|  |  |
| --- | --- |
| 3GPP TR 33.870 V0.4.0 (2022-10) | |
| Technical Report | |
| 3rd Generation Partnership Project;  Technical Specification Group Services and System Aspects;  Study of privacy of identifiers over radio access;  (Release 18) | |
|  | |
|  | 3GPP-logo_web |
|  | |
| The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP. The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented. This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification. Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices. | |

|  |
| --- |
|  |
| ***3GPP***  Postal address  3GPP support office address  650 Route des Lucioles - Sophia Antipolis  Valbonne - FRANCE  Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16  Internet  http://www.3gpp.org |
| ***Copyright Notification***  No part may be reproduced except as authorized by written permission. The copyright and the foregoing restriction extend to reproduction in all media.  © 2022, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).  All rights reserved.  UMTS™ is a Trade Mark of ETSI registered for the benefit of its members  3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners LTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners  GSM® and the GSM logo are registered and owned by the GSM Association |

Contents

Foreword 5

Introduction 6

1 Scope 7

2 References 7

3 Definitions of terms, symbols and abbreviations 7

3.1 Terms 7

3.2 Symbols 8

3.3 Abbreviations 8

4 Architectural considerations 8

5 Key issues 8

5.1 Key issue #1: Privacy aspects of variable length user identifiers 8

5.1.1 Key issue details 8

5.1.2 Security threats 9

5.2 Key Issue #2: Users Identified by Priority Access 9

5.X Key issue #X: 11

5.X.1 Key issue details 11

5.X.2 Threats 11

5.X.3 Potential security requirements 11

6 Solutions 11

6.1 Solution #1: Use of fixed length identifiers to protect against anonymity set reduction 11

6.1.1 Introduction 11

6.1.2 Solution details 11

6.1.3 Evaluation 12

6.2 Solution # 2: Padding SUPIs in NAI format with Random Length of Characters for non-null schemes 12

6.2.1 Introduction 12

6.2.2 Solution details 12

6.2.3 Evaluation 14

6.3 Solution #3: Pseudonym based solution for k-anonymity of SUPI/SUCI 14

6.3.1 Introduction 14

6.3.2 Solution details 14

6.3.2.1 General 14

6.3.2.2 Procedure 14

6.3.2.3 Guidance on pseudonym allocation 15

6.3.3 Evaluation 15

6.4 Solution #4: Limited length of SUPIs in NAI format 15

6.4.1 Introduction 15

6.4.2 Solution details 15

6.4.3 Evaluation 16

6.6 Solution #6: Padding SUPI in NAI format to conceal the username length 18

6.6.1 Introduction 18

6.6.2 Solution details 18

6.6.3 Evaluation 19

6.7 Solution #7: Concealing length of SUPIs in SUCIs by truncating the SUPIs 20

6.7.1 Introduction 20

6.7.2 Solution details 20

6.7.2.1 UE Side 20

6.7.2.2 Home Network Side 20

6.7.3 Evaluation 20

6.8 Solution #8: Use of fixed length “username” for NAI 20

6.8.1 Introduction 20

6.8.2 Solution details 21

6.8.3 Evaluation 21

6.9 Solution #9: Concealing length of SUPIs in SUCIs by padding the SUPIs 21

6.9.1 Introduction 21

6.9.2 Solution details 21

6.9.2.1 Solution Basics 21

6.9.2.2 Padding parameters 21

6.9.2.3 UE Side 22

6.9.2.4 Home Network Side 22

6.9.3 Evaluation 22

6.10 Solution #10: Concealing length of SUPIs in SUCIs by hashing the SUPIs 22

6.10.1 Introduction 22

6.10.2 Solution details 22

6.10.2.1 Solution Basics 22

6.10.2.2 Hashing parameters 23

6.10.2.3 UE Side 23

6.10.2.4 Home Network Side 23

6.10.3 Evaluation 23

6.A Solution #A: <Solution Title> 23

6.A.1 Introduction 23

6.A.2 Solution details 23

6.A.3 Evaluation 23

7 Conclusions 24

Annex A: List of 3GPP identifiers. 24

Annex <X> : Change history 25

# Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, modal verbs have the following meanings:

**shall** indicates a mandatory requirement to do something

**shall not** indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

**should** indicates a recommendation to do something

**should not** indicates a recommendation not to do something

**may** indicates permission to do something

**need not** indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

**can** indicates that something is possible

**cannot** indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

**will** indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**will not** indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**might** indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

**might not** indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

**is** (or any other verb in the indicative mood) indicates a statement of fact

**is not** (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

# Introduction

This clause is optional. If it exists, it shall be the second unnumbered clause.

# 1 Scope

The present document achieves the following objectives:

Analysis of 3GPP identifiers that represent either targets of privacy attacks themselves or may aid adversaries in privacy attacks.

