**3GPP TSG-RAN WG3 Meeting #111-eR3-211005**

**Online, January 25th – February 4th 2021**

Agenda Item: 13.3.1

Source: Ericsson (moderator)

Title: Summary of Offline Discussion on IAB Congestion Mitigation

Document for: Approval

# Introduction

This is the SoD for the following comeback: **CB: # 38\_IABcongestionMitigation**

The deadline for providing replies to Phase 1 is **Thursday, January 28th at 23.59 UTC.**

Relevant papers:

1. R3-210220 Discussion on CP-based and UP-based congestion mitigation in Rel-17 IAB (Samsung)
2. R3-210350 Enhancements to congestion control for IAB (Qualcomm Incorporated)
3. R3-210388 Congestion Indication to CU-CP (Intel Deutschland GmbH)
4. R3-210460 Discussion on CP-based approach for DL and UL congestion mitigation in IAB networks (Fujitsu)
5. R3-210490 Analysis on Congestion mitigation (Nokia, Nokia Shanghai Bell)
6. R3-210550 Discussion on IAB E2E flow control (Huawei)
7. R3-210614 Discussion on congestion mitigation for IAB (Lenovo, Motorola Mobility)
8. R3-210718 Discussion on DL E2E flow and congestion control in R17-IAB (ZTE)
9. R3-210724 Congestion Mitigation in IAB Networks (Ericsson)
10. R3-210781 IAB End-to-End Flow Control Feedback’ (Ericsson)
11. R3-210781 Issues on CP-based approach for DL congestion mitigation (LG Electronics)

# For the Chairman’s Notes

**Proposal 1-1: The CP-based congestion indication contains a choice between reporting per 1) BAP routing ID and 2) child link and/or BH RLC CH ID.**

**Proposal 1-2: The CP-based congestion indication reuses the F1AP GNB-DU Status Indication procedure.**

**Proposal 1-3: The CP-based congestion indication pertains to DL congestion.**

**Proposal 1-4: The trigger for sending the CP-based congestion indication is up to implementation.**

**Proposal 1-5: The CP-based congestion indication contains the “congested” and “not congested” indications.**

**Proposal 2-1: RAN3 considers the following two options for the UP-based approach to IAB congestion mitigation:**

* **No enhancements;**
* **Packet marking-based approach.**

**Proposal 2-2: RAN3 to discuss the potential concerns related to packet marking.**

**NOTE: the green color is for convenience of the Chairman, these proposals are still to be discussed.**

# Discussion

## CP-based congestion mitigation

Contributions [1]-[9] and [11] discuss CP-based congestion mitigation. The issues that can be derived from the contributions are:

