Recommendation

ITU-T Y.3142 (04/2024)

SERIES Y: Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities

Future networks

Requirements and framework for AI/ML-based network design optimization in future networks including IMT-2020

ITU-T Y-SERIES RECOMMENDATIONS

Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities

GLOBAL INFORMATION INFRASTRUCTURE	Y.100-Y.999
General	Y.100-Y.199
Services, applications and middleware	Y.200-Y.299
Network aspects	Y.300-Y.399
Interfaces and protocols	Y.400-Y.499
Numbering, addressing and naming	Y.500-Y.599
Operation, administration and maintenance	Y.600-Y.699
Security	Y.700-Y.799
Performances	Y.800-Y.899
INTERNET PROTOCOL ASPECTS	Y.1000-Y.1999
General	Y.1000-Y.1099
Services and applications	Y.1100-Y.1199
Architecture, access, network capabilities and resource management	Y.1200-Y.1299
Transport	Y.1300-Y.1399
Interworking	Y.1400-Y.1499
Quality of service and network performance	Y.1500-Y.1599
Signalling	Y.1600-Y.1699
Operation, administration and maintenance	Y.1700-Y.1799
Charging	Y.1800-Y.1899
IPTV over NGN	Y.1900-Y.1999
NEXT GENERATION NETWORKS	Y.2000-Y.2999
Frameworks and functional architecture models	Y.2000-Y.2099
Quality of Service and performance	Y.2100-Y.2199
Service aspects: Service capabilities and service architecture	Y.2200-Y.2249
Service aspects: Interoperability of services and networks in NGN	Y.2250-Y.2299
Enhancements to NGN	Y.2300-Y.2399
Network management	Y.2400-Y.2499
Computing power networks	Y.2500-Y.2599
Packet-based Networks	Y.2600-Y.2699
Security	Y.2700-Y.2799
Generalized mobility	Y.2800-Y.2899
Carrier grade open environment	Y.2900-Y.2999
FUTURE NETWORKS	**Y.3000-Y.3499**
CLOUD COMPUTING	Y.3500-Y.3599
BIG DATA	Y.3600-Y.3799
QUANTUM KEY DISTRIBUTION NETWORKS	Y.3800-Y.3999
INTERNET OF THINGS AND SMART CITIES AND COMMUNITIES	Y.4000-Y.4999
General	Y.4000-Y.4049
Definitions and terminologies	Y.4050-Y.4099
Requirements and use cases	Y.4100-Y.4249
Infrastructure, connectivity and networks	Y.4250-Y.4399
Frameworks, architectures and protocols	Y.4400-Y.4549
Services, applications, computation and data processing	Y.4550-Y.4699
Management, control and performance	Y.4700-Y.4799
Identification and security	Y.4800-Y.4899
Evaluation and assessment	Y.4900-Y.4999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T Y.3142

Requirements and framework for AI/ML-based network design optimization in future networks including IMT-2020

Summary

Due to the development of IMT-2020 networks, a large number of new applications and services have emerged, leading to explosive growth in network traffic and much more complicated networks, which results in increased difficulty in attaining network design optimization. Many services are changing rapidly, and network operators need to dynamically design and adjust their networks based on the traffic distribution and the traffic increment trends in order to guarantee the necessary quality of service (QoS). Artificial intelligence (AI) can be used for prediction and decision-making, which can make network design more intelligent and automated.

Recommendation ITU-T Y.3142 focuses on the use of artificial intelligence (AI)/machine learning (ML) technologies to improve network design mechanisms. This Recommendation describes how AI/ML can be integrated in order to optimize the design of network capacity, network topology, and routing to satisfy all of the demands of service level agreements (SLAs) in a cost-effective way, instead of only guaranteeing the SLAs without taking into account the overall cost.

History *

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T Y.3142	2024-04-29	13	11.1002/1000/15869

Keywords

AI, future networks, IMT-2020, ML, network design, optimization.

* To access the Recommendation, type the URL https://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T Y.3142

Requirements and framework for AI/ML-based network design optimization in future networks including IMT-2020

1 Scope

This Recommendation considers the comprehensive application of artificial intelligence (AI)/ machine learning (ML) technologies in network design optimization, including network topology, network capacity, and routing.

It specifies functional requirements, framework and process for AI/ML-based network design optimization in IMT-2020 networks and beyond.

The scope of this Recommendation includes:

– Requirements for the framework of AI/ML-based network design optimization;

– AI/ML-based network design optimization framework and its components;

– AI/ML-based network design optimization process.

Appendix I provides examples of network design optimization.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T Y.3102] Recommendation ITU-T Y.3102 (2018), *Framework of the IMT-2020 network.*

[ITU-T Y.3172] Recommendation ITU-T Y.3172 (2019), *Architectural framework for machine learning in future networks including IMT-2020.*

[ITU-T Y.3174] Recommendation ITU-T Y.3174 (2020), *Framework for data handling to enable machine learning in future networks including IMT-2020.*

[ITU-T Y.3181] Recommendation ITU-T Y.3181 (2022), *Architectural framework for machine learning sandbox in future networks including IMT-2020.*

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 machine learning (ML) [ITU-T Y.3172]: Processes that enable computational systems to understand data and gain knowledge from it without necessarily being explicitly programmed.

3.1.2 machine learning function orchestrator (MLFO) [ITU-T Y.3172]: A logical orchestrator that can monitor and manage the nodes in a machine learning pipeline.

3.1.3 machine learning pipeline [ITU-T Y.3172]: A set of logical nodes, each with specific functionalities, that can be combined to form a machine learning application in a telecommunication network.

