ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

Recommendation

ITU-T Y.3141 (12/2023)

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

Future networks

Energy efficiency management of virtual resources in IMT-2020 networks and beyond

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.3141

Energy efficiency management of virtual resources in IMT-2020 networks and beyond

Summary

Recommendation ITU-T Y.3141 specifies energy efficiency management (EEM) assisted by artificial intelligence of virtual resources in IMT-2020 networks and beyond, including:

- functional requirements;
- an architectural model;
- reference points;
- procedures.

History *

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T Y.3141	2023-12-14	13	11.1002/1000/15740

Keywords

AI-assisted, energy efficiency, virtual resources.

* 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.3141

Energy efficiency management of virtual resources in IMT-2020 networks and beyond

1 Scope

This Recommendation specifies energy efficiency management assisted by artificial intelligence (AI) of virtual resources in IMT-2020 networks and beyond, including:

- functional requirements;
- an architectural model;
- reference points;
- procedures.

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.3111] Recommendation ITU-T Y.3111 (2017), *IMT-2020 network management and orchestration framework.*

[ITU-T Y.3150] Recommendation ITU-T Y.3150 (2020), *High-level technical characteristics of network softwarization for IMT-2020.*

[ITU-T Y.3156] Recommendation ITU-T Y.3156 (2020), *Framework of network slicing with AI-assisted analysis in IMT-2020 networks.*

[ITU-T Y.3179] Recommendation ITU-T Y.3179 (2021), *Architectural framework for machine learning model serving in future networks including IMT-2020.*

[ETSI ES 203 228] ETSI Standard ETSI ES 203 228 V1.4.1 (2022), *Environmental engineering (EE); Assessment of mobile network energy efficiency.*

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 virtual resource [b-ITU-T Y.3011]: An abstraction of physical or logical resource, which may have different characteristics from the physical or logical resource and whose capability may be not bound to the capability of the physical or logical resource.

3.1.2 network function [b-ITU-T Y.3100]: In the context of IMT-2020, a processing function in a network.

NOTE 1 – Network functions include but are not limited to network node functionalities, e.g., session management, mobility management and transport functions, whose functional behaviour and interfaces are defined.

NOTE 2 – Network functions can be implemented on a dedicated hardware or as virtualized software functions.

NOTE 3 – Network functions are not regarded as resources, but rather any network functions can be instantiated using the resources.

3.1.3 virtualized network function [b-ITU-T Y.3321]: A network function whose functional software is decoupled from hardware, and runs on virtual machine(s).

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AI Artificial Intelligence

CPU Central Processing Unit

DV Data Volume

EC Energy Consumption

EE Energy Efficiency

EEM Energy Efficiency Management

EEO Energy Efficiency Orchestrator

ID Identifier

I/O Input/Output

IP Internet Protocol

MANO Management and Orchestration

ML Machine Learning

MN Mobile Network

NF Network Function

NFV Network Function Virtualization

RAN Radio Access Network

REEM Energy Efficiency Management of Resources

RP Reference Point

SDN Software-Defined Networking

VNF Virtualized Network Function

5 Conventions

In this Recommendation:

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

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

6 Overview

Virtualization technology builds infrastructure into a resource pool, including the three underlying resources of network, computing and storage, forming the integration of these resources. With the deployment and application of IMT-2020 networks, the continuous increase of new services has brought the complexity of resource management and control, how to flexibly schedule, integrate and release resources to improve the energy efficiency is a challenge. The introduction of AI technology will help to enable intelligent network management and improve energy efficiency of IMT-2020 network.

With the growth of IMT-2020 networks, it is expected that, of total network energy consumption (EC), by 2025: the radio access network (RAN) will account for 50.6%; data centres for 23.3%; edge and core networks, as well as the service core for 13.3%; and others for 12.8% [b-Lorincz]. AI-assisted energy saving methods for RAN equipment has been introduced in [b-ITU-T L.1390] and [b-ITU-T M.3381]. Besides energy efficiency of RAN, AI-assisted energy efficiency management (EEM) of virtual resources will also play an important role to reduce EC of IMT-2020 networks.

AI-assisted EEM needs to: monitor and collect data related to energy usage of virtual resources; detect resource utilization patterns with AI algorithms; make intelligent energy-saving decisions taking workload characteristics and energy usage into consideration; dynamically adjust the allocation and status of virtual resources; and optimize energy efficient policies based on the evaluation of energy-saving effect and changes in workload demands and environmental conditions.

7 Functional requirements of energy efficiency management of virtual resources in IMT-2020 networks and beyond

7.1 Requirements for energy efficiency-related data management

This clause describes high-level requirements of energy efficiency-related data management to enable EEM of virtual resources in IMT-2020 networks and beyond.

[REQ-DM-001]: Energy efficiency management of resources (REEM) is required to support the management of multiple data sets. The data sets are required to include network data (traffic data, resource configuration data, etc.) and EC data (current, voltage, power, etc.).

[REQ-DM-002]: REEM is recommended to support the collection and storage of different types of data by various collection frequencies.

