Computing-Aware Traffic Steering(CATS)
Internet-Draft
Updates: 2025-3-3 (if approved)
Intended status: Informational
Expires: 4 September 2025

T. Fu, Ed.
H. Zhang, Ed.
China Academy of Information and Communications Technology
J. Wang, Ed.
China Mobile

3 March 2025

Problem statements and requirements of Deterministic CATS on the Industrial Internet
draft-ftzhs-cats-industrial-requirement-03

Abstract

This draft illustrates use cases of traffic steering for Industrial Internet in terms of dynamic computing and networking resource status,together with the requirements and solutions for CATS(Computing-Aware Traffic Steering).Industrial production tasks are time-sensitive, which put forward high requirements on collaboration of networks and applications. Industrial management platforms need to unify network forwarding and computing tasks at the same time.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Copyright Notice

Table of Contents

1. Introduction

1.1. Backgroud

The Industrial Internet is a new infrastructure, application mode and industrial ecology with the deep integration among the new information technology, communication technology and the industrial economy.Industrial production tasks are time-sensitive, which put forward high requirements on networks and applications, and need to meet the deterministic requirements in terms of delay, jitter, reliability, etc. Industrial deterministic service refers to a closed loop composed of communication paths and control processes in which two or more applications participate.Industrial management platforms need to unify network forwarding and computing tasks for each deterministic service.

1.2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119][RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Definition of Terms

TBD.

3. Problem Statement of Industrial CATS

3.1. Industrial production service

In the Industrial Internet, the performance requirements of production processes are much higher than the Internet. Therefore, the industrial network naturally needs the support of compute-aware traffic steering. Application services are dynamically adapted to industrial scenarios, tasks, and resources. The computation that devices participate in evolves from simple control logic to complex big data decision-making. In the application layer of industrial Internet, deterministic service is the combination of network communication process and calculation process. It refers to a closed loop formed by one or more application communication links and control links. Industrial CATS needs to manipulate the all steps of the industrial production cycle from the service-initiating devices through remote devices with dependencies, such as edge computing or cloud services. In this process, the forwarding and calculation of data must comply with the performance requirements of the service-initiating device, such as delay, jitter, reliability, packet loss rate, etc. TBD.

3.2. Deterministic Industrial Production Service

In smart factories, there are mass production systems, edge computing, industrial clouds, remote communication relationships between various applications, and various services share network queues and computing resources. In order to satisfy the requirements of time-sensitive industrial production services, it is necessary to realize the deterministic management of computing power and network resources based on CATS. The concurrent processing of multiple services must ensure the strict requirements of delay, jitter, sequence, and reliability. At present, it is feasible to use deterministic network and edge computing to ensure multi-service load with millisecond delay through clock synchronization and resource reservation. By strengthening the bidirectional perception of computing power and network, all kinds of resources are uniformly adjusted, and the compromise between computing performance and network communication performance is achieved on various resource-competing devices.TBD.

3.3. Deterministic CATS

Deterministic CATS is adjusting network forwarding configurations according to the computing requirements. Taking the common visual detection scene in industry as an example, industrial robot arm and edge computing are interconnected by deterministic network, which involves the application of industrial robot-arm and image processing application of edge computing. The robot arm periodically collects high-definition images of the parts or products being machined and

sends the data to matched edge computing device. Then, edge computing device feeds the results back to the industrial robot-arm after processing. The traditional network can only control the round-trip transmission process. If the edge side blocks the visual detection task due to multiple services, the processing delay of the industrial robot arm will increase. In existing factories, one container or edge computing device is often configured for several tasks, resulting in a waste of entire plant resources. Deterministic CATS manages both the communication process and the calculation process, accurately ensuring the indicators of the entire visual inspection process from the perspective of an industrial robot arm, and then splits the overall indicators into the indicators of each step of the calculation process and the communication process according to the strategy. When multiple deterministic services run concurrently, this scheme can comprehensively schedule resources, do overall multi-objective optimization for multiple services, and appropriately adjust the communication cost and network cost ratio of each deterministic service.TBD.

