**3GPP TSG-RAN WG3 #112-e [R3-212680](C:\\Temp\\RAN3 docs\\112\\Inbox\\Drafts\\CB # 40_IAB_CongestionMitigation\\Inbox\\R3-212680.zip)**

**17-27 May 2021**

**Online**

Agenda Item: 13.3.1

Source: ZTE (moderator)

Title: Summary of Offline Discussion on IAB Congestion Mitigation

Document for: Approval

# Introduction

**CB: # 40\_IAB\_CongestionMitigation**

**- (E///)**

**specify the CP-based congestion indication with per child granularity.**

**discuss whether there are merits of reporting granularity per BH RLC CH ID and per BAP routing ID.**

**specify the packet marking-based approach for UP-based congestion mitigation.**

**- (CATT)**

**Support CP-based congestion mitigation per BAP routing ID and per BH RLC CH ID.**

**Reuse current DDDS for UP-based congestion mitigation.**

**- (Persp,CISA)**

**IAB-gNB-CU should exempt priority traffic (e.g. MPS) from throttling on a congested IAB-DU or at intermediate IAB-nodes up to the point where backhaul congestion mitigation cannot be achieved without throttling the priority traffic, otherwise a user authorized to receive MPS priority service might not receive MPS priority handling when in an IAB session during congestion**

**- (Nok)**

**CP-based congestion indication is only reported per BH RLC channel.**

**Reporting the CP-based congestion indication per BAP Routing ID is not required from an IAB node.**

**No enhancements are introduced to UP-based congestion mitigation.**

**- (SS)**

**CP-based congestion indication can be reported per BAP routing ID or per BH RLC CH + Child node BAP address.**

**IAB Congestion Indication IE should be present when the gNB-DU Overload Information IE is set to “not-overloaded”.**

**CP-based indication is triggered by IAB donor CU-CP polling, while the CU-CP polling is triggered by the congestion indication per GTP-U tunnel from the CU-UP.**

**“non-overloaded” indication is not needed**

**In Rel-17, legacy DDDS is used for UP-based E2E congestion mitigation.**

**- (ZTE)**

**The following three types of congestion indication are supported in CP-based congestion mitigation: 1) per BAP routing ID; 2) per child link; 3) per BH RLC CH ID. Which type of congestion indication to be reported could be configured by donor-CU.**

**Event-based reporting mechanism could be used for the trigger of CP-based congestion indication, e.g., IAB donor-CU could configure IAB node with the threshold.**

**Regarding the co-existence of IAB Congestion Indication IE and the gNB-DU Overload Information IE, we suggest to leave it to IAB-DU implementation.**

**choose the “do nothing” option, i.e. use current DDDS as it is for IAB DL end-to-end UP-based flow control.**

**- (Len,Moto)**

**Nothing needs to be enhanced for DDDS in the IAB DL E2E flow control.**

**Per child link level (or per child node level) reporting can be the baseline for CP based congestion indication.**

**Per BH RLC CH level reporting can be introduced upon per child link level in order for bearer mapping reconfiguration.**

**Per BAP routing ID level reporting is not used for CP based congestion indication.**

**- (LG)**

**CP-based congestion indication should contain reporting per BAP routing ID, per child link and BH RLC CH ID.**

**A choice structure can be considered so that the gNB-CU-CP applies only backhaul congestion mitigation actions when to receive the IAB Congestion Indication IE using the gNB Status Indication procedure.**

**- (HW)**

**Specify per child link level congestion report from parent DU to IAB-donor-CU-CP in R17.**

**introduce the congestion level more than 1 bit for CP-based congestion indication report.**

**reuse the existing DDDS solution and introduce no enhancement to the DL E2E flow control.**

**- Chair: on congestion indication, views are still split; Any additional enhancements needed for CP-based (e.g. 1889)? On UP-based approach, seems some consensus to reuse current DDDS?**

(ZTE - moderator)

Summary of offline disc [R3-212680](C:\\Temp\\RAN3 docs\\112\\Inbox\\Drafts\\CB # 40_IAB_CongestionMitigation\\Inbox\\R3-212680.zip)

The deadline for Phase 1 is Thursday May20th, 23:59 UTC.

