

RTGWG  
Internet-Draft  
Intended status: Standards Track  
Expires: 8 September 2022

D.H. Daniel  
B.T. Bin  
ZTE Corporation  
P.L. Peng  
China Mobile  
7 March 2022

Computing Delivery in Routing Network  
draft-huang-computing-delivery-in-routing-network-01

## Abstract

This document drafts a proposal of the architecture of Computing Delivery in Routing Network which incorporates both computing and networking metrics into the routing policies and enables the network sensing and scheduling computing services based upon traditional networking services. A mechanism of two-class computing status granularity and two segment routing is illustrated for end-to-end networking and computing service in the cloud sites, while major networking and computing actors is defined in terms of functionality. An example work flow is demonstrated, and both control plane and data plane solution consideration is proposed..

## Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 8 September 2022.

## Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

Internet-Draft

Abbreviated Title

March 2022

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the [Trust Legal Provisions](#) and are provided without warranty as described in the Revised BSD License.

## Table of Contents

<a href="#">1.</a>	Introduction . . . . .	<a href="#">2</a>
<a href="#">1.1.</a>	Requirements Language . . . . .	<a href="#">3</a>
<a href="#">2.</a>	Terminology . . . . .	<a href="#">3</a>
<a href="#">3.</a>	Computing delivery in routing network reference architecture . . . . .	<a href="#">4</a>
<a href="#">3.1.</a>	Hierarchical granularity routing scheme . . . . .	<a href="#">5</a>
<a href="#">3.2.</a>	Two-segment routing and forwarding . . . . .	<a href="#">6</a>
<a href="#">3.3.</a>	Cross-domain computing routing and forwarding . . . . .	<a href="#">6</a>
<a href="#">3.4.</a>	CSI routing . . . . .	<a href="#">7</a>
<a href="#">3.5.</a>	Traffic affinity . . . . .	<a href="#">7</a>
<a href="#">4.</a>	Computing delivery in routing network architecture work flow . . . . .	<a href="#">8</a>
<a href="#">4.1.</a>	Computing resource and service update work flow . . . . .	<a href="#">8</a>
<a href="#">4.2.</a>	Service flow routing and forwarding work flow . . . . .	<a href="#">8</a>
<a href="#">5.</a>	Control plane . . . . .	<a href="#">9</a>
<a href="#">5.1.</a>	Centralized control plane . . . . .	<a href="#">9</a>
<a href="#">5.2.</a>	Distributed control plane . . . . .	<a href="#">9</a>
<a href="#">5.3.</a>	Hybrid control plane . . . . .	<a href="#">9</a>
<a href="#">6.</a>	Data plane . . . . .	<a href="#">9</a>
<a href="#">6.1.</a>	CSI encapsulation . . . . .	<a href="#">10</a>
<a href="#">6.2.</a>	CSI for GCR, CUR and LCR . . . . .	<a href="#">10</a>
<a href="#">7.</a>	Summary . . . . .	<a href="#">10</a>
<a href="#">8.</a>	Acknowledgements . . . . .	<a href="#">10</a>
<a href="#">9.</a>	IANA Considerations . . . . .	<a href="#">10</a>
<a href="#">10.</a>	Security Considerations . . . . .	<a href="#">11</a>
<a href="#">11.</a>	Informative References . . . . .	<a href="#">11</a>
	Authors' Addresses . . . . .	<a href="#">11</a>

[1.](#) Introduction

Computing-related services have been provided in such a way that computing resources either are confined within isolated sites (data

centers, MECs etc.) without coordination among multiple sites or they are coordinated and managed within specific and closed service systems without fine-grained networking facilitation, while the industry develops into an era in which the computing resources start migrating from centralized data centers to distributed edge nodes.

Therefore substantial benefits in light of both cost and efficiency resulting from scale of economy, would be brought into multiple industries by intelligently and dynamically connecting the distributed computing resources and rendering the coordinated computing resources as a unified and virtual resource pool. On top of the cost and efficiency gains, applications as well as services would be served in a more sophisticated way in which computing and networking resources could be aligned more efficiently and agilely than conventional way in which the two are delivered in separate systems. Some impressive drafts such as [[I-D.liu-dyncast-ps-usecases](#)] and [[I-D.li-dyncast-architecture](#)] analyze the benefits of routing related solution, and give the reference architecture and preliminary test results. End applications could be served not only by fine-grained computing services but also fine-grained networking services rather than the best-effort networking services without routing network involved otherwise. The cost is the burden of maintaining and sensing computing resource status in the networking layer. The architecture proposal is designed to be as much smoothly compatible with the ongoing routing architecture as possible.