[REQ-DM-003]: REEM is required to support the real-time backup of key EC data and keyword querying of specified query conditions.

[REQ-DM-004]: REEM is required to support cleaning, auditing and validation of data collected from different data sources.

[REQ-DM-005]: REEM is required to label data based on data-labelling rules.

[REQ-DM-006]: REEM is recommended to provide visualization of EC information, temperature information and energy saving statistics.

7.2 Requirements for energy efficiency-related model management

This clause describes high-level requirements of energy efficiency-related model management of virtual resources in IMT-2020 networks and beyond.

[REQ-MM-001]: REEM is required to support energy efficiency-related model selection, model training and model interference [ITU-T Y.3179].

[REQ-MM-002]: REEM is required to support the iteration of the model based on network performance and EC.

[REQ-MM-003]: REEM is recommended to support construction of an energy management model library, in which the historical models stored can be selected for reuse in the same and similar energy management scenarios.

[REQ-MM-004]: REEM is recommended to encapsulate the inference model as an API service or image file for invocation.

[REQ-MM-005]: REEM is recommended to support version management and maintenance of AI models.

7.3 Requirements for network status awareness and prediction for energy saving

This clause describes high-level requirements of status awareness to enable EEM of virtual resources in IMT-2020 networks and beyond.

[REQ-NSAP-1]: REEM is required to continuously monitor network operation status (latency, jitter, bandwidth, resource usage, etc.) and be aware of the requirements for triggering energy saving.

[REQ-NSAP-2]: REEM is required to intelligently predict network traffic, service load and dynamically adjust energy-saving policies.

7.4 Requirements for energy efficiency-related policy management

This clause describes high-level requirements of policy management to enable EEM of virtual resources in IMT-2020 networks and beyond.

[REQ-PM-001]: REEM is required to support effective energy-saving policies on adjustment, evaluation, iteration and execution.

[REQ-PM-002]: REEM is required to support the manual adjustment based on the policy generated by the energy efficiency orchestrator (EEO).

[REQ-PM-003]: REEM is required to support the energy-saving effect evaluation before and after the execution of energy-saving policies.

[REQ-PM-004]: REEM is recommended to support building an energy-saving policy library to record the policies generated in all cases, and support addition, deletion, modification and checking of energy-saving policies in the policy library.

[REQ-PM-005]: REEM is recommended to support labelling the policies that have no energy-saving effects or may cause anomalies.

7.5 Requirements for energy-saving abnormal situation management

This clause describes high-level requirements of energy-saving abnormal situation management in IMT-2020 networks and beyond.

[REQ-ASM-1]: REEM is required to support real-time abnormality detection and pre-warning, update the energy-saving policy or recover the original policy (namely without using energy-saving policy) if the abnormal results are caused by the implementation of energy-saving policies.

[REQ-ASM-2]: Abnormal situation management is required to adjust the resource configuration of a virtualized network function (VNF) to ensure network performance and user experience if network congestion and other faults occur after the execution of the energy-saving policies.

[REQ-ASM-3]: REEM is required to support the statistics and analysis of alarms caused by energy-saving policies.

[REQ-ASM-004]: REEM is recommended to support construction of an alarm-processing policy library to handle the energy-saving abnormal situation in a timely and efficient manner.

7.6 Requirements for security management

This clause describes high-level requirements of security management to enable EEM of virtual resources in IMT-2020 networks and beyond.

[REQ-SM-1]: REEM is required to provide security protection and privacy preservation for data sets, model training, model aggregation and (machine-learning) ML services.

[REQ-SM-2]: REEM is required to support the authorization and authentication of consumers to invoke energy-saving services.

8 Architectural model of energy efficiency management of virtual resources in IMT-2020 networks and beyond

8.1 Architectural model and energy efficiency management service

The architecture model of EEM is depicted in Figure 8-1. The architecture model consists of four layers for resources, data service, EEM and EEM consumers.

The resource layer consists of software-defined networking (SDN) infrastructure, network function virtualization (NFV) infrastructure, cloud infrastructure and physical infrastructure [ITU-T Y.3150].

The data service layer collects network data and EC data from the resource layer.

The EEM layer is performed by the EEO, and can act as data service consumer and EEM provider. It consumes energy efficiency-related data provided by data service layer, and provides EEM services to the consumer layer.

The EEM consumer invokes services such as data analysis, energy efficiency models, energy-saving policies and energy efficiency evaluation services. It executes energy-saving policies to reduce the EC of the resource within its management domain.

