

Age-of-Information Aware In-band Network Telemetry for Better Network Predictability

Jetmir Haxhibeqiri, Ramyashree Venkatesh Bhat, Ingrid Moerman, Jeroen Hoebeke
 IDLab, Ghent University – imec, Ghent, Belgium
 Email: [name.surname]@ugent.be

Abstract—In-band network telemetry (INT) monitoring is getting traction in the recent years for its ability to offer high-granularity network information on per-hop, per-flow and end-to-end basis. Such detailed network monitoring helps in better network management and accurate network (re)configuration, supporting automated network management and control. However, until now the collected INT information is used as is, without any classification based on its provided information value. In this paper we derive the Age of Information (AoI) mathematical model to calculate the INT hop-time average AoI and the impact of reporting on the INT AoI. Such metric can be used to reduce the network overhead in term of INT reporting by not sending old AoI INT and determining the information accuracy. We also show that the INT hop-time average AoI for a layer three INT implementation for a wireless network depends heavily on the communication latency between each hop and end hop, and that it is bounded by the end-to-end communication latency averaged over the number of hops.

Index Terms—INT, AoI, wireless networks, agile management

I. INTRODUCTION

Management of wireless professional networks is getting complex due to the diversity of network technologies being used, their increased programability and the increased demand for Quality of Service (QoS) by applications. New advancements in wireless networking concepts and technologies, like Software Defined Networking (SDN) [1], Software Defined Radios (SDR) [2], 6G and WiFi6 [3], offer higher flexibility in network (re)configuration capabilities in order to offer better network performance in line with the applications' demands.

For better network management and control, detailed and timely network monitoring approaches are required. Moreover, for network performance verification against the individual application demands, per-flow monitoring techniques are required. In recent years there have been a number of advancements in network monitoring such as in-band network telemetry (INT) [4], that offers detailed monitored information on per-hop, per-flow and end-to-end basis. Such techniques are being extended also to wireless networks for achieving low overhead network monitoring [5]. Different wireless networking parameters on per-flow and per-hop basis can be collected together with data packets without introducing any additional contention or overhead in the wireless links [6].

The collection of detailed information on the network performance for each traffic flow at each hop is only one side of the medal. The other side of the medal is how to ensure that the collected information does not become obsolete, causing

network (re)configurations based on wrong and outdated data. This is especially important in wireless networks where node and environment dynamics can cause such issues. As such, for highly dynamic networks, real-time monitoring is becoming a necessity in order to have timely network status updates. Thus, information collected using INT should be assessed not only on the amount of information collected, but also based on the freshness of the provided information. The Age of Information (AoI) is a new metric that is used to determine such information freshness, showing the time elapsed since the information was generated until the information is received at the central observer [7].

To mitigate the usage of obsolete information reported by INT, a generalized model to determine AoI collected by INT has to be derived. In this paper we derive the AoI of multi-hop and end-to-end INT. In section II we discuss related work on INT solutions as well as the usage of the AoI metric in other research fields. In section III we derive a general expression to calculate the hop-time average AoI collected by INT on hop-by-hop and end-to-end basis. In addition AoI of INT reports is derived as well. In section IV we show AoI results for a multi-hop network setup using our layer 3 INT implementation. Section V presents conclusions and further ideas on how the AoI metric can be used to make INT more dynamic and agile.

II. RELATED WORK

In the following subsections, we will discuss research works related to in-band network monitoring and its design for wired and wireless networks, as well as research works focused on AoI metric modeling for different network type cases.

A. In-band network telemetry

Efficient network monitoring is a cornerstone for reliable network management and network performance verification. Accurate and timely monitoring of the network will impact the accuracy of management applications in SDN-based networks. The recently introduced in-band network telemetry technique [4] offers a detailed overview of the network on a per-flow basis, both at each network hop and end-to-end. In-band monitoring is achieved by adding additional monitoring info to the data packets on each hop.