Analysis of the feasibility of privacy attacks; the analysis should consider newer methodologies such as those involving AI/ML

Analysis of available countermeasures, including technical remedies, security guidance, to the identified and feasible privacy attacks; the analysis should consider newer methodologies such as those involving AI/ML

Recommendations to the identified and feasible privacy attacks. Recommendations may include but are not limited to non-technical remedies, architectural recommendations, and procedural fixes.

# 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)”.

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

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

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

[6] 3GPP TS 33.220: "Generic Authentication Architecture (GAA); Generic Bootstrapping Architecture (GBA)".

# 3 Definitions of terms, symbols and abbreviations

This clause and its three subclauses are mandatory. The contents shall be shown as "void" if the TS/TR does not define any terms, symbols, or abbreviations.

## 3.1 Terms

For the purposes of the present document, the terms given in 3GPP TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

**example:** text used to clarify abstract rules by applying them literally.

**k-anonymity:** is a property of anonymized data. The release of such data is said to have the k-anonymity property if the information for each subject contained in the release cannot be distinguished from at least k - 1 subjects whose information also appear in the release.

## 3.2 Symbols

For the purposes of the present document, the following symbols apply:

<symbol> <Explanation>

## 3.3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

<ABBREVIATION> <Expansion>

# 4 Architectural considerations

### 

# 5 Key issues

## 5.1 Key issue #1: Privacy aspects of variable length user identifiers

### 5.1.1 Key issue details

Networks can decide to allow user identifiers with variable length, e.g., in case SUPI of type NAI. If an attacker can learn something about the length, this will reduce the size of the anonymity set.

The length can become visible to an attacker in case a length preserving encryption scheme is being used for identifier concealment.

3GPP authentication schemes referred to in TS33.501[3] are: 5G-AKA and EAP-AKA', which are mandatory to support, as well as other key generating EAP methods, e.g., EAP-TLS and EAP-TTLS. All of these methods identify the subscriber using SUPI. As SUPI of type IMSI has a fixed length, this key issue is not applicable to SUPIs of type IMSI.

For NAI based SUPI types, the authentication method may leak the length of the SUPI even if identifier privacy mechanisms specified for the authentication methods are used.

These privacy mechanisms are:

- For 5G-AKA and EAP-AKA' the mechanisms are profile A, profile B, or proprietary SUCI calculation scheme.

- When some EAP based methods are used, e.g., EAP-TLS and EAP-TTLS, an anonymous SUCI can be used, and the actual SUPI is sent after an EAP secure channel is established, e.g. the TLS tunnel.

### 5.1.2 Security threats

An attacker on the air interface can identify and track subscribers with unusual lengths of the username field of variable-length SUPI in NAI format even if it is confidentiality protected (e.g., relatively short or long SUPIs).

Note: NAIs can be used for any EAP method.

If such an unusual length of the username field is unique to a single subscriber, an adversary might be able to uniquely attribute it to that subscriber.

If there is a group of subscribers with unusual lengths of username fields in their SUPIs, the attacker might be able to infer the membership of those subscribers in such a group.

When using EAP methods for authentication, it is not sufficient to protect the variable-length SUPI in NAI format against the above threat only in NAS messages (e.g., protecting the SUPI when sending it in a Registration Request). When such variable length SUPIs (i.e., username) are also used in EAP authentication methods (irrespective of whether the EAP authentication method is privacy preserving or not), an attacker may be able to identify and track subscribers at the EAP layer even if the user identifier is protected in the NAS layer against the above attack. This is because an attacker may be able to perform the above attack by identifying the confidentiality protected NAI within the EAP message that is sent over the air and then inferring the length of the NAI even if it is ciphered. The attacker can perform the same attack actively by sending an EAP Identity request to the UE. For example, in case of EAP-TLS or EAP-TTLS, even if the identifier is sent after TLS ciphering is turned on, the attacker may be able to infer the length of the EAP identifier of the UE by locating the ciphertext associated with the identifier. This attack is possible since TLS (both TLS 1.2 and TLS 1.3) leaves any padding to the application. Moreover, the EAP-TLS RFC does not specify any such padding (RFC 9190 recommends use of padding only for TLS record packets to hide the length of client certificates, c.f., section 5.8 of RFC 9190).

NOTE: The above threat of using EAP layer to infer the length of NAI is not applicable for 5G EAP-AKA’ specified in TS 33.501. In 5G EAP-AKA’, the UE always sends the same SUCI in the EAP layer. 5.1.3 Potential security requirements

The 5G system should protect against anonymity set reduction based on identifier length.