### [1.1](#). Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [[RFC2119](#)].

## [2](#). Terminology

- \* Global Computing-related Routing Node (GCR): routing node maintaining computing resource as well as service status from remote cloud sites, and executing the cross-site routing policies in terms of the aforementioned status as well as the identification of computing resource and service. GCR usually resides at the network edge and works as ingress of the end to end service flow.

- \* Local Computing-related Routing Node (LCR): routing node maintaining computing resource as well as service status from the geographically local cloud sites and being responsible for the last hop of the service flow towards the computing resource and service instance in the specific cloud site. LCAT usually resides at the network edge and works as egress of the end to end service flow.

- \* Computing Unaware Routing Node (CUR): routing node unaware of computing resource and service status and disregarding encapsulation of the identification of computing service. CUR usually resides between GCR and LCR and works as ordinary routing nodes and stays irrelevant of computing delivery.
- \* Global Computing Resource and Service Status (GCRS): General cloud site status of the computing resource and service which consists of overall resource occupation and types of computing service (algorithms, functions etc.) the specific cloud site provides. GCRS is maintained at GCR and expected to remain relatively stable and change in slow frequency.
- \* Local Computing Resource and Service Status (LCRS): fine-grained cloud site status of the computing resource and service which consists of status of each active computing service instance as well as its parameters which impact the way the instance would be selected and visited by LCR. LCRS is maintained at LCR and expected to stay quite active and change in high frequency.
- \* Computing Service Identification (CSI): a globally unique identification of a computing service with optional parameters, and it could be an IPv6-like address or specifically designed identification structure.
- \* Instantiated Computing Service (ICS): an active instance of a computing service identification which resides in a host usually purporting to a server, container or virtual machine.

### 3. Computing delivery in routing network reference architecture

Routing network is enabled sensing the computing resource and service from the cloud sites and routing the service flow according to both network and computing metrics by a computing delivery in routing network architecture as illustrated in figure 1. The architecture is a horizontal convergence of cloud and network, while the latter maintains the converged resource status and thus is able to achieve an end to end routing and forwarding policy from a perspective of cloud and network resource. PE1 maintains GCRS with a whole picture of the multiple cloud sites, and executes the routing policy for the network segment between PE1 and PE2 or PE3, namely between ingress and egress, while PE2 maintains LCRS with a focus picture of the cloud site where S1 resides, and establishes a connection towards S1. S1 is an active instance of a specific computing service type (CSI). On top of the role of LCR which maintains LCRS, PE2 and PE3 also fulfill the role GCR which maintains GCRS from neighboring cloud sites. P provides traditional routing and forwarding functionality for computing service flow, and remains unaware of any computing-

Daniel, et al.

Expires 8 September 2022

[Page 4]

Internet-Draft

Abbreviated Title

March 2022

related status as well as CSI encapsulations.

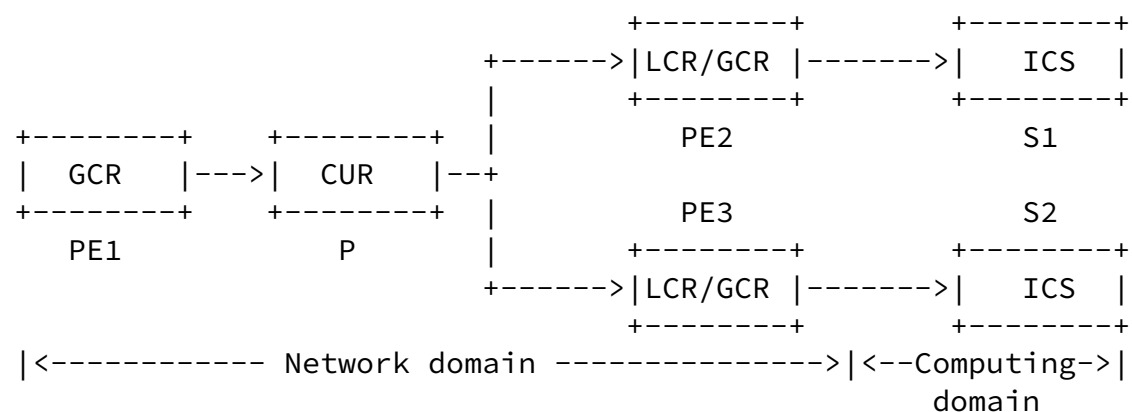


Figure 1

#### 3.1. Hierarchical granularity routing scheme

Status updates of computing resource and service in the cloud sites extend in a quite broad range from relatively stable service types and overall resource occupation to extremely dynamic capacity changes as well as busy and idle cycle of service instance. It would be a disaster to build all of the status updates in the network layer which would bring overburdened and volatile routing tables.