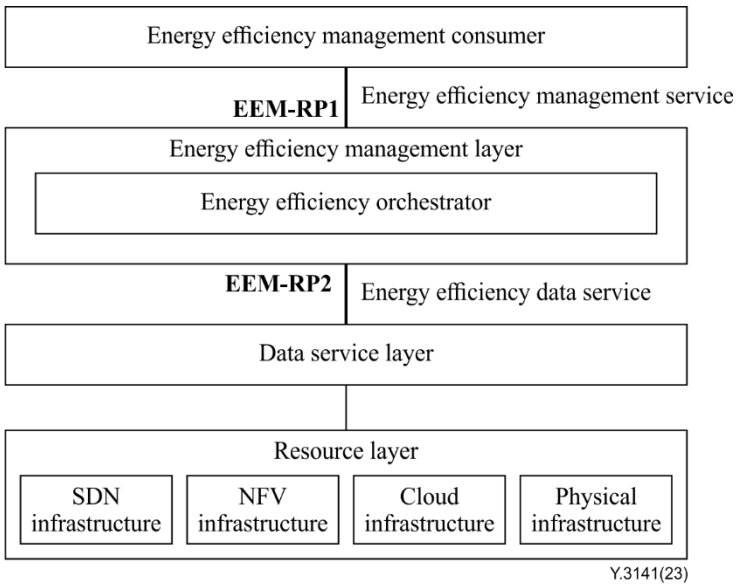

Figure 8-1 – Architectural model of energy efficiency management of virtual resources in IMT-2020 networks and beyond

The EEM layer communicates with the EEM consumer via reference point EEM-RP1, and it exchanges information with the data service layer via EEM-RP2. These two RPs are defined in detail in clause 9.

8.2 Detailed functional components of the energy efficiency orchestrator

EEM services are provided by the EEO, of which Figure 8-2 shows the functional components.

The EEO consists of five functions: data management; AI modelling and model management; energy efficiency policy management; energy efficiency evaluation; and abnormal situation management for EEM.

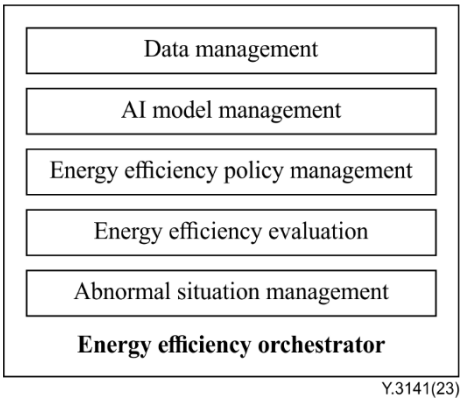

Figure 8-2 – Functional components of the energy efficiency orchestrator

8.2.1 Data management

The data management function supports the collection, storage and processing of multiple data sets.

1) Data collection of energy-consuming objects

Energy-consuming objects include virtual machines, server nodes that run a set of virtual machines and resource clusters located in a set of racks or a data centre. The data that need to be collected from these energy-consuming objects are listed in Table 1.

Table 1 – Data collection of energy-consuming objects

Data type	Information element
Configuration data of virtual machines	The configuration data of virtual machines can be resource identifier (ID), resource Internet protocol (IP), service type, host resource ID to which it belongs, central processing unit (CPU) core number, memory size, the number and size of network ports, disk capacity, active service port traffic, collection time, etc.
Operation data of virtual machines	The operation data of virtual machines can be resource ID, maximum CPU usage, minimum CPU usage, average CPU usage, maximum memory usage, minimum memory usage, average memory usage, maximum network traffic, minimum network traffic, average network traffic, maximum disk input/output (I/O), minimum disk I/O, average disk I/O, maximum hard disk usage, minimum hard disk usage, average hard disk usage, collection time, etc.
Configuration data of server nodes	The configuration data of server nodes can be resource ID, host type, host role, resource name, resource IP, service type, CPU model, CPU cores, memory size, disk capacity, rated power, active service port traffic, collection time, etc.
Operation data of server nodes	The operation data of server nodes can be resource ID, maximum CPU usage, minimum CPU usage, average CPU usage, maximum memory usage, minimum memory usage, average memory usage, maximum network traffic, minimum network traffic, average network traffic, maximum disk I/O, minimum disk I/O, average disk I/O, maximum hard disk usage, minimum hard disk usage, average hard disk usage, collection time, etc.
Configuration data of resource clusters	The configuration data of resource clusters can be cluster ID, data centre ID, cluster activation time, number of hosts, cluster type, collection time, etc.

Table 1 – Data collection of energy-consuming objects

Data type	Information element
Configuration data of services	The configuration data of services can be service ID, data centre ID, service activation date, number of hosts, service description, collection time, etc.
EC-related data	CPU load, CPU temperature, fan speed, active power of electricity
NOTE – The collection frequency of configuration data is recommended by day. The collection frequency of operation data is recommended by hour or minute.	

2) Data management and maintenance

The data management function provides the capabilities of data query, update and deletion.

The data management function provides data quality management such as cleaning, auditing and validation of data collected from different data sources, evaluates the current data quality based on some key data dimensions and provides analysis for data quality improvement.

NOTE – Data quality relates to the accuracy of data format, ranges, values, collection frequency, etc.

The data management function provides visualization of EC information, energy reduction information and machine temperature trend information.

8.2.2 AI model management

The AI model management function provides energy-saving model training, retraining and interference, as well as models for version management and energy management model library maintenance.

