ITUPublications

International Telecommunication Union
Standardization Sector

ITU Focus Group Technical Specification

(01/2024)

Focus Group on Autonomous Networks

(FG-AN)

Concepts and principles of trust for autonomous networks including IMT-2020 and beyond

Technical Specification ITU FG-AN

Concepts and principles of trust for autonomous networks including IMT-2020 and beyond

Summary

This Technical Specification provides the concepts, and basic principles for trust in autonomous networks.

Any material in this Technical Specification is intended not to define normative specifications of Recommendations, but to provide technical information that could be used for future possible ITU-T Recommendations on trust evaluation methods and/or other aspects.

Keywords

Architecture, autonomous networks, basic principles, metrics, trust.

Note

This is an informative ITU-T publication. Mandatory provisions such as those found in ITU-T Recommendations are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Contributors: Xiaojia SONG
China Mobile
P.R. China
Tel: +86 15011488067
E-mail: songxiaojia@chinamobile.com

Li YU
China Mobile
P.R. China
Tel: +86 15801696688
E-mail: yuliyf@chinamobile.com

Xi CAO
China Mobile
P.R. China
Tel: +86 13911364997
E-mail: caoxi@chinamobile.com

Gyu Myoung LEE
KAIST
Korea (Rep. of)
Tel: +82-42-350-6282
E-mail: gmlee@kaist.ac.kr

Leon WONG
Rakuten Mobile
Japan
E-mail: leon.wong@rakuten.com

Paul HARVEY
Rakuten Mobile
Japan
E-mail: paul.harvey@rakuten.com

Laurent CIAVAGLIA
Rakuten Mobile
Japan
E-mail: laurent.ciavaglia@rakuten.com

Pedro García PARRA
Telefónica
Spanish
E-mail: pedro.garciaparra@telefonica.com

This Technical Specification has been published as approved by the focus group, without any subsequent editorial review.

Table of Contents

Technical Specification ITU FG-AN

Concepts and principles of trust for autonomous networks including IMT-2020 and beyond

1 Scope

This Technical Specification provides concept, and basic principles for trust in autonomous networks (AN). The scope of this technical specification including:

- **Stage one (frozen, closed to be discussed, contributed or updated since 2021.11.11):**
  - Overview of trust in AN, including concepts and basic principles (clause 6).

2 References

[ITU-T Y.3051] Recommendation ITU-T Y.3051 (2017), *Basic principles of trusted environment in information and communication technology infrastructure.*

[ITU-T Y.3052] Recommendation ITU-T Y.3052 (2017), *Overview of trust provisioning in information and communication technology infrastructures and services.*

3 Definitions

None.3.1 Terms defined elsewhere

This Technical Specification uses the following term defined elsewhere:

3.1.1 trust [ITU-T Y.3052]: Trust is the measurable belief and/or confidence which represents accumulated value from history and the expecting value for future.

3.2 Terms defined in this Technical Specification

This Technical Specification defines the following terms:

3.2.1 trusted AN: The autonomous network which is trustworthy enough (i.e., working correctly as intended), so that the network itself can be partly or completely autonomous.

3.2.2 trust in AN: A measurable and quantifiable degree of the trustor's confidence in a network to let it be governed by itself with minimal to no human intervention.

3.2.3 trustor in AN: The one who/which has the authority to authorize a network and/or the relevant entity be governed by itself with minimal to no human intervention.

3.2.4 Trustee in AN: A network and/or the relevant entity with autonomy capabilities which is to be authorized to govern itself with minimal to no human intervention.

4 Abbreviations and acronyms

This Technical Specification uses the following abbreviations and acronyms:

AI Artificial Intelligence

AN Autonomous Network

CSP Communication Service Provider

ICT Information and Communication Technologies

5 Conventions

None.