Note: the following conditions are necessary for proper evaluation of a solution

* the solution needs to indicate which authentication mechanisms it works with and whether that authentication mechanism preserves SUPI length.
* the solution needs to be evaluated as to whether it is backwards compatible with SUPIs in NAI format, which might already be deployed.

## 5.2 Key Issue #2: Users Identified by Priority Access

5.2.1 Key Issue Details

During connection establishment, a UE selects an RRC establishment cause value according to its access identity and access category based on the rules specified in table 4.5.6.1 and table 4.5.6.2 in TS 24.501[2]. The establishment cause value is sent in the clear over-the-air in RRC Setup Request messages. Ues assigned access identities 11-15, will send establishment cause “highPriorityAccess”, which affords them admission benefits when accessing the network. NR also supports two new establishment causes, “mps-PriorityAccess” and “mcs-PriorityAccess”, which indicate that Ues assigned access identity 1 and 2 are permitted to use multimedia priority services and mission critical services, respectively. The priority access cause values are different and can be distinguished from the values used by ordinary Ues assigned access identity of 0. Ues with access identity 0 use establishment causes which include: “mt-Access”, “emergency”, “mo-Signalling”, “mo-SMS”, “mo-VoiceCall”, etc.

Similarly, when a UE resumes a suspended connection it sends an RRC resume cause in the RRC Resume Request message. The options for the resume cause values are the same as for the establishment cause values. The resume cause is also sent in the clear over-the-air.

The establishment cause can also be linked to other identifiers that appear during an RRC Connection. For example, the TMSI is sent in the same RRC Setup Request message as the establishment cause. This allows the attacker to associate the establishment cause to the TMSI. Additionally, there is an exploitable linkage between the establishment cause and the C-RNTI because after the C-RNTI is sent in the RAR, it is present in the MAC layer of the RRC Setup Request, which also contains the establishment cause IE. Using an uplink sniffer, an attacker can link the establishment cause to the C-RNTI until the UE releases its connection. The attacker can only track the C-RNTIs associated with the Pcells. The C-RNTIs for Scells are not sent in the clear.

As a result, priority users are easily distinguishable from other subscriber groups and can be tracked based on the RRC establishment cause. The exposed establishment cause and resume cause reveal private user information and introduce privacy threats. This information leakage makes it possible to infer the group membership of priority users, the general location of priority users (e.g., localize users to specific cells), the number of priority users (e.g., as distinguished by different TMSIs), and the type of priority users (e.g., as distinguished by different priority establishment/resume causes).

Priority access Ues can be tracked within and across cells using the establishment cause coupled with the C-RNTI. Additionally, RRC Connections can be linked together until the TMSI is reassigned as there is no relationship between a TMSI allocation timespan and an RRC Connection. For example, it is left to implementation to re-assign 5G-GUTI after a Service Request message from the UE not triggered by the network. Inevitably, the TMSI and C-RNTI will change, but if the establishment cause remains the same, it can be determined that the UE is one with high

priority. This is valid whether a UE stays within the same cell or moves across cells because the UE will likely complete the RRC connection setup procedure often, exposing the establishment cause, TMSI, and C-RNTI each time.

The threat varies depending on the number of priority users in the area tracked by an attacker. If there are a few priority users, it may be possible to track them individually across various connections using some assumptions (e.g., no new priority users are attaching, the same users are re-establishing connections, etc.). In a situation where there are many priority users, it may be difficult to single out and track a specific user, but the ability to track a group of priority users as they move through the network is a privacy threat, in and of itself.

In addition, the detection of priority access users may be a prelude to another (e.g., kinetic) attack on priority access users. In that case, the privacy attack allows inference of the group membership and is independent to the number of priority users.

5.2.2 Security Threats

UEs using priority access can be distinguished from other subscriber groups based on the RRC establishment cause. The establishment cause can also be linked to C-RNTI and TMSI identifiers that appear during an RRC Connection. UEs using priority access can be tracked until its RRC connection is released or until it is assigned a new or additional C-RNTI. RRC Connections may be linked together until the TMSI is reassigned as there is no relationship between a TMSI allocation timespan and an RRC Connection. In a situation where there are many priority users, it may be difficult to single out and track a specific user, but the ability to identify a group of UEs using priority access as they move through the network poses a privacy threat.

Editor’s Note: The validity of the threat depends on how often or when do high priority UEs use the "highPriorityAccess" establishment cause.

5.2.3 Potential Security Requirements

The 5GS should provide means to mitigate the privacy risk of UEs with high priority access.

## 5.X Key issue #X:

### 5.X.1 Key issue details

### 5.X.2 Threats

### 5.X.3 Potential security requirements

# 6 Solutions

## 6.1 Solution #1: Use of fixed length identifiers to protect against anonymity set reduction

### 6.1.1 Introduction

The solution addresses KI#1.

