# ETSI GR NGP 006 V1.1.1 (2018-06)


**Next Generation Protocols (NGP);**
**Intelligence-Defined Network (IDN)**



***Disclaimer***

The present document has been produced and approved by the Next Generation Protocols (NGP) 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/NGP-006

Keywords

framework, next generation protocol

*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

# Contents

# Intellectual Property Rights

## Essential patents

IPRs essential or potentially essential to normative deliverables 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.

## 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.

# Foreword

This Group Report (GR) has been produced by ETSI Industry Specification Group (ISG) Next Generation Protocols (NGP).

# 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.

# 1 Scope

The scope of the present document is to specify the self-organizing control and management planes for the Next Generation Protocols (NGP), Industry Specific Group (ISG).

# 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] https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/.

[i.2] https://www.thesun.co.uk/tech/4141624/facebook-robots-speak-in-their-own-language/.

[i.3] Reed S, Akata Z, Yan X, et al.: "Generative adversarial text to image synthesis", in ICML 2016.

[i.4] Oord A, Dieleman S, Zen H, et al.: "Wavenet: A generative model for raw audio", arXiv:1609.03499, 2016.

[i.5] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton: "Deep learning", in Nature 521.7553 (2015): 436-444.

[i.6] Kingma D P, Welling M.: "Auto-encoding variational bayes", in ICLR 2014.

[i.7] Goodfellow, Ian, et al.: "Generative adversarial nets", in NIPS 2014.

[i.8] Cisco White Paper.

NOTE: Available at https://www.cisco.com/c/en/us/products/collateral/routers/wan-automation-engine/white_paper_c11-728552.html.

[i.9] https://arxiv.org/abs/1701.07274.

[i.10] ETSI TR 121 905: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; Vocabulary for 3GPP Specifications (3GPP TR 21.905)".

[i.11] ETSI TS 136 401: "LTE; Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Architecture description (3GPP TS 36.401)".

# 3 Abbreviations

For the purposes of the present document, the abbreviations given in ETSI TR 121 905 [i.10] and the following apply to scenarios that include mobile network architectures:

| | |
|---|---|
| 3GPP™ | 3rd Generation Participation Project |
| AI | Artificial Intelligence |
| DHCP | Dynamic Host Configuration Protocol |
| E-W | East and West (direction) |
| IDN | Intelligence-Defined Network |
| IETF | Internet Engineering Task Force |
| IP | Internet Protocol |
| ISG | Industry Specific Group |
| ML | Machine Learning |
| NE | Network Element |
| NGP | Next Generation Protocols |
| NMS | Network Management System |
| N-S | North and South (direction) |
| OAM | Operation And Management |
| OSPF | Open Shortest Path First |
| QoE | Quality of Experience |

# 4 Overview

The Next Generation Protocols (NGP), ISG aims to review the future landscape of Internet Protocols, identify and document future requirements and trigger follow up activities to drive a vision of a considerably more efficient Internet that is far more attentive to user demand and more responsive whether towards humans, machines or things.

A measure of the success of NGP would be to remove historic sub-optimised IP protocol stacks and allow all next generation networks to inter-work in a way that accelerates a post-2020 connected world unencumbered by past developments.

The NGP ISG is foreseen as having a transitional nature that is a vehicle for the 5G community and other related communications markets to first gather their thoughts together and prepare the case for the Internet community's engagement in a complementary and synchronised modernisation effort.

Therefore NGP ISG aims to stimulate closer cooperation over standardisation efforts for generational changes in communications and networking technology.

The present document focuses on proposing a new Intelligence-Defined Network (IDN) architecture and a gap analysis of current architectures. The intelligence technologies can learn from historical data, and make predictions or decisions, rather than following strictly predetermined rules. On one hand, the IDN can dynamically adapt to a changing situation and enhance its own intelligence with by learning from new data. On the other hand, IDN can also aim at supporting human-based decision by pre-processing data and rendering insights to users through advanced user interfaces and visualisations. The integration with various network infrastructures, such as SDN, NFV&MANO, intelligence router, traditional router, is in the scope of the present document.

# 5 Background

## 5.1 Continuous Evolution of Network

The development of network is continuously evolving process. In different stages, the network faces to various and different complexity problems. Therefore, the operating and management methodologies are also various.