NOTE – The nodes are entities that are managed in a standard manner and can be hosted in a variety of network functions.

3.1.4 machine learning sandbox [ITU-T Y.3172]: An environment in which machine learning models can be trained, verified and their effects on the network analysed.

NOTE – A machine learning sandbox is designed to prevent a machine learning application from affecting the network, or to restrict the usage of certain machine learning functionalities.

3.1.5 machine learning model [ITU-T Y.3172]: Model created by applying machine learning techniques to data to learn from.

NOTE 1 – A machine learning model is used to generate predictions on new (untrained) data.

NOTE 2 – A machine learning model may be encapsulated in a deployable fashion in the form of a software or hardware component.

NOTE 3 – Machine learning techniques include learning algorithms (e.g., learning the function that maps input data attributes to output data).

3.1.6 machine learning underlay network [ITU-T Y.3172]: A telecommunication network and its related network functions which interfaces with corresponding machine learning overlays.

NOTE – An IMT-2020 network is an example of a machine learning underlay network.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AI	Artificial Intelligence
CAPEX	Capital Expenditure
IPFIX	IP Flow Information Export
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
ML	Machine Learning
MLFO	Machine Learning Function Orchestrator
OPEX	Operational Expenditure
QoS	Quality of Service
SLA	Service Level Agreement
SNMP	Simple Network Management Protocol
XML	Extensible Markup Language

5 Conventions

In this Recommendation:

The keywords "is required to" indicate a requirement which must be strictly followed and from which no deviation is permitted, if conformance to this Recommendation is to be claimed.

The keywords "is recommended" indicate a requirement which is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

The keywords "can optionally" indicate an optional requirement which is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option, and the feature can be optionally enabled by the network operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with this Recommendation.

6 Overview

A large number of new applications and services have emerged, leading to explosive growth in network traffic and much more complicated networks. Many services are changing rapidly, and network operators need to dynamically design and adjust their networks based on the traffic distribution and the traffic increment trends, in order to guarantee quality of service (QoS). In addition, construction and operating costs are areas that network operators are very concerned about and so cost-effective network design is a high-level priority. AI/ML-based network design optimization can help network operators to update their networks and schedule services efficiently, thus improving network utilization and reducing capital expenditure (CAPEX) and operational expenditure (OPEX).

To conduct network design optimization, network operators need to accurately perceive the network status and identify possible bottlenecks in the network. The use of AI/ML is viewed as beneficial in processing the large amounts of data necessary for extraction of service requirements, analysis of traffic trends and identification of network bottlenecks. Moreover, AI/ML can be used to make decisions concerning changes in the network topology, network capacity and service routing.

A number of Recommendations consider the application of AI/ML in networks:

– [ITU-T Y.3172] provides a high-level architectural framework and specifies generic requirements.

– [ITU-T Y.3174] describes a framework for data handling to enable ML in future networks and also identifies generic requirements regarding data collection, processing mechanisms and ML output.

– [ITU-T Y.3181] describes requirements and high-level architecture for ML sandbox.

These three Recommendations are high-level and generic with respect to AI/ML application in future networks, and can serve as references. Some generic requirements and components from the above-mentioned Recommendations are applicable to this Recommendation. In this Recommendation, the focus is on network design optimization and thus relevant requirements, framework components and procedures are specified to address the challenges of network design optimization.

In the framework described in this Recommendation, AI/ML is used to process service requirements data, traffic data and network key performance indicator (KPI) data, to infer traffic trends and KPI changes, to evaluate network status, and to globally optimize network topology, network capacity and service routing. By means of multiple AI/ML applications enabled by this framework in order to deal with the complexity and dynamics of future networks, higher level automation for network design optimization can be achieved. Network performance can be enhanced, and service level agreements (SLAs) can be satisfied with low cost by leveraging AI/ML to analyse data and make intelligent decisions for network design. AI/ML-based network design optimization makes it possible to optimize the utilization of network resources leading to cost reductions regarding network upgrades or expansions.

7 Requirements of AI/ML-based network design optimization

Network design optimization involves numerous aspects and necessitates accurate data on service requirements and network status. It involves effectively predicting future network conditions, analysing potential bottlenecks, and balancing diverse optimization objectives. As a result, it demands requirements for data collection and processing, traffic prediction, KPI prediction, network analysis and policy optimization. These requirements are presented in clauses 7.1 to 7.6.

7.1 Data collection

REQ-DC-01: The data collection is required to collect network data from multiple domains and sources [ITU-T Y.3174], e.g., network traffic, connection related data, network topology, link bandwidth, link utilization, link delay and network fault data.

REQ-DC-02: The data collection is required to collect the SLAs requirements of the services (service data) from users, e.g., bandwidth requirement, delay requirement, reliability guarantee, etc.

REQ-DC-03: The data collection is recommended to support service-related data collection, including service types (e.g., video, games, social interaction, etc.), real-time attributes (e.g., real-time, non real-time), and other traffic characteristics.

REQ-DC-04: The data collection is recommended to support multiple data collection methods, e.g., simple network management protocol (SNMP) [b-SNMP], IP flow information export (IPFIX) [b-IPFIX].

REQ-DC-05: The data collection is recommended to support multiple data collection formats, e.g., extensible markup language (XML) [b-XML], JavaScript object notation (JSON) [b-JSON].

REQ-DC-06: The data collection is recommended to support highly efficient data collection methods (e.g., using AI based techniques) to realize trade-offs between accuracy and overheads of collection.