Based on 23.003 Clause 2.2A, a SUPI type can be network specific identifier (NSI), and in such case the NSI takes the form of a Network Access Identifier (NAI) (i.e., username@realm) as defined in TS 23.003 Clause 28.7.2.

If the identifier in the username is variable length, then extremem differences in the length of the identify may give way for the threats discussed in the Key issue#1, therefore this solution proposes to configure and use an additional identifier with fixed length (for the NSI based SUPIs) to be used for the SUCI generation and related use for the network access.

### 6.1.2 Solution details

The UE can be configured by the operator with an additional fixed length identifier (i.e., a digital identifier) for the NAI SUPI (e.g., in the existing system, the Operator need to configure the UE with routing ID, and other information related to SUCI generation. So, similar methods can be reused for the digital identifier configuration). The fixed length digital identifier generation/assignment is upto the Operators implementation.

For NAI based SUPI, if the UE is provisioned with a fixed length digital identifier, then the UE can use the fixed length digital identifier as the username part of NAI for SUCI generation. The SUCI construction related to scheme Output can be same as described in TS 33.501, but the SUPI type should be set as digital identifier based NSI type.

The Home network on receiving the SUCI with SUPI type indicating ‘digital identifier based NSI type’, deconceals the SUCI as in TS 33.501, fetches the SUPI (i.e., NAI SUPI) related to a fixed length digital identifier and continues with the existing authentication procedure defined in 33.501.

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

### 6.1.3 Evaluation

Editor’s Note: Potential complexity in the CN (UDM) is FFS.

## 6.2 Solution # 2: Padding SUPIs in NAI format with Random Length of Characters for non-null schemes

### 6.2.1 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.

### 6.2.2 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 6.2.2-1: Authentication initiation using SUCI in NAI format with random padding

1. Using pre-configured padding parameters stored in USIM, 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 [5] and RFC 3269 [4] (i.e., not UTF-8 (see RFC 3629 [4]) 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.

The padding method indication may be included (e.g., appended to) in the cleartext ECIES input, resulting in confidentiality and integrity-protected padding method indication. This allows for the ECIES output with padding to be indistiguishable from the ECIES output without padding for an eavesdropper. Note that there is no impact on the ECIES functionality.

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

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]. If the the padding method indication is included in the cleartext ECIES input (see step 1) the result of the decryption will have padding method indication (e.g., appended to) the deconcealed padded SUPI.

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 [5] and RFC 3629 [4] (i.e., not a UTF-8 character set) to obtain the actual username part of the SUPI.

The USIM may be pre-configured by the operator with the supported padding method to be used. USIM 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. lmin/lmax values are used such as to ensure that resulting cleartext length is according to a normalized range across SUPIs after padding.

Editor's Note: How and how much privacy is achieved through random padding in the context of an IMSI catcher is FFS.Editor's Note: This solution may need to be updated to align with the KI once the ENs in the KI are resolved.

### 6.2.3 Evaluation

FFS.

## 6.3 Solution #3: Pseudonym based solution for k-anonymity of SUPI/SUCI

### 6.3.1 Introduction

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

Editor's Note: The k-anonymity analysis in the context of an IMSI catcher is FFS.

The solution addresses key issue 1. It is based on the use of pre-provisioned pseudonyms that when chosen carefully can guarantee k-anonymity (for a given k) for the SUPI/SUCI.

### 6.3.2 Solution details

#### 6.3.2.1 General

It is assumed that the UE can be pre-provisioned with a pseudonym for the SUPI. The pseudonym is allocated and managed by the operator. It is stored alongside the SUPI. The pseudonym is chosen to be unique to avoid collision with other pseudonyms or SUPIs. Clause 6.A.2.2 describes how such pseudonyms are used. Clause 6.A.2.3 describes how they can be allocated in order to guarantee a desired k-anonymity level for any given k.

The UE uses the pseudonym only if present and only instead of the SUPI when calculating a SUCI with a non-null encryption scheme. To signal the use of pseudonyms, the solution relies on the introduction of new protection scheme identifies. For example: 0x3 for Profile <C> where Profile C is identical to Profile <A> except that the pseudonym is used instead of the SUPI.

#### 6.3.2.2 Procedure

It is assumed that the UE can be preconfigured with a pseudonym and that the SIDF is preconfigured with a map from pseudonyms to SUPIs.

1. If the UE is preconfigured with a pseudonym and the UE is required to calculate a SUCI with other than the null encryption scheme, for example for an initial registration procedure, then the UE calculates the SUCI using the pseudonym and includes the corresponding new scheme identifier to indicate that SUCI was calculated using a pseudonym.
2. If the SIDF receives a SUCI including a scheme identifier signalling the use of a pseudonym, then after decryption of the SUCI, the SIDF uses the preconfigured map to recover the corresponding SUPI. If the included scheme identifier does not signal the use of pseudonym then the SIDF obtains the SUPI directly after decryption. In both cases, normal network operations can continue using the SUPI.