1) Energy-saving model training and retraining

The AI model management function can train different energy-saving models for energy-consuming objects based on the collected data, such as the energy-saving model for virtual machines in clause10.1 and energy-saving model for server nodes in clause 10.2.

Energy-saving models retrain if data quality is improved and model accuracy needs to be improved.

2) Energy management model library

The AI model management function provides evaluation (such as application and accuracy) of each model in the model library. Newly trained models including online training models and offline training models are recommended for update in the model library once model test validation is successful.

8.2.3 Energy efficiency policy management

The energy efficiency policy management function provides policy generation, execution and manual adjustment.

1) Energy efficiency policy generation

Energy efficiency policy generation can be triggered through timed tasks, a condition threshold or by force.

Timed tasks triggering: By configuring timed tasks, energy-saving policies are generated based on task execution frequency.

Condition threshold triggering: By configuring a condition threshold, energy-saving policies are generated if some energy-saving indicators (such as CPU usage, memory usage and disk usage) reach the threshold.

Forced triggering: Energy efficiency policy generation is triggered by personnel through interface operations.

2) Energy efficiency policy confirmation

When some energy-saving policies involve important services, the policies need to be manually verified before execution. Personnel can review policies, revise energy-saving policies or exclude adjustments from energy-consuming objects if the services performed on them cannot be affected by EC.

General energy-saving policies can be configured automatically.

3) Energy efficiency policy switch

The energy efficiency policy management function provides a switch between using a policy that is energy saving or one that is not.

8.2.4 Energy efficiency evaluation function

The energy efficiency evaluation function provides energy-saving metrics measurement and visualization, such as daily average energy saving, cumulative energy saving, annualized energy saving, average energy saving rate and energy efficiency.

1) Energy-saving value measurement

Energy-saving value measurement indicates the calculation of daily average energy saving, cumulative energy saving and annualized energy saving. The daily average energy saving can be measured by calculating the difference between EC before and after the execution of energy-saving policies and energy-saving duration in similar service condition. Cumulative energy saving and annualized energy saving can be measured based on daily average energy saving and total energy saving time duration.

2) Energy-saving rate measurement

Energy-saving rate measurement is energy-saving value divided by the EC in a non-energy-saving state.

3) Energy efficiency measurement

Energy efficiency of IMT-2020 networks relies on the following principles.

The energy efficiency in the mobile network by data volume, $EE_{MN,DV}$, is based on two high-level key performance indicators specified in [ETSI ES 203 228]:

$$EE_{MN,DV} = \frac{DV_{MN}}{EC_{MN}}$$

where DV_{MN} is data volume of the mobile network;

EC_{MN} is energy consumption of the mobile network.

$EE_{MN,DV}$ requires the collection of both data volume (DV) and EC of IMT-2020 network functions (NFs); measurement methods for VNFs lie outside the scope of this Recommendation.

8.2.5 Abnormal situation management function

The abnormal situation management function provides real-time abnormality detection, pre-warning and abnormal recovery.

1) Abnormal situation automatically handling

If the alarms are detected by implementation of the energy-saving policy, then the abnormal situation management function automatically switches to the original policy in a non-energy-saving state.

2) Alarm policy library

The abnormal situation management function builds an alarm policy library to provide solutions to handle energy-saving abnormal situations.

9 Reference points of energy efficiency management of virtual resources in IMT-2020 networks and beyond

In the architectural model of EEM of virtual resources in IMT-2020 networks and beyond, EEM-RP1 and EEM-RP2 are specified in clause 8. These two RPs need to support the capabilities for exchanging information and managing energy.

9.1 Requirements for EEM-RP1

[REQ-EEM-RP1-001]: The EEO is required to provide EEM services to the consumer via EEM-RP1.

[REQ-EEM-RP1-002]: The EEO is recommended to provide energy efficiency data analysis, energy efficiency models, energy-saving policies, and energy efficiency evaluation services to the consumer via EEM-RP1.

[REQ-EEM-RP1-003]: The EEO is recommended to provide comprehensive energy efficiency analysis services of different domains (e.g., core network, transport network and cloud) to the consumer via EEM-RP1.

[REQ-EEM-RP1-004]: It is recommended that the consumer subscribes to or requests EEM policies such as scaling down and ending up a VNF instance, from the EEO via EEM-RP1.

9.2 Requirements for EEM-RP2

[REQ-EEM-RP2-001]: The EEO is required to collect data from the data service layer via EEM-RP2.

[REQ-EEM-RP2-002]: It is recommended that the data service layer provide data from different domains (e.g., core network, transport network and cloud) to the EEO via EEM-RP2.

[REQ-EEM-RP2-003]: The EEO is recommended to subscribe to and query service information (e.g., service type and resource usage) via EEM-RP2.

[REQ-EEM-RP2-004]: The EEO is recommended to subscribe to and query configuration and usage information of cloud resources (e.g., virtual machine, bare metal server and cloud storage) via EEM-RP2.

[REQ-EEM-RP2-005]: The EEO is recommended to subscribe to and query network status and performance data of transport devices (e.g., virtual switches and routers) via EEM-RP2.