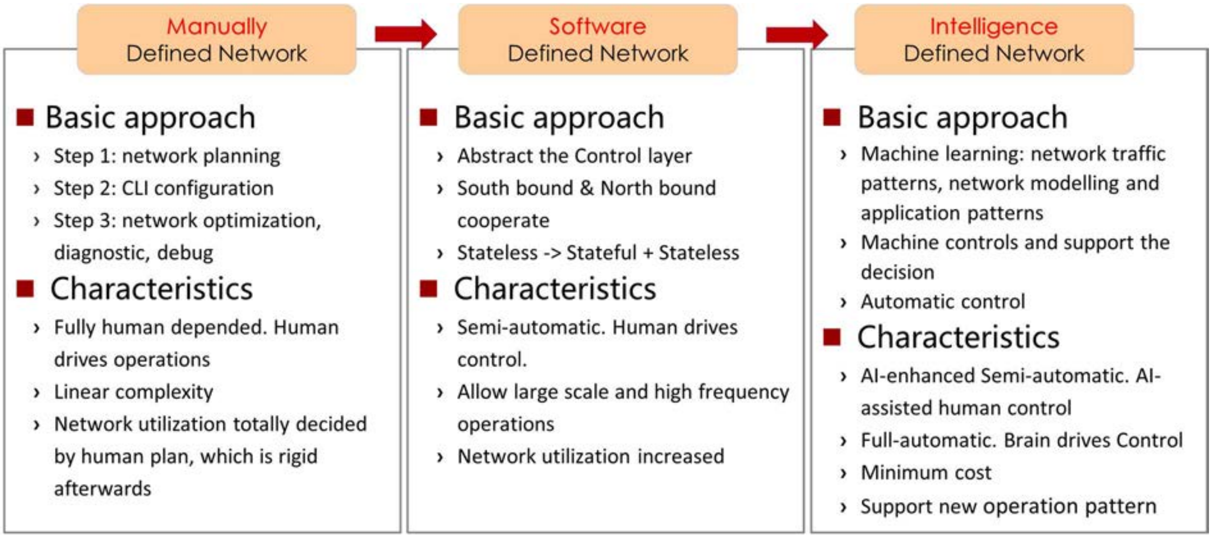


**Figure 1: Three Stages Development**

As Figure 1 shows, early networks are referred as Manually Defined Network. In such networks, the OAM basic approach is network planning, CLI configuration and network optimization. All the operations are fully human driven. Since the administrator needed to configure and control each device individually, the complexity and the cost of OAM was very high.

Along with the development, the scale of network and service became larger and larger. The OAM requirement has also increased. Due to the high degree of required control degree of requirement, a virtual control layer was developed. This layer supports batch operation of low layer devices, which improves the efficiency significantly. Because of the divide of control layer and forwarding layer, the configuration and controlling operation is implemented by south and north-bound cooperation. Southbound typically uses Netconf/YANG, OpenFlow, etc. to configure the network forwarding, policies, etc. Northbound abstracts the functionalities for application requirements, thus deriving the forwarding table and policies, etc. With this paradigm, the entire network was transformed to become semi-automatic. Many operations can be executed automatically and the administrator is only responsible for decision making.



Currently, the network undergoing a new transformation towards Intelligence-Defined Network (IDN). Since the network problems evolve to more complex, the traditional human decision-making can hardly support the requirements. Therefore, the AI methods, which can help for decision making and analysis, are introduced to solve OAM problems. The core of IDN is machine learning algorithms and models. The network, traffic and application patterns can be modelled by AI methods via learning from the existing data and experiences. It is expected as a full-autonomic system that can make decision itself, especially in the common and repeat events that do not need human to judge. This will decrease the OAM cost hugely in the future.

## 5.2 Functional and Systemic Requirement

IDN is seen as the next form of network evolution. Comparing with the current state, there are new requirements that declare the essential improvements of new approach.

The first part is functional requirements, which means the IDN approach should own the functions that the previous network approaches do not have. For IDN, some of the potential functional requirements are following:

- Real-time assessment. The IDN approach should provide a consolidated view of the current network status including traffic and running applications by providing aggregated and condensed insights.
- Prediction or inference. The IDN approach should have the ability to predict / infer the oncoming trend of network in multiple dimensions, such as inferring the QoS parameter according to the traffic matrix. This ability will support the intelligent system to implement proactive operations.
- Autonomic decision making. The network will not only execute the policies which produced by administrator but also autonomously make decision. This is one of the most important reasons that why AI technology is introduced into network.

- Verification. IDN is not only in charge of taking decisions but also (1) verifies that its own decision are properly applied and results in the expected states and (2) can be leveraged to verify that policies derived from multiple entities (concurrent IDN algorithms and even users) are coherent.