It should be reasonable to divide the wide range of computing resource and services into different categories with differentiated characteristics from routing perspective. GCRS and LCRS correspond to cross-site domain and local site domain respectively, and GCRS aggregates the computing resource and service status with low update frequency from multiple cloud sites while LCRS focuses only upon the status with high frequency in the local sites. Under this two-granularity scheme, computing-related routing table of GCRS in the GCR remains in a position roughly as stable as the traditional routing table, and the LCRS in the LCR maintains a near synchronized state table of the highly dynamic updates of computing service instances in the local cloud site. Nonetheless, LCRS focusing upon a single and local cloud site is the normal case while upon multiple sites should be exemption if not impossible.

### [3.2.](#) Two-segment routing and forwarding

When it comes to end to end service flow routing and forwarding, there is an status information gap between GCRS and LCRS, therefore a two-segment mechanism has to be in place in line with the two-granularity routing scheme demonstrated in 3.1. As is illustrated in figure 2, R1as ingress determines the specific service flow's egress which turns out to be R2 according to policy calculation from GCRS. In particular, the CSI from both in-band (user plane) and out-band (control plane) is the only index for R1 to calculate and determine the egress, it's highly possible to make this egress calculation in terms of both networking (bandwidth, latency etc) and computing Service Agreement Level. Nevertheless, the two SLA routing optimization could be decoupled to such a degree that the traditional routing algorithms could remain as they are. The convergence of the

SLA policies as well as the methods to make GCR aware of the two SLA is out of scope of this proposal.

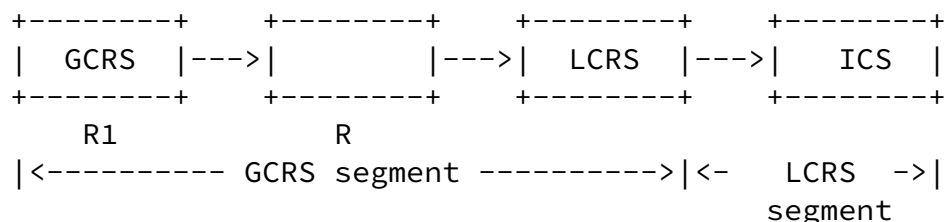


Figure 2

When the service flow arrives at R2 which terminates the GCRS segment routing and determines S1 which is the service instance selected according to LCRC maintained at R2. Again CSI is the only index for LCRC segment routing process.

### 3.3. Cross-domain computing routing and forwarding

Co-ordinated computing resource scheduling among multiple regions which are usually connected by multiple network domains, as illustrated in [section 1](#), is an important part of intended scenarios with regard to why computing-based scheduling and routing is proposed in the first place. The two-segment routing and forwarding scheme illustrated in 3.2 is a typical use case of cross-domain computing routing and forwarding and a good building block for the full-domain scenario solution. Computing metric information is brought into network domain to enable the latter scheduling routing policies beyond network. However, a particular scheme has to be put in place to ensure mild and acceptable impacts upon the ongoing IP routing scheme. A consistent CSI across terminal, network (multiple domains)

and cloud along with hierarchical CSI-associated computing resource and service status which corresponds with different network domains, is the enhanced full-domain routing and forwarding solution. Each domain maintains a corresponding computing resource and service status at its edge node and makes the computing-based routing for the domain-related segment which should be connected by the neighboring segments.

### [3.4.](#) CSI routing

CSI encapsulated in the headers and maintained in LCRS and GCRS indicates an abstract service type rather than a geographically explicit destination label, thus the routing scheme based upon CSI is actually a two-part and two-layer process in which CSI only indicates the routing intention of user's requested computing service type where routing does not actually materialize in forwarding plane and the explicit routing destination would be determined by LCRS and GCRS. Therefore the actual routing falls within the traditional routing scheme which remains intact.

