ETSI TR 103 304 V1.1.1 (2016-07)

TECHNICAL REPORT

CYBER; Personally Identifiable Information (PII) Protection in mobile and cloud services

Reference

DTR/CYBER-0002

Keywords

access control, privacy

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Copyright Notification

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Cyber Security (CYBER).

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Executive summary

ICT is moving towards a genuinely distributed and virtualized environment characterized by a rich set of mobile and cloud services available to users. In this context, it may be difficult to have a priori knowledge of who may need access to data, when and where this may happen and whether that data could be or contain Personally Identifiable Information (PII). The present document proposes a number of scenarios focusing on today's ICT and develops an analysis of possible threats related to PII in mobile and cloud based services. It also presents technical challenges and needs derived from regulatory aspects (lawful interceptions). The aim is to consolidate a general framework, in line with regulation and international standards, on top of which technical solutions for PII protection can be developed.

1 Scope

The present document proposes a number of scenarios focusing on today's ICT and develops an analysis of possible threats to Personally Identifiable Information (PII) in mobile and cloud based services. It also presents technical challenges and needs derived from regulatory aspects (lawful interceptions). It consolidates a general framework, in line with regulation and international standards, where technical solutions for PII protection can be plugged into.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ISO/IEC 29100:2011: "Information technology - Security techniques - Privacy framework".

[i.2] National Institute of Standards and Technology NIST SP 800-122: "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)".

[i.3] Regulation 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC.

[i.4] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[i.5] Directive 2002/21/EC of the European Parliament and of the council of 7 March 2002 on a common regulatory framework for electronic communications networks and services (Framework Directive).

[i.6] Directive 2002/22/EC of the European Parliament and of the Council of 7 March 2002 on Universal service and users' rights relating to electronic communications networks and services (Universal Service Directive - OJ L 108, 24.04.2002).

[i.7] Directive 1999/5/EC of the European Parliament and of the Council of 9 March 1999 on radio equipment and telecommunications terminal equipment and the mutual recognition of their conformity.

[i.8] Directive 1995/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

[i.9] US President's Council of Advisors on Science and Technology: "Report to the president. Big data and privacy: a technological perspective".

[i.10] ETSI TR 101 567: "Lawful Interception (LI); Cloud/Virtual Services for Lawful Interception (LI) and Retained Data (RD)".

[i.11] ETSI Cloud Standards Coordination: Final Report.

[i.12] ISO/IEC 11889:2009: "Information technology - Trusted Platform Module" (Parts 1-4).

[i.13] ISO/IEC 29191:2012: "Requirements for partially anonymous, partially unlinkable authentication".

[i.14] ISO/IEC 29115:2011: "Entity authentication assurance framework".

[i.15] ETSI TS 119 612: "Electronic Signatures and Infrastructures (ESI); Trusted Lists".

[i.16] ETSI TR 103 308: "CYBER; Security baseline regarding LI and RD for NFV and related platforms".

[i.17] ETSI TR 187 010: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Security; Report on issues related to security in identity imanagement and their resolution in the NGN".

[i.18] ISO/IEC 27040:2015: "Information technology - Security techniques - Storage security".

[i.19] ISO/IEC 17789:2014: "Information technology - Cloud computing - Reference architecture".

[i.20] ISO/IEC 9594-8:2014: "Information technology - Open Systems Interconnection - The Directory - Part 8: Public-key and attribute certificate frameworks".

[i.21] ETSI TS 101 331: "Lawful Interception (LI); Requirements of Law Enforcement Agencies".

[i.22] ETSI TS 101 671: "Lawful Interception (LI); Handover interface for the lawful interception of telecommunications traffic".

[i.23] ISO/IEC JTC 1/SC 38 CD 19944: "Information technology - Cloud computing - Data and their flow across devices and cloud services".

NOTE: Standard under development.

[i.24] ISO/IEC JTC 1/SC 37 AWI 20889: "Information technology - Security techniques - Privacy enhancing data de-identification techniques".

NOTE: Standard under development.

[i.25] J.A. Akinyele, C. U. Lehmanny et Al. Self-Protecting Electronic Medical Records: Using Attribute-Based Encryption. Cryptology ePrint Archive, Report 2010/565. 2010.

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

app: "software application", typically running on a user's device platform

anonymization: process that replaces an actual identifier with an attribute obtained by randomization or generalization in such a way that there is a reasonable level of confidence that no individual can be identified

Cloud Service Customer: individual or organization consuming one or more cloud services provided by a Cloud Service Provider

Cloud Service Partner: individual or organization providing support to the provisioning of cloud services by the Cloud Service Provider, or to the consumption of cloud service by the Cloud Service Customer

Cloud Service Provider: individual or organization providing cloud services to one or more Cloud Service Customers

Cloud Service user: individual consuming one or more cloud services using a particular device

consent: freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed

data breach: compromise of security that leads to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to protected data transmitted, stored or otherwise processed [i.18]

data consumer: entity accessing data for a given purpose

data fusion: process of combining multiple data sets into one improved data set in order to discover any information which cannot be derived from the original data sources

data subject: identifiable person, i.e. a person who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity

de-anonymization: any process in which anonymous data is cross-referenced with other sources of data to re-identify the anonymous data source

Device Platform Provider: Cloud Service Provider providing services necessary to support the device platform

generalization: process that reduces the degree of granularity (known as precision) of a set of attributes

identity theft: inappropriate use of someone else's credentials to commit fraud or crimes

lock-in: process which makes a customer dependent on a given service provider and unable to use another provider without substantial switching costs

metadata: data about the data, which can be structural or descriptive

mis-contextualization: process in which data from different personas is mixed and used inappropriately

over-collection: practice of collecting information unrelated to a stated purpose

persona: role played by an individual user in the context of a service

Personally Identifiable Information (PII): any information that (a) can be used to identify the PII principal to whom such information relates, or (b) is or might be directly or indirectly linked to a PII principal

NOTE 1: To determine whether a PII principal is identifiable, account can be taken of all the means which can reasonably be used by the privacy stakeholder holding the data, or by any other party, to identify that natural person [i.1].