[REQ-EEM-RP2-006]: The EEO is recommended to subscribe to and query configuration and performance data of VNF instances of the core network via EEM-RP2.

10 Procedures of energy efficiency management of virtual resources in IMT-2020 networks and beyond

10.1 Procedure of capacity management for virtual machine

Figure 10-1 illustrates the procedure for capacity management of a virtual machine, which avoids the waste of virtual machine resources in the idle state.

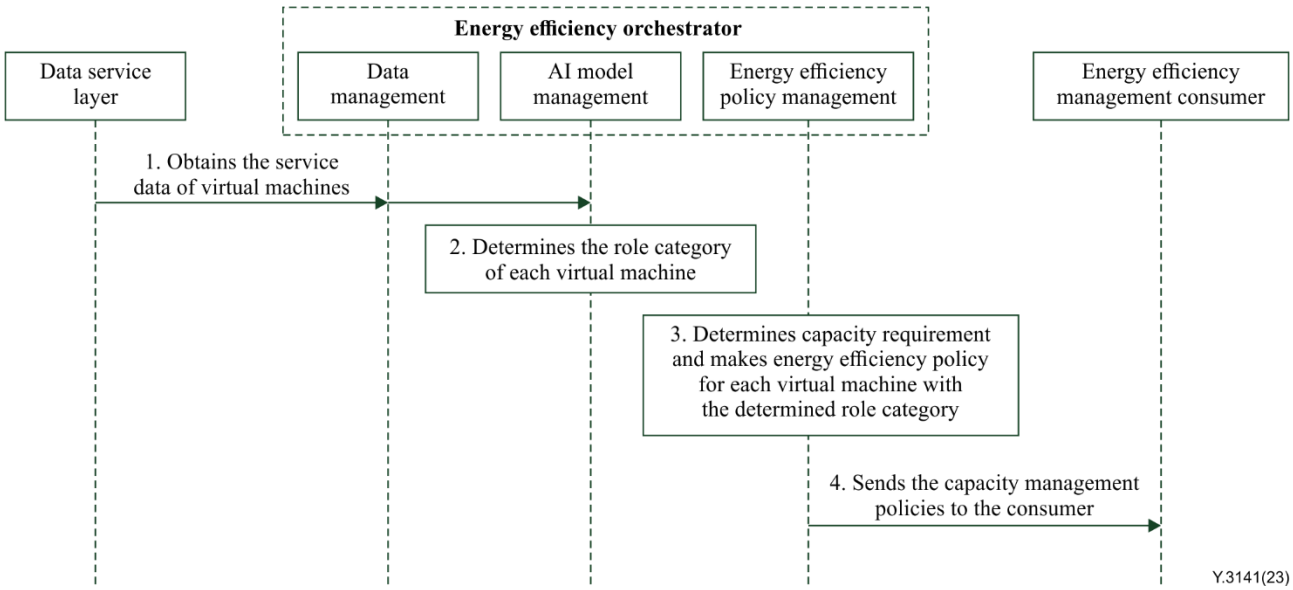

Figure 10-1 – Procedure of capacity management for virtual machine

1. The EEO obtains the service data of each virtual machine in the cloud server from the data service layer, and the service data refers to the configuration parameter information required for service execution.
2. The AI model management of the EEO uses the service data to train AI clustering models and to determine the role category of each virtual machine.

   NOTE 1 – The categories of role that virtual machines execute are indicated by service, such as that for traffic, disk, memory and CPU.

   NOTE 2 – The input of clustering models is service data of virtual machines and the output of clustering models for virtual machines includes multiple clustering sets. For each cluster, the virtual machine in the cluster can be divided into at least two subsets (such as a task subset and a non-task subset) based on whether the virtual machine executes the service of the cluster set, and the role category of the virtual machine is then determined based on subsets to which it belongs.
3. The energy efficiency policy management of the EEO analyses and determines the capacity requirements of each virtual machine according to the role categories, and makes energy efficiency policies (e.g., reducing CPU cores or memory size) of capacity management for each virtual machine with the determined role category.

   NOTE 3 – The priority list is pre-configured to avoid capacity scaling down for some virtual machines. If the role category of a virtual machine belongs to the priority list, then the capacity resource of the virtual machine can be adjusted but scaling down is excepted. If the virtual machine does not belong to the priority list, its capacity resource can be scaled down.
4. The EEO sends the energy efficiency policies for capacity management of virtual machines to the consumer who subscribes to or requests the services of energy efficiency analysis (e.g., cloud resource management system).

10.2 Procedure of on/off control for server nodes

Figure 10-2 illustrates the procedure of on/off control for server nodes in the idle state.

NOTE – The server nodes represent the infrastructure that carry some of the IMT-2020 NFs, IMT-2020 network controllers, IMT-2020 management systems, IMT-2020 services, etc.