There have been a number of extensive studies regarding the applicability of INT in wired networks [8]–[13]. All these studies have focused on enabling INT in wired networks and its evaluation either using P4 programmable switches [10],

[11] or a simulation environment [12], [13]. In our previous studies, we have presented an INT design and implementation for wireless networks [6] and how the INT information can be used for SDN-based management [5], [14]. In this paper, for evaluating the AoI of INT information, we will use our layer three implementation of INT.

B. Age of information

Over the past years, there is an increased interest in using AoI as a metric for evaluating real-time monitoring status update performance in different scenarios: single-hop networks (single-user [15] or multi-user [16]), multi-hop networks [17], multi-access networks [18], [19]. Also, different studies focus on using AoI in Internet of Things (IoT) scenarios for different optimization problems like power usage of low power end-device [20]–[22].

The concept of the Age of Information (AoI) was coined by Kaul et al. in [7] to describe the time elapsed since the generation of information at its source and the reception of that information as part a message exchange. The authors of [7] also derived the first generalized form to calculate AoI for service system updates using a single queue. In [18], the authors investigate AoI performance of a system with multiple monitoring sources and single destination, under either scheduled channel access or random channel access. The time average AoI is modeled as the sum of the expectations of random variables that capture the AoI, namely *system time* and *inter-arrival time update*. Differently, in [15], the authors look at the system update AoI of a single source-destination system where the source node is composed of a single first-in-first-out (FIFO) queue with different AoI sensitive applications. The authors formulate the AoI optimization problem as a Markov Decision Process. In [17], a multi-hop (two-hop) network scenario is considered with multiple data flows (two data flows) that are scheduled using different priorities based on the AoI requirements. Lastly, in [19], the authors provide an AoI-aware scheduling policy for 1-to-many sensor applications, based on metrics such as average AoI and average application response time.

Deriving the AoI of INT information differs from the previous research work in two aspects. First, the information is not collected at a single point in the network, rather it is collected on each network-hop, introducing different AoI for different INT information parts. Second, the addition of new INT information in another hop will impact the AoI of the previous hops, which is not the case with the previous studies.

III. MATHEMATICAL MODEL FOR INT AOI

In a multi-hop network, the Age of Information of the INT monitored parameters depends on the number of hops that the data packet has traversed. The further the source node from the destination node, the higher the AoI will be at the destination node for the monitored parameters on the first hop. The AoI of the first hop also depends on the processing of INT data on the next hops until the destination is reached. This means the initial AoI of data collected in the first hop

will increase based on the communication latencies of the other hops until the destination is reached. Still, in this case the communication latencies can be measured on per-hop bases using INT, which will help to calculate the AoI at the destination. If the collected INT data are passed to a central entity in the network, then the INT AoI will be increased by the delay of such communication. For accurate decision making and network reconfiguration the monitoring information should have a low average AoI metric.

We will consider the hop-time average AoI of INT information. Let N be the number of hops in the network and t_i the time instance when the INT enabled packet is processed at network hop i . The processing delay of INT enabled packets at network hop i is denoted as δ_i . In Figure 1 we show an example of how INT AoI changes over time for hop-by-hop INT. The initial AoI of the information collected on the first hop will be the processing time of the INT packet on that node, δ_1 . The AoI of the information collected on the first node will continue to increase linearly over time until the INT packet is processed by the second node in the network at time instant t_2 . Due to processing delays the AoI of the first hop will increase by δ_2 , which is also the initial value of AoI for information collected at second hop. Both AoI curves for information collected at the first and second hop will continue to increase linearly over time until they get processed at the next hop, where another processing delay will introduce an additional increase in AoI of both flows. Considering the time average AoI metric of monitored information on hop i , it can be expressed as the area under the function Δ_i , normalized by the communication latency between hop i and the last hop, noted as T_i :

$$\overline{\Delta}_i = \frac{1}{T_i} \int_{t_i}^{t_N} \Delta_i \quad (1)$$

while the hop-time average of all INT data will be:

$$\overline{AoI}_{INT} = \frac{1}{N-1} \sum_{i=1}^{N-1} \overline{\Delta}_i \quad (2)$$