NOTE 2: In the US, according to [i.2]: any information about an individual maintained by an agency, including any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.

PII controller: privacy stakeholder that determines the purposes and means for processing personally identifiable information (PII) other than natural persons who use data for personal purposes [i.1]

PII principal: natural person to whom the personally identifiable information (PII) relates [i.1]

PII processor: privacy stakeholder that processes personally identifiable information (PII) on behalf of and in accordance with the instructions of a PII controller [i.1]

portability: usability of the same software, data or metadata in different environments

processing of PII: operation or set of operations performed upon personally identifiable information (PII) [i.1]

NOTE: Examples of processing operations of PII include, but are not limited to, the collection, storage, alteration, retrieval, consultation, disclosure, anonymization, pseudonymization, dissemination or otherwise making available, deletion or destruction of PII) [i.1].

pseudonymization: process that replaces an actual identifier with an alias ensuring that it cannot be reverted by reasonable effort of anyone (other than the party providing them)

randomization: process that reduces the degree to which data reflects the true value of a set of attributes (known as accuracy)

ransomware: type of malware that restricts access to the infected device, demanding that the user pay a ransom to the malware operators to remove the restriction

re-identification: action performed on de-identified data with the purpose of re-linking the information to a person or group of persons

secure data deletion: irreversible destruction of electronic data so that no party is capable of recovering

spyware: type of malware that collects/intercepts/retrieves data from a (mobile) device and sends it to a remote (Command&Control) server

Terminal Equipment: product enabling communication or relevant component thereof which is intended to be connected directly or indirectly by any means whatsoever to interfaces of public telecommunications networks

traceability: ability to interrelate individuals in a way that is verifiable

trust: level of confidence in the reliability and integrity of an entity to fulfil specific responsibilities

unlinkability: act of ensuring that a user may make multiple uses of resources or services without others being able to link these uses together

3.2 Abbreviations

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

5G-PPP	5G Infrastructure Public Private Partnership
ABE	Attribute-Based Encryption
API	Application Programming Interface
AWI	Approved Work Item

NOTE: http://www.iso.org/iso/home/faqs/faqs_abbreviations.htm.

BYOD	Bring Your Own Device
CA	Certification Authority
CD	Committee Draft
CEO	Chief Executive Officer
CP-ABE	Ciphertext Policy Attribute-Based Encryption
CPU	Central Processing Unit
CSC	Cloud Service Customer
CSP	Cloud Service Provider
CSPa	Cloud Service Partner
Csu	Cloud Service user
DPP	Device Platform Provider
EC	European Community
EU	European Union
GPS	Global Positioning System
GSM	Global System for Mobile
ICT	Information and Communication Technology
IMEI	International Mobile Station Equipment Identity
IMSI	International Mobile Subscriber Identity
ISO	International Organization for Standardization
JTC	Joint Technical Committee
LEA	Law Enforcement Authority
LI	Lawful Interception
PC	Personal Computer
PII	Personally Identifiable Information
PIN	Personal Identification Number
PKI	Public Key Infrastructure
PLMN	Public Land Mobile Network
PNR	Passenger Name Record

PUA	Potentially Unwanted Application
RAM	Random Access Memory
SAREF	Smart Appliances REFerence ontology
SC	Subcommittee
SMS	Short Message Service
TE	Terminal Equipment
TEE	Trusted Execution Environment
TPM	Trusted Platform Module
TS	Technical Specifications
UMTS	Universal Mobile Telecommunications System
US	United States

4 Overview

An even growing number of human activities are today performed using Internet-based (and particularly, cloud-based) services. Information that can be used to identify a natural person or might be directly or indirectly linked to her, known in literature as Personally Identifiable Information (PII) may be potentially present in almost all these activities. While technology is apparently "disappearing" to naive eyes, as people are focusing on services regardless of the devices, terminals or platforms they actually use, awareness of data transaction and transparency about its use is decreasing. This may cause social and legal concerns when data transactions may involve PII.

Code of practices and regulatory aspects protecting PII were present since the advent of mobile communications in middle 1990s. Directive 95/46/EC (data protection) [i.8], directive 2002/58/EC (privacy) [i.7], and Directive 99/5/EC (radio equipments) [i.5], [i.6], for instance, state the legal obligations to preserve a user's control of their identity in electronic communication, as well as obligations intended to avoid frauds. Properly using identifiers and identity management as suggested in previous ETSI TR 187 010 [i.17] massively reduces the risk to exploit of PII in traditional communication signalling.

However, today the ICT is moving towards a genuinely distributed and virtualized environment characterized by a rich set of mobile and cloud services available to users. The eIDAS Regulation [i.3] first and the EU General Data Protection Regulation [i.4] then have provided a legal framework to address challenges raising from the digital age and its "app economy", in order to booster citizen's trust in the emerging Digital Single Market.

In fact, differently from previous telecom scenario where user data was mostly accessible from network functional elements, several kinds of information are today easily accessible from terminal equipments or end user devices, through open and specialized Application Programming Interface (API). Thus, it may be difficult to have a priori knowledge of who may need access to users' data, when and where this may happen and whether that data could be or contain PII.