The need of a preconfigured map on the network side depends on how the pseudonyms are generated. For methods that require keeping an association such as hashing, random generation, etc, then such a map is needed. For other methods such as padding, the use of special delimiters or padding characters would suffice, in which case a preconfigured map is not needed and the SIDF can simply recover the SUPI from the decrypted pseudonym by stripping the padding characters.

#### 6.3.2.3 Guidance on pseudonym allocation

Assume a bell-like shaped distribution of the SUPIs in function of the length as shown in Figure 6.3.2.3-1 below. A fixed k value (for a desired k-anonymity level) gives two length limits shown as lmin and lmax. All subscribers whose SUPI's length is less than lmin or greater than lmax are allocated pseudonyms.

Diagram

Description automatically generated with medium confidence

**Figure 6.3.2.3-1 Example of SUPI distribution**

One straightforward way to guarantee k-anonymity is that SUPIs that are shorter than lmin or longer than lmax all allocated pseudonyms of length between lmin and lmax. Observe that this is sufficient but not necessary because for example, if the total number of subscribers with short SUPIs (less than lmin) is greater than k, then it is enough if they are allocated pseudonyms of the same length, irrespective of lmin. That group will be of size greater than k and hence k-anonymity is realized. The same reasoning applies to longer SUPIs (longer than lmax).

Observe also that the pseudonym value is irrelevant for anonymity. Only the length is decisive. The only requirement on the value is that it is unique to avoid collisions and to enable efficient recovery of the original SUPI on the network side. For the pseudonym value itself, there are many ways it can be generated: padded SUPI, truncated SUPI, hash of SUPI, random fixed length value, fixed length counter, etc. This could be left to implementation as well.

### 6.3.3 Evaluation

Editor's Note: Impact on the UE, the SIDF, and the UDM is FFS.

## 6.4 Solution #4: Limited length of SUPIs in NAI format

### 6.4.1 Introduction

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

Editor's Note: The k-anonymity analysis in the context of an IMSI catcher is FFS.

The solution addresses key issue 1. It is based on the control of the length limit of SUPIs in NAI format that can provide k-anonymity (for a given k) for the SUPI/SUCI if chosen carefully.

### 6.4.2 Solution details

Assume a typical distribution of the SUPIs in function of the length as shown in Figure 6.x.2 below. A fixed k value (for a desired k-anonymity level) gives two length limits shown as lmin and lmax. The middle parts of distributions between lmin and lmax typically have much higher frequencies and no privacy disclosure issue in length.

Diagram

Description automatically generated with medium confidence

**Figure 6.4.2 Example of SUPI distribution**

One straightforward way to guarantee k-anonymity is to limit the length of SUPIs in NAI type between lmin and lmax, which can be allocated and managed by the operators based on the subscribers distribution in a specific realm, i.e. 6-16 characters limit for username, to ensure the SUPIs can not only be chosen to be unique so that they do not collide with each other, and also the desired k-anonymity level for any given k can be guaranteed.

Note: How to analyze and choose lmin and lmax could be left to implementation.

### 6.4.3 Evaluation

Editor's note: evaluation is ffs

6.1.5 Solution #5: Solution for Privacy aspects of variable length user identifiers

6.1.5.1 Introduction

According to clause 2.2A of TS 23.003[2], the 5G standard allows the use of Network Specific Identifiers (NSI) as SUPI. An NSI will take the form of a Network Access Identifier (NAI) as defined in clause 28.7.2 of TS 23.003 [2]. The NAI for SUPI can have the form username@realm. Username in NAI format is encrypted during SUCI generation for privacy reasons. Usually the username part of NAI is created based on real-world names. Hence any encoding of the realworld names can lead to predictable outcomes which could also be guessed. This may lead to same privacy issues.

Key Issue #1 identified in [1] describes the privacy concern due to variable length SUPIs in NAI format.

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

6.1.5.2 Solution details

Text

Description automatically generated with medium confidence

Figure 6.1.5.2-1: Message flow detailing the solution

Figure 6.1.5.2-1 illustrates the system level message sequence detailing this solution. The steps are described as follows:

1. NAI configuration is performed by HN or Operator in USIM.
2. Extension/padding of SUPI NAI is configured by the operator.

Note 1: Operators may have their own specific extension length (fixed or variable according to the USIM).

As part of this solution, an extension/padding after a configurable delimiter, for example, “!”, in username can be used in SUPI\_NAI. This padding can ensure that the length of each username for a specific Operator adds up to a fixed number of octets. This also ensures that for any given length of SUPI\_NAI, the input to SUCI generation is always having a fixed length. Also, the delimiter can be used to extract the actual SUPI\_NAI after de-concealing the username from SUCI at the home network.

