Workgroup: OPSAWG
Internet-Draft:
draft-song-opsawg-ifit-framework-19
Published: 24 October 2022
Intended Status: Informational
Expires: 27 April 2023
Authors: H. Song (Futurewei)
F. Qin (China Mobile)
H. Chen (China Telecom)
J. Jin (LG U+)
J. Shin (SK Telecom)

A Framework for In-situ Flow Information Telemetry

Abstract

As network scale increases and network operation becomes more sophisticated, existing Operation, Administration, and Maintenance (OAM) methods are no longer sufficient to meet the monitoring and measurement requirements. Emerging data-plane on-path telemetry techniques, such as IOAM and AltMrk, which provide high-precision flow insight and issue notifications in real time can supplement existing proactive and reactive methods that run in active and passive modes. They enable quality of experience for users and applications, and identification of network faults and deficiencies. This document describes a reference framework, named as In-situ Flow Information Telemetry, for the on-path telemetry techniques.

The high-level framework outlines the system architecture for applying the on-path telemetry techniques to collect and correlate performance measurement information from the network. It identifies the components that coordinate existing protocol tools and telemetry mechanisms, and addresses deployment challenges for flow-oriented on-path telemetry techniques, especially in carrier networks.

The document is a guide for system designers applying the referenced techniques. It is also intended to motivate further work to enhance the OAM ecosystem.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents

at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 27 April 2023.

Copyright Notice

Table of Contents

1. Introduction

Efficient network operation increasingly relies on high-quality data-plane telemetry to provide the necessary visibility into the behavior of traffic flows and network resources. Existing Operation, Administration, and Maintenance (OAM) methods, which include proactive and reactive techniques, running both active and passive modes, are no longer sufficient to meet the monitoring and measurement requirements when networks becomes more autonomous [RFC8993] and application-aware [I-D.li-apn-framework]. The complexity of today's networks and service quality requirements demand new high-precision and real-time OAM techniques.

The ability to expedite network failure detection, fault localization, and recovery mechanisms, particularly in the case of soft failures or path degradation is expected, and it must not cause service disruption. Emerging on-path telemetry techniques can provide high-precision flow insight and real-time network issue notification (e.g., jitter, latency, packet loss, significant bit error variations, and unequal load-balancing). On-Path Telemetry (OPT) refers to data-plane telemetry techniques that directly tap and measure network traffic by embedding instructions or metadata into user packets. The data provided by OPT are especially useful for verifying Service Level Agreement (SLA) compliance, user experience enhancement, service path enforcement, fault diagnosis, and network resource optimization. It is essential to recognize that existing work on this topic includes a variety of OPT techniques, including In-situ OAM (IOAM) [RFC9197], IOAM Direct Export (DEX) [I-D.ietf-ippm-ioam-direct-export], Marking-based Postcard-based Telemetry (PBT-M) [I-D.song-ippm-postcard-based-telemetry], Enhanced Alternate Marking (EAM) [I-D.zhou-ippm-enhanced-alternate-marking], and Hybrid Two-Step (HTS) [I-D.mirsky-ippm-hybrid-two-step], have been developed or proposed. These techniques can provide flow information on the entire forwarding path on a per-packet basis in real-time. The aforementioned OPT techniques differ from the active and passive OAM schemes in that they directly modify and monitor the user packets in networks so as to achieve high measurement accuracy. Formally, these OPT techniques can be classified as the OAM hybrid type I, since they involve "augmentation or modification of the stream of interest, or employment of methods that modify the treatment of the streams", according to [RFC7799].

On-path telemetry is useful for application-aware networking operations, not only in data center and enterprise networks, but also in carrier networks which may cross multiple domains. The techniques can provide benefits for carrier network operators in various scenarios. For example, it is critical for the operators who

offer high-bandwidth, latency and loss-sensitive services such as video streaming and online gaming to closely monitor the relevant flows in real-time as the basis for any further optimizations.