6 Overview of trust in AN

6.1 Introduction

As decision-making behaviour, trust is affected by past experience and associated predictions for the future. The study of trust in automated systems has been a topic of psychological study previously. However, artificial intelligence (AI) poses unique challenges for user trust, the AI user has to trust the AI, changing the interaction between a user and a system into a relationship. Trust is a complexity-reduction mechanism, whose importance increases the less we know about the technology. In information and communication technologies (ICT), trust has been studied and discussed since the application or usage of machine intelligence or AI. Mobile networks are evolving into the intelligence era with multiple application scenarios, features, services and operation requirements. Technologies including AI are expected to enable autonomous networks (AN) in areas such as network planning, deployment, operation, optimization, service deployment, and assurance. With the development of network systems and evolution of AI technology applications, operators are supposed to gradually handover their work and duties to network systems themselves which have self-X properties (the abilities to monitor, operate, recover, heal, protect, optimize, and reconfigure themselves), AN are becoming an inevitable trend of the network evolution.

From the perspective of operators who are interested in and plan to deploy AN, the following questions may arise:

- Should an operator trust the AN?
- How to make human operator trust AN, and willing to hand over the control authority of the network to the AN system?
- How much can human operators trust their AN?
- What level of trust is sufficient/required for different mechanisms/aspects in AN?
- What are the most important factors for AN to earn human operator's trust?
- How to evaluate human operator's trust in AN?

If the above questions cannot be answered properly, the following serious problems and challenges will happen: the AN solutions including reliable ones and risky ones are difficult to be distinguished and may not be applied in operators' real networks due to the lack of confidence from operators, especially when the operators are facing the pressure of network quality assessment and market competition, traditional but more familiar solutions may be preferred then. On the other hand, from the AN's perspective, due to the lack of sufficient trust, AN solutions cannot get enough application opportunities in the real network, and thus lose the chance to learn and evolve to a more advanced level.

To solve the above problems, it is recommended to investigate and study trusted AN and trust in AN to help the telecom industry including communication service providers (CSPs), vendors and other industry participators to reach a consensus and unified understanding of the concepts and evaluation method on trust in AN. It will be discussed and defined in this Technical Specification, including but not limited to the concepts, basic principles, and evaluation of trust in AN (metrics, evaluation methodology, evaluation models, use cases and gap analysis).

6.2 Concepts

Trust concept itself is a complicated notion with different meanings depending on both participators and situations and influenced by both measurable and non-measurable factors. There are various kinds of trust definitions leading to difficulties in establishing a common, general notation that holds,

regardless of personal dispositions or differing situations. Generally, trust is considered as a computational value depicted by a relationship between trustor and trustee, described in a specific context and measured by trust metrics and evaluated by a mechanism. [b-ITU-T TR Trust]

Trust is defined as "the measurable belief and/or confidence which represents accumulated value from history and the expecting value for future" in [ITU-T Y.3052].

Followings are the relevant concepts defined in this Technical Specification (see Figure 1):

- **Trusted AN**: the autonomous network which is trustworthy enough (i.e., working correctly as intended), so that the network itself can be partly or completely autonomous.
- **Trust in AN**: a measurable and quantifiable degree of the trustor's confidence in a network to let it be governed by itself with minimal to no human intervention.
- **Trustor in AN**: the one who/which has the authority to authorize a network and/or the relevant entity be governed by itself with minimal to no human intervention.
- **Trustee in AN**: a network and/or the relevant entity with autonomy capabilities which is to be authorized to govern itself with minimal to no human intervention.