PII in long term data records (e.g. in health, public administration, education, financial and legal domains) are dynamic and grow over the life of an individual. The set of actors/individuals/roles that need to access and amend it over a lifetime is potentially unlimited. It is also not reasonable to expect the record to be "a single document" rather to likely appear as a large set of data, retained in data centres located in many different national Countries and managed by various stakeholders with different levels of trust. In such records there may be a need to enable security controls of some complexity.

The present document proposes a number of scenarios focusing on today's ICT and develops an analysis of possible threats related to PII in mobile and cloud based services.

5 Threats to PII

5.1 Overview

This clause presents threats derived from the analysis of the scenarios reported in Annex A. The scenarios are not exhaustive rather they are representative of most common and relevant situations.

Threats sources may include accidents, natural disasters, humans authorized or unauthorized to access data and systems. A synopsis relating threats with risks and vulnerabilities is provided in table 5.1.

5.2 Data fusion and re-identification

Data from different sources, which are typically designed for specific and limited purposes, may be merged and analyzed for secondary use either within an organization or outside of it through several different techniques. The US report on big data and privacy [i.9] presents some examples of threats derived from data fusions.

Concentration of large amount of data on few service providers may encourage data fusion, although proper access control techniques should be put in practice to discourage this practice when undesired.

Re-identication may be achieved through various techniques (ISO/IEC JTC 1/SC 37 AWI 20889 [i.24]). In the context of web applications, tracking cookies (particularly third parties tracking cookies) may facilitate this process. De-anonymization may be one relevant consequence.

Data fusion and re-identification normally occur with authorized access to data and might not be evident to the PII principal.

5.3 Data breaches

Especially in Cloud environments, many providers operate through an agreement with a partner playing the role of PII processor in order to provide services to users. For instance, a service provider might choose to rely on a storage provider as a partner. Data braches may thus occur to service providers or to partners of providers processing data.

This process might not be evident to the PII principal, which may trust the service provider but not necessarily the partner, e.g. due to its location under a different legislation or to the partner's infrastructure (e.g. because of unauthorized personnel accessing the hardware/software infrastructure, data processing through rogue devices).

5.4 Service termination/inaccessibility

Remote (and cloud) storage offers several advantages in terms of availability (data is no longer dependent on the device used to create it and manage it). However, cloud computing storage at large data centres may increase the chance of temporary unavailability - due to network connection issues, shortages or server failures - or even large losses of data.

Normally terms of service contain clauses which address the aforementioned problems, providing recovery solutions. Even in this case, however, services might be terminated at any time due to non technological problems (e.g. bankruptcy of the provider) or legal reasons (e.g. file-sharing services infringing copyright laws).

Location of the service provider may be a relevant aspect, as the service may be under different regulations and laws.

5.5 Lock-in mechanisms

Ideally data availability should be granted for the data owner whenever and wherever and be immediate.

However, lock-in mechanisms may practically lead to data unavailability. Lock-in prevents portability (and sometimes interoperability) of customer's data across different service providers. As a result, when a customer decides to change provider, data in his or her account may simply be unavailable or lost.

5.6 Ransomware and Spyware

Data availability and confidentiality may be threatened by malicious software such as ransomware and spyware.

Typically, the ransomware silently encrypts contents inside a terminal equipment, a device or on a remote storage. As a result, documents, images, and other kind of files - which may be or may contain PII - are no more available to users nor to their service providers unless the users proceed with the payment of the ransom.

Installing vetted software from a trusted source is a general measure to prevent malware. Additionally access control mechanisms enabling only authorized processes to access files may apply as a specific protection measure.

5.7 Over-collection

Over-collection may lead to unwanted disclosure of PII. Several examples and cases of over-collections are described in the US report on big data and privacy [i.9].

Over-collection, might be present by default when the information arises from the physical world and is captured by sensors, due to the difficulty to filter out signals not related to the scope of the program (so called "noise"). Instead it is always intentional for information "born digital", i.e. data and meta-data created specifically for use by a computer or a digital system [i.9].

Over-collection is not necessarily clandestine. Potentially Unwanted Applications (PUA) are common applications that collect user data, which may include PII, and send it to remote but not necessarily malicious servers (e.g. for targeted advertisement purposes). Applications belonging to a category commonly known as "people as sensors applications" may exploit over-collection to provide additional services to users.

5.8 Mis-contextualization

Mis-contextualization occurs when data from different users or from the same user but in different role (i.e. a different "persona") is mixed and used inappropriately.

Mis-contextualization might be an unintentional event due to, e.g. a missing reset of a terminal equipment (end user device) or to a missed account switch when the same service subscription is used by more than one user.

Despite potentially supporting multiple user accounts, some terminal equipments, devices, or even single applications running on them do not provide an easy way for users to switch between different accounts. The terminal, device or application may contain PII and pointers to subscribed services which might not be erased by the original subscriber. As a consequence, if the device is not properly reset when it is sold - or even stolen - data related to the original subscription may be accessed by users other than the subscriber or the legitimate user, leading to disclosure of PII, and even to an inappropriate use of the subscriber's credentials to commit fraud or crimes (identity theft).

5.9 User Impersonation

An attacker may discover user identities by snooping authentication traffic identity. In some situations in current mobile networks (e.g. GSM, UMTS and in all networks during an emergency call setup) the IMEI or the IMSI is sent to the network in plain text. This opens the door to identity disclosure.

Such information can be used later by malicious user to pretend to be the legitimate user, even in order to commit fraud or crimes (identity theft).

5.10 Alteration of ownership or access rights

A service provider generally owns account data concerning their subscribers. The business model of some services may require subscribers to consent to a number of uses of the data. Thus, the owner relationship or rights over it may be altered when data is released to the service. As a consequence data itself may be made public, be altered, or even be indexed in search engines. Regulation however may provide use limitations, acknowledging principals to effectively retain some rights over PII.

