ETSI GR ENI 009 V1.1.1 (2021-06)

Experiential Networked Intelligence (ENI); Definition of data processing mechanisms

Disclaimer

The present document has been produced and approved by the Experiential Networked Intelligence (ENI) ETSI Industry Specification Group (ISG) and represents the views of those members who participated in this ISG.
It does not necessarily represent the views of the entire ETSI membership.

Reference
DGR/ENI-0017-Data_Process_Mech

Keywords
artificial intelligence, data collection, data management, data mechanism, data processing, data storage, network management

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 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

Users of the present document should be aware that the document may be subject to revision or change of status.
Information on the current status of this and other ETSI documents is available at
https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Notice of disclaimer & limitation of liability

The information provided in the present deliverable is directed solely to professionals who have the appropriate degree of experience to understand and interpret its content in accordance with generally accepted engineering or other professional standard and applicable regulations.
No recommendation as to products and services or vendors is made or should be implied.
No representation or warranty is made that this deliverable is technically accurate or sufficient or conforms to any law and/or governmental rule and/or regulation and further, no representation or warranty is made of merchantability or fitness for any particular purpose or against infringement of intellectual property rights.
In no event shall ETSI be held liable for loss of profits or any other incidental or consequential damages.

Any software contained in this deliverable is provided "AS IS" with no warranties, express or implied, including but not limited to, the warranties of merchantability, fitness for a particular purpose and non-infringement of intellectual property rights and ETSI shall not be held liable in any event for any damages whatsoever (including, without limitation, damages for loss of profits, business interruption, loss of information, or any other pecuniary loss) arising out of or related to the use of or inability to use the software.

Copyright Notification

Contents

Intellectual Property Rights 5
Foreword 5
Modal verbs terminology 5
Introduction 6
1 Scope 7
2 References 7
2.1 Normative references 7
2.2 Informative references 7
3 Definition of terms, symbols and abbreviations 8
3.1 Terms 8
3.2 Symbols 9
3.3 Abbreviations 10
4 Overview 11
4.1 Background 11
4.2 Data Precondition 11
5 Data Mechanism 11
5.1 Introduction 11
5.1.1 Data Mechanism Overview 11
5.2 Data Characteristics 12
5.2.1 Configuration Data 12
5.2.2 Sequential Data 12
5.2.3 Data Format 13
5.3 Data Source 13
5.3.1 Introduction 13
5.4 Data Collection 14
5.4.1 Introduction 14
5.4.2 Data Acquisition Modes 14
5.4.3 Data Collection Techniques 15
5.4.3.1 Introduction 15
5.4.3.2 Data carried out in functional planes protocols 15
5.4.3.2.1 Data carried out in the Forwarding/User Plane 15
5.4.3.2.2 Data carried out in the Control Plane 15
5.4.3.2.3 Data carried out in the Management Plane 16
5.4.3.3 Specific data used to deploy telemetry 16
5.4.3.3.1 Network Telemetry 16
5.4.3.3.2 Resource Telemetry 18
5.4.3.3.3 Fault Telemetry 19
5.4.3.3.4 Streaming Telemetry 20
5.5 Hierarchical data storage 21
5.6 Data Processing 21
5.6.1 Data Correlation 21
5.6.2 Data Cleansing 22
5.7 Data Sharing 22
5.8 Data Management 23
5.8.1 Overview 23
5.8.2 Metadata Management 23
5.8.3 Data Security Management 23
5.8.4 Data Quality Management 23
6 Example Scenarios to Illustrate Data Mechanisms 23
6.1 AI-enabled Traffic Classification Use Case 23
6.1.1 Introduction 23
6.1.2 Data Acquisition 23

6.1.3 Data Processing for traffic classification 24
6.2 Network Fault Root-Cause Analysis and Intelligent Recovery Use Case 24
6.2.1 Introduction 24
6.2.2 Data Acquisition 24
6.2.3 Data Processing 25
6.3 Intelligent Service Experience Evaluation Use Case 25
6.3.1 Introduction 25
6.3.2 Data Acquisition 26
7 Recommendations 26
Annex A: Change History 28
History 29

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are 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 Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, 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.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Foreword

This Group Report (GR) has been produced by ETSI Industry Specification Group (ISG) Experiential Networked Intelligence (ENI).

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.

Introduction

The present document outlines a high-level reference framework that describes technical methods for producing high-quality actionable data efficiently and in a timely manner.

The organization of the present document is as follows:

- Clause 1 defines the scope of the present document.
- Clauses 2 and 3 provide informative references, terms, symbols and abbreviations.
- Clause 4 describes an overview of the data mechanism, including its motivation and challenges.
- Clause 5 defines components in the high-level framework of the data mechanism in terms of data acquiring and data processing.
- Clause 6 presents the data mechanisms in some example scenarios proposed in ETSI GR ENI 001 [i.1], Use Case specification.
- Clause 7 concludes possible contributions to other ENI group specifications of the present document.

Data Telemetry is used as an example for data mechanisms description and analysis.

1 Scope

The present document describes some technical methods to support data-driven intelligent network scenarios. The realization of intelligent networks depend on extracting value from Big Data using AI algorithms. Therefore, effective data acquisition, processing and management is extremely important as described in this context.