The deadline for Phase 2 is Tuesday, May 25, 12:00 UTC.

# For the Chairman’s Notes

**...**

# Discussion

## CP-based congestion mitigation

CP-based approach for DL congestion mitigation was agreed to be supported in Rel\_17 IAB in RAN3#111-e meeting. In CP-based approach, the congested IAB-node could send a congestion indication to the donor-CU-CP via F1AP signaling. And then the donor-CU-CP could adjust the route of the corresponding F1-U GTP tunnel based on the congestion indication. The granularity of the congestion indication was discussed in last meeting, and it was agreed to consider three types of congestion indication, i.e. per BAP routing ID, per child link and per BH RLC channel. Reporting per BAP routing ID can directly help the donor CU-CP to determine which F1-U GTP tunnel needs to be adjusted. However, if congestion occurs due to the link quality degradation in the backhaul link, it is better to report per child link congestion to donor CU-CP. Besides, if congestion only exists in a specific BH RLC channel, it is more appropriate to report the corresponding BH RLC channel to the IAB donor CU. Contributions [6] proposes to support these three types of congestion indication and the type of congestion indication to be reported by the IAB-node can be configured by donor-CU. In sum, the following 4 options could be considered about the granularity of the congestion indication:

* Option 1: per BAP routing ID
* Option 2: per child link
* Option 3: per BH RLC channel ID
* Option 4: configured by the donor-CU on which type of congestion indication is reported, i.e. BAP routing ID, or child link or BH RLC channel ID.

Companies are invited to provide their views on the above 4 options.