4. Use Cases

4.1. Computing-Aware Industrial robots

The automatic manufacturing of soft materials has always been a difficult problem in industrial digitization. For the sake of unpredictable deformation of materials, it brings difficulties to the traditional equipments, unless it can recognize real-time states of the deformation for producets. Furthermore, productlines need to accurately perform operations and correct the negative effects of the deformation. With the improvement of industrial intelligence, the industrial robot can replace the manual handling of soft materials. In this case, double robotic arms plan the operation simultaneously and can judge the status of the flexible material in real time. During the whole folding operation, the offline algorithm updates the neural network periodically to optimize the parameters of the model. The online algorithm periodly recognizes the image, and the robot arm will continuously feed back the folding effect to the neural network, and judge whether to enter the next cycle on the basis of predicting the operation result and judging the folding effect. The total time of this use case is mainly limited by the recognition accuracy and recognition delay. If the computing power resources are enabled through edge computing, and the network-edge communication is ensured through application-oriented control. The processing complexity, folding speed and accuracy of flexible objects can be further improved by coordinate network and computing power. The new production line of flexible material processing can gradually replace the production line workers, and the intelligence level and

cooperation level are increased through the cloud business collaboration within the network and edge.

4.2. Computing-Aware vCloud terminial

Cloud services are control functions migrated from physical devices to Industril cloud, datacenter or edge, dramatically reducing production line costs. Traditional networks do not have the capability of application-oriented adjustment, usually directly by the engineer to estimate the processing and calculation time, subtract the estimated time with the control cycle, and finally determine the network requirements, and the calculation part is actually inaccurate or overreserved estimates. It will cause a waste of network or computing resources. The requirements of CATS are proposed to reflect the exact needs from end to end at the application level, so as to realize the remote closed-loop control between control and execution. Meet the application’s high performance requirements such as communication delay, bandwidth, and cloud computing power. Typical scenarios for cloud X services include on-site Machine control, automatic guided vehicle control, 5G PLC, etc. It has the characteristics of multi-network integration, broadening the acquisition channel, industrial equipment reusability, and improving the robustness of production network. PLC logic control has a fixed control cycle, assuming that the application needs to complete an IO data reading, processing and writing operations within a 10ms cycle. Motion control is a precision control business involving robots, servo motors and other equipment, which requires high requirements of delay and jitter, with end-to-end delay to be controlled within 10ms and jitter less than 100us. Machine vision quality inspection is an intelligent quality inspection service involving image processing and analysis. It has high requirements for uplink bandwidth capability, and the network should provide uplink bandwidth greater than 80Mbps. Power differential protection is a key service related to the safety and stable operation of the power grid. It has high requirements for delay and reliability. The service delay requirement is less than 15ms, and the reliability reaches 99.999%.

4.3. Computing-Aware Multi-Application collaboration

With the increasing degree of networking and digitalization of industrial enterprises, limited by the 1-3 layer network certainty, the combination of machine learning, big data and other technologies with the production line is mainly direct deployment and simple interconnection, and the role of advanced algorithms in resource optimization allocation in the life cycle of industrial production is still not fully reflected. If the network with high certainty accuracy is equipped with IT services that try their best to deal

with IT, it is not enough. There is a risk that some links will time out, making it difficult to accurately guarantee the complex production tasks of multi-equipment collaboration. CATS can solve this problem through multi-application collaborative global scheduling. Take the intelligent processing line in Figure 2 as an example. The edge equipment carrying IT technology runs a large number of new algorithms and models, and introduces new computing resources into the workshop network architecture through edge cloud facilities. In order to ensure the certainty of the entire production business, it is necessary to globally manage a series of applications and network transmission according to constraints to ensure that the overall service quality meets the needs of users. The global scheduling of multi-application collaboration needs to support the information interconnection between various systems at the data application level of the whole network, and regulate the real-time and reliability capabilities of all devices, so as to strictly meet requirements of various applications. The certainty between applications actually includes logic, computing, network transmission and other links, and it is difficult to achieve the overall certainty guarantee by using only one level of certainty guarantee technology. In the future, the certainty of the application layer should be arranged from the network, calculation, logic and other aspects to achieve the overall certainty of the overall high service quality assurance. The certainty of multi-application collaboration can be comprehensively controlled for complex business, which is an important way to realize unmanned production line. Automatic application association analysis helps you connect service planning and device configuration, saving the cost of service configuration and change.