5.11 Alteration of persistence

To ensure business continuity and prevent data loss, generally data centres provide by default security policies such as backup and replication over different nodes. Compared to traditional "local" storage devices (e.g. a local hard disks), this feature may introduce an alteration of natural data persistence.

An irreversible destruction of electronic data so that no party is capable of recovering (so called "Secure" data deletion) might be difficult to ensure without the ability to resort to the destruction of the hardware.

NOTE 1: Alteration of persistence makes evident that data stored on a remote location may present additional risks compared to a local storage, although cryptographic techniques may provide means to enforce confidentiality. These risks might not be evident to the service subscriber.

NOTE 2: At the present time, no technical option seems suitable to ensure information extinction after its disclosure (a feature referred to as "the right to be forgotten" in some legal contexts). The safest assumption is that data, once released and disclosed, will be persistently present in the Cloud [i.9]. A weaker version of the "right to be forgotten", known as "right to erasure" has been recently introduced in EU regulation.

5.12 Synopsis

This clause provides a synopsis of threats found in the scenarios reported in Annex A. The analysis of these scenarios leads to the identification of three main categories of risks: data disclosure, data manipulation and data unavailability (corresponding to violations of the three classical security properties: confidentiality, integrity and availability). Furthermore, each scenario is mapped into one of the two general use cases (UC1, UC2) identified in clause 7 and addressing architectural aspects.

Table 5.1: threats, risks and vulnerabilities

Scenarios	Threats	Vulnerabilities	Security Properties	Risk Category	Use Case
Social Networking App, In-car blackbox, Self-quantifying	Over-collection	Permissions	Confidentiality	Disclosure	UC2
Installing untrusted apps	Ransomware	Permissions	Availability	Unavailability	UC2
BYOD, Social Networking App, Self-quantifying	Data sharing	Permissions, no way to extend access control outside TE	Confidentiality, Accountability, Authentication	Disclosure	UC2
Social Networking App, In-car blackbox, Self-quantifying	Data fusion related threats	Changed scope/use	Accountability, Confidentiality, Authentication (Unlinkability, untraceability)	Disclosure	UC2
Cloud Unavailability	Service termination or inaccessibility, lock-in mechanisms	Remote Storage	Availability	Unavailability	UC1
Social Networking App, Self-quantifying	Alteration of access rights, indexing in search engines	Loose ownership, Changed scope/use	Confidentiality, Integrity, Availability, Accountability (right to be forgotten)	Disclosure, Unavailability, Manipulation	UC2
Social Networking App, In-car blackbox	Alteration of persistence	Remote Storage, replication	Accountability, Confidentiality (right to be forgotten)	Disclosure	UC2
Medical Scenario, PNR	Data breaches	Untrusted infrastructure	Confidentiality, Integrity, Availability, Accountability (location control)	Disclosure, Unavailability, Manipulation	UC1
BYOD, In-car blackbox	Mis-contextualization, identity theft	Missing device reset or account switch	Authentication, Confidentiality, Integrity, Availability, Accountability	Disclosure, Unavailability, Manipulation	UC2
Fake or untrusted access mobile networks	User Impersonation	Untrusted infrastructure	Confidentiality, Integrity	Disclosure, Manipulation	UC2

6 Technical aspects

6.1 Principles from ISO/IEC 29100

In previous clauses a number of criticalities in PII processing have been identified. The present clause first provides a summary of general protection principles that have been defined in ISO/IEC 29100 [i.1]; then, it describes a number of related technical aspects in designing services processing PII.

ISO/IEC 29100 [i.1] recommends service providers to apply the following principles without or with very limited exceptions, despite jurisdictional differences in national countries. What follows is a summary of the principles; ISO/IEC 29100 [i.1] contains full details.

- (Informed) consent and choice whenever applicable, including the ability for the PII principal to withdraw their consent and special provisions for individuals not legally able to express their consent and procedures by government authorities.

NOTE: Even if consent is withdrawn, the PII controller may need to retain data for a given period for legal or contractual obligation.

- Purpose legitimacy and specification, including special provisions for sensitive PII which may be subject to specific legal constraints.
- Principle of collection limitation: organizations should not collect PII indiscriminately, and should inform PII principals when collection of additional information (not specifically related to the provision of the main service) is optional.
- Data minimization, including removal of unnecessary processing and unnecessary access to PII and deletion of PII when no longer strictly needed for the provision of a service (or any legal obligation thereof).
- Limitation of use, retention and disclosure.
- Periodic verification of accuracy of the information and quality of the processing, especially when inaccurately collected or processed data could result in harm to the PII principal.
- Principle of openness and transparency, including notices on the options available to the PII principal for accessing, correcting and removing or limiting processing of information; and notices on what PII is being requested, purpose of processing, details of processing including types of authorized persons able to access the records, mechanisms of collection, storage, communication, retention and disposal procedures.
- Individual participation and access: ability of the PII principal to access, review and provide any correction to their PII or request of removal (subject to applicable law for specific cases).
- Principle of accountability of processing and measures taken for protecting PII from privacy breaches and compensation in case of identity theft, reputation damages, PII misuse or accidental mistakes in the processing of PII.
- Implement information security to protect confidentiality, integrity and availability of PII from risks such as unauthorized access, destruction, use, modification, disclosure or loss throughout the whole data lifecycle.
- Compliance with privacy law.