Operator can ensure that the choice of delimiter and maximum SUPI\_NIA length configurations are also made in UDM.

1. SUCI NAI is generated as described in Annex C.3.2 of TS 33.501.
2. After ME requests for SUPI\_NAI request, it reads the EF file of SUPI\_NAI which has anonymity configured username from USIM as described in Step 4.
3. ME requests for SUCI\_NAI and USIM shares the generated SUCI.
4. ME sends the SUCI\_NAI to HN.
5. After de-concealment of SUCI\_NAI, UDM will retrieve the SUPI\_NAI as “username!any\_non\_null\_string@realm”. UDM ignores the content after “!” (configurable delimiter) and considers only the username part in both fixed or variable NAI cases.

Note 2: Each user within same operator can have pre-defined or configured maximum length of username part. This will make a uniformity between different users of same operator. Operators can configure different delimiters and maximum lengths.

Editor’s Note: Provisioning of fixed or random padding method and parameters is FFS.

6.1.5.3 Evaluation

TBD

## 6.6 Solution #6: Padding SUPI in NAI format to conceal the username length

### 6.6.1 Introduction

This solution addresses the key issue #1: padding SUPI in NAI format to conceal the username length .

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

### 6.6.2 Solution details

To conceal the username length leaked by SUCI and make it harder for an attacker to distinguish SUCIs based on their lengths, it is proposed to pad the plaintext before encryption with variable-length of padding octets behind or before the username.

There are a variety of padding schemes such as block-length, random length padding, etc. Details of the SUPI padding mechanism may depend on the network operator and other deployment preferences.

Editor’s note: Details about padding, padding parameters provisioning, or use are FFS.



Figure 6.6.2-1: authentication procedure when SUPI padding is used.

If UE and the network decide to use SUPI padding method to conceal the username length in NAT format. The original SUPI and plaintext are pre-configured in both USIM and UDM.

1. The UE sends the Registration Request message to the AMF/SEAF containing SUCI, and the SUCI includes SUPI Type, Home Network Identifier, Routing Indicator, Protection Scheme Identifier, Home Network Public Key Identifier and Scheme Output. The Cipher value text in Scheme Output of SUCI is the encryption of SUPI in NAI format and plaintext.

2. The SEAF invokes the Nausf\_UEAuthentication service by sending a Nausf\_UEAuthentication\_Authenticate Request message containing the SUCI to the AUSF.

3. The Nudm\_UEAuthentication\_Get Request containing SUCI is sent from AUSF to UDM.

4. Upon reception of the Nudm\_UEAuthentication\_Get Request, the UDM invokes SIDF (Subscriber Identity De-concealing Function) to de-conceal the SUCI to obtain (e.g. determine) the SUPI with plaintext. If the SUPI is found in the database of the UDM, the UDM can compare the plaintext to get the username of SUPI without padding..

5. If SUPI with plaintext are found in the database of the UDM, the UDM selects the authentication method according to the SUPI. Then, the UDM generates the authentication data including authentication vector and sends it to AUSF in the Nudm\_UEAuthentication\_Get Response message with "200 OK". If SUPI is not found in the database, the UDM returns "404 Not Found" with "USER\_NOT\_FOUND" in the Nudm\_UEAuthentication\_Get Response message.

6. Upon reception of "200 OK", , the AUSF sends "201 Created" to AMF/SEAF with UEAuthentictionCtx containing authentication vector in the Nausf\_UEAuthentication\_Authenticate Response message. Upon reception of "404 Not Found", the AUSF sends "404 Not Found" to AMF/SEAF with "USER\_NOT\_FOUND".

7. The AMF/SEAF sends RAND and AUTN to the UE in the Authentication Request message in the case of "201 Created". Otherwise, the AMF/SEAF sends the Registration Reject message with Cause#3 to the UE in the case of "404 Not Found".

### 6.6.3 Evaluation

Edtor’s Note: evaluation is FFS.

## 6.7 Solution #7: Concealing length of SUPIs in SUCIs by truncating the SUPIs

### 6.7.1 Introduction

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

This is a solution to KI #1, using truancation of SUPIs.

### 6.7.2 Solution details

Edtor’s Note: The exact way that this solution addresses requirements in KI#1 needs to be elaborated in detail.

#### 6.7.2.1 UE Side

UE shall truancate the username portion before encrypting it using ECIES.

Truancation of SUPIs in NAI format shall be performed by the same component, either USIM or ME that performs the calculation of SUCI in the following manner:

* Encrypt username portion using byte-encode, e.g. ASCII.
* Choose one kind of the bytes (e.g. 0-F in [hexadecimal](javascript:;)) by the random number generator.
* Delete the corresponding byte chosen before on the corresponding positions and record.
* Encrypt the truancated SUPI using ECIES.