REQ-DC-07: The data collection is required to interact with an ML simulation network environment, e.g., ML sandbox [ITU-T Y.3181], to collect relevant simulation data [ITU-T Y.3174].

REQ-DC-08: The data collection is recommended to collect cost related data, e.g., the cost of expanding link capacity, the cost of adding a new link, etc.

REQ-DC-09: The data collection is required to collect related data with adaptive granularity, ensuring that the collected data align with specific needs, e.g., in short-term traffic prediction, data with granularity at minute level may be needed.

REQ-DC-10: The data collection is required to support multiple data collection interfaces, including smartphones, wearable sensors, Internet-based data sources, etc.

REQ-DC-11: The data collection can optionally support the third parties' applications for collecting the data related to the connectivity performance of the applications, e.g., delay, jitter, and packet loss.

NOTE 1 – Third parties may use data to execute AI/ML analysis and learn the relationship between connectivity performance and network status, including for possible support to operators in order to facilitate the network design.

REQ-DC-12: The data collection is recommended to support synthetic data generation and integration if there is a lack of data in the actual data collection operations.

REQ-DC-13: The data collection is required to provide security and privacy features in order to protect sensitive information and ensure the integrity of the collected data.

NOTE 2 – This includes support of encryption mechanisms for sensitive data to ensure confidentiality and integrity.

7.2 Data processing

REQ-DP-01: The data processing is required to process data by fully satisfying all the requirements specified in [ITU-T Y.3172] and [ITU-T Y.3174].

REQ-DP-02: The data processing is required to aggregate related data with adaptive granularity, e.g., in short-term traffic prediction, data with granularity at minute level may be needed.

REQ-DP-03: The data processing is required to support continuously re-starting the AI process over time to consider major changes in data.

REQ-DP-04: The data processing is recommended to support data timestamping.

REQ-DP-05: The data processing is required to support AI/ML-based techniques for data privacy and security while still meeting the required performance in future networks.

REQ-DP-06: The data processing is required to support AI techniques that can overcome faults and failures quickly, thereby ensuring system resilience and service continuity.

REQ-DP-07: The data processing is recommended to be capable of connecting to any network function and utilizing any network function data.

REQ-DP-08: The data processing is recommended to be hosted at different levels (e.g., user level, access level, and core level) in the network.

7.3 Traffic prediction

Accurate and reasonable network design optimization requires forecasting of traffic changes in advance. The output of the traffic forecast is one of the inputs to the network design. The temporal and spatial correlation between traffic flows should be considered and used to further improve the prediction accuracy.

REQ-TP-01: The traffic prediction is required to have prediction capability for different horizons, e.g., short-term prediction and long-term prediction.

REQ-TP-02: The traffic prediction is required to have the capability of spatiotemporal correlation analysis in order to extract spatiotemporal characteristics of traffic data.

REQ-TP-03: The traffic prediction is recommended to have the capability of multiple AI prediction algorithms in order to enable multi-model integration.

NOTE 1 – Multi-model integration means integrating multiple different models or algorithms, e.g., tree-base algorithms, recurrent neural network, etc., in order to achieve better prediction accuracy.

REQ-TP-04: The traffic prediction is required to have prediction capability for the different network elements, e.g., the traffic of a link, the traffic matrix of all network nodes, etc.

REQ-TP-05: The traffic prediction is required to have prediction capability for different services and applications.

NOTE 2 – This can allow identification of which services and applications are heavily consuming network resources and potentially causing congestion or security risks.

REQ-TP-06: The traffic prediction is required to be validated and tested in order to improve its accuracy and reliability.

NOTE 3 – Feedback loop [b-Ian] is essential for improving prediction models.

7.4 KPI prediction

AI/ML should allow for the prediction of KPIs if the network configuration stays unchanged. Such prediction will allow proactive reconfigurations before the KPIs fall below the assumed level. The KPI prediction may take as input not only the actual or historical data, but also the network traffic prediction.

NOTE – Typical KPIs are delay, packet loss, jitter, utilization, and throughput.

REQ-KP-01: The KPI prediction is recommended to have prediction capability for different services.

REQ-KP-02: The KPI prediction is recommended to have prediction capability for different network elements, e.g., a link's KPI, a network device's KPI and a sub network's KPI.

REQ-KP-03: The KPI prediction is required to have prediction capability for different timing durations.

REQ-KP-04: The KPI prediction is recommended to have the capability of spatiotemporal correlation analysis in order to extract spatiotemporal characteristics of multiple KPIs.

7.5 Network analysis

Network design involves multiple domains and layers. Before conducting specific design optimization under given conditions, it is necessary to comprehensively perceive the network status, including traffic status, to find out the existing or potential problems of the running network and identify which domains, and/or which aspects (e.g., network topology, network capacity, routing), of the network should be optimized. On the other hand, it is necessary to take possible failure scenarios into account to make the network robust.

REQ-NA-01: The network analysis is required to analyse the traffic and KPI changes in the network and identify whether it is necessary to conduct network design optimization.

REQ-NA-02: The network analysis is recommended to have the capability to correlate traffic and KPIs to find out some correlations.

REQ-NA-03: The network analysis is required to have the capability to identify the network domains or network aspects to be optimized.

REQ-NA-04: The network analysis is required to have the capability to take into account typical failure scenarios in network design optimization.

REQ-NA-05: The network analysis is required to have the capability to conduct security analysis in order to identify vulnerabilities, anomalies, and potential security threats.