6.2 Degree of link-ability

By definition, PII is any information that can be used to identify or might be directly or indirectly linked to an individual (PII principal). Data may be qualified according to the degree to which data can be linked to an individual. ISO/IEC JTC 1/SC 38 CD 19944 [i.23] reports the following classification:

- identified data provides identity of the individual;

- pseudonymized data contains aliases (aka pseudonyms) on behalf of the actual identifiers. Aliases can be obtained by encryption or hashing and cannot be reverted by reasonable effort of anyone other than the party providing them. The term "securely pseudononymed data" is used when aliases cannot be reverted by reasonable effort of anyone, including the party providing the aliases;
- anonymized data contains attributes on behalf of the actual identifiers which are randomized or generalized in such a way that there is a reasonable level of confidence that no individual can be identified. The randomization process reduces the degree to which the aliases reflect the true value of attributes (accuracy). The generalization process reduces the degree of granularity (precision) of the attributes;
- aggregated data does not contain individual level entries.

Services requiring authentication may link user's data transactions to the user's identity in the context of that service, or of those services provided by the same service provider or even of those provided by the provider's partners and/or third parties. ISO/IEC 29191 [i.13] provides requirements for partially anonymous, partially unlinkable authentication. From the provider's perspective there may be different levels of assurance that the entity claiming a particular identity is in fact the entity to which that identity was assigned. The choice of which level is appropriate for a given service depends on a variety of factors. ISO/IEC 29115 [i.14] analyses this aspect in details.

6.3 Trust

Trust is the level of confidence in the reliability and integrity of an entity to fulfil specific responsibilities. The identification of involved entities or at least the verification of their attributes is a prerequisite to achieve trust.

A Trusted Execution Environment (TEE) may be present in network entities (ETSI TR 103 308 [i.16]), terminal equipments and devices to provide a secure hardware and software area where integrity of trusted applications may be executed and confidentiality of data there processed may be ensured. A TEE offers an execution space that provides a higher level of security than the rest of a platform or infrastructure, but it may be expensive and not viable for every applications. Current efforts toward the definition of a standardized TEE are intended to deliver a standard interface in form of Application Programming Interfaces (APIs) which client applications may use to implement trusted transactions.

A current standard implementation of a TEE by ISO/IEC is the Trusted Platform Module (TPM) defined in ISO/IEC 11889 [i.12].

There may be various levels of trust that an entity has for another. Trust relationship is rarely symmetric and may be relative, not absolute, as an entity may trust a second one more than a third one, without trusting it absolutely.

In traditional Public Key Infrastructure (PKI) digital certificates and signatures (ISO/IEC 9594-8 [i.20]) are used to implement relational trust. Certificates are verified using a chain of trust. Each Certification Authority (CA) issues a digital certificate binding a public key to a particular distinguished name. The Root Certification Authority represents the topmost authority (the anchor) along the path on the chain and its root certificate, being not verifiable by any higher authority, is typically distributed to the clients by physical delivery.

Trust may vary over the time and the level of dynamism may also vary between different relationships. Typically certificates expire and Certificate Revocation Lists (CRLs) are a way to achieve revocation for non expired certificates. Other methods are based on online verification.

Trusted electronic transactions rely on trust services which aim primarily at issuing digital certificates and supporting the validation of electronic signatures. ETSI TS 119 612 [i.15] provides technical details on how to implement trust service directories compliant with eIDAS regulation [i.3].

6.4 Awareness of data transaction

Transparency about which data is being collected and for what purpose (purpose legitimacy and specification according to ISO/IEC 29100 [i.1]) and what actions are possible on data after it has been given (individual participation in ISO/IEC 29100 [i.1]) is desired by users and regulators but may be difficult to achieve at any time and place and through any service/device.

Some modern services on novel kind of devices (e.g. wearable or IoT devices) might not make users fully aware of data transactions, especially when performing actions apparently familiar to them (e.g. driving, running), but involving data transactions for the purpose of providing a service. Achieving the informed consent and choice principle (ISO/IEC 29100 [i.1]) on such devices may prove to be harder than usual when they lack of an ad-hoc user interface.

Terms of service should contain clauses that refer to transactions and data uses, though sometimes they might be not evident to the user, especially when she is not the subscriber - and as such, possibly, not fully aware of these terms. In addition, providers should give users the ability to access, correct and remove their information according to the principles of individual participation.

Malicious data transactions may follow data breaches occurred to service providers or their partners. As part of the principle of accountability, providers should implement mechanism to provide users with notifications of breaches.

6.5 Semantics

A primary technical aspect to enable description and understanding of data processing by a machine is the need of assigning common and unambiguous identifiers for different kinds of data resources (including those that may be PII).

Operating systems designed for modern device platforms typically have built-in security mechanisms that prevent new installed programs to automatically access certain kinds of resources accessible from the platform - unless user's explicit consent normally provided through "permission forms". In order to do so, these operating systems are able to distinguish and manage different classes of resources - including data resources that may be PII or that contain PII - and to provide identifiers for them. Identifiers for these classes of data resources may be provided in forms of Uniform Resource Identifiers to facilitate interoperability between different systems.

In general, standard data models may provide identifiers for different classes of data in various domain. In the context of ETSI standardized machine-to-machine architecture the Smart Appliances Reference ontology (SAREF) (available at http://ontology.tno.nl/saref.ttl and the documentation on: http://ontology.tno.nl/saref/) provides such an example. The ontology uses identifiers in form of URIs for different kind of information obtained from devices and sensors in households, common public buildings or offices.

Taxonomies defined at a higher level of abstraction are valuable as well, in that they reduce the number of identifiers enabling shorter and clearer description. ISO/IEC JTC 1/SC 38 CD 19944 [i.23] describes a taxonomy branching into the four basic data categories: customer data/content, derived data, cloud service provider data and account data. Each category is further specialized into sub-sets of related data objects.