Based on Figure 1, the area under the Δ_i curves for each hop information is composed of an isosceles right-angle triangle (shown as light blue for the first hop curve) with base T_i , a rectangle with base T_i and height δ_i (shown as dark blue for the first hop curve) and parallelogram(s) with base T_x and height δ_x , for $i < x < N-1$ (shown as white parallelograms for the first hop curve). Putting everything together, the area under curve Δ_i can be calculated as follows:

$$\int_{t_i}^{t_N} \Delta_i = \frac{T_i^2}{2} + \sum_{j=i}^{N-1} T_j \delta_j \quad (3)$$

All terms in equation 3 depend on T_i that is proportional with the transmission delay between hops, while the second term also depends on the additional INT processing delays introduced on each hop, δ_i . So, the time average AoI will be minimized by minimizing the area under each Δ_i curve. This

is achieved by minimizing the INT processing delays on each node that will minimize the AoI collected at the first hop and improve the AoI of other hops. By substituting equations 1 and 3 in equation 2 we get (the full derivation can be found in Annex I):

$$\begin{aligned} \overline{AoI_{INT}} &= \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{1}{T_i} \left(\frac{T_i^2}{2} + \sum_{j=i}^{N-1} \delta_j T_j \right) \\ &= \frac{\mathbb{E}[T]}{2} + \mathbb{E}[\Delta] + \frac{1}{N-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N-1} \frac{\delta_j T_j}{T_i} \end{aligned} \quad (4)$$

where $\mathbb{E}[\cdot]$ is the expectation operator, and T and Δ are the random variables for communication latencies between each hop and the end destination and INT processing delay in each hop, respectively, and N is large. The first and second term in equation 4 depend only on the communication latency and INT processing delay of the respective hops, whereas the last term shows the impact of INT processing delays introduced by the preceding hops on the AoI collected on previous hops. The upper bound for the INT hop-time average AoI will be the end-to-end communication latency averaged by the number of hops.

In order to decrease the INT overhead in the network, INT data can be initialized at the source node probabilistically. This means that for each packet the source node will decide probabilistically to initialize INT data or not. In addition to this, INT data can be added using a certain probability on each hop in the network, reducing even further the network overhead. In such a case, the expression in equation 4 will take the form:

$$\begin{aligned} \overline{AoI_{INT}} &= \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{1}{T_i} \left(\frac{T_i^2}{2} + \sum_{j=i}^{N-1} p_j \delta_j T_j \right) \\ &= \frac{\mathbb{E}[T]}{2} + \mathbb{E}[\Delta] + \frac{1}{N-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N-1} \frac{p_j \delta_j T_j}{T_i} \end{aligned} \quad (5)$$

where p_i is the probability that an intermediate node i will add INT data to the packet, and thus, introducing INT delay δ_i .

The equation 4 is used to determine only the hop-time average AoI of hop-by-hop INT data. The time-average AoI of the end-to-end information collected by INT packets will be as the one of the first hop:

$$\overline{AoI_{E2E}} = \frac{1}{T_1} \Delta_{E2E} = \frac{1}{T_1} \Delta_1 = \frac{T_1}{2} + \delta_1 + \frac{1}{T_1} \sum_{i=2}^{N-1} T_i \delta_i \quad (6)$$

The first and second term of equation 6 depend only on the end-to-end communication latency and the INT processing delay of the first hop, while the last term shows the impact of delays in other hops into the time average AoI of the end-to-end collected INT information.

The INT monitoring information can be used by the application itself on the end node for verifying the current network