***Q1: Which option(s) do you prefer about the granularity of congestion reporting?***

|  |  |  |
| --- | --- | --- |
| **Company** | **Option** | **Comments** |
| ZTE | 4 | In our view, CP-based congestion indication could be used to mitigate long-term congestion by adjusting the route of the corresponding F1-U GTP tunnel. Reporting per BAP routing ID can directly help the donor CU-CP to determine which F1-U GTP tunnel needs to be adjusted. However, if congestion occurs due to the link quality degradation in the backhaul link, it is better to report per child link congestion to donor CU-CP. If congestion only happens to a specific BH RLC channel, it is more appropriate to report the corresponding BH RLC channel. Beside, reporting per BH RLC channel ID can achieve the finest granularity, and reminds the donor CU-CP the potential QoS impact. Each of the above three options is beneficial to help donor CU-CP to mitigate congestion by adjusting the routing configuration of the corresponding F1-U GTP tunnel. So we suggest that the above three types of congestion indication are all supported. And IAB node could report the specific type of congestion indication based on donor-CU configuration. |
| Lenovo | 2 and 3 | Since the IAB-donor-CU-CP is responsible for centralized resource, topology and route management for the IAB network. And these management functionalities are performed based on a per-link criterion.  In addition, per BH RLC CH level can be also introduced upon per child link level for bearer mapping reconfiguration. |
| Nokia | 3 | Based on the following observations in our contribution [4]:  Observation 2: On a given child link, some BH RLC channels can be congested while others are not.  Observation 3: Congestion impacting a given Routing ID equally impacts all the Routing IDs sharing the same BH RLC channel. |
| Samsung | 1 and 3 | This two granularities are the same as HbH flow control, which mean that the buffer at each IAB node is maintained at both granularities.  We understand the CP-based solution aims at congestion control as well, and it mainly used when the HbH flow control and E2E flow control cannot work. Thus, we didn’t say much difference.  On the other hand, if per BH RLC CH and per BAP routing ID are used, per child link can be achieved. |
| **Ericsson** | **2** | **Point 1:** If we report per BH RLC CH ID, we also need to report the corresponding child ID (BAP address). If we report per BAP routing ID, we also need to report the corresponding BH RLC CH ID from the routing table entry, meaning that we also need to report the child ID.  **Point 2:** We also need to account for the **case when the entire link between two nodes is affected**, and it would be really bad to have to report all BH RLC CH IDs, instead of just reporting the child ID.  **Point 3:** We also feel that reporting per BAP routing ID induces a large overhead, especially if buffers per BAP routing ID need to be supported.  So, we propose to **go forward with reporting per child link** and **preclude reporting per BAP routing ID.** |
| LG | 1, 2 and 3 | Reporting per BAP routing ID can directly help the donor CU-CP to determine which F1-U GTP tunnel needs to be adjusted. Through reporting per child link, the IAB-donor-CU-CP can know that a parent IAB-node transmitting the congestion indication has the congestion link with which child IAB-node. Also, in case a parent IAB-node has several BH RLC channels with the child IAB-node, it is able to detect whether each of its BH RLC channels with the child IAB-node is congested or not. |
| Qualcomm | 3 | Option 3. Buffer is allocated per BH RLC channel. Therefore load report should have BH RLC CH granularity. |
| CATT | 1 and 3 | CU-CP configures the BH information which including the BAP routing ID and egress BH RLC CH List. And IAB-DU detects the congestion which is per BH RLC CH. Therefore, BH RLC channel and BAP routing ID is the most useful information for CP.  Furthermore, if consider the BH RLC CH is not unique within a donor (maybe), CU would be confused when only report congestion per BH RLC. The BAP routing ID could help. |
| Fujitsu | 4 | We think each reporting granularity has its usage on the congestion mitigation, such as routing reconfiguration, BH RLC CH reconfiguration, etc. Different congestion indication type may be used under different scenario. We support all these granularities and support CU configuration on the congestion indication type. |
| Huawei | 2 | In our view, per child link feedback can also be seen as per child node feedback. With such feedback, the IAB-donor-CU-CP knows the congestion status of each BH link of the whole topology, so it can be aware of which routing path suffers congestion naturally, and then it can adjust the routing configuration or the resource allocation by implementation.  Per BH RLC CH report can provide finer [granularity](https://dict.youdao.com/search?q=%5B%E5%B2%A9%5D%20granularity%0D%0A&keyfrom=fanyi.smartResult" \t "_blank) of feedback. But in real deployment scenario, it is unlikely that a link is congested due to some reason (e.g. the link status deteriorates, or the link resource is not enough), but only part of the BH RLC CH are congested. So in most congestion cases, the congestion status of all the BH RLC channels may need to be reported to the IAB-donor-CU, which is an over-kill design and will cause very large overhead.  Similarly, if some BH link is congested, the per BAP routing ID report will also cause large overhead since all the BAP routing IDs which share the congested link should be reported, and this is also not necessary.  So we think the per child link feedback is enough, the other two granularity will result in superfluous design for the congestion feedback. |
|  |  |  |
|  |  |  |
|  |  |  |

Contribution [5] and [6] discuss the triggering of CP-based congestion indication. Contribution [5] proposes the CP-based congestion indication can be triggered by IAB-donor CU-CP polling, while such polling is triggered by the CP-UP indication. Contribution [6] states that event-based reporting mechanism could be used for the trigger of CP-based congestion indication. For example, IAB donor-CU could configure IAB node with the available buffer size threshold. When the threshold is satisfied, the congestion report should be triggered and sent to donor CU.

Companies are invited to provide their views on the following questions.