REQ-NA-06: The network analysis is required to have the ability to integrate forecasting results in order to determine the overall traffic demand.

REQ-NA-07: The network analysis is required to have the ability to determine network design optimization objectives. That is, based on the prediction results, the SLAs of services and the network performance, the network analysis is required to have the ability to identify the priority of the optimization objectives.

7.6 Policy optimization

Beyond the prediction of traffic changes, AI technologies can be used to analyse and evaluate existing network resources, analyse the targets to be optimized, and make overall optimal network design decisions, such as determining which links' capacity needs to be expanded or which links need to be added.

Optimization policies can be obtained by AI/ML models based on the related network and service data, prediction results, objectives, and constraints. Optimization policies include specific network updates to improve the whole network performance and service quality, e.g., which links' capacity should be expanded, how to modify the network topology, and/or how to change the traffic routing.

Optimization policies can be used to modify the AI/ML-underlay network or change traffic routing in order to achieve better performance in terms of cost, utilization, delay, etc. For example, the delay of a given service traffic might be reduced by adding a new link in the current network, or the link utilization might be improved without causing congestion by appropriately adjusting the traffic routing. Basically, from the whole network perspective, policy optimization involves optimization on the network topology, network capacity, and traffic routing.

REQ-PO-01: The policy optimization is required to have the ability to access the corresponding traffic prediction results.

REQ-PO-02: The policy optimization is required to have the ability to access network data, e.g., network topology, link bandwidth, link utilization, link delay, link jitter, and network fault data.

REQ-PO-03: The policy optimization is required to allow optimization objectives be specified by network operators and users be able to participate in defining objectives in some scenarios.

REQ-PO-04: The policy optimization is required to have the capability of multiple AI solving techniques, such as reinforcement learning, graph neural networks, generative adversarial networks, etc.

REQ-PO-05: The policy optimization is required to support different mechanisms of network design optimization, such as design optimization of topology, capacity, and routing.

REQ-PO-06: The policy optimization is required to have the ability of multi-objective optimization, such as joint optimization of multiple KPIs, e.g., delay, link utilization and packet loss, etc.

REQ-PO-07: The policy optimization is required to have logical network design and physical network design capability, and have the overall optimization capability. There exists a mapping relationship between logical networks and physical networks. To optimize the overall network design, both optimization for logical network and physical network should be considered simultaneously.

REQ-PO-08: The policy optimization is required to keep the history of optimisation-related reconfiguration and use the fallback procedure if a new configuration leads to KPI degradation.

REQ-PO-09: The policy optimization is required to give the network operator access to reconfiguration and KPI data and allow the network operator to optimise the network manually.

REQ-PO-10: The policy optimization is required to have the ability to use the predicted KPIs to trigger the optimisation procedures.

REQ-PO-11: The policy optimization is required to have the ability to coordinate AI/ML optimization functions focused on different goals and resolve configuration conflicts if necessary (direct configuration parameter conflict or indirect via environment feedback).

REQ-PO-12: The policy optimization is required to have the ability to cooperate with security and fault management functions to avoid optimisation until the fault management function provides network recovery or security functions provide attack mitigation. An exception is the optimization procedure coordinated with the mentioned functions.

REQ-PO-13: The policy optimization is recommended to use explainable AI [b-Explainable AI] in support to network optimization.

REQ-PO-14: The policy optimization is recommended to automate the network management and configuration tasks, which in turn lowers the OPEX by reducing human-machine interaction.

8 Framework of AI/ML-based network design optimization

Based on the requirements specified in clause 7, this clause describes the framework and the related components of AI/ML-based network design optimization.

The framework is based on the high-level architecture for ML in future networks including IMT-2020 specified in [ITU-T Y.3172], and it is compatible with the data handling framework specified in [ITU-T Y.3174] and the ML sandbox framework specified in [ITU-T Y.3181].

Figure 1 shows the high-level framework for network design optimization with AI/ML.

