e-meeting, 27 June - 1 July 2022**

**Source: Interdigital**

**Title: New solution for Key issue #1**

**Document for: Approval**

**Agenda Item: 5.4**

# 1 Decision/action requested

***This document proposes a new solution to key issue #1: Privacy aspects of variable length user identifiers. SA3 is kindly requested to approve this doc.***

# 2 References

[1] 3GPP TR 33.870 v0.1.0

# 3 Rationale

This new solution proposes a padding mechanism to protect the privacy of variable length SUPIs in NAI format.

# 4 Detailed proposal

\*\*\*\* START OF CHANGE 1 \*\*\*\*

## 6.X Solution # X: Padding SUPIs in NAI format with Random Length of Characters for non-null schemes

### 6.X.1 Introduction

Key issue #1, Privacy aspects of variable length user identifiers, states that some networks may decide to allow user identifiers with variable length, e.g., in case of NAI type SUPI. The length can become visible to an attacker in case a length preserving encryption scheme is being used for identifier concealment. If an attacker can learn something about the length, such knowledge will reduce the size of the anonymity set.

The proposed solution aims to address Key issue #1 by adding padding and unpadding mechanisms (Steps 1 and 7 in Figure 6.X.2-1) with complementing functionalities before and after the existing processes specified in clauses 5.8.2, 6.12, and Annex C of TR 33.501 [aa]. Padding is performed in the UE and un-padding in the UDM/SIDF as shown in Figure 6.X.2-1.

### 6.X.2 Solution details

This solution proposes a padding mechanism to protect the privacy of variable length SUPIs in NAI format. In this solution, the UE pads the username with a random length padding. The length of the random padding depends on the length of the original username length to maximize the k-anonymity value and minimize the complexity of the deployed privacy protection solution.

The solution reuses the existing ECIES based de/concealment mechanism as described in TS 33.501 [aa]. The proposed padding mechanism provides backward compatibility with legacy UEs by using an optional padding method indication included in the SUCI output.

The text below describes the steps needed to pad the SUPI’s username with special characters:



Figure 6.X.2-1: Authentication initiation using SUCI in NAI format with random padding

1. The UE pads (e.g., by append, prepend) the cleartext username part of NAI, with a randomly selected length of special characters that cannot be used for a username based on IETF RFC 7542 [zz] and RFC 3269 [yy] (i.e., not UTF-8 (see RFC 3629 [yy]) character set)

To support random padding while supporting legacy UEs, and to accommodate future concealing/padding methods, the UE includes a padding method indication as part of the final ECIES output so that the SIDF can detect whether and how to unpad de-concealed SUCI.

2. The UE performs ECIES-based encryption on the resulting username padded with special characters to generate the ciphertext used to form the final SUCI output (e.g., Pub key || ciphertext of username with special chars padding || MAC tag)

3. UE sends the resulting SUCI to the network

4. SEAF forwards the SUCI containing SUPI in NAI format to the AUSF

5. AUSF forwards the SUCI containing SUPI in NAI format to the UDM/SIDF

6. UDM/SIDF performs ECIES-based decryption of the ciphertext to deconceal (padded) SUPI in NAI format as per TS 33.501 [aa].

7. If UDM/SIDF receives a padding method indication with the SUCI, UDM/SIDF unpads SUPI in NAI format based on the padding method indication. From the resulting cleartext padded username UDM/SIDF filters out special characters that cannot be used for a username based on IETF RFC 7542 [zz] and RFC 3629 [yy] (i.e., not a UTF-8 character set) to obtain the actual username part of the SUPI.

The UE may be pre-configured by the operator with the supported padding method to be used. UE may be pre-configured with other parameters to be used during padding such as padding character set, min-max values of added padding, or encoding scheme (e.g., append, prepend).

NOTE: if lmin and lmax values are too small, then an attacker might still be able to infer something of the distribution of lengths after padding

Editor's Note: It is FFS if and how, by analysis of the SUCI output, an attacker can detect if the padding method is applied or not, and then use this knowledge to the adverse effect.

### 6.X.3 Evaluation

FFS.

\*\*\*\* END OF CHANGE 1 \*\*\*\*

\*\*\*\* START OF CHANGE 2 \*\*\*\*

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[aa] 3GPP TR 33.501: " Security architecture and procedures for 5G system”.

[yy] RFC 3629: “UTF-8, a transformation format of ISO 10646".

[zz] IETF RFC 7542: "The Network Access Identifier".

\*\*\*\* END OF CHANGE 2 \*\*\*\*