4.4. Industrial Digital Twins

Digital workshop is an information workshop designed and constructed by applying lean production, lean logistics, visual management, standardized management, green manufacturing and other advanced production control theories and methods. The application of a deterministic digital twin on the end side provides a comprehensive understanding of the details of the production environment situation. The generation of massive data is often conducive to the realization of big data decisions, but different types of data are often distributed in several independent information systems or in different parts of the production process. Through industrial Internet sensors and other data acquisition devices, valuable business data is continuously collected in different forms and on an unprecedented scale, and then uploaded to digital twin applications, and real-time graphical display and predictive analysis results can be effectively responded to emergency situations. The digital twin business composed of several links needs to guarantee deterministic

parameters at the application level to ensure accurate and reliable projection of the whole plant in the information space. The utility of end-to-side data integration mainly comes from the massive application data collected, the diversity of application data and the accuracy of the collected data. The traditional data integration generally uses the industrial Ethernet private network to collect information, and then converges and processes it to the upper data center through the controller. The whole application to application link is the conversion of multiple system protocols and the interaction between multiple layers of applications. Among them, the industrial system detection, control, implementation of high real-time, industrial production site data has a large volume, and has real-time demand, some scenes real-time requirements within 10ms. At the same time, data collected from endpoints can introduce failure information, duplicate information, and other types of problems. The certainty in traditional data integration scenarios is usually guaranteed by the network layer, but the certainty between applications actually includes multiple links such as sequential processing, computing, and network transmission, and it is difficult to rely on a single network level guarantee capability to support the application-to-application data integration requirements.

4.5. Customized Production Lines

The new model of Customers to Production Lines (C2PL) applies the new technology of artificial intelligence to the layout of factory production activities, and pulls customized orders and flexible production with intelligent large models as the core. Support users to customize their favorite products on the official website, and freely play in terms of appearance, material, size, etc. Users complete the customization, ordering and payment of high-end products on the Internet side, and the joint business arrangement of IT system and OT system is completed by industrial management software. Automatic formation of a temporary overall certainty, disassembly into multiple processing links and execution. C2PL realizes unmanned production line scheduling and can be widely applied to products with simple structure to improve user participation and demand collection efficiency. The existing ChatGPT can support voice, link, picture and other forms of data. After the system gives the effect diagram, the product form can be improved interactively. In this scenario, AI connects IT and OT workflows, and CATS requires global quality of service definition, control, monitoring and evaluation. Support a variety of demand expression forms, voice, pictures, links and other needs can be understood by the system and present the sample effect; Through AI large model engine, order feature interpretation, production process creation, resource scheduling, etc. Ensure the industrial Internet certainty under the condition of product diversification; The overall control is disassembled into a set of

processing steps of IT design link and OT production link through AI large model engine. Implement OT infrastructure to provide unified defined manufacturing services to the control system. After realizing standardized production links, CATS can predict, adjust and monitor control results more accurately. Allocate temporary resources in the factory resources to complete the production task to ensure that the processing process can be completed in real time.

5. Requirements

5.1. Requirements for Deterministic CATS Service

a. The industial system needs a specific controller to unify scheduling of network resources and computing resources for deterministic services;
b. It needs to establish independent network integration diagram for each deterministic service to accurately reflect "application-network" correlation;
c. The performance indicators of deterministic services need to be converged within expected boundaries, such as the overall service completion delay, overall jitter, bandwidth, packet loss rate, etc.
d. Industrial equipments, such as OT devices and IT devices, need information model to uniformly define the deterministic parameters of deterministic devices;
e. All deterministic devices need to support and enable deterministic application and network deterministic control protocols;
f. The deterministic management and control of deterministic computing tasks need to be supported, and the integrated scheduling policy of the computing network should be split into the resource allocation of deterministic devices and delivered to the target devices through the deterministic southbound interface.
g. The industial system requires deterministic execution of computing tasks . After receiving the deterministic index parameters of the deterministic computing tasks being executed, the deterministic device ensures the execution speed and output result quality of the computing tasks by means of scheduling priority, elastic allocation of computing resources, isolation of computing units, etc.
h. The deterministic controller needs to support deterministic control of deterministic computing tasks, convert the integrated scheduling strategy of the computing network into resource allocation for deterministic devices, and issue it to the target device through southbound interfaces.
i. It is necessary for deterministic controller to support the deterministic execution of deterministic computing tasks. After receiving the deterministic indicator parameters of the ongoing

deterministic computing task, deterministic devices ensure the execution speed and output quality of the computing task through scheduling priorities, elastic allocation of computing resources, isolation of computing units, and other methods.
j. The service needs to provide users with synchronization functions between applications, support the coordination of time sequence between applications, and ensure consistency in the entire deterministic system for applications.k.TBD.