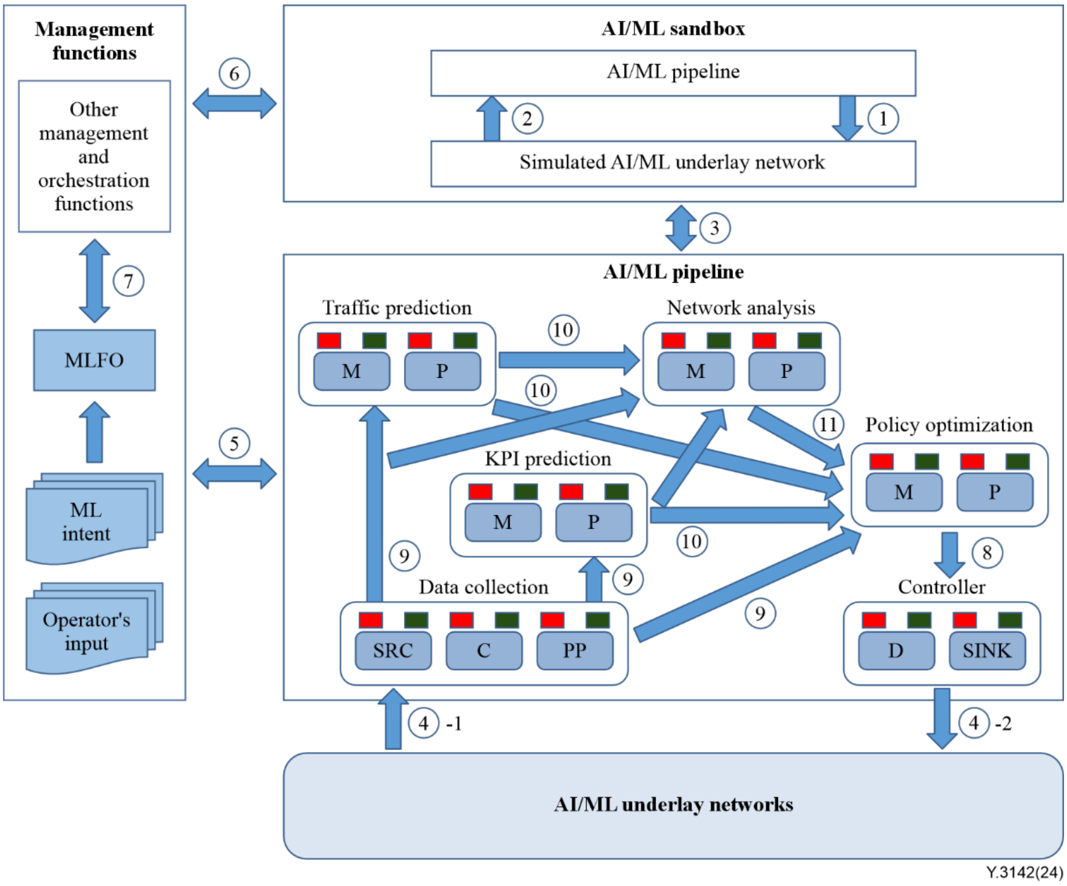

Figure 1 – Framework of AI/ML-based network design optimization

There are four subsystems in this framework, in alignment with [ITU-T Y.3172], including management subsystem, AI/ML sandbox, AI/ML pipeline, and AI/ML underlay networks.

In the AI/ML pipeline subsystem, there are several AI/ML pipeline components, as shown in Figure 1 and as described below, with dedicated functionalities for network design optimization. These specific AI/ML pipeline components are supported by basic nodes defined in [ITU-T Y.3172].

In the management subsystem, in addition to AI/ML intents, the machine learning function orchestrator (MLFO) takes operator's inputs into account. An operator may specify some objectives or constraints for design, e.g., the network design objective may be the minimization of cost and delay, or keeping some links' utilization below a certain threshold.

The reference points shown in Figure 1 are briefly described as follows:

– (1) and (2): These are the data handling reference points between simulated underlay networks and an AI/ML pipeline in the AI/ML sandbox subsystem. They are used to transfer relevant data or control information between simulated underlay networks and the AI/ML pipeline in the AI/ML sandbox subsystem.
– (3): This is the reference point between the AI/ML sandbox subsystem and the AI/ML pipeline subsystem. The data and AI/ML models can be transferred through this reference point.
– (4)-1 and (4)-2: These are the data handling reference points between AI/ML underlay networks and the AI/ML pipeline subsystem, and are similar to (1) and (2).
– (5) and (6): These are the reference points between the management subsystem, and the AI/ML pipeline subsystem and the AI/ML sandbox subsystem, respectively.
– (7): This is the reference point between the MLFO and other management and orchestration functions of the management subsystem.

NOTE – Reference points (1) to (7) are the same as specified in [ITU-T Y.3172].

– (8), (9), (10) and (11): These are the reference points between different nodes of the AI/ML pipeline with different components. (8) is used to transfer optimization policy from the policy optimization component to the controller component. (9) is used to transfer relevant data from the data collection component to traffic prediction, KPI prediction, and network analysis and policy optimization components. (10) and (11) are used to transfer one AI/ML model's output to another AI/ML model as input, i.e., the traffic prediction and KPI prediction components' outputs are transferred through (10), and the network analysis component's output is transferred through (11) to the policy optimization component, respectively.

The overall AI/ML-based network design optimization is based on various specific components in the AI/ML pipeline. The AI/ML pipeline in Figure 1 consists of six functional components: data collection; traffic prediction; KPI prediction; network analysis; policy optimization, and controller. Their functions are described below.

Each of the traffic prediction, KPI prediction, network analysis and policy optimization components can be considered as an AI/ML pipeline itself. They contain some basic nodes defined in [ITU-T Y.3172] but have customized functionalities for network design optimization. For example, the basic nodes named as "M" in [ITU-T Y.3172] are machine learning model nodes, while in the traffic prediction component of this framework, these nodes are specific machine learning models for time series prediction. The chaining of these AI/ML pipelines form an overall AI/ML pipeline to achieve overall network design optimization.

8.1 Data collection

This component collects related network data and service data from different domains and sources.

It processes the required data by data cleaning, data imputation, data normalization, data correlation etc., and provides the processed data to other components for training the AI/ML models.

8.2 Traffic prediction

This component uses AI algorithms, e.g., the long short-term memory model (a widely adopted model [b-LSTM]), to create a traffic prediction model. It uses data to train the AI/ML models from the data collection component. To achieve higher prediction accuracy, more complex and advanced algorithms can be used to extract the spatial-temporal correlation from traffic data. Based on the traffic prediction, the future traffic distribution and trend can be obtained, which can facilitate network design optimization.

Accurately predicting the future traffic of links is crucial in network design. There may exist spatio-temporal correlations in the traffic of different links. To improve the accuracy of traffic prediction for a specific link (target link), historical traffic data from multiple related links can be used for training. Links that exhibit a correlation with the target link's traffic greater than a predefined threshold are considered as correlated links, therefore, the correlated links for the target link can be determined and obtained. Based on the historical traffic of multiple time points of the target link and its correlated links, data sequences of historical traffic are generated, including data for training and data for prediction. The labels of these training data are the historical traffic value of the target link after one or more time points, and these data are used to train a neural network model corresponding to the target link. Once the training is completed, a predictive model for the target link is obtained. The data sequences of historical traffic intended for prediction are input into the trained model to obtain the future traffic for the target link.

