**3GPP TSG-SA3 Meeting #108e-AdHoc *S3-222821-r1***

**e-meeting, 10 - 14 October 2022**

**Source: Ericsson**

**Title: Hash-based solution to the leakage of the length of SUPI through SUCI**

**Document for: Approval**

**Agenda Item: 5.4**

# 1 Decision/action requested

***In this box give a very clear / short /concise statement of what is wanted.***

# 2 References

[1] 3GPP TS 23.003: "Numbering, addressing and identification".

[2] 3GPP TS 33.501: "Security architecture and procedures for 5G System".

# 3 Rationale

According to TS 23.003 [1], subscribers’ permanent identifiers, called SUPIs, are allowed to be in Network Access Identifier (NAI) format (username@realm), which can have variable lengths.

Using the Elliptic Curve Integrated Encryption Scheme (ECIES), a user device and the home network agree on a shared key by leveraging the public key of the home network. The user device uses the shared key in a symmetric encryption scheme (AES in counter mode) to encrypt SUPIs, into concealed identifiers, called SUCIs [2].

In the symmetric-key setup, security notions like real-or-random, left-or-right, or semantic security are defined in the context where plaintexts have the same lengths [3]. Though AES counter mode is secure according to these notions, direct use of it is not sufficient to serve an intended purpose of SUCIs -- indistinguishability of SUCIs. This is because SUPIs can have different lengths, and in counter mode, the length of the plaintext and the corresponding ciphertext is the same. Therefore, when two SUPIs have different lengths, their ciphertexts are distinguishable from each other, causing the reduction of the associated anonymity set.

# 4 Detailed proposal

\*\*\* 1st CHANGE \*\*\*

# 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".

[2] 3GPP TS 24:501: “Non-Access-Stratum (NAS) protocol for 5G System (5GS)”.

[XX] 3GPP TS 33.501: "Security architecture and procedures for 5G System".

[XY] 3GPP TS 33.220: "Generic Authentication Architecture (GAA); Generic Bootstrapping Architecture (GBA)".

\*\*\* 2nd CHANGE \*\*\*

## 6.A Solution #A: Concealing length of SUPIs in SUCIs by hashing the SUPIs

### 6.A.1 Introduction

This is a solution to KI #1. The solution uses pre-encryption hashing of SUPIs.

Editor’s Note: This solution may need to be updated to align with the KI once the ENs in the KI is resolved.

### 6.A.2 Solution details

#### 6.A.2.1 Solution Basics

The basics of the solution are:

- This solution uses hashing of SUPIs to protect against the anonymity set reduction.

- The solution uses the hashes of SUPIs, instead of SUPIs themselves to compute SUCIs.

- The UDR maintains an injective map between the SUPIs and their unkeyed hashes.

- The necessary hashing parameters, which can be the home operator's choices, are stored in USIM and UDM.

NOTE 1: In this solution, SUPI in IMSI format is not hashed because it is fixed length, and SUCI cannot be attributed to a particular SUPI based on length.

#### 6.A.2.2 Hashing parameters

Padding parameters comprise the name of a hash function and the desired length value. An example of a suitable hash function could be the 3GPP key derivation function (KDF) specified in TS 33.220 [XY] with a dummy key, e.g., all zeros. The output of the KDF could be truncated to the desired length.

#### 6.A.2.3 UE Side

- The UE computes an unkeyed hash of the NAI format SUPI and encrypts the hash of the SUPI, instead of the SUPI itself, into the concealed subscription identifier part of a SUCI.

- The UE also includes a signal for the UDM in the final SUCI so that the UDM can know that the concealed subscription identifier part of the SUCI is computed from the hash of the SUPI, not the SUPI itself. This signaling can be done, for example, by using a new protection scheme identifier.

- Everything else regarding SUCI computation remains the same. Hashing of SUPIs in NAI format is performed by the same component, either USIM or ME, that performs the calculation of SUCI.

#### 6.A.2.4 Home Network Side

The UDR maintains an injective map between the SUPIs and their unkeyed hashes. Therefore, the length of the hash function has to be chosen in a way so that probability of collision is astronomically small. Once the SUCI arrives at the UDM, the following computations happen:

- On the network side, the UDM gets the SUCI decrypted with the help of ARPF and SIDF and obtains the deconcealed subscription identifier.

- The UDM checks the signal (e.g., protection scheme identifier, if used) set by the UE to know if the deconcealed subscription identifier is a SUPI or the hash of the SUPI.

- If the deconcealed subscription identifier is signaled to be a hash of the SUPI, then the UDM sends the hash of the SUPI to the UDR.

- The UDR retrieves the SUPI and sends it to the UDM.

### 6.A.3 Evaluation

Editor’s Note: assessment of the potential impact on the UDR is FFS.

TBD

\*\*\* END OF CHANGES \*\*\*