5.2. Requirements for Deterministic Networks

a. In the case of data transmission with high throughput, it is required to implement multi-channel deterministic transmission through the deterministic transport layer protocol;
b. Network is required to support fast session establishment, connection migration (wireless network), elastic congestion control;
c. Network controller needs application-oriented subflow management;
d. It needs cross-layer configuration and consistency of key parameters such as clock, cycle, data unit and priority;
e. Deterministic networks need supporting cross-domain scheduling of data flows, which involves crossing networks at different levels and crossing boundary devices (devices that modify protocols such as gateways);
f. Network controller is required to achieve model-based predictability for cross-layer, large-scale and heterogeneous networking, etc. (Predictable function is useful for automated configuration)
g. The network needs to have time synchronization function. On the one hand, it provides synchronization support for upper layer application services. On the other hand, through the synchronization of network devices and terminals, it can support time slot scheduling of data traffic and improve the deterministic ability of network transmission.
h. In the transmission of high reliability services, the network needs to have the function of multi-path redundant transmission, which can support the transmission of business data in multiple paths, to ensure that when the determinacy of one link is difficult to meet, other links can still meet the transmission requirements of the business.
i.The network can provide deterministic transmission capabilities based on application business requirements, including latency, bandwidth, reliability, etc., to allocate network resources and scheduling strategies for different business traffic needs, achieving on-demand transmission guarantee.
j. TBD.

5.3. Requirements for Internal factory computing

a. Computing resources imformation, such as basic information, computing information, load information, task list, etc. are required to upload actively after computing equipments are registered to CATS controller;
b. The computing equipment needs to regularly upload real-time information;
c. CATS controller needs to support calculation force and valuation function, used to evaluate the scheduling results of own resources;
d. CATS controller support billing and query of external computing resources;
e. CATS controller needs supporting precise isolation for multi-core hardware, and support the mapping of some certain computing units to a deterministic computing task;
f. Resource reservation for a deterministic computing task should be supported on computing devices;
g. CATS controller support elastic resource expansion and contraction of containers;
h. A deterministic information model for identifying deterministic services needs to be supported.
i. Deterministic controller needs to support the cross domain interconnection of computing power and the long-distance lossless data transmission of computing power data.
j. Deterministic controller needs to support public network computing power perception and network active perception of computing power network application status;
k. Deterministic controller needs to support computing power management and operation: fine-grained data collection capability, hyper visual monitoring capability, automated operation and maintenance capability, and intelligent collaborative optimization;
l. Deterministic controller needs to support intelligent scheduling and optimization of computing resources: meet the QoS service guarantee requirements of applications, flexibly optimize networking and computing resource allocation strategies, and achieve deterministic computing resource supply for business;
m. Deterministic controller needs to support trusted computing power trading services: trusted real-time dynamic computing power service supply and demand configuration, efficient and reliable data transmission, secure and reliable data protection, fair and equal distribution of benefits;
n.TBD.

5.4. Requirements for External Factory Computing

TBD.

5.5. Requirements for Global Management

a. CATS controller support differentiated assurance and control methods for deterministic computing tasks;
b. CATS controller support deterministic service management of two-way sensing on the cloud side of the network;
c. CATS controller support controlled service migration and service change downtime;
d. All deterministic devices need to support deterministic northbound interfaces and deterministic southbound interfaces;
e. OPC-UA protocol should be adopted to transmit deterministic requirements, monitoring information, and deterministic configuration among application-oriented deterministic management and control systems, deterministic computing devices, and deterministic network devices.f. CATS controller support the use of OPC-UA protocol to transmit deterministic requirements, monitoring information, and deterministic configurations between application-oriented deterministic control systems, deterministic computing devices, and deterministic network devices.

6. IANA Considerations

This memo includes no request to IANA.

7. Security Considerations

This document should not affect the security of the Internet.

8. References

8.1. Normative References

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

8.2. Informative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[exampleRefMin] Surname, Initials., "Title", 2006.

[exampleRefOrg] Organization, "Title", 1984, <http://www.example.com/>.

Appendix A. Appendix 1

TBD.

Acknowledgements

TBD.

Contributors

Thanks to all of the contributors.

Authors' Addresses

Fu Tao (editor)
China Academy of Information and Communications Technology
Huayuanbei No.52
beijing
beijing, 100191
China
Email: futao@caict.ac.cn

Zhang Hengsheng (editor)
China Academy of Information and Communications Technology
Huayuanbei No.52
beijing
beijing, 100191
China
Email: zhanghengsheng@caict.ac.cn

Wang Jing (editor)
China Mobile
beijing
beijing, 100191
China
Email: wangjingjc@chinamobile.com