8.3 KPI prediction

This component is similar to traffic prediction and also uses AI algorithms, e.g., the long short-term memory model, to create a KPI prediction model. The KPI data may involve different levels and different components, including services, devices, elements, and systems. This component uses related data from the data collection component in order to train the AI/ML models to achieve higher prediction accuracy, more complex and advanced algorithms can be used to extract the spatial-temporal relation from multiple KPI data and multiple KPI predictions can be simultaneously considered. Based on the KPI prediction, the future KPI variation can be obtained, which can be used as input to the network analysis component and the policy optimization component.

8.4 Network analysis

This component integrates the results and data from the components of data collection, traffic prediction and KPI prediction. It can comprehensively analyse traffic and KPIs, and assess the overall current network status. In addition, it provides the capability to identify the priority optimization objectives given the current network status, prediction results and service requirements. In addition, the network analysis component needs to take into account typical failure scenarios based on current traffic and network status.

8.5 Policy optimization

This component supports different kinds of capabilities of network design optimization.

This component employs AI/ML models, and uses the results of traffic and KPI predictions, network analysis, and other relevant data. It determines capacity, topology and routing, separately or simultaneously, in order to satisfy the SLA requirements of services at low cost from a whole network perspective. After model selection, training and evaluation, the optimization policy result is output to the controller component for deployment and implementation on the underlying network.

There are basically three types of optimization policy: capacity optimization, topology optimization, and routing optimization.

Capacity optimization: This optimization policy is inferred from AI/ML models in order to modify the specific network capacity, e.g., a link's capacity, a network element's capacity, etc.

Topology optimization: This optimization policy is inferred from AI/ML models in order to modify the specific network topology, e.g., adding or deleting a connection between two nodes.

Routing optimization: This optimization policy is inferred from AI/ML models in order to modify the routing for traffic, e.g., finding a new path for a traffic flow or adjusting the traffic split of a service between multiple paths.

8.6 Controller

The controller receives the optimization policy from the policy optimization component.

This component, supported by the distributor (D) and sink (S) nodes [ITU-T Y.3172], identifies the relevant sinks, distributes to them the optimization policy, and implements these policies on the sink nodes.

Basically, the controller uses optimization policies to adjust capacity, topology and routing. This may involve dynamically adjusting bandwidth allocation, optimizing data transfer paths, and reconfiguring the network topology.

The controller interacts with the underlay networks to implement the corresponding network modifications.

8.7 MLFO

The machine learning function orchestrator (MLFO) [ITU-T Y.3172] manages and orchestrates the AI/ML nodes in the AI/ML sandbox subsystem and in the AI/ML pipeline subsystem based on the AI/ML intents and operator's inputs. The operator' inputs support the operator to specify the priority for network design optimization or explicitly impose some constraints on network design optimization. MLFO manages and controls the phases of training, evaluation and deployment of the model. Model selection and reselection for each AI/ML pipeline is also managed by MLFO.

9 Network design optimization with AI/ML

As the complexity and dynamics of future networks and services are increasing, AI/ML is required to achieve a cost-effective network design solution to optimize network performance and satisfy services requirements. This clause describes the network design optimization with AI/ML. Network design optimization includes several important fundamental aspects, such as topology optimization, capacity design and routing planning. Figure 2 shows the network design optimization–related AI/ML pipeline components, which are data collection, traffic prediction, KPI prediction, network analysis, policy optimization and controller.

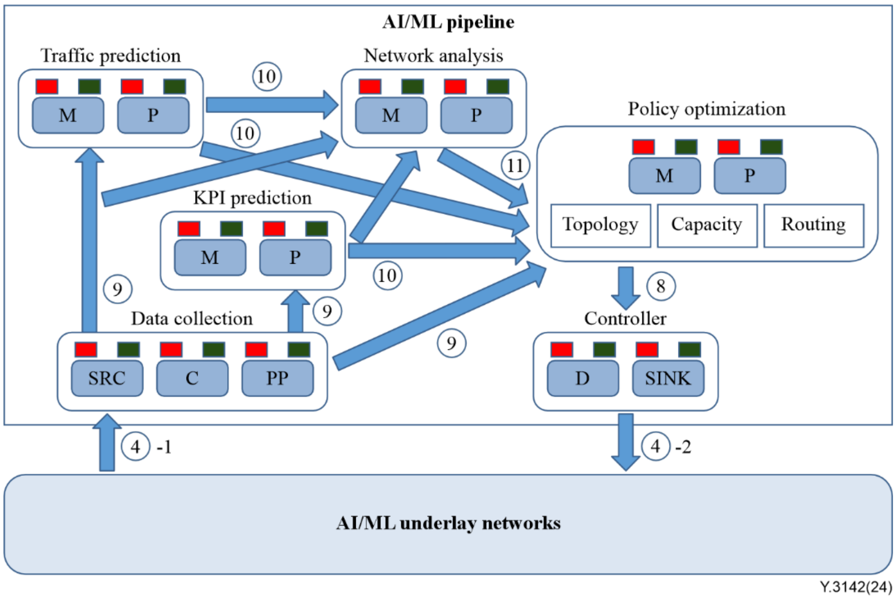

Figure 2 – Network design optimization with AI/ML

The data collection component receives traffic-related data, network elements, network links' status-related data and network resources-related data from the AI/ML underlay networks through point (4)-1. After data pre-processing, data processed in specified formats is provided to the traffic prediction component, KPI prediction component, network analysis component, and policy optimization component.

