FindINT: Detect and Locate the Lost In-band Network Telemetry Packet

Lizhuang Tan, *Student Member, IEEE*, Wei Su, Jingying Miao and Wei Zhang, *Member, IEEE*

***Abstract*—In-band network telemetry (INT) is a new network measurement technology that uses business packet to collect network information hop by hop. Since business packets may be lost due to various reasons, INT telemetry information will inevitably be lost. Unfortunately, telemetry system is not aware of this loss. In this letter, we present FindINT, an INT packet coding scheme with two marking strategies, to measure per-flow packet loss rate and location. FindINT is inspired by alternative marking method, which is triggered by loss event instead of polling.Experimental results show that FindINT has extremely low overhead and high detection accuracy.**

***Index Terms*—In-band Network Telemetry, Network Measurement, Loss Detection, Loss Localization.**

I. Introduction

NETWORK telemetry is a new network measurement method that can quickly collect and integrate network status data to monitoring service quality. Network telemetry can be divided into in-band network telemetry (INT) and out-band network telemetry (ONT). The characteristic of INT can be summarized as using the business packet to carry the status information of the hop-by-hop switching device. Because INT can realize the end-to-end traffic visualization of network services, it has attracted the attention of academia and industry[1].

As is shown in the upper part of Figure 1, the principle of INT is as follows: When a user packet enters the first switching device (*INT Source Node*), *INT Source Node* encapsulates an INT header in the packet to define the INT telemetry instruction, and fills in the *INT metadata* with the network device information to be collected after the INT header. Then, the packet is forwarded to the next switching device (*INT Transit Hop Node*), and *INT Transit Hop Node* continues to add the INT metadata according to telemetry instruction. After packet pass through all *INT Transit Hop Node* and are forwarded to the last switching device (*INT Sink Node*), *INT Sink Node* removes INT header and all INT metadata, and sends them to *Telemetry Server*. *Telemetry Server* can collect switch meta information, including switch/port ID, port utilization, and queue to calculate the delay and congestion experienced by network services.

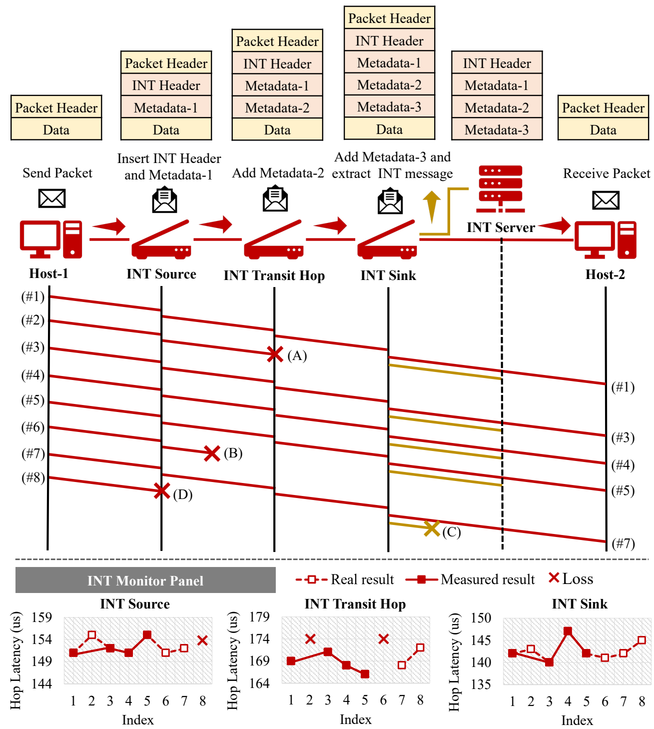

Fig. 1. The INT and monitoring result when telemetry packet is lost.