NOTE 1: SUPI in IMSI format is not truncated because it is fixed length.

NOTE 2: The UE shall not truancate the SUPI in NAI format when using the null scheme.

#### 6.7.2.2 Home Network Side

The UDM invokes the SIDF to de-conceal the SUCI to the truancated SUPI. The UDM restores the original SUPI according to the record.

### 6.7.3 Evaluation

TBD

## 6.8 Solution #8: Use of fixed length “username” for NAI

### 6.8.1 Introduction

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

This solution addresses Key Issue #1.

A SUPI that is in NAI format (i.e., username@realm) may be used in some 5GS deployments (e.g., SNPN deployments or 5GS deployments that do not require interworking with EPS).

In such deployments, if the “username” part is of variable length, it may be subject to the threats described in Key Issue #1. Since the SUPI assignment is under the control of home network operator, this solution proposes that the home network operator assigns SUPIs such that the “username” part of the NAI is always of fixed length.

This solution is motivated by the fact that in 5GS, SUPIs do not need to be comprehensible or handled by human users. GPSIs (including MSISDN) are used for such purposes, which of course can be based on real world names of the actual subscribers.

### 6.8.2 Solution details

In this solution, the home network operator configures “username” for their SUPIs such that the “username” for all their subscribers is of fixed length. The SUPI with fixed length for “username” part of the NAI is configured on the UE as well as on the network (e.g., UDM) as part of the operators existing process for configuring subscription credentials.

Editor’s Note: Storage details for fixed length SUPI on the UE is FFS.

The fixed length that is chosen for the username can be left to the decision of the home network operator. However, the home network operator needs to select the fixed length such that it is greater than the maximum length of the username for all possible subscribers.

Note that this solution does not prevent the home operator from assigning “username” based on subscriber’s real world names when necessary. In such case, the home operator can use a padding scheme to generate a fixed length “username” and then configure the UE and the UDM with a fixed length SUPI. The padding scheme used by the home network operator can also be left to the decision of the home network operator.

### 6.8.3 Evaluation

TBD.

## 6.9 Solution #9: Concealing length of SUPIs in SUCIs by padding the SUPIs

### 6.9.1 Introduction

This is a solution to KI #1. It uses pre-encryption padding 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.

### 6.9.2 Solution details

#### 6.9.2.1 Solution Basics

For SUPIs taking the form of a NAI, the subscription identifier part of the SUPI includes the "username" portion of the NAI as defined in NAI RFC 7542 [5]. With an exception for the null scheme, the UE pads the username portion before encrypting it using ECIES.

NOTE 1: The null scheme does not provide SUPI privacy in the first place; therefore, padding will only increase bandwidth without improving any privacy.

NOTE 2: In the present document, SUPI in IMSI format is not padded because it is fixed length, and SUCI cannot be attributed to a particular SUPI based on length.

The necessary padding parameters, which are the home operator's choices, is stored in USIM.

#### 6.9.2.2 Padding parameters

Padding parameters comprise a list of pLen. Each pLen in the list indicates the number of octets in the padded username for a certain number of octets in the unpadded username.

The HN stores the whole pLen. The USIM stores only one element of the pLen that indicates the number of octets in the padded username for the unpadded username associated with the USIM.

Editor’s Note: The complexity of recalculating effective padding parameters when the length distribution of SUPIs changes is FFS.

#### 6.9.2.3 UE Side

Padding of SUPIs in NAI format is performed by the same component, either USIM or ME, that performs the calculation of SUCI in the following manner:

- If the number of octets in the unpadded username is indicated in the list of pLen, the username is padded to the corresponding pLen. In this case, the username is prepended with the necessary numbers of octet value 0x20.

- If the number of octets in the unpadded username is not indicated in the list of pLen, the username is not padded.

- The UE does not pad the SUPI in NAI format when using the null scheme.

#### 6.9.2.4 Home Network Side

With an exception for the null scheme, when the de-concealed SUPI is in NAI format and padded by the UE, the SIDF unpads the username portion. The SIDF removes 0x20 octet from the beginning of the username until a non-0x20 octet value is identified.

### 6.9.3 Evaluation

Editor's Note: The k-anonymity analysis in the context of an IMSI catcher is FFS.

## 6.10 Solution #10: Concealing length of SUPIs in SUCIs by hashing the SUPIs

### 6.10.1 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.

### 6.10.2 Solution details

#### 6.10.2.1 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.

#### 6.10.2.2 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 [6] with a dummy key, e.g., all zeros. The output of the KDF could be truncated to the desired length.