The traffic prediction function exploits AI/ML models to train the data to predict the future traffic trends and distribution. The traffic prediction can be implemented on different levels and timescales. For example, it may predict traffic status on links (link load) or between network nodes (traffic matrixes). It may also predict traffic status at the day's level or minute's level. The traffic prediction results are provided to the network analysis component through reference point (10) for further analysis.

The KPI prediction function is similar to the traffic prediction function, and also uses AI/ML models to train the data to predict the future KPI conditions. The KPI prediction can also be implemented on different levels and timescales. For example, it may predict KPIs for services, network elements, or links. The KPI prediction result is provided to the network analysis component through reference point (10) for further analysis.

The network analysis component needs to use AI/ML models to further perceive the traffic and KPI prediction results. It provides comprehensive analysis on the overall traffic and network performance and evaluates the network status. It can identify the network optimization priority based on these prediction results, which will facilitate the network design optimization. For example, correlating and analysing these predictions, it may figure out that the network needs to increase some link capacity and that it is not necessary to optimize the topology, therefore, the network design will focus on the capacity design and choose appropriate models. In addition, the network analysis component needs to take into account typical failure scenarios based on current traffic and network status. The network analysis component provides the output results to be used in the policy optimization component through reference point (11). The network analysis component may also receive data from the data collection component through reference point (9).

The policy optimization component uses AI/ML models to figure out how to update the network regarding topology, capacity and routing. It uses the results from data collection, traffic prediction, KPI prediction, and network analysis components. It finally figures out the optimal topology design, capacity design or routing planning. The optimization policy is sent to the controller through reference point 4-(2) and the controller executes updates or adjustments in the underlay networks according to the optimization policy.

Appendix I provides examples concerning specific types of network updates related to network design optimization.

10 Details on the AI/ML-based network design optimization process

AI/ML-based network design optimization involves training and coordinated processing of AI/ML models across multiple components. By predicting future traffic and KPI conditions, it identifies potential bottlenecks and determines the priority for optimization in the network based on service requirements and existing network capabilities. Ultimately, it derives an optimized network design solution that meets the specified constraints.

The basic sequential steps of the AI/ML-based network design optimization process are the following:

1. The data collection component collects, processes, and provides the required data to the traffic prediction, KPI prediction, network analysis and policy optimization components.
2. The traffic and KPI prediction components choose appropriate AI/ML models to train the data and predict the traffic distribution and trends, and potential KPI changes, in the next prediction interval.
3. Based on the operator's inputs, network status, traffic and KPI forecast results, the network analysis component uses AI/ML models to evaluate the existing network globally, take into account typical failure scenarios, and determine the priority of policy optimization.
4. Based on traffic prediction, KPI prediction, network analysis results and other relevant data, the policy optimization component uses AI/ML models to determine the capacity, topology and routing to meet SLAs at low cost. The policy optimization component outputs the optimization policy to the controller to implement the modifications on the underlay networks accordingly.

11 Security considerations

This Recommendation describes requirements and framework for AI/ML-based network design optimization on the basis of the high-level architecture of machine learning in future networks including IMT-2020 [ITU-T Y.3172].

The security considerations of [ITU-T Y.3172] are also applicable to this Recommendation.

On the other hand, this Recommendation also involves AI/ML sandbox [ITU-T Y.3181] and data handing [ITU-T Y.3174], and the security considerations related to AI/ML sandbox and data handing should also be taken into account.

Appendix I

Network design optimization examples

(This appendix does not form an integral part of this Recommendation.)

The operator needs to monitor the network traffic and network status, and, based on the service requirements, identify appropriate network updates on the underlay networks accordingly. There are three main types of network updates related to network design optimization: topology optimization, capacity optimization, and routing optimization.

I.1 Topology optimization

Topology optimization means changing the connection relations between network elements. The topology structure impacts many aspects of network performances, such as link utilization, throughput and latency. For example, an unreasonable topology design may lead to uneven traffic load. Therefore, in order to achieve traffic load balancing, it may be needed to add some new links between network elements in order to split some traffic. Based on the traffic and KPI prediction results, and other constraints or data from the network analysis component, an AI/ML model is used to find out the optimized topology. In some cases, it is not possible to improve the network performances without changing the network topology. However, due to complex network connections and status, it might be challenging to appropriately implement the changes. Therefore, AI/ML can help to make an intelligent decision on which links should be added or deleted. In topology optimization, the capacity of links and routing of traffic may also be reconfigured.

Figure I.1 illustrates a topology optimization example on a simplified network topology. Based on the traffic distribution, network status, service requirements, prediction results and network analysis results, AI/ML identifies that a new link needs to be added between the network nodes A and E.