- Dynamical configuring ability. For stability consideration, typically operators try to minimize changes on the devices. However, one of the purposes of introducing AI technology is to modify the configuration so that adapt for the variation of network traffic and state.

The second part is systemic requirements, which means the IDN approach should own the system level abilities in the low layer (or say primordially) that the previous network approaches do not have or cannot easily complement. For IDN, the potential systemic requirements are following but not limited.

- Inherent data collection and orchestration. The current measure method is driven by external command. Namely, all the network data is a response (or feedback) of a specific command. The network devices do not widely support the actively data upload functions. This leads to at least two problems. The first one is the cost. When the measured data volume is large, there is nearly half of the signal messages and transfer time are wasted because one data feedback should be potentially triggered by one measure signal. This external trigger mechanism may not satisfy the requirements of huge network data collection. The second one is the complexity. The current measured factors are few (delay, jitter, loss). Even if in this case, it is hard to obtain the accurate data according to simple operations. The intelligent system may handle not only the existing factors but also other complex data types. Some of the factors may be hard to measure, such as if the queuing length is wanted to know. Furthermore, IDN decision algorithms could also rely on external data for which particular connectors are required. This potentially becomes one of the key systemic requirements.

- Data pre-processing. As multiple sources of data will be leveraged, normalisation techniques it its large sense should be used (including data alignment, sanitization). It also concerns the establishment of proper metrics (distance, similarities, and dissimilarities) which are in the core of ML algorithms whereas some collected data may not be easily mapper to a metric space by nature.

- Map algorithm to network. There is a huge gap between the current AI algorithms input/output and network policy. The former is pure mathematical expression while the latter one tends to be a kind of programmed language. If the intelligent system is seen as the mapping of physical network, it is very important to build up the "bridge" between the network semantics and algorithm semantics. Different with the process of data orchestration, the core problems here is how to generate and delivery the network policies based on the mathematical input/output of the algorithm.



## 5.3 Rapid Development of Machine Learning Technologies

Even though the use of Machine Learning technologies is still in its infancy in most fields of networking, it will become a much thought for opportunity to enhance network operations and performance in the coming years. This is mostly due to the rapid development of Machine Learning (ML) and associated Artificial Intelligence (AI) technologies in other fields.

ML/AI in picture/video/speech recognition as well as big-data analytics in areas such as e-commerce and search have evolved to a point where many of the methods and components of building solutions are well enough understood to apply them to novel fields - such as networking.

The ability of developers to rapidly build systems with ML/AI was vastly improved in the last few years through common tools such as TensorFlow that took most of the novel and unique complexity of building ML/AI solution into those expert built tools/libraries. The layers above those common libraries now become areas of development where more and more the subject matter experts (such as networking engineers) will be able to collaborate with data analysts to build those ML/AI solutions.

The performance of both AI/ML learning/training as well as the execution of the trained neural networks has been improved radically in the past years and it is expected lot more of these recent developments to proliferate into products.

GPUs (Graphic Processor Units) such as those from NVidia (as leader in the market) have evolved to be equal good high-performance parallel execution units for ML/AI training and inference. Algorithms to improve performance of execution by more than a factor of 1 000 have been developed in the past years.

Low-end ML/AI neural network inference hardware is now being released on products. Product means that these are hardware building block that can for example be added to existing low-end CPU chips such as ARM CPUs for cellphones and low-end network devices. This hardware can only execute neural networks (this is called inference), but not train those neural networks. These accelerators can do inference at minute fractions of the power needed in GPUs. The likely first big area where these will be used is speech recognition and translation on mobile phones.

# 6 Benefits of Introducing AI into Network

## 6.1 Towards Fully Autonomic Network

A fully autonomic network means that the network contains a closed-loop of "Measure-Analyse-Decide" which can realize the whole process autonomically. By means of AI-based learning and optimization techniques, the goal of IDN architecture is to learn about its behaviour, the fundamental relation between traffic load, network configuration and the resulting performance, understand the target policy set by the network administration, and configure that policy efficiently and fully autonomously. The advantage of a fully autonomic network is realizing the closed-loop of "Measure-Analyse-Decide", which will minimize the requirements for human administrators.