This document describes a reference framework, named as In-situ Flow Information Telemetry (IFIT), which outlines the system architecture for applying the OPT techniques to collect and correlate performance measurement information from the network. This framework document is intended to guide system designers attempting to use the referenced techniques as well as to motivate further work to enhance the telemetry ecosystem. It highlights requirements and challenges, outlines important techniques that are applicable, and provides examples of how these might be applied for critical use cases.

The document scope is discussed in Section 1.3.

1.1. Classification and OPT Modes

The operation of OPT differs from both active OAM and passive OAM as defined in [RFC7799]. It does not generate any active probe packets or passively observe unmodified user packets. Instead, it modifies selected user packets in order to collect useful information about them. Therefore, the operation is categorized as the hybrid OAM type I method per [RFC7799].

This hybrid OAM type I method can be further partitioned into two modes [passport-postcard]. In the passport mode, each node on the path can add telemetry data to the user packets (i.e., stamps the passport). The accumulated data trace is exported at a configured end node. In the postcard mode, each node directly exports the telemetry data using an independent packet (i.e., sends a postcard) while the user packets are unmodified. It is possible to combine the two modes together in one solution. We call this the hybrid mode.

Figure 1 shows the classification of the OPT techniques.

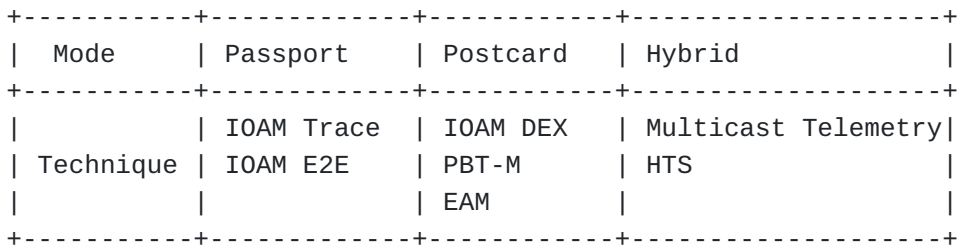

Figure 1: OPT Technique Classification

IOAM Trace and E2E options are described in [RFC9197].

EAM is described in [I-D.zhou-ippm-enhanced-alternate-marking].

IOAM DEX option is described in [I-D.ietf-ippm-ioam-direct-export].

PBT-M is described in [I-D.song-ippm-postcard-based-telemetry].

Multicast Telemetry is described in [I-D.ietf-mboned-multicast-telemetry].

HTS is described in [I-D.mirsky-ippm-hybrid-two-step].

The advantages of the passport mode include:

* It automatically retains the telemetry data correlation along the entire path. The self-describing feature simplifies the data consumption.

* The on-path data for a packet is only exported once so the data export overhead is low.

* Only the head and tail nodes of the paths need to be configured for header insertion and removal, so the configuration overhead is low.

The disadvantages of the passport mode include:

* The telemetry data carried by user packets inflate the packet size, which may be undesirable or prohibitive.

* Approaches for encapsulating the instruction header and data in transport protocols need to be standardized.

* Carrying sensitive data along the path is vulnerable to security and privacy breach.

* If a packet is dropped on the path, the data collected are also lost.

The postcard mode complements the passport mode. The advantages of the postcard mode include:

* Either there is no packet header overhead (e.g., PBT-M) or the overhead is small and fixed (e.g., IOAM DEX).

* The encapsulation requirement may be avoided (e.g., PBT-M).

* The telemetry data can be secured before export.

* Even if a packet is dropped on the path, the partial data collected are still available.

The disadvantages of the postcard mode include:

- Telemetry data are spread in multiple postcards so extra effort is needed to correlate the data.

- Every node exports a postcard for a packet which increases the data export overhead.

- In case of PBT-M, every node on the path needs to be configured, so the configuration overhead is high.

- In case of IOAM DEX, the transport encapsulation requirement remains.

The hybrid mode either tailors for some specific application scenario (e.g., Multicast Telemetry) or provides some alternative approach (e.g., HTS). A postcard can be sent per segment of a path or the telemetry data can be carried in a companion packet following each monitored use packet. The hybrid mode combines the advantages of both the passport mode and the postcard mode, but it may incur extra processing complexity.