**Issue 1:** Granularity of congestion reporting, i.e. per: Child link, BH RLC channel, BAP routing ID, configurable?

***Q1-1: What should be the granularity of congestion reporting?***

---------------------

**Issue 2:** F1AP procedure used: GNB-DU Status Indication, F1AP Notify, a new F1AP procedure?

***Q1-2: Which F1AP procedure should be used for CP-based congestion reporting?***

---------------------

**Issue 3:** Support for DL only or both UL and DL?

***Q1-3: Should CP-based congestion reporting be supported only on DL or on both UL and DL?***

|  |  |
| --- | --- |
| **Company** | **Answer and motivation** |
| Ericsson | **Q1-1:** we think that reporting per child IAB node or per BAP routing ID is enough.  **Q1-2:** we proposed GNB-DU Status Indication, but we can consider other alternatives, as well.  **Q1-3:** only DL congestion indication should be supported.  Regarding the UL indication, we do not see a practical need for it. Also, an implication of the RAN3#110-e agreement that ‘An IAB node at the parent side of a congested backhaul link may send a congestion indication to the IAB-donor-CU-CP.’ is that DL-only congestion indication is supported. If UL congestion indication is to be supported, should the parent send the indication? |
| Intel | **Q1-1:** Child IAB node and BAP routing ID. Make it configurable  Q2-2: We are open  Q1-3: DL indication is sufficient |
| Lenovo | Q1-1: **Per child link level (or per child node level)** reporting can be the baseline for CP based congestion reporting, in order for the IAB-donor-CU to reconfigure the resource and routing.  Q1-2: **GNB-DU Status Indication is preferred**, and it can be enhanced to report per child link level.  Q1-3: **Only support DL congestion indication.** Since it has been agreed that End-to-end UL flow control is deprioritized in Rel17. |
| ZTE | **Q1-1:** Reporting per BH RLC channel or per BAP routing ID. In addition, per child link can also be supported to indicate the specific congestion IAB nodes.  **Q1-2:** we prefer to use GNB-DU Status Indication.  **Q1-3:** both UL and DL congestion indication can be supported.  If UL congestion occurs, since the CU-CP has full control of the network, it is reasonable to feedback the information to CU-CP for network optimization. |
| Samsung | Q1-1: per BH RLC CH and per BAP routing ID should be supported. This is aligned with the hop-by-hop flow control. Moreover, the congestion mitigation policy at the IAB donor CU side can be performed at such granularity.  Q1-2: GNB-DU Status Indication  Q1-3: DL only |
| CATT | Q1-1: prefer per child link.  Q1-2: GNB-DU Status Indication  Q1-3: only DL. Since Uplink congestion has little to do with CU-CP. UE/IAB node MT may require more SR/BSR which can be discussed in RAN2. Or simply release some UE/IAB node |
| Fujitsu | **Q1-1:** we think the reporting granularity can be configurable. Different report granularity can support different means of congestion mitigation, such as routing path reconfiguration, BH RLC channel remapping, etc. Donor-CU can configure IAB node to report congestion per child link, BH RLC channel, or routing ID through polling or a separate configuration procedure.  **Q1-2:** since the congestion reporting is closely related to IAB BAP configuration, and a polling may be used, we propose to use a new F1AP procedure.  **Q1-3:** we support for both DL and UL. Both RAN3 and RAN2 have discussed congestion indication for UL. UL congestion mitigation was not supported for UP-based approach mainly due to the implementation complexity. For CP-based approach, the reporting and mitigation methods are basically the same for DL and UL. It is better to support both from the beginning, with almost no extra effort. |
| LG | Q1-1: prefer BH RLC CH because a parent IAB node can detect whether each of its BH RLC CHs is congested or not  Q1-2: gNB-DU Status Indication  Q1-3: DL only |
| QC | Q1-1: The IAB-DU can be assumed to have a separate buffer per BH RLC channel and congestion is therefore specific to BH RLC CH. Providing additional information, e.g., per BAP routing ID may help the CU-CP, e.g., to perform efficient rerouting.  Q1-2: gNB-DU Status Indication is fine for UL reporting. For congestion polling, other message type is needed.  Q-13: DL and UL. UL for the following reason: This is not conventional flow control but load reporting so that the CU-CP can optimize resource allocation, routing, topology, etc. |
| AT&T | Q1-1: Per BH RLC channel and BAP routing ID  Q1-2: gNB-DU Status Indication  Q1-3: Downlink only should be sufficient |
| Huawei | Q1-1: **Per child IAB node**. With such feedback, the CU-CP know the congestion status of each BH link of the whole topology, and then it can be aware of which routing path suffers congestion naturally. And can adjust the routing configuration or the resource allocation by implementation  Q1-2: gNB-DU Status Indication  Q1-3: Downlink only |
| Nokia | **Q1-1:** BH RLC channel  **Q1-2:** We propose F1AP Notify, because it reports a status of BH RLC Channel, which aligns with the purpose of Notify procedure that is currently reporting the status of a DRB.  **Q1-3:** We propose support for both DL and UL. For the latter, the parent node has information from buffer-status reports received from child nodes. |
|  |  |