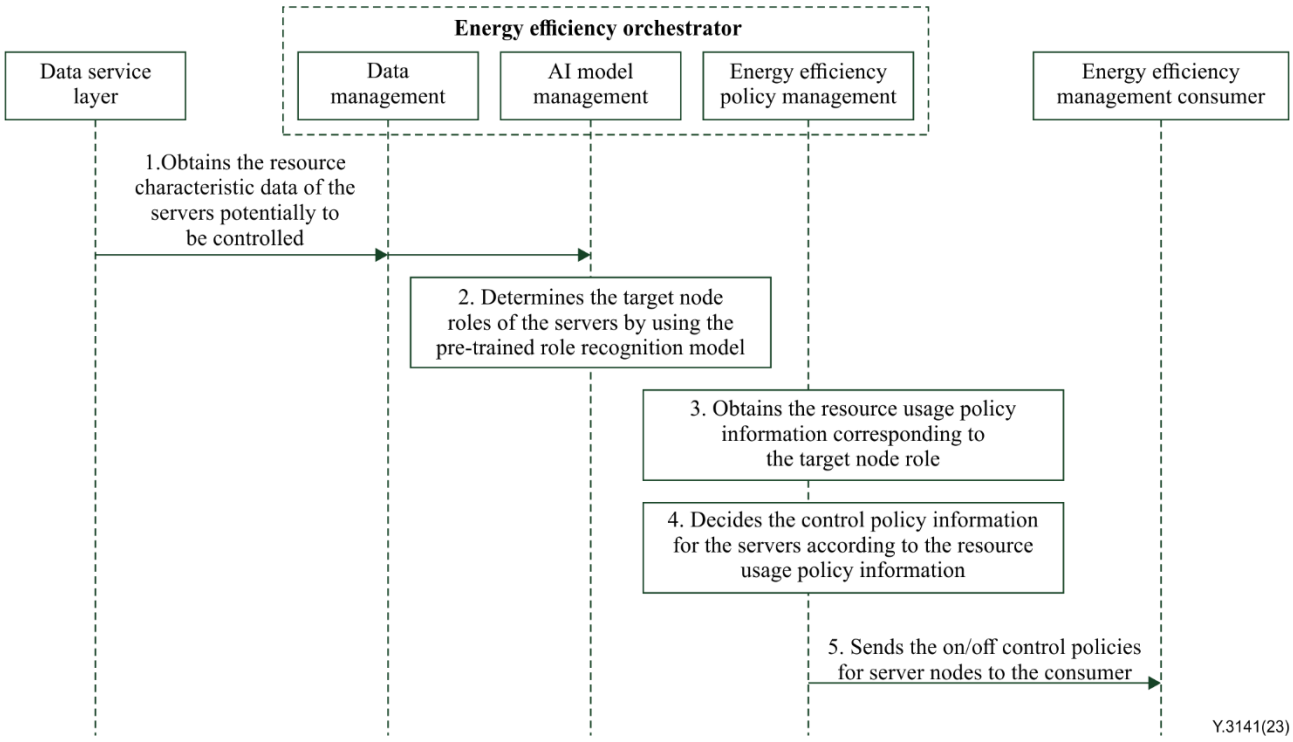

Figure 10-2 – Procedure of on/off control for server nodes

1. The EEO obtains the resource characteristic data (e.g., maximum, minimum or average: CPU usage; memory usage; and network traffic) of the servers potentially to be controlled from the data service layer.
2. The AI model management of the EEO uses the pre-trained role recognition model to process the characteristic data in order to determine the target node roles of the servers to be controlled,

   NOTE 1 – The role recognition model is pre-trained using the characteristic data in the collection period and the node role label that can be an ID of a node role (e.g., standby node ID and data node ID).
3. The energy efficiency policy management of the EEO obtains resource usage policy information corresponding to the target node role based on the node information database. The resource usage policy information indicates the resource usage threshold when server nodes have on/off behaviour.

   NOTE 2 – The node information database is built by the node role label and resource state information (e.g., CPU and memory resource usage) and configuration information of servers in the sampling period.
4. The energy efficiency policy management of the EEO decides the control policy information (e.g., switch control on to off) for the servers potentially to be controlled, according to the resource usage policy information corresponding to the target node role and the current resource state information.

   NOTE 3 – If the target node role belongs to a preset role (e.g., necessary task role) in the target cluster, control policy information of the server is determined as maintaining running state.
5. The EEO sends the energy efficiency policies for on/off control of server nodes to the consumer (e.g., bare metal server management system).

10.3 Procedure of virtual machine migration based on load balancing

Figure 10-3 illustrates the procedure for virtual machine migration based on load balancing, which ensures the dynamic load balance in the cluster, and improves the overall utilization rate.