Recently, numberous INT-based network measurement solutions have been proposed. They can measure one-way delay[2], tail latency[3], available bandwidth[4], queue depth[5], etc. Many advanced network controls based on INT have also significantly improved network efficiency, including congestion control[6], routing decisions[7], abnormal detection[8] and path tracing[3]. However, according to the survey of above-mentioned research works and our practical experience in deploying INT, we discovered the following facts:

1) There is no feasible packet loss measurement solution for INT[1]. Some SDN packet loss measurement and localization schemes have been proposed, but none of them have been integrated into INT. [12] and [13]

This work was supported in part by the National Key R&D Program of China (No.2018YFB1800305), the Fundamental Research Funds for the Central Universities of China (No. 2020YJS013), the National Natural Science Foundation of China (No. 61802233) and the Natural Science Foundation of Shandong Provincial (No. ZR2020LZH010). *(Corresponding Author: Wei Su)*

L. Tan, W. Su and J. Miao are with School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing 100044, P. R. China. (e-mail: lzhtan; wsu; 19120098@bjtu.edu.cn)

W. Zhang is with Shandong Provincial Key Laboratory of Computer Networks, Shandong Computer Science Center (National Supercomputer Center in Ji'nan), Qilu University of Technology (Shandong Academy of Sciences), Ji'nan 250000, P. R. China. (e-mail: wzhang@sdas.org)

[1]The latest version of INT specification[9] has divided INT into three types: INT-XD, INT-MX and INT-MD. The first two draw on the ideas of iFIT[10] and IOAM[11], and export metadata hop by hop, so they can be directly used for loss measurement. We only discuss the third type of INT here, that is, how to measure the loss of telemetry packets when only the Sink Node exports metadata.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LNET.2021.3067343, IEEE Networking Letters

require strict synchronization of network time. [14] is an active measurement solution, not compatible with INT. [15] and [16] are representative of the alternate-marking performance measurement (AM-PM), they only need a few bits to achieve loss measurement. However, their marks are updated periodically, which requires strict time synchronization too. But, [15] and [16] give us inspiration. If AM-PM is applied to INT measurement, we can know when packet loss has occurred. It is worth noting that [17] use INT to measure packet loss rate, but it is active measurement, measuring the probe rather than the per-flow packet loss rate. At the same time, using multiple probes to locate the INT packet loss location requires huge network overhead. Therefore, INT urgently needs a packet loss detection and localization solution.

2) The working principle of INT make it insensitive to packet loss. As shown in the lower half of Figure 1, telemetry packets #2, #6, #7, and #8 are lost due to different reasons, but the *Telemetry Server* does not know about these packet losses. since INT reports only the good news but not the bad, we cannot observe real network status. Those existing INT applications do not consider the impact of network packet loss on the INT measurement results. Although the *Telemetry Server* can continuously receive INT report packets, these packets are incomplete in fact. The telemetry information of critical network failures is lost along with user packets. We counted the performance of anomaly detection in [8] under different packet loss rates. The result is shown in Figure 2. The RNN model trained in an environment where the link packet loss rate is zero cannot accurately classify abnormal traffic in a lossy network. The reason is that packet loss destroys the input-data characteristics (duration, hop latency, flow latency and queue occupancy) of the RNN model. If some special supplementary schemes are not adopted, the *Telemetry Server* does not know that the network flow is experiencing packet loss.

In summary, it is our goal to design an efficient INT packet loss measurement solution and apply it to INT applications to improve the monitoring performance. In general, the innovations of this letter include:

1) We propose an AM-based INT loss detection and localization mechanism, FindINT, compliant with available standards and internet drafts[9], enabling INT to perceive packet loss events in actual network.
2) We propose two *Loss Bit* coding schemes, SAM and MCM. The former occupies less bandwidth, and the latter has high accuracy.
3) We implement and open source the above mechanism on P4 switch[18]. The experimental results show that it can measure the per-flow packet loss with low overhead and high accuracy.

II. Scheme Design

Alternate-marking performance measurement (AM-PM) method[19] has been standardized. It realizes efficient measurement of delay and packet loss by periodically changing the 1 bit or 2 bit coloring flag. In AM-PM, every packet of the monitored flow carries 1 or 2 marking bits that are used for signaling and coordinating measurement events across the measurement points. For packet loss measurements, a periodically alternating marker, Color Bit, is dividing the traffic into consecutive blocks of data. By counting the number of packets in each block and comparing the values measured by different network devices along the path, it is possible to measure packet loss occurred in any single block between any two points[20]. Applying the AM-PM method to INT, we proposed FindINT and designed the protocol format, marking strategy and loss localization mechanism.