**Summary**:

**Q1-1:**

* 6 out of 12 companies support congestion reporting per child link;
* 7 out of 12 companies support congestion reporting per BH RLC channel;
* 7 out of 12 companies support congestion reporting per BAP routing ID;

In fact, reporting per child link is a superset of reporting per BH RLC channel, since BH RLC CH ID space is per child node, meaning that, together with the BH RLC channel ID indication, the IAB-DU must send the indication of the child ID. One way forward is to:

* Introduce the possibility to report per BH RLC CH ID, where the list of BH RLC CH IDs would be optional, while the child ID would be mandatorily present.
* Introduce the possibility to report per BAP Routing ID.

**Proposal 1-1: The CP-based congestion indication contains a choice between reporting per 1) BAP routing ID and 2) child link and/or BH RLC CH ID.**

**Q1-2:**

* 9 out of 12 companies propose to use the F1AP GNB-DU Status Indication procedure, 1 company prefers F1AP Notify and one company proposes to define a new procedure.

**Proposal 1-2: The CP-based congestion indication reuses the F1AP GNB-DU Status Indication procedure.**

**Q1-3:**

* 8 out of 12 companies think only DL congestion indication should be specified, while 4 companies think this should be specified for both DL and UL.

**Proposal 1-3: The CP-based congestion indication pertains to DL congestion.**

**Issue 4:** Congestion report triggering: Threshold-based, polling based, reporting rate set by donor CU-CP, up to implementation?

***Q1-4: Should the report triggering be controlled by the donor CU-CP (threshold-, polling-, maximum rate-based) or up to IAB-DU implementation?***

---------------------

**Issue 5:** Additional information in the congestion indication

***Q1-5: Should the following be reported:***

* ***Degree of congestion?***
* ***“Congestion over” indication?***

|  |  |
| --- | --- |
| **Company** | **Answer and motivation** |
| Ericsson | **Q1-4:** We should not standardize any thresholds or triggers. When the reporting is activated, the IAB-DU should be able to report when it thinks it is necessary, since it has the best knowledge of its own situation. This is the principle followed in legacy DUs. So, our answer is: **up to IAB-DU implementation.**  **Q1-5:** **disagree to “degree of congestion” indication**; in legacy DUs we only have indication of congestion (i.e. DU overload), so the same reasons apply here. For IAB, in Q1-1 we are discussing additional granularity of reporting, and that should be enough to cover the gap between the legacy DU and IAB-DU. “**Congestion over” indication is acceptable**, as something similar is supported in the legacy already. |
| Intel | **Q1-4:** We think the CU should be able to control the reporting  **Q1-5:** “**Congestion over” indication seems sufficient** |
| Lenovo | Q1-4: Like legacy gNB-DU Status Indication, the reporting is **up to IAB-DU’s implementation.**  Q1-5: **A specific congestion degree** can be used for CU-CP to finer reconfiguration than only congestion or not indication. |
| ZTE | **Q1-4:** Agree with Ericsson. We think it is **up to IAB-DU implementation.**  **Q1-5:** Agree with Ericsson. |
| Samsung | Q1-4: we need define polling scheme. RAN3 agreed that “UP-based and CP-based approaches for DL congestion mitigation in IAB networks are complementary.”. In our understanding, the CP-based solution is applied when the UP-based solution cannot solve the problem, this is only known by the IAB donor CU. Thus, the donor CU can poll IAB-DU when the UP-based solution does not work. If we leave it to IAB-DU implementation, it may cause that the congestion reporting is sent out when IAB donor CU still works on UP-based solution.  Q1-5: “congestion over” may not need. If IAB donor CU can identify the relief of congestion via UP-based information, such indication is not needed. |
| CATT | Q1-4: up to IAB-DU implementation.  Q1-5: we are open for “Degree of congestion”  “Congestion over” indication may be required for CU-CP |
| Fujitsu | **Q1-4:** we think both the threshold-based and polling-based triggers should be supported. Donor-CU-CP configures the threshold, reporting interval, etc., when necessary. The threshold can be the maximum number of BAP flow control feedback instances.  **Q1-5:** we think the degree of congestion is helpful for donor-CU-CP to judge the severity of the congestion. The congestion indication can include for example available buffer size or desired data rate.  “Congestion over” indication is also supported. |
| LG | Q1-4: up to IAB-DU implementation  Q1-5: For congestion over indication, we share view with Samsung. “Degree of congestion” may not need. Legacy DU provides only whether it is overload or not to CU. So, the same mechanism should be applied to IAB case. |
| QC | **Q1-4:** This may cause a **signaling explosion** if left up to implementation! To avoid such signaling explosion, the CU should be able to limit the report frequency or to obtain congestion report based on polling. Note that CP congestion reporting should be occasional events opposed to UP congestion feedback.  **Q1-5:** Degree of congestion and/or “Congestion over” may not be necessary. |
| AT&T | **Q1-4: We are OK to follow the legacy principle of gNB Status Indication and keep up to IAB-DU implementation.**  **Q1-5: Congestion over indication should be sufficient to allow the CU-CP to trigger corrective actions.** |
| Huawei | **Q1-4:** Agree with Ericsson. Just same principle for legacy gNB-DUs, it is **up to IAB-DU implementation.**  **Q1-5:** The baseline can be [overloaded, not overloaded], similar to the traditional gNB-DU’s overload indication.  Regarding the degree of congestion, it may be helpful for the CU-CP to know more details about the congestion status, but the granularity should be proper to saving the overhead.  The “congestion over” indication seems same as the “not overloaded”, if so, it is needed. |
| Nokia | **Q1-4:** We propose to leave this up to implementation.  **Q1-5:** We propose to report only “congested/no longer congested”. |
|  |  |