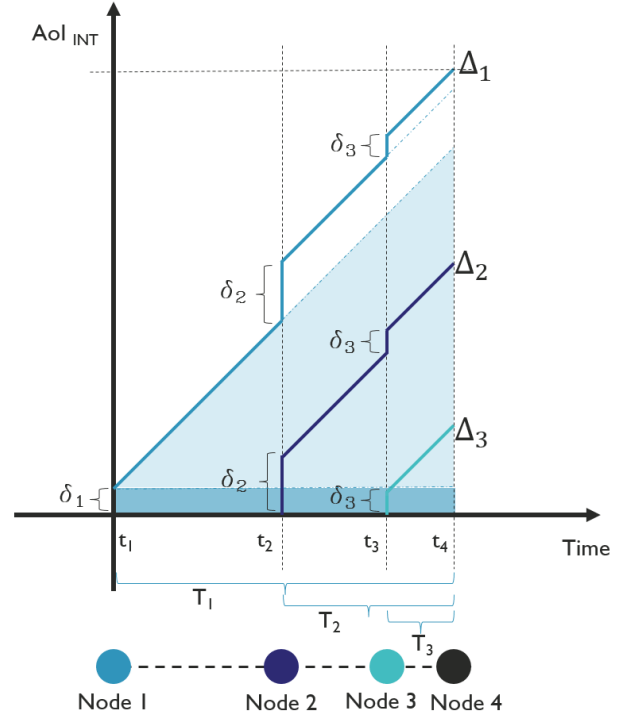


Fig. 1. Age of Information change for hop-by-hop INT monitored information.

performance. Alternatively, it can be sent to a central network entity and used as input for further network (re)configurations. In the second case, the AoI of the monitored parameters will increase further and will become:

$$\overline{AoI_{INTC}} = \overline{AoI_{INT}} + \frac{1}{T_1 + T_c} \left(\frac{T_c^2}{2} + T_c \delta_c \right) \quad (7)$$

where T_c is the delay between the moment the INT report is generated at the end node and the moment the INT report is received at the central network entity, and δ_c is the processing delay at the end node to extract INT data from the packet and to generate the INT report.

IV. RESULTS

The in-band network telemetry is implemented at layer three of the OSI layer structure. The IPv6 extension header is used to carry the INT data, that will be read by each hop in the network. If any of the nodes in the network is not an INT enabled node, then the data packet can still pass through the router without breaking the communication, contrary to a layer 2 implementation of INT [6]. To determine the age of information collected by layer three INT, we create a measurement setup in Mininet¹, a network simulator, as well as in a real-world scenario in the w-iLab.t² test-bed. In both cases we used a multi-hop setup.

¹<http://mininet.org/>

²<https://doc.ilabt.imec.be/ilabt/wilab/>

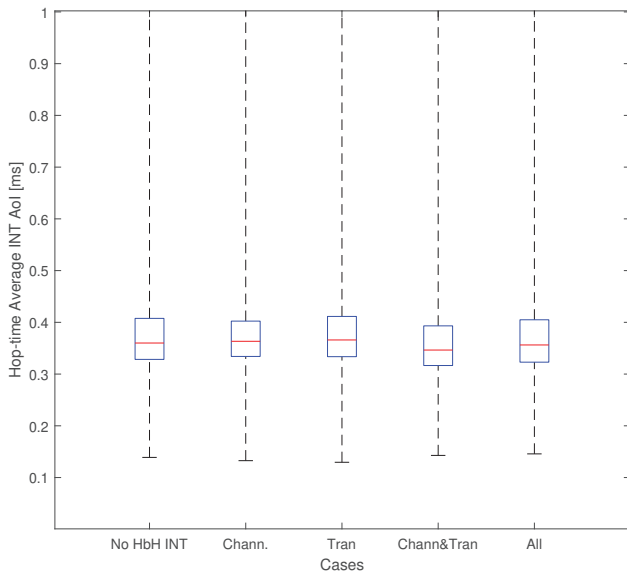


Fig. 2. INT hop-time average AoI for different information collected on each hop in simulated network.

A. Results from Mininet emulator

We created a network topology with 4 hops in Mininet as depicted in Figure 1. Only the wireless links are emulated in the Mininet, where the communication delays depend on the modulation and coding scheme (MCS) used on that particular link. The processing delay in each node is random. We evaluate the impact of different collected parameters on INT hop-time average AoI, as well as probabilistic addition of INT data on intermediate hops.

1) *Impact of number of monitored parameters:* INT monitoring can collect both end-to-end and hop-by-hop information for each traffic flow. The collected information on each hop depends on the *trace type* field included by the INT source header [4]. Each bit of the *trace type* field determines the addition of new information at each communication hop. In our implementation of wireless INT, we can collect different wireless parameters such as: channel info (RSSI value, channel used and retransmission flag), transmission info (MCS and data rate) and hardware queue space left.