***Q2: Do you agree to define trigger condition for CP-based congestion indication?***

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes or No** | **Comments** |
| ZTE | Yes | In CP-based congestion mitigation method, donor-CU could adjust the route of the corresponding F1-U GTP tunnels. As a result, the topology controlled by the CU may be updated, and BH configuration of the IAB-nodes may be modified. So it is better for donor-CU to control the trigger of CP-based congestion mitigation mechanism, and IAB-node reports congestion indication to donor-CU only when some condition is satisfied. |
| Lenovo | No | It can be left to implementation.  For CP based congestion report, it has been agreed to reuse the F1AP GNB-DU Status Indication procedure, However, there is also no defined trigger condition for the F1AP GNB-DU Status Indication procedure. |
| Nokia | No | No standardized triggers needed in our opinion. This can be up to the implementation. |
| Samsung | Yes | We have define several congestion control schemes. Those schemes should be applied in different situations:   * HbH flow control: short-term congestion * E2E flow control: a relative long-term congestion * CP-based: all above schemes do not work   So, the CP-based scheme can be triggered when UP-based scheme cannot work.  In summary,   * **CP-based signalling is triggered by polling from the CU-CP** * **The polling is triggered when CU-UP indicates the congestion.** |
| **Ericsson** | **No** | Agree with Nokia. |
| LG | No | Share view with Lenovo and Nokia. |
| Qualcomm | No | However, it is not straightforward to properly define trigger conditions in absence of a congestion metric. We don’t want to spend to discuss congestion metrics. |
| CATT | Tend to agree | We have the common understanding that CP based congestion mitigation can be triggered when UP based cannot work. But the question is who has the ability to trigger CP-based, IAB node (buffer size) or UP (DDDS cannot work)? UP as a trigger condition tends to up to implementation |
| Fujitsu | Yes | The CP-based congestion mitigation approach may affect the other IAB nodes than the reporting IAB node. For example, donor-CU may reroute traffic from a congested path to another path, which may cause high load to the other path. It is better to standardize the trigger of the CP-based congestion indication, to let the CU has a fair view of the congestion status of the whole network and take proper congestion mitigation approach. |
| Huawei | No | Agree with Lenovo and Nokia |
|  |  |  |
|  |  |  |

***Q3: If your answer to Q2 is yes, please provide the preferred trigger mechanism, event-based or polling or other mechanisms?***

|  |  |
| --- | --- |
| **Company** | **Comments** |
| ZTE | In our view, event-based reporting mechanism could be used for the trigger of CP-based congestion indication. For example, IAB donor-CU could configure IAB node with the available buffer size threshold. When the threshold is satisfied, the congestion report should be triggered and sent to donor CU. |
| Samsung | Polling scheme is enough. |
| Fujitsu | We think both event-based and polling can be supported. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Contribution [9] proposes, in order to help IAB-donor-CU-CP make better decision for congestion mitigation, it is better that the congestion report can reflect the congestion level, e.g. the congestion level may be low, medium or high. Companies are invited to provide their views on the following question.

***Q4: Do you agree to introduce the congestion level for CP-based congestion indication report?***

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes or No** | **Comments** |
| ZTE | No | There are many issues of the congestion level, such as how to measure the congestion level, what level of congestion can be regarded as low or medium or high, is this configured by donor-CU or is this up to IAB-node implementation? This requires additional standard impact. |
| Lenovo | Yes | Congestion level is beneficial for the IAB-donor-CU to perform finer reconfiguration. |
| Nokia | No | Measuring congestion level seems to come with similar ambiguity issues as event-based reporting. |
| Samsung | No for now |  |
| **Ericsson** | **No** | Agree with Nokia. |
| LG | No | Similar view with ZTE. |
| Qualcomm | No | We would have to define a congestion metric first. We don’t want to do that. |
| CATT | No | When CP based congestion migration is performed, it is more likely already be highly congested |
| Fujitsu | See comment | We think that the report can indicate the available buffer size, which reflects the congestion level, similar to the HbH flow control indication. |
| Huawei | Yes | The congestion level will be helpful for CU-CP to know the congestion status for the BH link, and can provide more precise reaction to mitigate such congestion. |
|  |  |  |
|  |  |  |
|  |  |  |

A CR which captures the agreements on the CP-based congestion indication was approved during last RAN3 meeting [10]. Considering that the *gNB-DU* *Overload Information* IE is mandatory, the co-existence of the *IAB Congestion Indication* IE and the *gNB-DU* *Overload Information* IE was brought out during last meeting. Since no consensus was achieved, an FFS issue was left in the CR:

|  |
| --- |
| Editor’s NOTE: The handling with respect to simultaneous presence of *IAB Congestion Indication* IE and the *gNB-DU* *Overload Information* IE is FFS. |