6.6 Portability

Portability is the usability of the same data and meta-data under different services, typically by different providers. Portability is a desirable feature which avoids unavailability of data as a consequence of change of provider (data lock-in). When combined with protection, portability of protection settings may ensure that proper protection is still in place despite data is moved or replicated across nodes belonging to different providers or located in different countries.

6.7 Access control

Data contained in an isolated and physically protected device is only available to the holder(s) of the device. A common assumption is that the holder of the device is the PII principal. If this assumption is not guaranteed to hold, then the principal may inhibit access though firmware or software mechanisms (e.g. unlock pin or equivalent protection).

When a remote or Cloud storage is used information security should be implemented in order to provide assurance that only the owner/holder and authorized parties are able to access the data. Policy rules may be used to govern access to data resources. To this end various kinds of access control mechanisms have been developed leveraging on Identity-Based Access Control, Role-Based Access Control and Attribute-Based Access Control. Mechanisms combining access control with encryption (Attribute-Based Encryption) have been specifically considered for Cloud services.

6.8 Log and auditing

According to the principle of accountability of processing, access to data should be logged to allow auditing, anomaly detections and alerting.

Metadata to be recorded in a log file may include the identifier of the data resource, the date and time of access, the kind of operation that is performed, one or more identifiers related to the user, the device or the network service that is accessing the data.

Log files should be retained for a period of time (normally months or years); during the retention period their integrity should be ensured. Log files should be protected in order to prevent unauthorized access to them.

6.9 Embedded sensors and devices

Most PII will be generated through embedded sensors or devices. When PII processing has to be enforced as soon as the data is generated (e.g. in processing using end-to-end encryption), any adopted technical solution should be compatible with embedded memory size constraints both for RAM and Flash. Industrial RAM and Flash can be pretty expensive, especially if industrial temperature range is required (-40 °C, +85 °C).

CPU can be also heavily mobilized during PII processing. As CPU are in general constrained for embedded devices (even though most of them are, at the time of writing, running at 500 MHz or more), it is critical to size the mechanism for embedded sensor and device.

Embedded devices have very strong constraints when dealing with telecom stack (bandwidth, message size, distance, et cetera). Any processing of PII access control mechanism should be compatible with these telecom constraints.

Power consumption is a consequence of memory, telecom and CPU usage. Most of embedded device or sensors are battery-powered. Thus, it is critical to size the overall PII processing in order to be compatible with acceptable battery duration or size.

Finally, design and production cost are key success factors when large rollouts come. As a consequence, in addition to the technical constraints described above, it will be critical to design very cost-sensitive solutions.

6.10 Lawful interception

Compliance with legal and regulatory constraints related to law enforcement represent a principle (ISO/IEC 29100 [i.1]) to be respected in any specific solution. Data protection and privacy legislations typically include exceptions for disclosing of PII with law enforcement agencies, and define obligations for service providers. Standardized LI solutions exist for voice, conferencing, IMS-based services, messaging (SMS, e-mail, etc), and Internet access. Additional solutions may be created for other services including file sharing, telepresence, social media, online games, etc. Therefore, any solution should be as minimally invasive as possible and proportionate to the requirement.

Cloud based services present specific challenges which have been identified in ETSI TR 101 567 [i.10], and are hereafter recalled.

It is a fundamental requirement for the Law Enforcement Authority (LEA) to be able to identify and communicate with the service provider(s) responsible for the communications involving specific targets. Extra help may be needed in cloud environments because the relevant providers are often not subject to registration, regulatory, or CSP partnership requirements that facilitate discovery of the identity of partners.

Media, data and metadata may be encrypted by many parties when transferred or stored (encryption challenge). The service providers who initiate encryption should provide intercepted telecommunications en clair, or if they cannot remove the encryption, provide the LEA with the keys and other information needed to access the information where such keys are available to the service provider [i.21].

NOTE: The CSP typically has no ability to decrypt material already encrypted by subscribers of over the top services prior to transferring it to the provider (end-user encryption problem).

The target of lawful interception may have several different network or service identities, depending on the network or service provider and type of interception being accomplished. The identifiers used by the service provider and access network operator are typically different and the LEA may need extra help to maintain the binding of the identifiers/addresses between the domains.

The subscriber's ability to access their data and services from any device, especially devices with no known association (e.g. not owned) with the interception subject complicates an LEA's ability to initiate an access level lawful interception in a timely manner.

ETSI TS 101 671 [i.22] defines location information as "information relating to the geographic, physical or logical location of an identity relating to an interception subject." Extra help may be needed to the location at which users are using cloud based services an assured manner.

7 Use cases, actors and roles

7.1 Overview

The present clause introduces two general use cases. The use cases are "general", in what they focus on architectural aspects while abstracting from specific threats. Fitting today's mobile and cloud services business, the two use cases here presented exploit the definition of actors and roles introduced by the Cloud Coordination Working Group [i.11] and by ISO/IEC 17789 [i.19].

The main distinction between the two use cases lays on the role of the Device Platform Provider, a Cloud Service Provider offering an identity management service and a functional interface allowing third party service providers to pull and process Cloud Service Customers' or Cloud Service users' data.

7.2 Actors and roles

Actors are individuals or organizations which can play one or more of the following roles:

- Cloud Service Provider (CSP): The Cloud Service Provider role consists of those individuals or organizations providing cloud services to one or more Cloud Service Customers.
- Cloud Service Partner (CSPa): The Cloud Service Partner role consists of those individuals or organizations providing support to the **provisioning** of cloud services by the Cloud Service Provider, or to the **consumption** of cloud service by the Cloud Service Customer.
- Cloud Service Customer (CSC): The Cloud Service Customer role defined in consists of those individuals or organizations consuming one or more cloud services provided by a Cloud Service Provider.