**Summary**:

**Q1-4:**

* 8 out of 12 companies propose to leave the report triggering up to implementation;
* 4 out of 12 companies propose that the CU controls the reporting.

**Proposal 1-4: The trigger for sending the CP-based congestion indication is up to implementation.**

**Q1-5:**

* 8 out of 12 companies propose to include “congestion over” indication and 3 companies are against.
  + Having in mind that the congestion indication is also to be reported, the Moderator agrees with Huawei that this is the same as we have in legacy i.e. “overloaded” / ”not overloaded” indication.
* 5 out of 12 companies are explicitly in favour of indicating the degree of congestion, while 4 out of 12 companies are against.
  + No proposal can be derived.

**Proposal 1-5: The CP-based congestion indication contains the “congested” and “not congested” indications.**

Note: this is related to Issue 2.

## UP-based congestion mitigation

Contributions [1]-[2] and [5]-[10] discuss UP-based congestion mitigation. The set of potential solutions includes:

* **Opt1:** Highest PDCP SN received from parent node;
* **Opt2:** Bitmap of PDUs transmitted to lower layers out of sequence;
* **Opt3:** Packet marking;
* **Opt4:** Received volume and Receiving data rate.
* **Opt5:** “do nothing” option, i.e. use current DDDS as it is

With respect to the above options, submitted papers propose the following:

* Paper [1] proposes to consider the reporting of **received date volume** and **received data rate**.
* Papers [2] and [9] propose **marking of DL packets that experience an increasing queuing delay** at intermediate nodes.
* Paper [5] proposes that, if needed, **for UM mode, the DDDS is enhanced with a bitmap of PDUs transmitted to lower layers out of sequence**.
* Paper [6] proposes an **indication of the highest PDCP SN received from parent** node.
* Papers [7] and [8] propose the “do nothing” option.