Contribution [8] proposes to introduce a choice structure so that the gNB-CU-CP applies only backhaul congestion mitigation actions when receiving the *IAB Congestion Indication* IE using the gNB Status Indication procedure. The CHOICE structure in [8] is shown in below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **IE/Group Name** | **Presence** | **Range** | **IE type and reference** | **Semantics description** | **Criticality** | **Assigned Criticality** |
| Message Type | M |  | 9.3.1.1 |  | YES | reject |
| Transaction ID | M |  | 9.3.1.23 |  | YES | reject |
| CHOICE *gNB-DU Status* | M |  |  |  | YES | reject |
| >*Overload* |  |  |  |  |  |  |
| >>gNB-DU Overload Information | M |  | ENUMERATED (overloaded, not-overloaded) |  | - |  |
| >*IAB Congestion* |  |  |  |  |  |  |
| >>IAB Congestion Indication | M |  |  |  |  |  |

Contributions [2], [5], [6] indicate if GNB-DU STATUS INDICATION is only triggered by IAB congestion, IAB-DU can set the *gNB-DU* *Overload Information* IE as “not overloaded” considering that the *gNB-DU* *Overload Information* IE is mandatory. In this way, donor-CU could know only IAB congestion happens to the IAB-DU, and adjust the route of the corresponding F1-U GTP tunnel. In moderator’s opinion, how to set these two IEs depends on IAB-DU implementation. Considering the CHOICE structure cannot handle the case that both IAB congestion and traditional overload occur, **the moderator suggests that the handling with respect to simultaneous presence of *IAB Congestion Indication* IE and the *gNB-DU* *Overload Information* IE is up to implementation.**

Companies are invited to provide their views on the following question.

***Q5: Do you agree to leave the co-existence of IAB Congestion Indication IE and the gNB-DU Overload Information IE to IAB-DU implementation?***

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes or No** | **Comments** |
| ZTE | Yes |  |
| Lenovo | Yes |  |
| Nokia | Yes |  |
| Samsung | Yes |  |
| **Ericsson** | **Yes** |  |
| LG | Yes |  |
| Qualcomm | Yes |  |
| CATT | YES |  |
| Fujitsu | Yes |  |
| Huawei | Yes |  |
|  |  |  |
|  |  |  |

Contribution [3] states during IAB operation under congestion, the MPS traffic could be subject to IAB congestion mitigation mechanisms. To satisfy requirements per TS 22.153 and as implemented in TS 29.274, contribution [3] proposes normative text against TS 38.473 to take those requirements into account in the IAB congestion mitigation mechanisms. The text is copied in below.

|  |
| --- |
| If the *IAB Congestion Indication* IE is contained in the GNB-DU STATUS INDICATION message, the gNB-CU shall, if supported, take it into account for backhaul congestion mitigation. Depending on regional/national requirements and network operator policy, priority traffic (e.g. MPS) at the IAB-DU and at intermediate IAB-nodes shall be exempted from overload reduction policy throttling at the gNB-CU up to the point where the backhaul congestion mitigation cannot be achieved without throttling the priority traffic. |

Companies are invited to provide their views on the following question.

***Q6: Do you agree to capture the above text provided by contribution [3] into TS 38.473?***

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes or No** | **Comments** |
| ZTE | No | We think the MPS traffic is also subject to gNB-DU overload mitigation mechanisms. But there is no related description about MPS in the current GNB-DU STATUS INDICATION procedure. So it is no need to specify it for IAB congestion. |
| Lenovo | No | The detailed congestion mitigation policy is up to CU implementation. |
| Nokia | No | We think the exact congestion-mitigation actions taken by the donor CU are anyway up to implementation. |
| Samsung | No | Congestion mitigation is implementation issue. |
| **Ericsson** |  | Up to implementation |
| LG | No | Up to implementation. |
| Qualcomm | No |  |
| CATT |  | Up to implementation. |
| Fujitsu | No | Agree with ZTE. |
| Huawei | No |  |

## UP-based congestion mitigation

According to RAN3#111-e meeting, the following IAB DL end-to-end UP-based flow control enhancement solutions were proposed and which one would be selected should be further discussed.

Solution 1: Packet marking;

Solution 2: “do nothing” option, i.e. use current DDDS as it is.