Additionally, the following role are defined:

- PII controller: The PII controller role (defined in ISO/IEC 29100 [i.1]) consists of those individuals or organizations determining the purposes and means for processing personally identifiable information (PII). PII controller role is played as well by individuals who use their PII for personal purposes.
- PII processor: The PII processor role (defined in ISO/IEC 29100 [i.1]) consists of those individuals or organizations processing personally identifiable information (PII) on behalf of and in accordance with the instructions of a PII controller.
- Law Enforcement Authority (LEA): The Law Enforcement Authority aka Government Authority role (defined in [i.11]) consists of those individuals or organizations interacting with providers, customers and partners for the purpose of regulation, law enforcement, national security, inspection, economic stimulation, et cetera.
- Device Platform Provider (DPP): The Device Platform Provider role (defined in ISO/IEC JTC 1/SC 38 CD 19944 [i.23]) is a sub-role of CSP where the CSP provides services necessary to support the device platform. These services typically include identity management services for the user of the device (Cloud Service user) and a marketplace service for the device applications.
- Cloud Service user (CSu): Cloud Service user (defined in ISO/IEC JTC 1/SC 38 CD 19944 [i.23]) is a sub-role of Cloud Service Customer and consists of those individuals consuming one or more cloud services using a particular device.

7.3 Use case UC1

CSC is an individual consuming the end user services offered by a CSP.

The CSP does not perform or does not perform exclusively processing of CSC's account data, rather pushes data to a CSPa, providing that processing under a Service Level Agreement with the CSP.

CSPa acts as PII processor (i.e. CSPa provides support to the **provisioning** of cloud services by the CSP).

The present use case does not exclude that the CSP and the CSPa (depending on the Service Level Agreement) may act, in addition, as PII controllers, processing CSC's account data in order to achieve their own purposes.

7.4 Use case UC2

CSu is an individual consuming the end user services offered by a DPP.

The DPP **acts as PII processor** and provides processing of PII.

CSPa may pull CSu's account data from the DPP typically under a Service Level Agreement with the DPP (i.e. CSPa provides support to the **consumption** of cloud service by the DPP). **CSPa acts as PII processor.**

NOTE 1: Though typical, it is not necessarily. In some contexts the CSu's may be allowed to use services by a third parties unknown to the DPP.

The present use case does not exclude that the DPP and the CSPa (depending on the Service Level Agreement) may act, in addition, as PII controllers, processing account data in order to achieve their own purposes.

NOTE 2: A common instance of UC2 may occur when third party service providers, acting as CSPa, may use the functional interface provided by the DPP to process CSu's account data.

NOTE 3: A second common instance of UC2 may occur when the role of CSPa is performed by a LEA. The LEA may exploit specific functional interface provided by the DPP (or it may require the DPP to provide data offline).

Annex A:
Scenarios

A.1 Medical scenario

Alice is a patient of a famous hospital. Recently, the hospital has worked with a research centre on an open source software solution to keep confidential patient's medical information, yet guarantee high availability of the information to authorized medical personnel of every hospital in the country. The solution is provided in form of an app for wearable devices, and exploits the device's third party cloud storage functionality. Alice installs the application on her smart glasses. Only Alice and authorized medical personnel will be able to access her confidential medical records, despite being stored on a potentially untrusted Cloud storage infrastructure.

NOTE: This scenario is based on original scenarios reported in J.A. Akinyele, C. U. Lehmanny et Al. Self-Protecting Electronic Medical Records: Using Attribute-Based Encryption. Cryptology ePrint Archive, Report 2010/565. 2010 [i.25].

A.2 Flight Passenger Name Record

The Passenger Name Record (PNR) plan is under debate. The plan is intended to implement a unified database of every passenger flying within Europe as well as flights in and out of Europe.

Collected data for every passenger would include several types of data which are clearly PII (name, surname, gender, date of birth, nationality and passport details, payment information, address and contacts) as well as other data which could be PII, such as: details on booking, accompanying persons, no-show/go-show history, bag tag history, special service requests.

The plan has raised privacy concerns from the European Court of Justice and civil liberty associations, who believe that blanket collection of personal data without detailed safeguards is a severe incursion on personal privacy. As a consequence, several aspects of this proposal are still under debate - including retention, processing, use of pseudonymized data, and applicability based on different situations (serious crimes or terrorism).

A.3 Bring Your Own Device (BYOD)

John works in the travel office of a medium size company. He is in charge of planning and arranging business trips for employees and the management board. Recently, the company has implemented a Bring Your Own Device policy, allowing employees to bring personally owned mobile devices (laptops, tablets, and smart phones) to the workplace, and to use those devices to access privileged company information and applications. John took the chance to buy a new tablet PC and decided to use it both for work and at home.

John normally accesses the administrative App of the company to check the new travel requests. This week he has to book four flights for the CEO and some overnight stays. John is used to checking flights and hotel rooms on the Internet using a popular website. When the best combination of flights and hotel is reached, John buys the package on behalf of the traveller using the company virtual credit cards.

At home, during the weekend, John is planning winter holidays for him and his family. Together they start to evaluate different possibilities. Using his new tablet PC John connects to his favourite travel portal (the same he uses for his work), which is offering very interesting deals. When the family finalizes their holyday plan, John proceeds to the purchase. He knows the online procedure very well. The browser auto-form-filling makes the procedure even easier. In fact, the browser saves the most frequently used data proposing the options for each field. Unfortunately this shortcut leads John to finalize his purchase without realizing that the system is using his company's credit card as the payment method. As a consequence, the CEO of his company receives an automatic alert from the travel portal. On Monday John will be informed of this problem.

A.4 Fake or untrusted access mobile networks

For the emerging Next Generation System, or 5G networks, privacy is one of the main aspect to be taken into consideration, having an important social impact. It is one of five major security enablers for the 5G system.