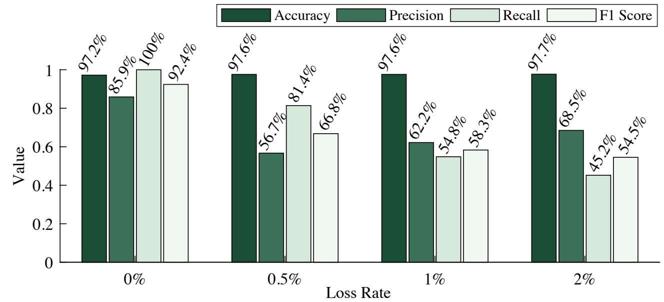

Fig. 2. Performance of the RNN model proposed in [8] under different packet loss rates.

A. Overview

The process of FindINT is shown in Figure 3. After inserting the telemetry header, the *INT Source Node* also marks this telemetry packets. The simplest way of marking is alternate marking. In other words, for the first packet, the *INT Source Node* sets the loss flag bit to 1. And the second is set to 0, the third is set to 1... (For more reliable and efficient marking schemes, please see II-C) *INT Sink Node* reports the telemetry information to the *Telemetry Server*. *Telemetry Server* calculates the loss rate based on the successively received loss flags. If the *Telemetry Server* does not receive the marking information it should receive, packet loss has occurred (for some special cases, please see II-E). When loss is detected, the *Telemetry Server* can notify SDN controller to collect more abundant loss information or perform some necessary operations. If it is necessary to detect the position of packet loss, all *INT Nodes* need to participate in marking (see II-D).

B. Protocol Design

As is shown in Figure 4, FindINT can use 1 or more bits (L), we call it *Loss Bit*, in the reserved field of the INT-MD protocol header to mark telemetry packet.

Figure 5 analyzes FindINT from the perspective of functional design. If only need to detect loss rate, *Loss Bit* is only marked once by *INT Source*. In the entire network, only one counter is needed for each flow, which can be integrated into the flow-table counter.

If need to detect the loss rate and loss localization, All *INT Nodes* maintain a counter for each flow and add their own

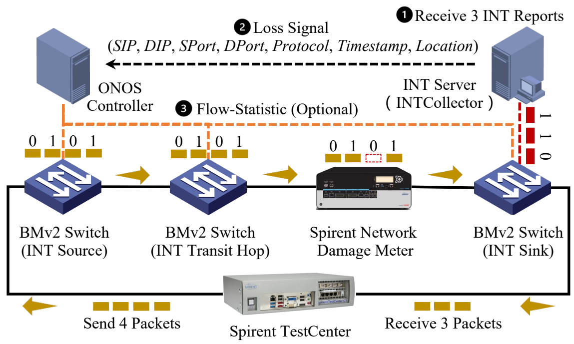

Fig. 3. The process of FindINT.

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Ver = 2|D|E|M|L|       Reserved       | Hop ML  |RemainingHopCnt|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Instruction Bitmap       |       Domain Specific ID      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        DS Instruction         |           DS Flags            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| INT Metadata Stack (Each hop inserts Hop ML * 4B of metadata) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              ...                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Last INT metadata                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Fig. 4. INT-MD metadata header format.

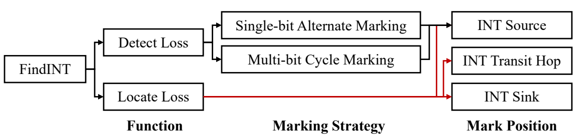

Fig. 5. The design idea of FindINT.

loss flags in turn. *Loss Bit* can be specified by Instruction Bitmap and stored in the INT metadata of each hop in the INT Metadata Stack.

C. Marking Strategy

We propose two packet loss marking strategies, Single-bit Alternate Marking (SAM) and Multi-bit Cycle Marking (MCM).