We run an UDP traffic flow in the 4-hop network with 5 Mbps achievable throughput and add INT information to each packet. We measure for 1 minute, resulting in ~ 6500 data points. Measurements are done using different *trace type* field parameters in the INT source header.

In Figure 2 statistic plots of INT hop-time average AoI are shown for different information collected at INT intermediate nodes (other combinations like channel info queue info, transmission info queue info and queue only info, all perform the same in terms of AoI average and are excluded from the graph). It is noticed that no matter what INT information is collected on each hop, the hop-time average AoI in its mean value is similar, being smaller than 0.4 ms for a 4-hop network, while its 75th percentile in each case is around 0.4 ms. As

TABLE I
99th PERCENTILE OF INT HOP-TIME AVERAGE AoI (\overline{AoI}_{INT}) AND E2E LATENCY (T_1) FOR DIFFERENT COLLECTED INFORMATION.

	No. HbH INT	Chan. info	Trans. info	Chan. & trans.	All
\overline{AoI}_{INT} [ms]	5.7	5.82	5.75	5.78	5.87
T_1 [ms]	25.14	25.31	25.2	22.6	25.65

explained in section III, the upper bound of (\overline{AoI}_{INT}) is the end-to-end communication latency averaged by the number of hops. Its 99th percentile is shown in Table I and for each of the cases it is smaller than the $\frac{1}{N-1}$ of the end-to-end communication latency, which gives the upper bound for the hop-time average INT AoI.

2) *Impact of INT processing delays:* Based on equation 4, the hop-time average AoI, \overline{AoI}_{INT} , also depends on the INT processing delays, $\mathbb{E}[\Delta]$, introduced on each hop. This is the only variable that depends directly on the INT layer design and can be decreased by considering probabilistic INT addition. In Figure 3, statistics for the introduced INT delays on the same network with 4 hops are presented. The mean AoI addition due to INT processing delays is insignificant, being between 25 and 35 ns. Also, the 99th percentile of the INT processing delays is smaller than ~ 70 ns, when INT data is added to each packet. This will reduce based on the probability of INT additions at each hop (Figure 3b).

B. Results from test-bed deployment

We created a test setup in w-iLab.t test-bed with a network topology that consisted of 4 hops. The communication included nodes with different processing capabilities such as: ARM-based with 1.4 GHz processing speed and INTEL-based with 2.67 GHz processing speeds. As we noticed from the results taken in Mininet that the hop-time average does not depend on the amount of information collected by INT nodes we run the experiments by including all possible INT information on each packet on each network hop. The INT hop-time average AoI statistics are shown in Figure 4. Due to the realistic scenario, the 99th percentile of \overline{AoI}_{INT} is higher than in case of simulated network, however, still it does not pass the upper bound of end-to-end latency (99th percentile 42.03 ms) averaged by the number of hops.

Similarly, the $\mathbb{E}[\Delta]$ is higher than in the measurements collected in the Mininet simulator, with its 99th percentile being as high as 96.8 ns, when INT data is added to each data packet. Similarly, for lower INT addition probability than 1 the $\mathbb{E}[\Delta]$ will be smaller too.

V. CONCLUSION AND FUTURE WORK

In this paper, we showed a new metric for evaluating the freshness of in-band network telemetry data. The mathematical model to determine the Age of Information for the collected INT data is derived. The hop-time average AoI is used to determine the added value of the information collected by INT. Also, we showed the time average AoI of the end-to-end INT data and the impact of INT reporting on the INT AoI average.