NOTE: See 5G Ensure project, Deliverable D2.1 Use Cases. Available at: http://www.5gensure.eu/sites/default/files/Deliverables/5G-ENSURE_D2.1-UseCases.pdf.

In order to get access to a mobile network, users need first to be identified. Anyway users stumble across fake networks. Fake networks may impersonate legitimate networks but which actually intercept or even change intended mobile device communications. A variety of sensitive information, which might be PII, can be exposed contextually to the user access network procedure, including, but not limiting to user identity, user subscribed services (voice, messaging, data et cetera), location information, traffic, network usage. This way information about subscribers might be exposed and used to impersonate the user or to cause denial of service. Such information requires enhanced protection to be ensured also in a roaming scenario.

John is visiting a foreign country and switches on his mobile. The visited PLMN requests John's user identity (e.g. the IMSI) in order to authenticate him. John implicitly relies on the assumption that the visited PLMN is a trusted network. Unauthorized disclosure of sensitive PII can happen if interconnected networks are not "legitimate" as expected to be.

A.5 Untrusted app scenario

John is at home with his son, Josh. John decides to go downstairs to do some homework, leaving Josh upstairs playing videogames on his tablet PC. After a while Josh gets bored with the game, so he decides to download a new game from the app store. He automatically connects to the store using his father's account. However, to perform a new buy the system asks for an additional password. Josh does not know the password, so he decides to download a free game. Once installed the game asks for a set of permissions. Josh agrees and the game get installed on his father's tablet PC.

Some days later, a popup appears on John's tablet PC screen, saying most of his files have been encrypted and are no longer accessible. To unlock these files, John needs a special cryptographic key that will be sent to him after a payment, in bitcoins, is sent to an anonymous account. What Josh installed was actually a ransomware. John decides to pay. After that, a key is sent to him. Unfortunately John discovers that the key is useful to unlock only a small number of files. To unlock more he is asked to pay a second ransom, etc.

A.6 Social networking

Profiling users is a practice which has its origins in managing fraud and access control in financial and ICT systems, and in law enforcement and national security communities' attempts to deal with serious crime and terrorism. It has also been used by the marketing sector, in consumer reporting services for targeted advertisements and to provide a broad array of free services to users to enable "free at point of delivery" services. Most service models - especially free ones - are non-negotiable. When users subscribe to new services by installing applications, service providers collect and store user's data and metadata in their service infrastructure and may obtain valuable information from the analysis of both. A significant data-ecosystem may exist between the user and the service providers, some of them funded by commercial use of the data flowing through the services.

Jane, John's wife, works as a nurse in a famous hospital of her city. Every day she drives from home to her workplace. Recently she installed a free messaging application to get in touch with her family members, her friends and her colleagues. The application asked her a number of permissions including access to her GPS position and her contact list, but she was not able to understand the terms of use and thus blindly accepted all conditions. Jane's colleagues, friends and family members were made aware of the app as well, because Jane consented to a feature of the app telling her contacts she had installed and was actually using the app. Jane discovered that her colleagues had joined a "group" to discuss work issues, problems, or simply to enjoy the group chat. As she is very social she decided to join the group. One day she even posted a picture of herself with some colleagues at work.

Some weeks after, the app started suggesting new contacts she might know. Jane was surprised, as actually she knew some of those people: they were co-workers and people living near her home. In addition, the app started displaying advertisements of medical products; some products were of little interest for her, but others were interesting. As Jane started watching some of the suggested commercials, she discovered the products were from the competitors of the manufacture her hospital usually dealt with and were being offered with a larger discount.

Months later she was even more surprised to find on the web an online banner showing a group of nurses promoting some medical products. She found the picture was vaguely familiar to her.

A.7 In-car blackbox

Blackboxes are in-car devices which track time, position, speed, steering, braking, mileage and help an insurer set premiums. Drivers refusing to install this device could pay higher premiums or be denied insurance cover. A number of organizations in the eco-system are (legitimately) interested in accessing and using blackbox data: insurers, law enforcement officers, rescuers and medical staff, and other third party.

Advances in this technology will make soon possible to monitor and control for safety purposes in-car connectivity, calls, radio and multimedia devices. Customer wants to make sure that many privacy principles are correctly applied: e.g. collection of data is reasonably limited, anonymity is respected for certain uses, driver's profiling happens correctly (given that a car may be used by many drivers), data is not used to track driver's habits which are not relevant: for instance, drivers want neither to be automatically fined for speed excess for example, nor any other direct and automatic application of law enforcement based on their private driving habits.

A.8 Cloud unavailability

Cloud or remote storage has many advantages for users: data is no longer dependent on the device used to create it and manage it and is "always" accessible. However, cloud computing storage at large date centres may increase the risk of temporary inaccessibility or even of large losses of personal data. In 2011 a hosting service company was offline due to a power lost after a fire broke out in one of their Uninterruptible Power Supply rooms. For a whole morning their users and customers of their users were unable to connect and use the hosted services.

A.9 Self-quantifying

Self-quantified service is proposed to people who want to track their lifestyle activities and to take care about their health and/or to compete against other people based on sport activities. This service can be applied to sport (running for example), to diet, etc. Collected data might be very private and sensitive, but might be shared by the owner (the sensor holder) to other individuals or public or private organizations.

Customers of this kind of service want to make sure that data are used in the right way by the right person. For example, the sensor holder might accept to share daily her heart pulsation rate with her family doctor but not with hospital or public social services. Another example could be the weight. The sensor holder could agree to share anonymously with other participants to a social network dedicated to sport and fitness.

History

Document history		
V1.1.1	July 2016	Publication