#### 6.10.2.3 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.

#### 6.10.2.4 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.

### 6.10.3 Evaluation

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

TBD

## 6.A Solution #A: <Solution Title>

### 6.A.1 Introduction

### 6.A.2 Solution details

### 6.A.3 Evaluation

# 7 Conclusions

Annex A:  
List of 3GPP identifiers.

The following table provides a non-exhaustive list of 3GPP identifiers and parameters transmitted over the air. These identities are provided for information only (e.g., inclusion neither suggests that the identity is in the scope of study nor that there is a privacy issue with that identity).

|  |  |  |  |
| --- | --- | --- | --- |
| No | Name of 3GPP Identifier | Description | Specified in 3GPP document |
| 1 | SUCI | SUbscription Concealed Identifier | TS 23.003 [aa], TS 23.501 [bb] |
| 2 | S-NSSAI | Single Network Slice Selection Assistance Information | TS 23.003 [aa], TS 23.501 [bb] |
| 3 | 5G-GUTI | 5G Globally Unique Temporary Identifier  5G-GUTI provides an unambiguous identification of the UE that does not reveal the UE or the user's permanent identity.  5G-GUTI has two main components:  - one that identifies the AMF(s) which allocated the 5G-GUTI; and  - one that uniquely identifies the UE within the AMF(s). | TS 23.003 [aa] |
| 4 | CAG Identifier | A Closed Access Group (CAG) within a PLMN is uniquely identified by a CAG-Identifier | TS 23.003 [aa], TS 23.501 [bb] |
| 5 | C-RNTI | Cell Radio Network Temporary Identifier  C-RNTI is a unique identifier dedicated to a particular UE and used for identifying RRC Connection and scheduling. C-RNTI can be reallocated when a UE accesses a new cell with the cell update procedure. | TS 38.300 [cc], TS 38.321 [dd] |
| 6 | Establishment Cause | RRC establishment cause value maps to an access identity. This value is sent in RRC Setup Request messages when establishing a connection. | TS 24.501 [ee] |
| 7 | Resume Cause | RRC resume cause value maps to an access identity. This value is sent in RRC Resume Request messages when resuming a suspended connection. | TS 38.331 [ff] |
|  |  |  |  |
|  |  |  |  |

Annex <X> :  
Change history

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Change history** | | | | | | | |
| **Date** | **Meeting** | **TDoc** | **CR** | **Rev** | **Cat** | **Subject/Comment** | **New version** |
| 2022-02 | SA3#106-e | S3-220514 |  |  |  | Skeleton | 0.0.1 |
| 2022-02 | SA3#106-e | S3-220515 |  |  |  | Scope | 0.0.1 |
| 2022-02 | SA3#106-e | S3-220516 |  |  |  | Annex A | 0.0.1 |
| 2022-05 | SA3#107-e | S3-221180 |  |  |  | Key Issue #1: Privacy aspects of variable length user identifiers | 0.2.0 |
| 2022-07 | SA3#107-e Ad Hoc | S3-221642 |  |  |  | New key issue on users identified by Priority Access | 0.3.0 |
| 2022-10 | SA3#108-e Ad Hoc | S3-222991 |  |  |  | Updates to Key Issue #2 | 0.4.0 |
| 2022-10 | SA3#108-e Ad Hoc | S3-223044 |  |  |  | PCR for KI #1: Privacy aspects of variable length user identifiers | 0.4.0 |
| 2022-10 | SA3#108-e Ad Hoc | S3-223005 |  |  |  | Solution to address KI#1 | 0.4.0 |
| 2022-10 | SA3#108-e Ad Hoc | S3-223045 |  |  |  | New solution for Key issue #1 | 0.4.0 |
| 2022-10 | SA3#108-e Ad Hoc | S3-223103 |  |  |  | New solution for key issue 1 | 0.4.0 |
| 2022-10 | SA3#108-e Ad Hoc | S3-223017 |  |  |  | New solution for key issue 1 | 0.4.0 |
| 2022-10 | SA3#108-e Ad Hoc | S3-223124 |  |  |  | New solution for privacy prevention of SUPI in NAI format | 0.4.0 |
| 2022-10 | SA3#108-e Ad Hoc | S3-223085 |  |  |  | SUPI padding solution on Key issue #1 | 0.4.0 |
| 2022-10 | SA3#108-e Ad Hoc | S3-223014 |  |  |  | New solution on Key issue #1 | 0.4.0 |
| 2022-10 | SA3#108-e Ad Hoc | S3-223011 |  |  |  | Solution for KI#1 | 0.4.0 |
| 2022-10 | SA3#108-e Ad Hoc | S3-223065 |  |  |  | Padding-based solution to the leakage of the length of SUPI through SUCI | 0.4.0 |
| 2022-10 | SA3#108-e Ad Hoc | S3-223066 |  |  |  | Hash-based solution to the leakage of the length of SUPI through SUCI | 0.4.0 |