Apart from the indication of computing service routing intention, CSI could also indicates a specific network service requirements by associating the networking service policy in GCRS which would therefore schedule the network resources such as an SR tunnel, guaranteed bandwidth etc. at egress.

Therefore, GCRS and LCRS in control plane along with CSI encapsulation in user plane enables an logical computing routing sub-layer which is able to be aware of the computing from cloud sites and forward the service flow in terms of computing services as well as computing resources. Nevertheless, this logical sub-layer remains only relevant at ingress and egress nodes and is simply about computing nodes selection rather than executing the real forwarding and routing actions.

### [3.5.](#) Traffic affinity

CSI holds the only semantics of the service type that could be deployed as multiple instances within specific cloud site or across multiple cloud sites, CSI in the destination field is not explicit enough for all of the service flow packets to be forwarded to a specific destination. Traffic affinity has to be guaranteed at both GCR and LCR. Once the egress is determined at GCR, the binding relationship between the egress and the service flow's unique identification (5-tuple or other specifically designed labels) is maintained and the subsequent flow could be forwarded upon this binding table. Likewise LCR maintains the binding relationship between the service flow identification and the selected service



Traffic affinity could be guaranteed by mechanisms beyond routing layer, but they will not be in the scope of this proposal.

#### 4. Computing delivery in routing network architecture work flow

#### 4.1. Computing resource and service update work flow

The full range of computing resource and service status from a specific cloud site is registered at LCR which maintains LCERS in itself and notifies the part of GCRS to remote GCRs where GCRS would be thus maintained and updated. As is illustrated in figure 3, GCR in R1 from site1 and site 2 is updated by R2 and R3, while LCERS of site 1 in R2 is updated by S1 and LCERS of site 2 in R3 is updated by S2. GCRS in R2 and R3 is updated by each other. Edge routers associating with local cloud site establish a mesh fabric to update the according GCRS among the whole network domain, the computing resource and services in distributed cloud sites thus are connected and could be utilized as a single pool for the applications rather than the isolated islands.

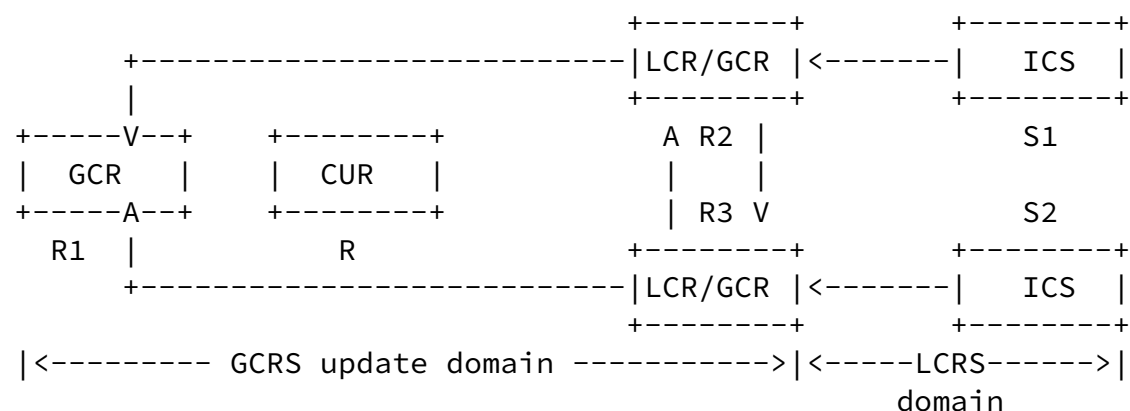


Figure 3

#### 4.2. Service flow routing and forwarding work flow

From perspective of the service work flow, more details have actually been demonstrated in 3.2 and 3.3. Rather than the traditional destination-oriented routing mechanism and the segment routing in which the ingress router is explicitly aware of a specific destination, CSI as an abstract label without semantics of physical address works as the required destination from viewpoint of the user in computing delivery in routing network architecture. Therefore the service flow has to be routed and forwarded segment by segment in which the two segment destinations are determined by GCRS and LCRS

respectively.