1) Single-bit Alternate Marking (SAM): Each bit of the telemetry packet is precious. In the case that the *Loss Bit* has only 1 bit, the set rule of *INT Source Node* is triggered by the arriving packet. *INT Source Node* change the *Loss Bit* from 0 to 1 or from 1 to 0 every $p/2$ telemetry packets. For example, while marking period $p = 10$, *INT Source Node* can mark *Loss Bit* of first 5 INT packets as 1 and then successively mark *Loss Bit* of last 5 INT packets as 0.

Assume that there is no disorder in the network, and each packet loss is independent. When the actual network packet loss probability is extremely small, the alternate marking ($p = 2$) can work normally. When the actual packet loss on the network is large, the probability of losing 2 packets in a row becomes larger. Alternate marking ($p = 2$) does not recognize that the telemetry flow has lost two consecutive packets. This causes the measured loss rate to be less than the real loss rate.

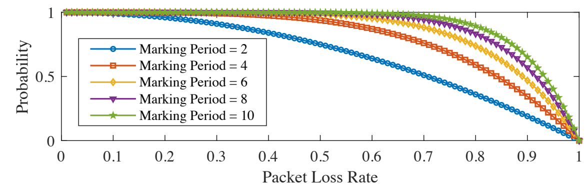

Fig. 6. The probability of error-free identification of packet loss under different marking periods and different network packet loss rates.

Assuming that network has n hops, and the random packet loss rate of each hop is $\epsilon_i, i = 1, ..., n$. The end-to-end random packet loss rate ϵ of telemetry packets is

$$\epsilon = 1 - \prod_{i=1}^{n} (1 - \epsilon_i). \tag{1}$$

The probability $\mathscr{P}$ of consecutive loss of p telemetry packets is

$$\mathscr{P} = \epsilon^p. \tag{2}$$

Because the SAM can identify the consecutive loss of no more than p telemetry packets, a larger marking period can improve the confidence of telemetry result. As is shown in Figure 6, We analyzed the probability that the solution can identify the continuous packet loss of the network during different marking periods. The high probability means that the detected packet loss rate is closer to the real rate. When the network packet loss rate is less than 20%, the probability of continuous packet loss is very small. The measurement robustness under the five marking periods is very good (the difference is less than 4%). So, we can improve the detection accuracy by increasing the marking period.

We use the time interval T between the occurrence of packet loss and the detection of packet loss to evaluate the detection sensitivity. Obviously, detection sensitivity is related to network traffic, topology size, packet loss location, etc. Assuming that the arrival of telemetry packets obeys the Poisson distribution with λ. Hop-by-hop delay is $d_i, i = 1, ..., n$. The transmission and processing delay between the INT Sink Node and Telemetry Server is $d_{Sink2Server}$. The detection interval of the last telemetry packet loss at the last hop before the marking strategy change represents the minimum detection interval T_{min}. The detection interval of the last telemetry packet loss at the last hop before the marking strategy change represents the minimum detection interval T_{min}. The detection of the loss of the first telemetry packet at the first hop after the marking strategy change represents the maximum detection interval T_{max}.

The distribution function of arrival interval t of two adjacent packets is

$$F(t; \lambda) = 1 - e^{-\lambda t}, t \geq 0. \tag{3}$$

The expectation of T_{min} and T_{max} can be expressed as

$$\mathbb{E}\{T_{min}\} = \mathbb{E}\{T\} = \frac{1}{\lambda} \tag{4}$$

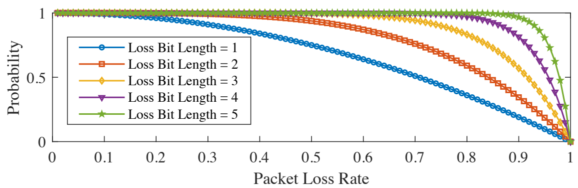

Fig. 7. The probability of error-free identification of packet loss under different loss bit lengths and different network packet loss rates.

$$\mathbb{E}\{T_{max}\} = \frac{p}{2}\mathbb{E}\{T\} + \sum_{i=1}^{n} d_i + d_{Sink2Server} \\ = \frac{p}{2\lambda} + \sum_{i=1}^{n} d_i + d_{Sink2Server}. \tag{5}$$

According to Equation 5, the larger the marking period, the lower the detection sensitivity. Therefore, although only 1 bit is used, SAM needs a suitable marking period to balance detection robustness and sensitivity.