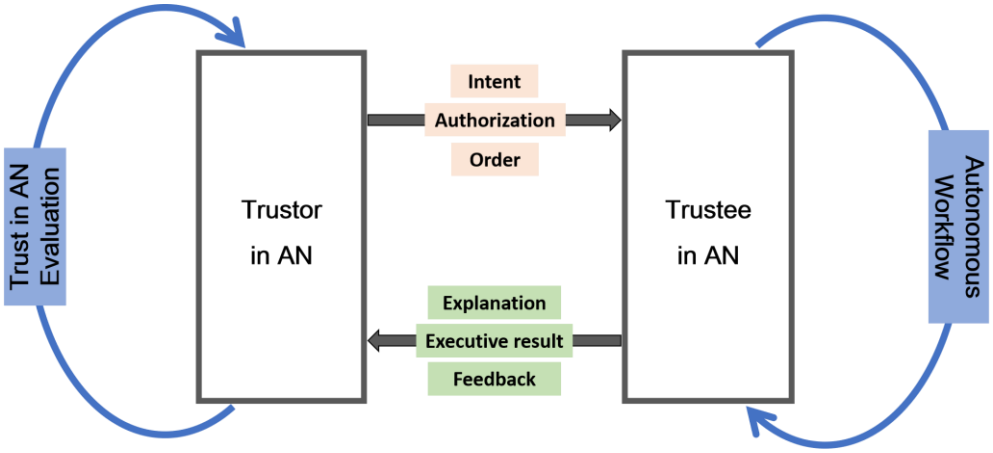

Figure 1 – A conceptual model of trusted AN

In the above conceptual model of trusted AN:

- Trustor in AN should do evaluate trust in AN and decide the following procedures, e.g., intent, authorization, order, etc.
- Trustee in AN should feedback the relevant issues (such as explanation, executive results, feedbacks, etc.) according to the period or time it has been set.
- The evaluation of trust in AN may be initiated and completed by trustor in AN.
- A close loop autonomous workflow may process with trustee in AN itself.

6.3 Basic principles

6.3.1 Current research situation of trust principles

In [ITU-T Y.3051], basic principles for creating a trusted environment in ICT infrastructure that provides information and communication services have been specified. The basic principles of trusted AN are specified carefully based on [ITU-T Y.3051].

In [ITU-T Y.3052], some descriptions about trustworthiness attributes are generally listed in Table II.1, and all of the attributes have been taken into consideration when provide the following basic principles of trusted AN.

In the meantime, the topic of "trusted AI principles" has been initiated as a working group to be discussed and studied in Linux Foundation. The (R)REPEATS acronym captures AI principles of Reproducibility, Robustness, Equitability, Privacy, Explainability, Accountability, Transparency, and Security. These principles were derived after over a year of deliberation which included parsing through the various industry, non-profit, and partner company's AI principles, guidelines, contributions, and principles, while always keeping the community and social impact front and center. In addition to member companies' and non-profit groups' input, guidelines from OECD, EU, SoA, ACM, IEEE, DoD were also referenced. The key criteria balanced competing interests across the industry and companies with the need for open and innovative technology built with trust and accountability.

NOTE – "LF AI & Data Announces Principles for Trusted AI" can be found on the website: https://lfaidata.foundation/blog/2021/02/08/lf-ai-data-announces-principles-for-trusted-ai/.

6.3.2 Basic principles of Trusted AN

Accountability, equitability, explainability, robustness and safety should be the basic principles for trusted AN and all of them shall be taken into consideration. The followings are the illustration of basic principles of trusted AN:

- **Accountability** requires AN and its provider(s) or vendor(s) to explain, justify and take responsibility for any decision and action made by the AN.
- **Equitability** for AN and its provider(s) or vendor(s) should take deliberate steps – in the AN life-cycle – to avoid intended or unintended bias and unfairness that would inadvertently cause any harm, damage or loss.
- **Explainability** is the ability to describe how AN work, i.e., make decisions and actions. Explanations should be produced regarding both the procedures followed by the AN (i.e., their inputs, methods, models, algorithms and outputs, etc.) and the specific decisions and actions those are made. These explanations should be accessible to people with varying degrees of expertise and capabilities including the public.