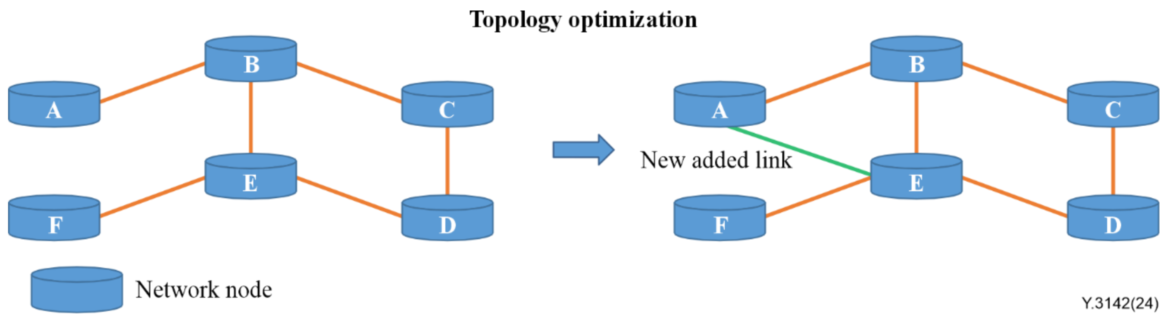

Figure I.1 – Topology optimization example

I.2 Capacity optimization

In capacity optimization, there are no changes in the connection relations. AI/ML is used to figure out the optimization policy for the capacity of links and routing of traffic. Figure I.2 illustrates a capacity optimization example on a simplified network topology. Based on the traffic distribution, network status, service requirements, prediction results and network analysis results, AI/ML identifies that the capacity of various links needs to be increased.

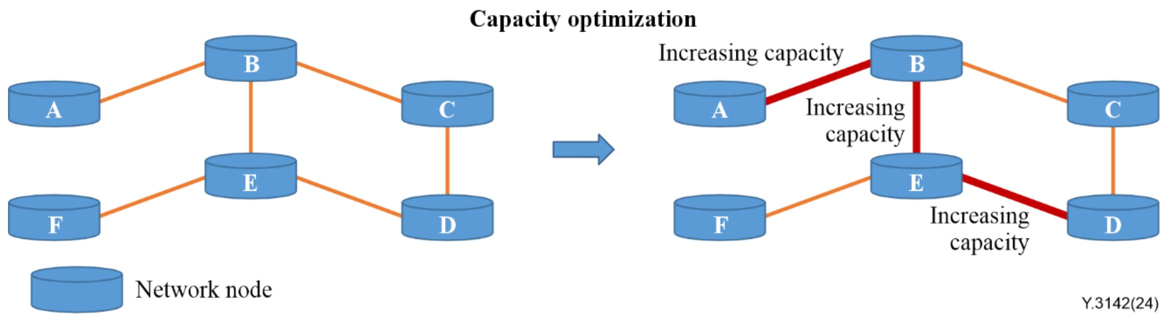

Figure I.2 – Capacity optimization example

I.3 Routing optimization

In routing optimization, the topology and capacity of the network remain unchanged. AI/ML is used to figure out the optimization policy routing of traffic. Figure I.3 illustrates a routing optimization example on a simplified network topology. Based on the traffic distribution, network status, service requirements, prediction results and network analysis results, AI/ML identifies that the original path for the traffic needs to be changed to a new path.

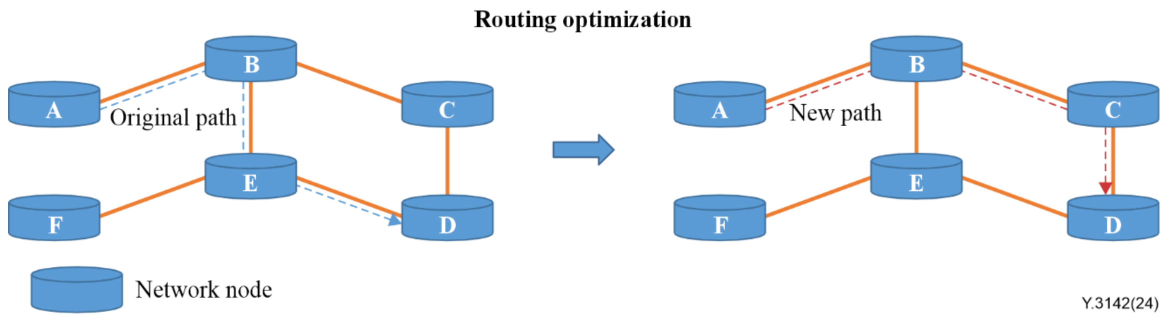

Figure I.3 – Routing optimization example

Bibliography

[b-ITU-ML5G-PS-001] ITU-ML5G-PS-001 (2021), *Graph Neural Networking Challenge 2021 – Creating a Scalable Network Digital Twin*, https://challenge.aiforgood.itu.int/match/matchitem/31

[b-Explainable AI] Ericsson White Paper, April 2021, *Explainable AI – how humans can trust AI*, https://www.ericsson.com/en/reports-and-papers/white-papers/explainable-ai--how-humans-can-trust-ai

[b-Ian] Ian Goodfellow, Yoshua Bengio, Aaron Courville (2016), *Deep Learning*, MIT Press.

[b-IPFIX] IETF RFC 7011 (2013) *Specification of the IP flow information export (IPFIX) protocol for the exchange of flow information.*

[b-JSON] IETF RFC 8259 (2017), *The JavaScript Object Notation (JSON) Data Interchange Format (JSON).*

[b-LSTM] Y. Hua, et at., *Deep Learning with Long Short-Term Memory for Time Series Prediction*. IEEE Communications Magazine, vol. 57, no. 6, pp. 114-119, 2019.

[b-SNMP] IETF RFC 1098 (1989). *Simple network management protocol (SNMP).*

[b-XML] W3C Recommendation (2008), *Extensible Markup Language (XML) 1.0 (Fifth Edition).*

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	**Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities**
Series Z	Languages and general software aspects for telecommunication systems

Published in Switzerland
Geneva, 2024