The present document covers the following aspects:

1) Data classification in terms of the data sources producing the data (e.g. network management system, network elements, servers, terminals and external environment data), data characteristics (e.g. configuration or sequential data), and data format.

2) Data operation including data collection, data storage, data processing, data sharing and data management:

   a) Data collection including description about collection modes (e.g. pull (query/request response) and push (publish/notify)), and data collection techniques, such as telemetry.

   b) Data storage recommendations.

   c) Data processing, including data cleansing and data correlation.

   d) Data sharing.

   e) Data management, including metadata management, data security management and data quality management.

3) Data acquisition and processing methods of selected use cases proposed in ETSI GR ENI 001 [i.1] for ENI systems executing intelligent tasks.

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] ETSI GR ENI 001 (V3.1.1): "Experiential Networked Intelligence (ENI); ENI use cases".

[i.2] ETSI GR ENI 004 (V3.1.1): "Experiential Networked Intelligence (ENI); Terminology for Main Concepts in ENI".

[i.3] ETSI GS ENI 005 (V2.1.1): "Experiential Networked Intelligence (ENI); System Architecture".

[i.4] IETF RFC 7011: "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information".

[i.5] IETF RFC 7950: "The YANG 1.1 Data Modeling Language".

[i.6] IETF RFC 4656: "A One-way Active Measurement Protocol (OWAMP)".

[i.7] IETF RFC 5357: "A Two-Way Active Measurement Protocol (TWAMP)".

[i.8] IETF I-D.ietf-ippm-ioam-data-11: "Data Fields for In-situ OAM".

[i.9] IETF RFC 8321: "Alternate-Marking Method for Passive and Hybrid Performance Monitoring".

[i.10] IETF RFC 8889: "Multipoint Alternate Marking method for passive and hybrid performance monitoring".

[i.11] IETF RFC 7799: "Active and Passive Metrics and Methods (with Hybrid Types In-Between)".

[i.12] Recommendation ITU-T Y.1731: "OAM functions and mechanisms for Ethernet based networks".

[i.13] IETF RFC 6241: "Network Configuration Protocol (NETCONF)".

[i.14] IETF RFC 4271: "A Border Gateway Protocol 4 (BGP-4)".

[i.15] IETF RFC 7854: "BGP Monitoring Protocol (BMP)".

[i.16] IETF I-D.draft-kumar-rtgwg-grpc-protocol-00: "gRPC Protocol".

[i.17] IETF I.D.draft-zhou-ippm-enhanced-alternate-marking-05: "Enhanced Alternate Marking Method".

[i.18] IETF I.D.draft-song-ippm-postcard-based-telemetry-08: "Postcard-based On-Path Flow Data Telemetry using Packet Marking".

[i.19] IETF RFC 793: "Transmission Control Protocol (TCP)".

[i.20] IETF RFC 768: "User Datagram Protocol (UDP)".

[i.21] VNF Event Stream (VES).

NOTE: Available at https://wiki.opnfv.org/display/ves/VES+Home.

[i.22] IETF RFC 3416: "Version 2 of the Protocol Operations for the Simple Network Management Protocol (SNMP)".

[i.23] IETF RFC 959: "File Transport Protocol (FTP)".

[i.24] The Atlan Data wiki definition of structured data.

NOTE: Available at https://wiki.atlan.com/structured-data/.

[i.25] The Atlan Data wiki definition of unstructured data.

NOTE: Available at https://wiki.atlan.com/unstructured-data/.

[i.26] IETF RFC 4560: "Definitions of Managed Objects for Remote Ping, Traceroute, and Lookup Operations".

[i.27] Prometheus open source.

NOTE: Available at: https://prometheus.io/.

3 Definition of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the terms given in ETSI GR ENI 004 [i.2], ETSI GS ENI 005 [i.3] and the following apply:

column-oriented database: database that organizes data by field

NOTE: This type of database keeps all of the data associated with a field next to each other in memory, and is optimized for online analytical processing. They are optimized for reading and computing on columnar data. Examples include Snowflake and BigQuery.

data lake: centralized storage repository that stores raw data that are in the form of structured, semi-structured and unstructured format

data mart: subset of a data warehouse focused on a particular line of business, department, or subject area

data warehouse: repository used to connect, analyse, and report on historical and current data from heterogeneous sources

NOTE: A data warehouse is designed for query and analysis as opposed to transaction processing. It analyses and reports on data from operational systems as used in decision-support systems.

hadoop distributed file system: distributed fault-tolerant file system that stores data on commodity machines and provides high throughput access

massively parallel processing: use of a large number of processing nodes that perform a set of coordinated tasks in parallel using a high-speed network

NOTE: The processing nodes typically are independent, and do not share memory, and typically each node runs its own instance of an operating system.