  NOTE – For the explainability principle to take effect, the AN engineering discipline should be sufficiently advanced such that technical experts possess an appropriate understanding of the technology, development processes, and operational methods of its AN systems, including the ability to explain the sources and triggers for decisions through transparent, traceable processes and auditable methodologies, data sources, and design procedure and documentation.
- **Robustness** refers to the stability, resilience, adaptability, recency and performance of the systems and machines dealing with changing ecosystems. AN should function robustly throughout its life cycle and potential risks should be continually assessed and managed.
- **Safety** of AN should be tested and assured across the entire life cycle within an explicit and well-defined domain of use. In addition, any AN should be designed to also safeguard the data, infrastructure, relevant hardware and software which are impacted.

In the above basic principles, none of them is with higher priority than any other, and each of them are related to each other.

6.4 Initiation and continuation of trust for trusted AN

6.4.1 General process of trusted AN

Trust can be divided into the objective part and the subject part. For AN, trust should be considered the subjective part from trustor's perspective and the objective part from trustee's perspective. For trusted AN, both initiation of trust and continuation of trust are the essential factors for networks working and evolving normally. The general process of trusted AN has been illustrated in following Figure 2.

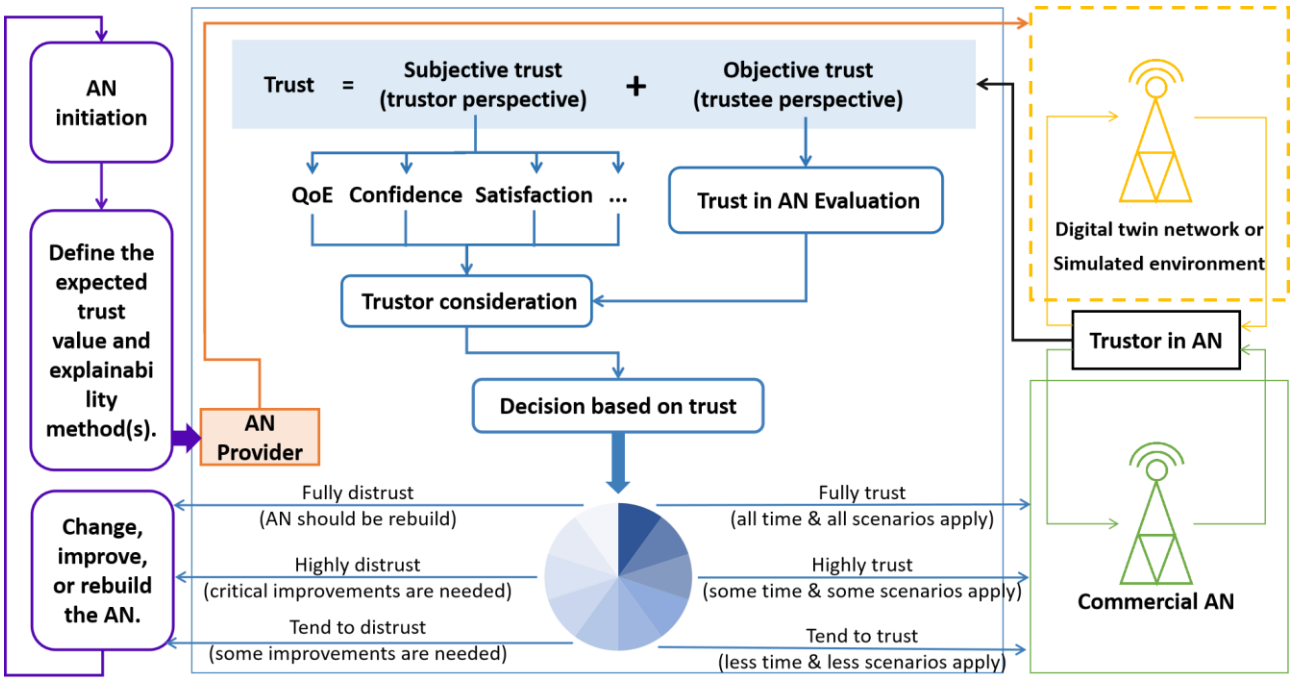

Figure 2 – General process of trusted AN

6.4.2 Initiation of trust for trusted AN

In order to initiate trust for trusted AN, trustor in AN can give an order to start the evaluation of trust in AN and the relevant process, after trust in AN evaluation, trustor in AN should take the results into consideration to make decision to AN.

