Instant Congestion Assessment Network (iCAN) for Data Plane Traffic Engineering
draft-liu-ican-00

Abstract

iCAN (instant Congestion Assessment Network) is a set of mechanisms running directly on network nodes:

- To adjust the flows paths based on real-time measurement of the candidate paths.

- The measurement is to reflect the congestion situation of each path, so that the ingress nodes could decide which flows need to be switched from a path to another.

This is something that current SDN and TE technologies can hardly achieve:

- SDN Controller is slow and far from the data plane, it is neither able to assess the real-time congestion situation of each path, nor able to assure the data plane always go as expected (especially in SRv6 scenarios). However, iCAN can work with SDN perfectly: controller planning multi-path transmission, and iCAN does the flow optimization automatically.

- Traditional TE is not able to adjust the flow paths in real-time.

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."
1. Introduction

Traditional IP routing is shortest path based on static metrics, which can fulfill basic requirement of connectivity. MPLS-TE brings the capability of utilizing non-shortest paths, thus traffic dispatch
is doable; however, MPLS-TE is only a complementary mechanism because of the scalability issue. Segment routing provides even more flexibility that paths could be easily programmed; and along with the controller, it could be scaled.

However, the above mentioned mechanism all run in the control plane, which implies that they are not able to sense the data plane situation in real-time, thus they are mostly for relative static planning/controlling (minuets, hours or even day-level) of network traffic and not able to adapt to the microscopic traffic change in real-time (e.g. mili-second level). So, in real bearer networks (metro, backbones etc.), it is always underload so that the redundant resources could tolerant the traffic burst, results in a significant waste of network resources.

This draft proposes the iCAN (Instant Congestion Assessment Network) architecture to achieve autonomous adapt to traffic changes in real-time in terms of switching flows between multiple forwarding paths. iCAN includes following things:

- A mechanism between ingress and egress nodes to assess the path congestion situation in RTT level speed, to recognize which paths are underload and which are heavy loaded.
- Recognizing big flows and small flows in the device, in real time
- Ingress node dispatches flows to multiple paths, to make load balance, or to guarantee SLA for specific flows

This draft also discusses use cases and implementation scenarios of iCAN.

2. iCAN Architecture and Key Technical Requirements

2.1. Architecture
As above figure shows, there are 3 entities:

1. **Controller**
   - Responsible for planning multiple paths for a set of flows that could be aggregated to a pair of Ingress/Egress routers.
   - After delivering the planned paths to the ingress router, the controller would need nothing to do.

2. **Ingress router**:
   - Serves as a local "controller" for the iCAN system.
   - Responsible for triggering the path congestion assessment, which is coordinated with the egress router through a measurement protocol.
   - After getting the assessment results, the ingress router would calculate which flows need to be switched to a different path, in order to make the paths load balanced or to assure the transport quality of a certain of important flows.
   - In order to do the path switching calculation, the ingress router needs to recognize the TopN flow passing by it, since switching the big flows would make the most effort.
3. Egress router:
   - Only needs to coordinate with the ingress router to do the path assessment.

2.2. Key technical requirements

2.2.1. Path quality assessment

   o Req-1: the assessment MUST reflex the congestion status of the paths. (Note: a candidate congestion metric is proposed as: [I-D.dang-ippm-congestion].)

   o Req-2: the assessment SHOULD be done within a RTT timeslot. Since iCAN is to adapt the traffic change in real-time, the assessment needs to be done very fast.

   o Req-3: the assessment MUST be done for multiple paths between the same ingress/egress routes simultaneously. (Note: a candidate congestion metric is proposed as: [I-D.dang-ippm-multiple-path-measurement].)

2.2.2. Recognition and statistic of flows in devices

   o Req-1: the device SHOULD be able to recognize TopN big flows within a timeslot.

   o Req-2: the device MAY need to statistic all flows’ amount within a timeslot.

2.2.3. Flow switching between paths

   o Req-1: the device SHOULD be able to recognize flow let. The flow switching is done from the next flow let.

   o Req-2: the device MAY need to actively generate gap to artificially create flow let. If the flow needs to be switched immediately, then the device would need to make the gap, to avoid out-of-order packets arriving to the destination through multiple paths.

   o Req-3: the device SHOULD avoid oscillation of frequently switching flows from one to another.

   o Req-4: multiple ingress devices SHOULD be able to coordinate so that they won’t switch flows to the shared path at the same time, to avoid potential congestion in the shared path.
3. Use Cases of iCAN

3.1. Network load balancing

Background problem: traffic is not balanced in current metro network.

While some links are heavily loaded, others might be still lightly loaded: unbalance could lows down the service quality (e.g. SLA could not be guaranteed in the heavily loaded links/paths); unbalance could lows down the network utilization ratio (normally with 30%, e.g. a 100G physical capacity network can only bear at most 30G traffic, a huge waste of network infrastructure).

iCAN could be used for load balance among the multiple paths between a pair of ingress/egress nodes. Once the network is balanced, the real throughput of the network could be elevated significantly.

3.2. SLA assurance

Since iCAN could switch flow in real-time, it can guarantee a set of important flows. Once the path which carries the important flows is to be congested, the other flows could be switched to alternative paths, and the important flows would stably running in the original path.

(More content TBD)

3.3. Fine-Granularity reliability

Traditional reliability protocols such as BFD, can only assess the link on or off. With the path congestion assessment ability, iCAN could also asses the quality degradation.

(More content TBD)

4. Implementation Scenarios

4.1. iCAN with SRv6

- SR Multiple Explicit Paths

For example, there are 3 paths between the ingress and egress nodes, and the multi-path is defined as a SR-List containing LSP1/2/3.

The probe message detects the congestion status of the three SR-list paths. The edge device adjusts the load balancing between the three paths according to the congestion status of the three
SR-lists, and switch the flows from the path with a high congestion to the path with a low congestion.

- SR Multiple Explicit+Loose Paths

In loose path scenario, there needs to be an additional approach to probe the specific paths of a SR tunnel. After that, operations on the probed paths are the same as explicit path scenario.

4.2. iCAN with VxLAN

TBD.

4.3. iCAN with MPLS/MPLS-TE

TBD.

5. Standardization Requirements

1. Multi-path Planning (North Interface between Controller and devices)

2. Path Congestion Assessment (Horizontal Interface between devices), mostly regarding to Req-1&2&3 described in Section 2.2.1.

3. Flow Switching Negotiation (Horizontal Interface between devices), mostly regarding to Req-3&4 described in Section 2.2.3.

(More content TBD.)

6. Security Considerations

TBD.

7. IANA Considerations

TBD.

8. Acknowledgements

Very valuable comments were from Shunsuke Homma, Mikael Abrahamsson and Bruno Decraene.

A commercial router hardware based prototype had been implemented to prove the machinisms discussed in the document are workable.
9. References

9.1. Normative References


9.2. Informative References


Author’s Address

Bing Liu
Huawei Technologies
Q14, Huawei Campus
No.156 Beiqing Road
Hai-Dian District, Beijing 100095
P.R. China

Email: leo.liubing@huawei.com