2) Multi-bit Cycle Marking (MCM): Because the SAM strategy can only make INT get the packet information of the packet loss but not the sequence information. When there is disorder in the network, the SAM is difficult to distinguish between continuous loss and out-of-order in two marking periods. FindINT can use more *Loss Bits* and monotonically increases the value of *Loss Bit* so that telemetry reports carry sequence information.

Assuming that the length of the *Loss Bit* of MCM is l. Similar to Equation 2, the probability $\mathscr{P}$ of consecutive loss of 2^l telemetry packets is

$$\mathscr{P} = \epsilon^{2^l}. \tag{6}$$

Figure 7 shows that the longer the *Loss Bit*, the better it can adapt to a worse packet loss. When the network packet loss rate is less than 10%, the recognition accuracy of the five lengths are all greater than 99%. MCM can alleviate but cannot completely overcome the error caused by disorder.

Since each telemetry report carries sequence information, the telemetry server can know which telemetry packet is lost according to the received Loss Bit. So, the expectation of detection sensitivity is $[\frac{1}{\lambda}, \frac{1}{\lambda} + \sum_{i=1}^{n} d_i + d_{Sink2Server}]$.

D. Packet Loss Localization

In order to locate packet loss, FindINT deploys the above detection method to every INT node on the path. All INT Nodes that detects loss localization maintain a counter for each flow and add their own *Loss Bit* in turn. *Loss Bit* can be specified by Instruction Bitmap and stored in the INT metadata of each hop in the INT Metadata Stack. Those counter can be integrated into the INT operation with little overhead.

Telemetry Server builds the matrix shown in Figure 8 (SAM) and Figure 9 (MCM) according to the received telemetry reports. The telemetry server generates a full-path *Loss Bit* measurement result based on the received telemetry report and locates the location of the packet loss. Both SAM and MCM can determine the localization of packet loss. As shown by the dashed box, MCM is more accurate.

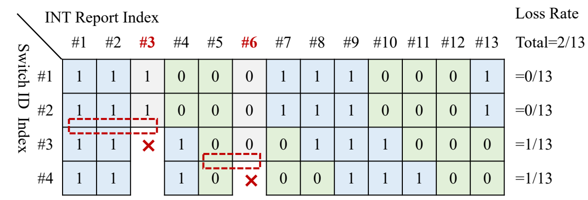

Fig. 8. Loss Bit encoding method (SAM) for loss localization.

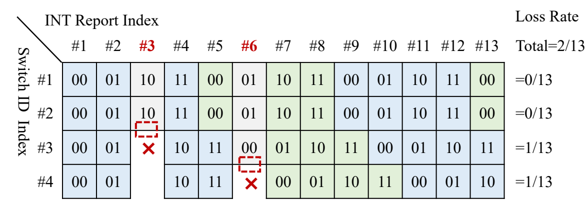

Fig. 9. Loss Bit encoding method (MCM) for loss localization.

E. Multipath and Out of Order

When the network has multi-path and QoS routing, the *Telemetry Server* may not receive the telemetry report in the marking order. How to distinguish between disorder and loss is an important content that FindINT needs to consider. Since INT often carry Switch ID or Timestamp, they can help *Telemetry Server* to identify the true routing path and sequence of INT packets. After reordering the received telemetry reports, *Telemetry Server* can accurately distinguish packet loss and disorder, and calculate more accurate measurement results. In summary, FindINT can at least combat disorder in three ways: (1) Increase the marking period for SAM, (2) Increase the length of *Loss Bits*, (3) Increase the statistics period.

III. Experimental Results

We built the testbed shown in Figure 3 to verify the measurement accuracy, delay and network overhead of FindINT. The testbed is composed of three BMv2 switches, a Spirent TestCenter, a Spirent Network Damage Meter, an ONOS controller and an INT remote server. The network bandwidth is 1Gbps. The Spirent TestCenter sends and receives packets, and counts the loss rate as a reference. The Spirent network damage meter achieves random packet loss. The ONOS controller version is 2.2 (Sparrow). The source code is open source[18], and it also supports Mininet environment.