- Trustor in AN should give some order or process to start the evaluation of trust in AN for the trustee in AN.
- The result of trust in AN evaluation is recommended to outcome as percentage.
- Trustor in AN shall take the result of trust in AN evaluation into consideration, alone with the subjective factors of trust, to make the following authorization decision(s):
  - Fully trust: the commercial AN can be applied in all the scenarios all the time before next evaluation.
  - Highly trust: the commercial AN can be applied in some scenarios in some specific time before next evaluation.
  - Tend to trust: the commercial AN can be applied just for a few scenarios in some specific time before next evaluation.
  - Tend to distrust: the AN provider(s) should make some improvement(s) for current AN, in order to gain more trust in following evaluation(s).
  - Highly distrust: the AN provider(s) should make some critical and essential improvements for current AN, in order to gain more trust in following evaluation(s).
  - Fully distrust: the AN provider(s) should rebuild current AN, in order to regain trust in following evaluation(s).
- The evaluation of trust in AN is recommended to take place in the digital twin network or some simulated environment which are both mirrored from commercial AN.

Trusted AN can be initiate when trustor in AN decide to trust, and then the trustee in AN will authorized to govern themselves with minimal to no human intervention.

6.4.3 Continuous trust for trusted AN

After trusted AN being initiated, continuous or periodic evaluation shall be taken place to monitor the trust's fluctuation, in order to provide reference to trustor in AN to adjust the authorization decisions.

The AN provider(s) should continuously improve the AN, in the meantime, AN may do evolve themselves, so that to earn continuous trust to maintain trusted AN.

As the essential and objective part of trust, trust in AN is mainly discussed in the follow-up contents, including but not limited the evaluation of trust in AN (including metrics, evaluation methodology and evaluation models) and architecture for enabling trust in AN.

Bibliography

[b-ITU-T Y.3053 Amd1] Recommendation ITU-T Y.3053 – Amendment 1 (2018), *Framework of trustworthy networking with trust-centric network domains.*

[b-ITU-T TR Trust] ITU-T Technical Report (2017), *Trust in ICT*

[b-Draft NISTIR 8312] P. Jonathon Phillips, Carina A. Hahn, Peter C. Fontana, David A. Broniatowski, 8 Mark A. Przybocki, Draft NISTIR 8312, *Four Principles of Explainable Artificial Intelligence.*

[b-Chen] Chen L., Weisi G., Schyler C.S., Saba A. and Antonios T. (2020), *Trustworthy deep learning in 6G-enabled mass autonomy, IEEE Vehicle Technology Magazine*, December 2020, pp. 112-121.

[b-Keng] Keng Siau and Weiyu Wang (2018), *Building Trust in Artificial Intelligence, Machine Learning, and Robotics, Cutter Business Technology Journal*, Vol. 31, No. 2, March 2018, pp. 47-53.

[b-KPMG] KPMG Report (2018), *Trust in Artificial Intelligence – Transform your business with confidence*.

[b-LF AI] Jacqueline Z.C., LF AI & Data Announces Principles for Trusted AI (2021), https://lfaidata.foundation/blog/2021/02/08/lf-ai-data-announces-principles-for-trusted-ai/.

[b-NISTIR 8330] Brian Stanton, Theodore Jensen, NIST IR 8330 (2020), *Trust and Artificial Intelligence*.

[b-ZSM 010] ETSI GR ZSM010, Group Report "*Zero-touch network and Service Management (ZSM); General Security Aspects*", July 2021.