## [5.](#) Control plane

### [5.1.](#) Centralized control plane

LCRS's volatility makes it infeasible to be maintained and controlled in a centralized entity, GCRS is the chief computing resource and service status information to be collected and managed in the controller when it comes to centralized control plane with regard to computing delivery in routing network architecture. Routing and forwarding policies from GCRS calculated in the centralized controller, as is demonstrated in 3.2, apply only to the segment from ingress and egress, while the second segment routing policy from egress to the selected service instance in the cloud site is determined by LCRS at egress.

Hierarchically centralized control plane architecture would be strongly recommended under the circumstances of nationwide network and cloud management.

### [5.2.](#) Distributed control plane

GCRS is updated among the edge routers which have been connected in a mesh way that each pair of edge routers could exchange GCRS to each other, while LCRS will be unidirectionally updated from cloud site to the associated edge router in which LCRS is maintained and its update process is terminated.

Protocol consideration upon which GCRS and LCRS is updated is out of the scope of this proposal and will be illustrated in another draft.

### [5.3.](#) Hybrid control plane

It should be more efficient to update the GCRS by a distributed way than a centralized way in terms of routing request and response in a limited network and cloud domain, but be the opposite case in a nationwide circumstance. This is how hybrid control plane could be deployed in such a scheme that overall optimization is achieved.

## [6.](#) Data plane

Internet-Draft

Abbreviated Title

March 2022

### [6.1.](#) CSI encapsulation

Computing service identification is the predominant index across the entire computing delivery in routing network architecture under which a new virtual routing sub-layer is employed with CSI working as the virtual destination. Data plane indicates the routing and forwarding orientation with CSI by inquiring GCRS and LCRS at GCR and LCR respectively. CSI encapsulation could be achieved by extending the existing packet header and also achieved by designing a dedicated shim layer, which along with the specific structure of CSI are out of the scope of this proposal and will be illustrated in another draft.

### [6.2.](#) CSI for GCR, CUR and LCR

GCR encapsulates CSI in a designated header format as a proxy by translating the user-originated CSI format, and makes the first segment routing policy and starts routing and forwarding the service traffic. CUR ignores CSI and simply forwards the traffic as usual. LCR decapsulates CSI and makes the second segment routing policy and completes the last hop routing and forwarding.

## [7.](#) Summary

It would significantly benefit the industry by connecting and coordinating the distributed computing resources and services and more so by further converging networking and computing resource. Uncertainty and the potential impacts over the ongoing network architecture is the main reason for the community to think twice. By classifying the end to end routing and forwarding path into two segments, the impacts from computing status and metrics are to be reduced to a degree they would be as acceptable and comfortable enough as they are as networking status and metrics. In particular, employment of CSI in computing delivery in routing network architecture enables a new service routing possibility perfectly compatible with the ongoing routing architecture.

## [8.](#) Acknowledgements

To be added upon contributions, comments and suggestions.

## 9. IANA Considerations

This memo includes no request to IANA.

Daniel, et al.

Expires 8 September 2022

[Page 10]

---

Internet-Draft

Abbreviated Title

March 2022

## 10. Security Considerations

As information originated from the third party (cloud sites), both GCRS and LCRS would be frequently updated in the network domain, both security threats against the routing mechanisms and credibility and security issues of the computing services should be taken into account by architecture designing. The detailed analysis as well as solution consideration will be proposed in the updated version of the draft.

## 11. Informative References

[I-D.li-dyncast-architecture]

Li, Y., "Dynamic-Anycast Architecture", February 2021,  
<<https://datatracker.ietf.org/doc/draft-li-dyncast-architecture/>>.

[I-D.liu-dyncast-ps-usecases]

Liu, Peng., "Dynamic-Anycast (Dyncast) Use Cases and Problem Statement", February 2021,  
<<https://datatracker.ietf.org/doc/draft-liu-dyncast-ps-usecases/>>.

[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>>.

## Authors' Addresses

Daniel Huang  
ZTE Corporation

Nanjing  
Phone: +86 13770311052  
Email: huang.guangping@zte.com.cn

Bin Tan  
ZTE Corporation  
Nanjing  
Phone: +86 13918622159  
Email: tan.bin@zte.com.cn

Peng Liu  
China Mobile  
Beijing  
Phone: +86 13810146105

Daniel, et al.	Expires 8 September 2022	[Page 11]
----------------	--------------------------	-----------

---

Internet-Draft	Abbreviated Title	March 2022
----------------	-------------------	------------

Email: liupengyjy@chinamobile.com