A. Detection Accuracy and Localization Accuracy

As shown in Figure 10, we measured the accuracy of the two marking strategies. $P = 6$ and $L = 2$ are the recommended values for parameter P in SAM and L in MCM for real usage. In general, the loss rate measured by FindINT is very close to the real loss rate. The average measurement deviation between SAM and MCM and Spirent TestCenter is 3.62% and 1.17%. It is worth noting that FindINT measures the actual packet loss experienced by INT flow, so this deviation can be regarded

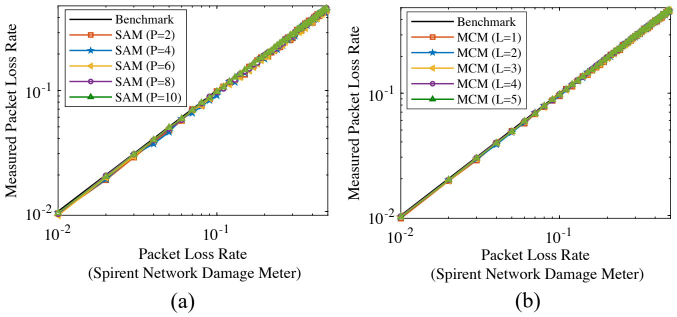

Fig. 10. Measurement accuracy result. (a) is SAM and (b) is MCM

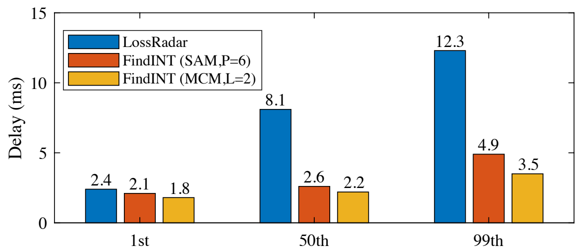

Fig. 11. Comparison of detection latency between FindINT and LossRadar.

as a probabilistic deviation. During the experiment with the packet loss rate [0%, 20%], FindINT locates the 100% packet loss position.

B. Detection Latency

The detection latency is affected by the path length, network rate, packet interval, marking period and *Loss Bit* length. *Telemetry Server* only can diagnose packet loss only after eliminating the influence of disorder. As is shown in Figure 11, we compared the detection latency of FindINT and LossRadar[13], and the latency of FindINT is more stable. The reason is that FindINT is triggered by continuous *Loss Bit*, so there is no need for excessive waiting time. Moreover, the query time of LossRadar is usually 10ms. Compared with SAM, MCM diagnoses packet loss more quickly. Because the diagnosis cycle of MCM (L=2) is shorter than SAM (P=6).

C. Overhead

Compared with OpenNetMon[12], LossRadar[13] and INT_DETECT[17], the overhead of FindINT is almost negligible. Because FindINT does not poll switches or construct probes, it only need a few bits of INT telemetry packets to detect loss. Compared with INT-XD and INT-MX, the INT report packets reported by FindINT is only $1/n$ of them. n is the number of switches participating in telemetry.

FindINT maintains a register for each flow, and the number of bits of the register depends on the period of SAM and MCM. FindINT needs to perform 100% accuracy statistics for each flow in each period. Taking SAM as an example, when the period is 8, the length of each register is 3 bits. 1KB of memory supports 2730 flows. 4KB memory supports 10922 flows. The memory requirement in a INT node only grows linearly with the number of telemetry flows instead of all the telemetry packets.

IV. Conclusion

In this letter, we propose a packet loss detection and localization mechanism, FindINT, for in-band network telemetry, which is derived from the AM-PM method. FindINT makes up for the defect that INT cannot measure network packet loss by marking telemetry packets, and can improve the performance of telemetry applications. Experimental results show that FindINT has extremely low overhead and high detection accuracy, and can be applied in a production environment.