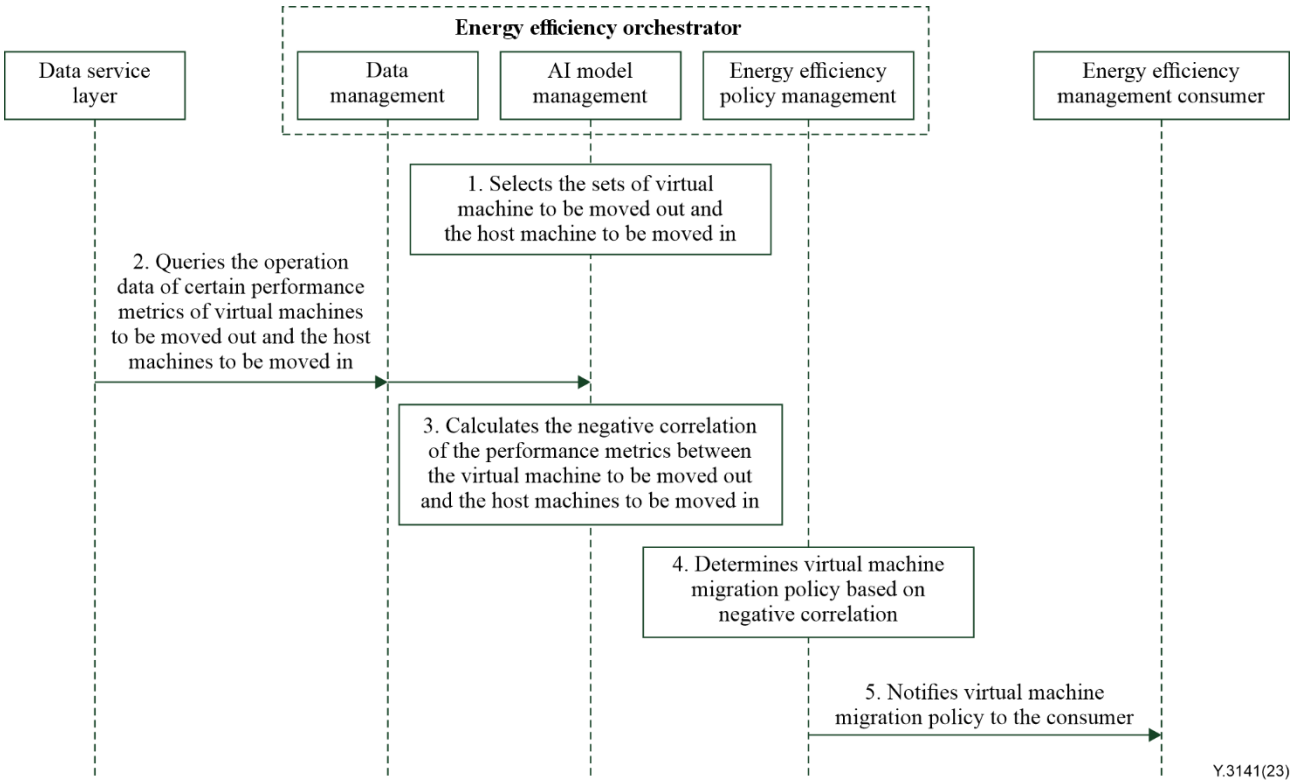

Figure 10-3 – Procedure of virtual machine migration based on load balancing

1. The EEO selects the sets of virtual machines to be moved out and the host machines to be moved in from the host cluster. The EEO calculates the number of hosts to be moved out based on the configuration of host and virtual machine resource, and the EEO selects hosts according to the disequilibrium of CPU and memory resource usage of the host and resource requirement priority information, thus all virtual machines on these selected hosts need to be moved out. The set of host machines to be moved in is selected to meet the resource configuration requirement of the virtual machine previously mentioned.

   NOTE 1 – Step 1 is triggered by the EEO itself based on pre-configured polices for load balancing.

2. The EEO queries the operation data of certain performance metrics (e.g., CPU and memory usage ratio) during collection periods of virtual machines to be moved out and the host machines to be moved in from the data service layer.

3. The EEO calculates the negative correlation between the virtual machine to be moved out and the host machines to be moved in according to the respective operation data of the performance metrics, and negative correlation calculation could focus on one main performance metric if this type of virtual machine resource is scarce.

   NOTE 2 – The EEO builds time series based on the respective operation data of the virtual machine to be moved out and the host machines to be moved in, and then calculates the negative correlation.

   NOTE 3 – The length and interval of the time series can be changed based on the actual situation, the length of the time series can be weeks, months, days, intervals of the time series can be 1 min, 5 min, 15 min, etc.

4. The EEO determines virtual machine migration policies, including selection of the virtual machine that has the highest priority from the set of virtual machines to be moved out and then migrates it into the potential host machine that can meet the resource requirement and has the highest negative correlation.

5. The EEO notifies the consumer who subscribes to the load-balancing analysis and energy-saving management service.

10.4 Procedure of energy efficiency management for VNF instances

Figure 10-4 illustrates the procedure of EEM for VNF instances. If VNF instances meet energy-saving requirements, energy management will be triggered to release the resources in the idle and low usage states.