Currently, the process of measure, analysis and decision are mainly independent and the cooperation of such processes typically relies on humans. The limitation is caused by the lack of analysis ability of network, which is precisely AI technology performs really well. While in operation, the IDN architecture will react autonomously to relevant events (e.g. a failure, a spike in the traffic, etc.) and change the configuration accordingly. The core of AI technology is extracting the patterns (or knowledge) from complex data, in other words, discovering the rules then applying. In current, the forwarding process has achieved stateless or stateful full automation while most of the controlling process, such as the configuration and optimization, are still manual. The roadmap should be gradual, which starts from the local area autonomic to large area and finally to global. As if the development of self-driving, the automatic transformation is realized step by step, from such as auto-shift and auto-break. As yet, the AI technology is the one of the most possible ways to realize the whole process. During the development, the introducing of the AI technology gradually implement the closed-loop of network controlling so that reduce the unnecessary manual operation including coding, configuring, simple inference, etc. A fully autonomic network potentially decreases the cost of carriers during management and control. It will be benefit for the income in the long term.



## 6.2 Response to the challenge of complexity

AI and most notably Machine Learning (ML) techniques play a central role in the future architecture of networks. By means of ML mechanisms, the network behaviour can be learnt to obtain a ML-based model. This model can account for any arbitrary network characteristic of interest. As examples the models can characterize the energy consumption of the network or understand the relation between the traffic load and external factors such as popular sports events.

Traditionally network modelling has been done by means of simulation, however ML provides many advantages in this regard. First, ML scales very well with complexity and it is able to understand and model non-linear (complex) issues, indeed deep neural networks are able to account for multi-dimensional non-linear problems. On the contrary, simulations require costly development to model complex behaviour. Second, although training the neural networks is a CPU/GPU intensive process, once trained the neural network is very lightweight and fast, actually just a multi-dimensional function. However, both developing and running simulations is a costly process. This is relevant particularly when using network simulations to optimize the network performance, since each run is CPU intensive. And third, ML is able to understand the network (or parts of it) as a black-box and model behaviours even in the presence of hidden information. This provides important advantages in the simplification of the measurement process, since even in the presence of uncompleted information, ML can produce efficient models. On the other side, simulation cannot work with hidden information.

Along with the growth of scale, very large scale networks become unmanageable without intelligence. Because it is hard to formulate or design a rule universal for all.

Take the allocation of link resource as an example. According to the calculation with a traditional analytical model, the optimal solution for a specific question can be obtained with a long time computing. More often, the computing time is perhaps longer than transmitting the data via the worst link, which is meaningless. The Machine Learning (ML) based method can use the trained model and quickly output an approximate optimal solution according to various network parameters as input. This operation can continually adjust and optimize the solution. In general, any questions can be calculated and obtain a theoretical optimal solution by accurate and comprehensive calculation. However, the cost may be far smaller than the benefit, for example, the time for calculating an optimal route for a flow maybe far longer than the delay detouring few of congesting links. The birth of ML method is aiming to solve the complex issues according to analysing the huge amount of historic data and extracting the hidden rules behind the issues. The complexity actually lies on two aspects: the complexity of data to handle in terms of volume, heterogeneity, accessibility and even veracity and the complexity of problems to solve which are nowadays multi-faceted and try to accommodate multiple and sometimes partially antagonist objectives. The traditional network modelling and optimization techniques may perform unsatisfactorily since they are not inherently designed for such big data scenario. However, the introducing of AI technology potentially becomes a sharp weapon to deal with the complex network issues.

## 6.3 Response to the challenge of variation

The AI technology brings learning-based adaptability and flexibility. Actually, the machine learning based algorithm upgrade the controlling logic more adaptive. Comparing with experienced parameters from expert, the result of ML is more flexible, especially in real-time operation of the network. The supervised training can be used in decision-making or classification problems while the unsupervised training is good at extracting the patterns that might be hard for human to find. Meanwhile, the experienced parameters (also including decisions or policies) that obtained from experts is replaced with adaptive parameters which are controlled and updated by learning algorithm. The ML based intelligence technology can make the policy flexible. According to the specific environment, the administrator or managing system can train its own model so that to satisfy the diversity. Meanwhile, whether there are any changes in the hardware layer of the network, the requirement of user and service will change always, especially behaving in traffic character. Base on the machine learning algorithm, the intelligent system can capture data (both network data and content data) in real-time and then obtain the feature of the network by distributed or centralized training process. Finally, it will modify the parameters to match the requirement of current service. The adaptive ability makes network devices configuration match the distribution character of traffic and service so that to utilize the bandwidth flexible.