***Q2: Please state your preference with respect to Opt1-5.***

|  |  |
| --- | --- |
| **Company** | **Answer and motivation** |
| Ericsson | We support Opt3 i.e. **packet marking**.  Packet marking is **the only candidate solution that**:   * **Keeps buffer levels low**, and thus prevents congestion from happening. All other approaches are reactive. * Is **stable to traffic rate variations** (which may be intensive in short periods of time). * Is **resilient to DDDS delay**, since the indication is provided way before congestion occurs.   Regarding the **concerns** raised by some companies regarding:   * Specification impact *(“it is a cross-layer solution”):* the impact on BAP is only header 1 bit. BAP is the only layer that is accessible by intermediate nodes and is hence the only tool for the intermediate nodes to provide indications to the access node. * *How to set a delay threshold*? This is up to implementation of an IAB-DU. **It is funny that this same group of people is discussing CP-based congestion indication, but yet no one asks how to define a threshold for congestion.** * *“Cannot reflect the real-time congestion”*: the intention is not to detect real-time congestion, but to identify a trend in queuing delay increase in intermediate nodes, which is an early sign of potential congestion.   Regarding the remaining solutions:  **Received volume and Receiving data rate:** the rate varies all the time, so these indications bring no informational value to the donor. Moreover, the delivery of DDDS to the donor CU will take some time, during which both the received rate and the received data volume may change significantly, meaning that the reported information is outdated by the time it reaches the donor.  **Indication of the highest PDCP SN received from parent node:** today in DDDS we have the NR-U SN that indicates the delivery status on the transport network. The delta with respect to NR-U SN needs to be clarified. This approach is essentially reactive.  **Bitmap of PDUs transmitted to lower layers out of sequence**: the reporting and processing bitmap-based indications is computationally expensive. This approach is essentially reactive. |
| Intel | We like packet marking, since it is the only one that provide the earliest feedback to CU regarding congestion.  We are also fine with no changes. |
| Lenovo | All the Opt1 – Opt4 can help the IAB-donor-CU-UP to determine the congestion is occurred in access link or backhaul link, but the location where congestion occurs in the BH link is still unperceived by IAB-donor-CU-UP.  In addition, according to the information in the legacy DDDS frame, IAB-donor-CU-UP can already 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.  Therefore, **nothing needs to be enhanced for legacy DDDS**. |
| ZTE | We support **Opt5:** “do nothing” option, i.e. use current DDDS as it is  **For Opt1 and Opt4,** “Highest PDCP SN received from parent node” and “Received volume and Receiving data rate” aim at reporting the receiving status of the access IAB-node to Donor CU-UP which will help CU-UP to judge whether backhaul link is congested or not. However, based on the content included in the current DDDS(i.e. the “Highest NR PDCP PDU sequence number” & “Number of successfully delivered out of sequence PDCP Sequence Number range” for RLC AM and the “Highest transmitted NR PDCP Sequence Number” for RLC UM), CU-UP is able to roughly know the receiving status of DU.  **For Opt2:** “Bitmap of PDUs transmitted to lower layers out of sequence” solutions is not suitable for RLC UM. The transmission of RLC UM bearer is not based on ARQ. Once the transmitting IAB-DU sends the packet to the lower layer, it thinks the packet is successfully delivered to UE on Uu. In this case, the “Highest transmitted NR PDCP Sequence Number” included in DDDS can reflect all the packet received by DU from CU-UP. Thus, CU-UP is able to know the pending PDUs between CU-UP and DU.  **For Opt3:** “Packet marking” solution may not reflect the real-time congestion, and the delay thresholds may not be uniformly configured. |