References

[1] 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.
[2] Y. Lin, Y. Zhou, Z. Liu, K. Liu *et al.*, "NetView: Towards on-demand network-wide telemetry in the data center," *Computer Networks*, vol. 180, p. 107386, 2020.
[3] R. Ben Basat, S. Ramanathan, Y. Li, G. Antichi *et al.*, "PINT: Probabilistic In-Band Network Telemetry," in *SIGCOMM '20*. Virtual Event, USA: ACM, July 2020, pp. 662–680.
[4] N. Kagami, R. I. T. da Costa Filho, and L. P. Gaspary, "CAPEST: Offloading Network Capacity and Available Bandwidth Estimation to Programmable Data Planes," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 175–189, 2020.
[5] T. Pan, E. Song, Z. Bian, X. Lin *et al.*, "INT-path: Towards optimal path planning for in-band network-wide telemetry," in *INFOCOM '19*. Paris, France: IEEE, April 2019, pp. 487–495.
[6] Y. Li, R. Miao, H. H. Liu, Y. Zhuang *et al.*, "HPCC: High Precision Congestion Control," in *SIGCOMM '19*. Beijing, China: ACM, August 2019, pp. 44–58.
[7] A. Karaagac, E. De Poorter, and J. Hoebeke, "Alternate Marking-based Network Telemetry for Industrial WSNs," in *WFCS '20*. Porto, Portugal: IEEE, April 2020, pp. 1–8.
[8] S. Nam, J. Lim, J.-H. Yoo, and J. W.-K. Hong, "Network Anomaly Detection Based on In-band Network Telemetry with RNN," in *ICCE-Asia 2020*. Busan, Republic of Korea: IEEE/IEIE, November 2020, pp. 1–5.
[9] "In-band Network Telemetry (INT) Dataplane Specification v2.1," 2020. [Online]. Available: https://github.com/p4lang/p4-applications/blob/master/docs
[10] B. Lu, L. Xu, Y. Song, L. Dai *et al.*, "IFIT: Intelligent Flow Information Telemetry," in *SIGCOMM Posters and Demos '19*. Beijing, China: ACM, August 2019, p. 15–17.
[11] "IETF IOAM," https://datatracker.ietf.org/wg/ioam, 2020.
[12] N. L. Van Adrichem, C. Doerr, and F. A. Kuipers, "OpenNetMon: Network monitoring in OpenFlow Software-Defined Networks," in *NOMS 2014*, Krakow, Poland, May 2014, pp. 1–8.
[13] Y. Li, R. Miao, C. Kim, and M. Yu, "LossRadar: Fast Detection of Lost Packets in Data Center Networks," in *CoNEXT '16*. Irvine, California, USA: ACM, 2016, pp. 481–495.
[14] X. Zhang, Y. Wang, J. Zhang, L. Wang, and Y. Zhao, "RINGLM: A Link-Level Packet Loss Monitoring Solution for Software-Defined Networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 8, pp. 1703–1720, 2019.
[15] M. Cociglio, G. Fioccola, G. Marchetto, A. Sapio, and R. Sisto, "Multipoint Passive Monitoring in Packet Networks," *IEEE/ACM Transactions on Networking*, vol. 27, no. 6, pp. 2377–2390, 2019.
[16] P. Loreti, A. Mayer, P. Lungaroni, S. Salsano, R. Gandhi, and C. Filsfils, "Implementation of Accurate Per-Flow Packet Loss Monitoring in Segment Routing over IPv6 Networks," in *HPSR '20*, Newark, NJ, USA, May 2020, pp. 1–8.
[17] C. Jia, T. Pan, Z. Bian, X. Lin *et al.*, "Rapid Detection and Localization of Gray Failures in Data Centers via In-band Network Telemetry," in *NOMS 2020*. Budapest, Hungary: IEEE, 2020, pp. 1–9.
[18] "FindINT," https://github.com/lzhtan/FindINT, 2020.
[19] G. Fioccola, A. Capello, M. Cociglio, L. Castaldelli *et al.*, "Alternate-Marking Method for Passive and Hybrid Performance Monitoring," *RFC 8321*, 2018.
[20] T. Mizrahi, G. Navon, G. Fioccola, M. Cociglio, M. Chen, and G. Mirsky, "AM-PM: Efficient Network Telemetry using Alternate Marking," *IEEE Network*, vol. 33, no. 4, pp. 155–161, 2019.