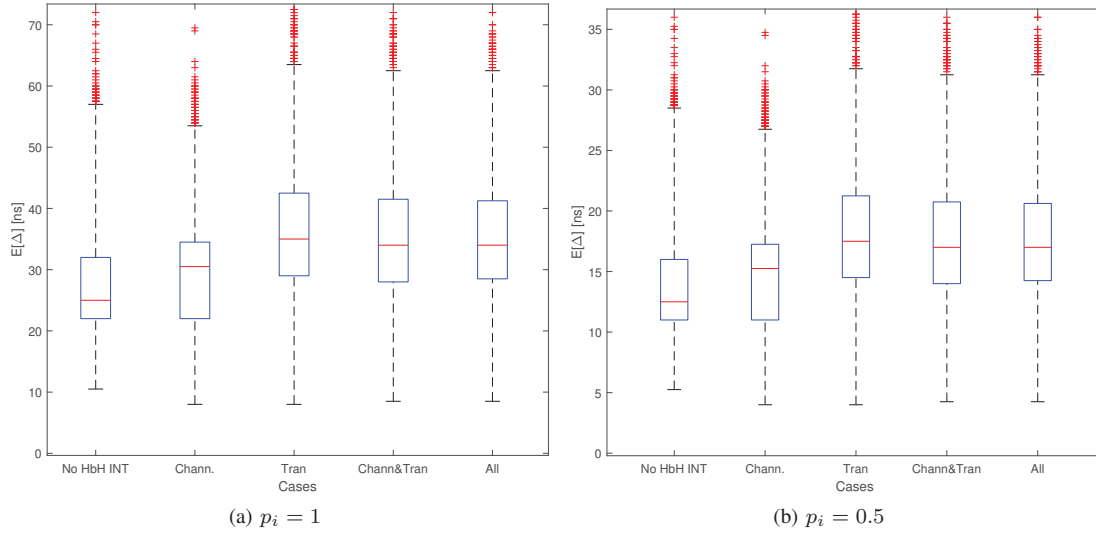


Fig. 3. INT processing delays for each different case and different INT addition probability.

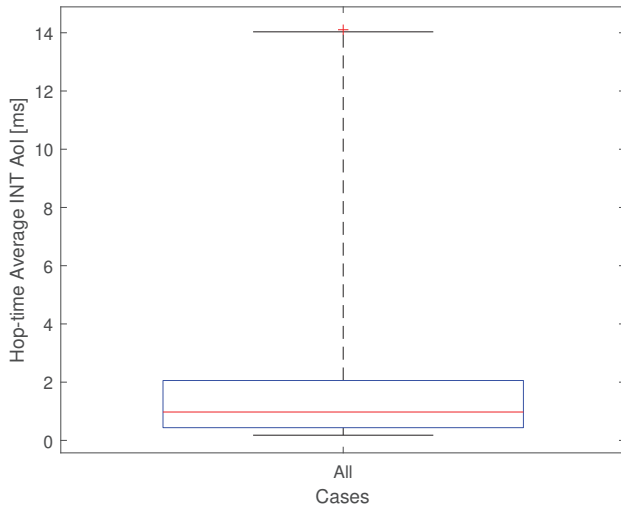


Fig. 4. INT hop-time average AoI for all information collected on each hop in real test-bed network.

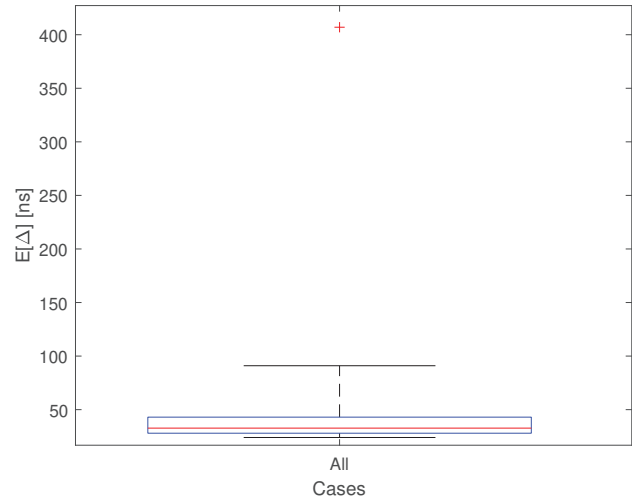


Fig. 5. INT processing delays for measurement in test-bed network.