Contribution [1] indicates that solution 1 enables early detection of potential congestion at the very place where the congestion occurs, i.e. at intermediate nodes, as opposed to the “do nothing” solution, which is reactive i.e. kicks in when the congestion has already manifested itself on packet losses or large delays.

Contributions [2], [4], [5], [6], [7] and [9] propose to support solution 2 and point out solution 1 is complex, e.g. imposing too much standard impacts, requiring cross-layer interaction between BAP and GTP, increasing overload for not only RAN3 but also RAN2 .

Companies are invited to provide their views on the following question.

***Q7: Please state your preference with respect to solution 1 and solution 2.***

|  |  |  |
| --- | --- | --- |
| **Company** | **Solution 1 or solution 2** | **Comments** |
| ZTE | Solution 2 | Solution 1 is complex and requires huge RAN2/3 work. |
| Lenovo | Solution 2 | IAB-donor-CU-UP can deduce the congestion is occurred in access link or backhaul link based on the combination of the highest NR PDCP PDU sequence number delivered to the UE and the desired buffer size. So, nothing needs to be enhanced for DDDS in the IAB DL E2E flow control. |
| Nokia | Solution 2 |  |
| Samsung | Solution 2 |  |
| **Ericsson** | **1** | *“IAB-donor-CU-UP can deduce the congestion* ***is occurred*** *in access link or backhaul link”*  E///: Note the “occur**ED**” – when congestion occurs, it may be too late. In other words, **the DDDS cannot indicate early signs of congestion**, which is quite desired in IAB network where the backhaul consists of multiple wireless hops. **DDDS rather indicates the consequence** – the packets are already being delayed due to e.g. congestion. Packet marking provides **an early warning that the buffers are growing.**  *“IAB node should maintain a timer for each packet”*  E///: Not really - the node can **simply observe that buffer is growing and mark** the corresponding packets.  *"the volume of marked packet cannot reflect the real congestion situation”*  E///: It **can be discussed** what exactly is reported. This **does not have to be the volum**e, it could be the fraction of marked packets since the last report, for example, or **a simple indication that buffer queues for the packets on this route are growing**.  *“RAN2 and/or RAN3 would need to specify the exact definition and measurement of delay at an IAB node, and based on this definition, triggers for packet marking.”*  E///: **Neither a definition of delay nor threshold definition are necessary**. It is **up to node implementation** to decide if the buffers are growing, which should be straightforward. Moreover, even today we have **“overload” indication in F1AP spec, without defining what an overload is**. Why cannot the same principle be applied here?  We understand that many companies voted against, but we need to have **a proper discussion**, without ignoring each other’s’ arguments. |
| Qualcomm | See comment | Solution 1 might provide some value. There is some specification overhead affecting RAN3 and RAN2. It is not obvious if the benefit justifies this effort, especially in light of the rather overwhelming opposition. |
| CATT | NO | We ack the benefit of solution 1. But we do not need to introduce that much work for early detective congestion by UP. We already have HBH flow control. Moreover, “**early**” congestion report also depends on the threshold setting by UP. |
| Fujitsu | No strong view. |  |
| Huawei | Solution 2 | Agree Lenovo |
|  |  |  |

# Part II…[if needed]

If needed

# References

1. R3-211727 Congestion Mitigation in IAB Networks (Ericsson)
2. R3-211802 Congestion Mitigation for CP-based and UP-based (CATT)
3. R3-211889 (TP for IAB BLCR 38.473) IAB Congestion Mitigation MPS exemption (Perspecta Labs, CISA ECD)
4. R3-211894 Analysis on Congestion mitigation (Nokia, Nokia Shanghai Bell)
5. R3-211943 Discussion on CP-based and UP-based congestion mitigation in Rel-17 IAB (Samsung)
6. R3-212040 Discussion on congestion control in R17-IAB (ZTE)
7. R3-212166 Discussion on congestion mitigation for IAB (Lenovo, Motorola Mobility)
8. R3-212393 Issues on CP-based congestion indication (LG Electronics)
9. R3-212416 (TP for NR\_IAB\_enh BL CR for TS 38.473): IAB E2E congestion mitigation (Huawei)
10. R3-211360 CP-based Congestion Indication for IAB Networks.