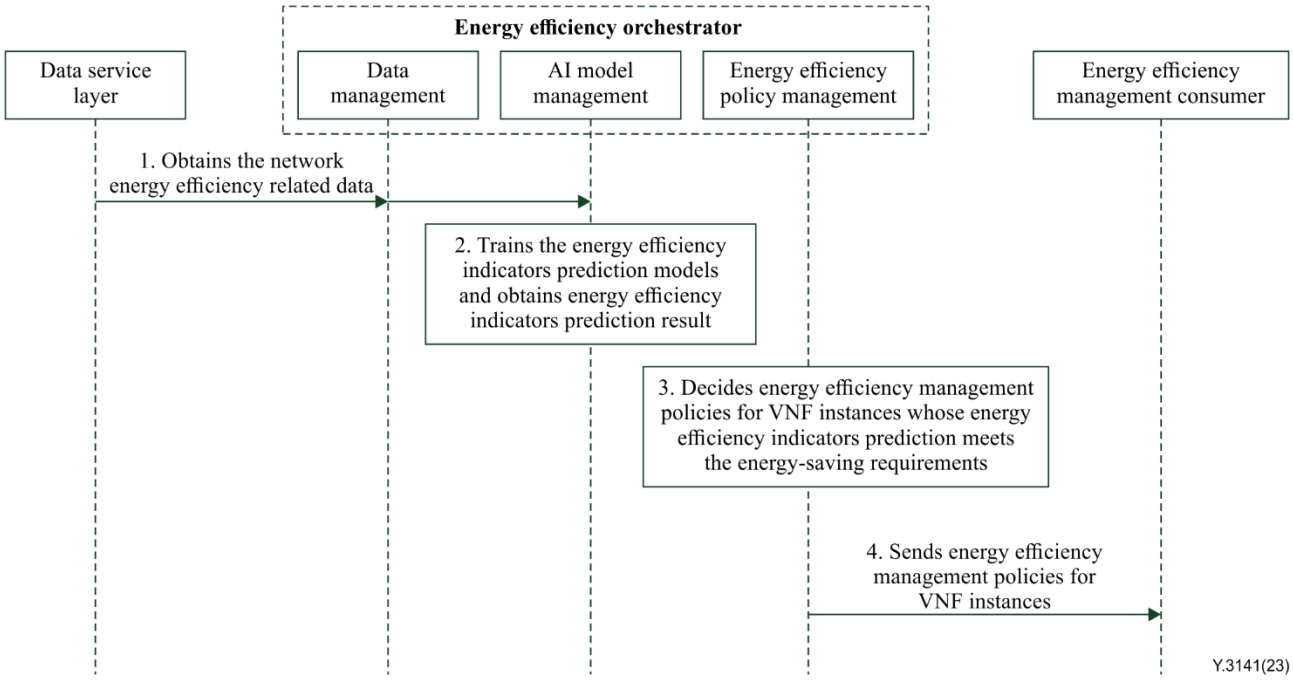

Figure 10-4 – Procedure of energy efficiency management for VNF instances

1. The EEO obtains network energy efficiency-related data, which includes dynamic operation data (e.g., energy cost, VNF instance throughput and VNF instance resource usage data) and static data (e.g., configuration information) from the data service layer.
2. The AI model management of the EEO trains energy efficiency indicator prediction models (e.g., neural network models) and obtains their results.
3. The energy efficiency policy management of the EEO decides EEM policies for VNF instances by deactivating or terminating those whose predicted energy efficiency indicators reach the energy-saving threshold.
4. The EEO sends EEM policies for VNF instances to the consumer (e.g., management and orchestration (MANO) system).

11 Security consideration

This Recommendation specifies AI-assisted EEM of virtual resources in IMT-2020 networks and beyond. It is assumed that security considerations in general are based on the security of IMT-2020 network MANO [ITU-T Y.3111] and IMT-2020 network slicing with AI-assisted analysis [ITU-T Y.3156]. Specifically, the major security requirements for EEM of virtual resources in IMT-2020 networks and beyond are listed as follows.

[REQ-SM-1]: REEM is required to provide security protection and privacy preservation for data sets, model training, model aggregation and ML services.

[REQ-SM-2]: REEM is required to support the authorization and authentication of consumers to invoke the energy-saving services.

Bibliography

[b-ITU-T L.1390] Recommendation ITU-T L.1390 (2022), *Energy saving technologies and best practices for 5G radio access network (RAN) equipment*.

[b-ITU-T M.3381] Recommendation ITU-T M.3381 (2022), *Requirements for energy saving management of 5G radio access network (RAN) systems with artificial intelligence (AI)*.

[b-ITU-T Y.3011] Recommendation ITU-T Y.3011 (2012), *Framework of network virtualization for future networks.*

[b-ITU-T Y.3100] Recommendation ITU-T Y.3100 (2017), *Terms and definitions for IMT-2020 network.*

[b-ITU-T Y.3321] Recommendation ITU-T Y.3321 (2015), *Requirements and capability framework for NICE implementation making use of software-defined networking technologies.*

[b-Lorincz] Lorincz, J., Capone, A., Wu, J. (2019). Greener, energy-efficient and sustainable networks: state-of-the-art and new trends. *Sensors*, **19**(22), 4864, 29 pp. DOI: 10.3390/s19224864

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