Not only to adapt to the change of network, the AI technology but also potentially bring more powerful ability to modify the network. For example, in the scenario that virtual topology is defined by software, the AI technology can control and modify the virtual network topology so that to satisfy the traffic change and user requirement. The under layer technology can be provided by slicing. Due to the introducing of AI technology, the network will be not only static but also dynamic, which enables more abilities for the network. Elasticity of networks promoted by SDN and NFV can be thus fully enabled by AI-based decisions.

## 6.4 Insights of the Network and Improve the Utilization

The rapid construction and change of the network makes itself complex than ever. The network is developing to a dynamically changing black box for the carriers. It is more and more difficult to learn what, where and when the fault happened, and eventually trace the root causes and responsible entities. This is paramount of importance since defining an appropriate responses, such as a counter-measures, would necessitate the most specific characterization (of the fault) to be efficient. The AI technologies might be helpful to this question. By data analysing and visualization, the intelligent system may help administrator to monitor and analyse the internal events and try to make them visual and comprehensible. Along with the increasing of network scale, the network administrator feels puzzled with the abnormal behaviour. Such a visualized system can help administrator to monitor, analyse and understand the internal events and even their relations with external indicators, which will improve the network insights dramatically.

The operation always faces to the fundamental trade-off between utilization and stability. Face to the uncontrollable and complex environment, because of the lack of methods for uncertain and latent elements, big amount of physical resource needs to be reserved to prevent the uncertain event happened. The redundancy depends on experience and risk control strategy. For example, for ensuring the service level agreement, carriers usually reserve an excess of bandwidth resource which may far larger than its real demand to prevent the accident. In other word, exchange utilization for stability. The intelligent system can deeply analyse the network in foundation level and utilize the knowledge to implement refine control and management. The learned knowledge is helpful for mining the latent factors that influence the network so that decrease the uncontrollability, which may help the administrator or decision system avoiding excessive waste. Along with the increasing of network scale, the intelligent system can essentially save the physical resource and improve the utilization of network and the mix between extending the network capacity with new dedicated hardware or accessing to shared cloud-based infrastructures. As an example, defeating link-flooding attacks can rely on increasing link capacity or buying a cloud-specific service, each of these solutions having their own costs depending on their utilization.

## 6.5 To Be Predictive

Prediction, may be the most important attribute that AI brings since it enables proactive behaviour. According to the model learnt from the network data, the prediction-based methods tend to foresee the evolution, such as the traffic load, link congestion, failure and all kinds of factors in advance. Take the risk of failure as an example. Whatever how precise and smart of calculation, including human brains, the deployed policy will face to the risk of failure. What is worse, the increasing number of network policies and their overlap will aggravate the risk. The predictability of network (e.g. the prediction of traffic) will enable the evaluation ability of network policy for network management system or decision system. By sensing the network state, the AI algorithm can derive the development of network and forecast the potential problems, such as traffic peak, congestion, device failure, etc. This will bring significant change for network management. The network administrator will always like to predict the problems and pre-process them, which will reduce the fault rate and save the repairing cost.

Another benefit of AI is that it can model the network and provide predictions of the performance before applying a particular configuration. This provides many advantages since configurations that reduce the performance and/or are unreliable can be avoided before applying it onto the real infrastructure.

ML methods have many applications in the field of prediction, since they are well-suited to model the dependencies of multi-dimensional non-linear behaviours. In this context, ML can help predicting important network parameters. Three relevant examples are listed following:

- Traffic characteristics. An ML-based algorithm can predict the traffic load in the network at different scales. First, different spaces can be considered: by individual links, by individual nodes (routers/switches), by end-point pairs (traffic matrices), by autonomous systems pairs, etc. The time dimension also affects the granularity (per hour, week, month, etc.). This can help preparing beforehand the network to account for such load and providing better service. With long-term prediction, this can help network operators to anticipate when the load will exceed the capacity and plan ahead future network hardware upgrades. This prediction can be based on external events, such as the weather, popular sports events, time-of-day, the load of certain services, etc. The same reasoning applies for other traffic characteristics such as latency or jitter.

- Failures or attacks. It is impossible to propose a system (hardware or software) to work without being faced with a failure or an attack. Predicting them is a valuable information to actually prepare proactive-plan by increasing redundancy among network equipment. For instance, predictive maintenance is thus also helpful to prepare hardware replacement in network. From a security perspective, analysing external indicators can be used to focus analysis on certain part of the network or services and thus do predictive security by adjusting in real time the security configurations.

- Prediction can also play a central role in low-level configurations parameters of the data-plane, for instance weights of routing protocols can be predicted to achieve optimal performance.