We used the derived mathematical model to determine the hop-time average of a layer three INT implementation. The evaluations were performed in a four hop network topology both in a simulator environment and in a real network setup. In both cases we showed that the hop-time average INT AoI is dependent on the communication latency between each hop and the destination, while the impact of the INT processing delays in each hop is negligible.

This research can be extended further leading to other interesting research topics related to INT and AoI. As it was presented in this research the INT hop-time average AoI depends on the communication latency. As INT data is used for detecting also the cases when communication latency is high, requiring low average AoI for such INT data will be not-realistic in such cases. In such cases, another approach can

be to extract data on each hop towards the central controller in order to reduce the INT average AoI. Further research on trade-offs between the achieved lower AoI and more loads in the control plane can extend the current research work.

In certain cases the INT data cannot be added to current data packets but will be added to the next data packet (e.g. think about low-level information such as contention window that is available only after the packet has been transmitted). As such the average AoI of the INT data will not only depend on the communication latency of the next follow up data packet, but also on the inter-packet generation time, increasing further the averaged AoI. The impact of this on the achieved averaged AoI should be further studied.

Another research aspect is the modeling of the decision-making freshness. Decision-making at a network central controller is often not based on a single INT report, but rather on

series of INT reports. Thus, the freshness of a series of INT reports need to be considered and trade-offs between control load and freshness of decision making in more advanced used cases need to be evaluated.

ACKNOWLEDGMENT

This research was partially funded by the Flemish FWO SBO S003921N VERI-END.com (Verifiable and elastic end-to-end communication infrastructures for private professional environments) project, the Flemish Government under the ‘‘Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen’’ program, and from the FWO-Flanders, under grant agreement G055619N.