Prometheus: open-source systems monitoring and alerting toolkit

NOTE: This open source is originally built at SoundCloud. Since its inception in 2012, many companies and organizations have adopted Prometheus, and the project has a very active developer and user community. It is now a standalone open source project and maintained independently of any company. To emphasize this, and to clarify the project's governance structure, Prometheus joined the Cloud Native Computing Foundation in 2016 as the second hosted project, after Kubernetes.

protocol buffers (protobuf): language-neutral, platform-neutral, extensible mechanism for serializing structured data

reinforcement learning: See ETSI GR ENI 004 [i.2] and ETSI GS ENI 005 [i.3].

row-oriented database: database that organizes data by record

NOTE: This type of database keeps all of the data associated with a record next to each other in memory, and is optimized for online transaction processing. An example is MySQL.

semi-structured data: information that does not conform to a formal data model, but does have some organizational properties that define key data (e.g. tags) that enable data to be self-describing

software defined hardware: software programmable hardware that is able to be reconfigured at runtime to enable near ASIC performance without sacrificing programmability for data-intensive algorithms

structured data: information organized in a predetermined way (a fixed format, data model or schema) within a record or a file

NOTE 1: As defined in [i.24].

NOTE 2: Structured data enables all elements to be individually addressable, and conform to a data model.

unstructured data: information that does not have a pre-defined data model, and does not contain properties that provide any organization or structure to its elements

NOTE: Unstructured data needs to be processed in order to find information by domain-specific applications.

video stalling: process during the video playback, the video is paused and waits for the buffer due to dragging or other reasons

3.2 Symbols

Void.

3.3 Abbreviations

For the purposes of the present document, the abbreviations given in ETSI GR ENI 004 [i.2], ETSI GS ENI 005 [i.3] and the following apply:

5G	Fifth Generation
AI	Artificial Intelligence
AS	Autonomous System
BGP	Border Gateway Protocol
BMP	BGP Monitoring Protocol
BSS	Business Support Systems
CPU	Central Processing Unit
CRM	Customer Relationship Management
EAM	Explicit Address Mapping
ENI	Experiential Networked Intelligence
FTP	File Transport Protocol
gNMI	gRPC Network Management Interface
IETF	Internet Engineering Task Force
IMS	Integrated Management System
IOAM	In-band OAM
IP	Internet Protocol
IPFIX	IP Flow Information eXport
IPFPM	IP Flow Performance Measurement
IPPM	IP Performance Metrics
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication standardization sector
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
MS	Monitoring System
NE	Network Element
NMS	Network Management System
OAM	Operation, Administration and Maintenance
OMC	Operations and Maintenance Centre
OSS	Operations Support Systems
OWAMP	One-Way Active Measurement Protocol
PBT	Postcard-Based Telemetry
QoS	Quality of Service
SDN	Software-Defined Networking
SDK	Software Development Kit
SLA	Service Level Agreement
SNMP	Simple Network Management Protocol
SQL	Structured Query Language
SR-IOV	Single Root I/O Virtualization
TCP	Transmission Control Protocol
TWAMP	Two-Way Active Measurement Protocol
UDP	User Datagram Protocol
VES	VNF Event Stream
VNF	Virtual Network Function
XML	Extensible Markup Language
YANG	Yet Another Next Generation

4 Overview

4.1 Background

Exploiting network data for intelligent network applications and use has been increasing in recent years. By combining AI and machine learning algorithms, network data is able to provide insights that help network operators better manage and optimize the network. Therefore, the quality of available sample data, for instance, time validity, diversity, volume, accuracy, plays an important role in learning from data. One challenge is that large amounts of data as well as data that meets the demands is able to be acquired. Additionally, the data collected from network equipments from different vendors varies in the aspect of name, format, calculation rules, etc. Thus a large amount of time is often be spent to do the data normalizing, cleansing, and engineering before those data could be used to train the model. This blocks the deployment of actionable decisions, which are meant to improve ENI System performance and User Experience.

The present document describes data acquisition, sharing and processing mechanisms, as well as supports for data privacy in AI-enabled network Operation, Administration and Management (OAM). The present document identifies the sources and data to be extracted, however it does not intend to explain how the mechanisms work, or how data is processed in order to became used. This could be addressed in a later release.

4.2 Data Precondition

Different types of data are able to be analysed only and interpreted correctly in particular contexts. The following are examples of some of the types of data that the present document focuses on.

Real-time data: Typically, network data has to be continually monitored and dynamically processed in real-time. Example processing includes filtering, correlation, and cleansing. This is typically down locally and then aggregated results are distributed for further processing.

Continuous data: In some cases, continuous data over a long time span is required for analysis or model training. For example, historical traffic data are used to predict future traffic trends. In general, the longer the time span, the more representative it is, but the larger the data volume. Therefore, a way of efficiently processing and managing continuous data is needed.

NOTE: More consideration on "historical data" will be described in a future version in a later release.

5 Data Mechanism

5.1 Introduction

5.1.1 Data Mechanism Overview

This clause defines components in a high-level overview for data acquisition and processing. Furthermore, this clause classifies different types of data in terms of their data sources, as well as describes data processing mechanisms, in order to support AI enabled network OAM and service management.

The Data Mechanism supports different data acquisition and processing mechanisms for data from different sources and for use by different network applications.

As shown in Figure 5-1, the data mechanism overview is able to be partitioned into the following components.