| Samsung | * Locating the congestion point is not the requirement of UP-based solution   CP-based solution is used to locate the congestion point, and it is implemented when UP-based solution does not solve the problem. Moreover, all the UP-based solutions on the table cannot identify the congestion point.   * E2E congestion mitigation is a RAN3-only issue (For legacy F1-U, we never consult RAN2 before)   Opt 3 is the only solution requiring addition enhancements except DDDS. Although only one bit is added in the BAP header, it causes additional processing at the IAB node, which equivalently increases the transmission delay of the packet. Moreover, the benefit on top of DDDS-based solution is not clear.  Opt 1&2&4&5 are solely focusing on DDDS. This is the direction we want to proceed:   * Opt4 is our preference. For the concerns raised by companies:   + **Variation of receiving rate**: the receiving data is derived based on the receiving packets. If we are concerning its variation, we are still concerning the variation of the number packets received during a certain time. The information in Opt1~5 closely relies on the receiving packets. So, variation should be applied to each options rather than opt4. On the other hand, we think receiving rate is a statistic value averaged over a timing window. It cannot be variable very much (if there is too much variation, the network is not stable)   + **Delay of DDDS**: all the solutions rely on DDDS. Why delay is the only issue for opt4. Actually, DDDS is for flow control in legacy network, the delay of DDDS cannot be a factor impacting the efficiency for the flow control. In IAB network, we should ensure the timely transmission of DDDS as well. Otherwies, it will impact the legacy DDDS.   + **Legacy DDDS reflects information of Opt4**: in our contribution, we explain that legacy DDDS can only reflect accessing link. In IAB, we need some information for the BH link, which is reflected by opt 4. * **Opt 1&2**: opt1 reflects BH link status at some extent, it can be combined with option 4. Opt2 seems to consider a general issue rather than IAB-specific.   In summary, our proposal is:   * The UP-based solution aims at potential enhancement of DDDS only. * Opt4 + Opt1 can be considered as a package for further discussion * Opt 2 can be discussed as a general issue rather than IAB-specific |
| CATT | The reason why we consider the enhancement of DDDS is that the current DDDS only reflect the access link. The congestion in backhaul link cannot be reflect. In my understanding, the intermediate node may also are connected with some other UEs, which means that the intermediate node may be access node for other link.  Besides, if we assume that there is no UE connect to the intermediate node. The above option still do not address the issue that intermediate node cannot report congestion expect option 3 packet marking.  Specifically, option 1 and 4 are focus on shown the congestion which the current DDDS also can be achieved, e.g., “Highest successfully delivered NR PDCP Sequence Number”. If the “Highest successfully delivered NR PDCP Sequence Number” to UE is smaller than highest sequence number which IAB-donor-CU had transmitted. The CU can find the congestion.  For option 3, the spec impact too large. I am not sure we are worth to do it due to the UE may connect to intermediate node.  In a word, we prefer option 5 “do nothing” if no better enhancement are provided. |
| Fujitsu | We support Opt3 i.e. packet marking or Opt5 “do nothing”. The remaining solutions (Opt1, 2, 4) add little value over the current DDDS while causing more overhead. |
| QC | We support Option 3. We agree with Ericsson’s summary.  It is not clear if and how options 1, 2 and 4 can provide benefit over option 3 or option 5 (do nothing). |
| Huawei | We support option 1 to provide additional information to let CU-UP deduce how many packets are still on-the-flight. But also fine to accept option 5 “do nothing”, CU-UP can control the per-DRB packet transmission according to the current DDDS, |
| Nokia | Option 2 or Option 5  For Option 3, this is an option involving multiple protocols/layers, i.e. the measure of the delay at RLC (?), the marking would take place at BAP and reporting of marked packets over NR-UP. Hence it would involve both RAN2 and RAN3. It also seems to be the least clear option when it comes to details. For the reporting to be meaningful to CU-UP, 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. |

**Summary**: Ten companies responded to the survey and two most popular options are:

* Opt5: do nothing, receiving 6 votes;
* Opt3: packet marking, receiving 4 votes.

The Moderator thinks that the next step should be to discuss the concerns related to Opt3, and makes the following proposals:

**Proposal 2-1: RAN3 considers the following two options for the UP-based approach to congestion mitigation:**

* **No enhancements;**
* **Packet marking-based approach.**

**Proposal 2-2: RAN3 to discuss the potential concerns related to packet marking.**