REFERENCES

- [1] J. A. Wickboldt, W. P. De Jesus, P. H. Isolani, C. B. Both, J. Rochol, and L. Z. Granville, ‘‘Software-defined networking: management requirements and challenges,’’ *IEEE Communications Magazine*, vol. 53, no. 1, pp. 278–285, 2015.
- [2] ‘‘Openwifi. available online: <https://orca-project.github.io/openwifi-tmp/openwifi-public.pdf>,’’ Accessed on 28.11.2019.
- [3] Aruba, ‘‘Ieee 802.11ax,’’ White Paper, 2019.
- [4] F. Brockners, S. Bhandari, and T. Mizrahi, ‘‘Data fields for in-situ oam, draft-ietf-ippm-iom-data-12,’’ 2021.
- [5] J. Haxhibeqiri, P. H. Isolani, J. M. Marquez-Barja, I. Moerman, and J. Hoebeke, ‘‘In-band network monitoring technique to support sdn-based wireless networks,’’ *IEEE Transactions on Network and Service Management*, 2020.
- [6] J. Haxhibeqiri, I. Moerman, and J. Hoebeke, ‘‘Low overhead, fine-grained end-to-end monitoring of wireless networks using in-band telemetry,’’ in *CNSM2019, the 15th International Conference on Network and Service Management*, 2019, pp. 1–5.
- [7] S. Kaul, R. Yates, and M. Gruteser, ‘‘Real-time status: How often should one update?’’ in *2012 Proceedings IEEE INFOCOM*. IEEE, 2012, pp. 2731–2735.
- [8] S. Tang, J. Kong, B. Niu, and Z. Zhu, ‘‘Programmable multilayer int: An enabler for ai-assisted network automation,’’ *IEEE Communications Magazine*, vol. 58, no. 1, pp. 26–32, 2020.
- [9] L. Tan, W. Su, W. Zhang, J. Lv, Z. Zhang, J. Miao, X. Liu, and N. Li, ‘‘In-band network telemetry: A survey,’’ *Computer Networks*, vol. 186, p. 107763, 2021.
- [10] A. Gulenko, M. Wallschläger, and O. Kao, ‘‘A practical implementation of in-band network telemetry in open vswitch,’’ in *2018 IEEE 7th International Conference on Cloud Networking (CloudNet)*. IEEE, 2018, pp. 1–4.
- [11] N. Van Tu, J. Hyun, G. Y. Kim, J.-H. Yoo, and J. W.-K. Hong, ‘‘Intcollector: A high-performance collector for in-band network telemetry,’’ in *2018 14th International Conference on Network and Service Management (CNSM)*. IEEE, 2018, pp. 10–18.
- [12] N. Van Tu, J. Hyun, and J. W.-K. Hong, ‘‘Towards onos-based sdn monitoring using in-band network telemetry,’’ in *2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. IEEE, 2017, pp. 76–81.
- [13] J. Hyun, N. Van Tu, and J. W.-K. Hong, ‘‘Towards knowledge-defined networking using in-band network telemetry,’’ in *NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2018, pp. 1–7.
- [14] P. H. Isolani, J. Haxhibeqiri, I. Moerman, J. Hoebeke, J. M. Marquez-Barja, L. Z. Granville, and S. Latré, ‘‘An sdn-based framework for slice orchestration using in-band network telemetry in ieee 802.11,’’ in *2020 6th IEEE Conference on Network Softwarization (NetSoft)*. IEEE, 2020, pp. 344–346.
- [15] G. Stamatakis, N. Pappas, and A. Traganitis, ‘‘Optimal policies for status update generation in an iot device with heterogeneous traffic,’’ *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5315–5328, 2020.
- [16] E. Fountoulakis, N. Pappas, M. Codreanu, and A. Ephremides, ‘‘Optimal sampling cost in wireless networks with age of information constraints,’’ in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2020, pp. 918–923.
- [17] O. Vikhrova, F. Chiariotti, B. Soret, G. Araniti, A. Molinaro, and P. Popovski, ‘‘Age of information in multi-hop networks with priorities,’’ *arXiv preprint arXiv:2009.12657*, 2020.
- [18] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, ‘‘Age of information performance of multiaccess strategies with packet management,’’ *Journal of Communications and Networks*, vol. 21, no. 3, pp. 244–255, 2019.
- [19] L. Corneo, C. Rohner, and P. Gunningberg, ‘‘Age of information-aware scheduling for timely and scalable internet of things applications,’’ in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 2476–2484.
- [20] M. A. Abd-Elmagid, H. S. Dhillon, and N. Pappas, ‘‘Aoi-optimal joint sampling and updating for wireless powered communication systems,’’ *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 14 110–14 115, 2020.
- [21] M. A. Abd-Elmagid, N. Pappas, and H. S. Dhillon, ‘‘On the role of age of information in the internet of things,’’ *IEEE Communications Magazine*, vol. 57, no. 12, pp. 72–77, 2019.
- [22] M. Moltafet, M. Leinonen, M. Codreanu, and N. Pappas, ‘‘Power minimization in wireless sensor networks with constrained aoi using stochastic optimization,’’ in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2019, pp. 406–410.

ANNEX I

Let \mathbf{T} and $\mathbf{\Delta}$ be the random variables that describe the communication latency between each hop and end destination and processing INT delays on each hop, respectively. Let N be a large number of hops in the network. Each of the outcomes of \mathbf{T} and $\mathbf{\Delta}$ are independent and equiprobable, thus the expected value $\mathbb{E}[\cdot]$ will be the arithmetical average of the all individual outcomes. Thus, equation 4 is derived from 2 and 3 as follows:

$$\begin{aligned}
 \overline{AoI_{INT}} &= \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{1}{T_i} \left(\frac{T_i^2}{2} + \sum_{j=i}^{N-1} \delta_j T_j \right) \\
 &= \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{T_i}{2} + \frac{1}{N-1} \sum_{i=1}^{N-1} \sum_{j=i}^{N-1} \frac{\delta_j T_j}{T_i} \\
 &= \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{T_i}{2} + \frac{1}{N-1} \sum_{i=1}^{N-1} \delta_i \\
 &\quad + \frac{1}{N-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N-1} \frac{\delta_j T_j}{T_i} \\
 &= \frac{\mathbb{E}[\mathbf{T}]}{2} + \mathbb{E}[\mathbf{\Delta}] + \frac{1}{N-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N-1} \frac{\delta_j T_j}{T_i}
 \end{aligned} \tag{8}$$