1.2. Requirements and Challenges

Although OPT is beneficial, successfully applying such techniques in carrier networks must consider performance, deployability, and flexibility. Specifically, we need to address the following practical deployment challenges:

- C1: OPT incurs extra packet processing which may cause stress on the network data plane. The potential impact on the forwarding performance creates an unfavorable "observer effect" (where the actions of performing OPT may change the behavior of the traffic being measured). This will not only damage the fidelity of the measurement, but also defy the purpose of the measurement.

- C2: OPT can generate a considerable amount of data which may claim too much transport bandwidth and inundate the servers for data collection, storage, and analysis. For example, if the technique is applied to all the traffic, one node may collect a few tens of bytes as telemetry data for each packet. The whole forwarding path might accumulate telemetry data with a size similar to or even exceeding that of the original packet.

- C3: The collectible data defined currently are essential but limited. This, in turn, limits the management and operational techniques that can be applied. Flexibility and extensibility of data definition, aggregation, acquisition, and filtering, must be considered.

*C4: Applying only a single underlying OPT technique may miss some important events or lead to incorrect results. For example, packet drop can cause the loss of the flow telemetry data and the packet drop location and reason remains unknown if only the In-situ OAM trace option is used. A comprehensive solution needs the flexibility to switch between different underlying techniques and adjust the configurations and parameters at runtime. Thus, system-level orchestration is needed.

*C5: We must provide solutions to support an incremental deployment strategy. That is, we need to support established encapsulation schemes for various predominant protocols such as Ethernet, IPv6, and MPLS with backward compatibility and properly handle various transport tunnels.

*C6: The development of simplified OPT primitives and models for configuration and queries is essential. Telemetry models may be utilized via an API-based telemetry service for external applications, for end-to-end performance measurement and application performance monitoring. Standard-based protocols and methods are needed for network configuration and programming, and telemetry data pre-processing and export, to provide interoperability.

1.3. Scope

Following the Network Telemetry Framework (NTF) [RFC9232], this document focuses on OPT, a specific subclass of data-plane telemetry techniques, and provides a high-level framework, IFIT, which addresses the challenges for deployment listed in Section 1.2, especially in carrier networks.

This document aims to clarify the problem space, essential requirements, and summarizes best practices and general system design considerations. This document provides examples to show the novel network telemetry applications under the framework.

As an informational document, it describes an open framework with a few key components. The framework does not enforce any specific implementation on each component, neither does it define interfaces (e.g., API, protocol) between components. The choice of underlying OPT techniques and other implementation details is determined by application implementers. Therefore, the framework is not a solution specification or a mandatory recommendation for OPT applications.

The standardization of the underlying techniques and interfaces mentioned in this document is undertaken by various working groups. Due to the limited scope and intended status of this document, it has no overlap or conflict with those works.

1.4. Relationship with Network Telemetry Framework (NTF)

[RFC9232] describes a Network Telemetry Framework (NTF). One dimension used by NTF to partition network telemetry techniques and systems is based on the three planes in networks (i.e., control plane, management plane, and forwarding plane) and external data sources. IFIT fits in the category of forwarding-plane telemetry and deals with the specific on-path technical branch of the forwarding-plane telemetry.

According to NTF, an OPT application mainly subscribes to event-triggered or streaming data. The key functional components of IFIT described in Section 2.2 match the general components in NTF with the details on the more specific functions. "On-demand Technique Selection and Integration" is an application layer function, matching the "Data Query, Analysis, and Storage" component in NTF; "Flexible Flow, Packet, and Data Selection" matches the "Data Configuration and Subscription" component; "Flexible Data Export" matches the "Data Encoding and Export" component; "Dynamic Network Probe" matches the "Data Generation and Processing" component.

1.5. Glossary

This section defines and explains the acronyms and terms used in this document.

OPT: On-Path Telemetry. Remotely acquiring performance and behavior data about network flows on a per-packet basis on the packet's forwarding path. The term refers to a class of data-plane telemetry techniques, including IOAM, PBT, EAM, and HTS. Such techniques may need to mark user packets, or insert instruction/metadata into the headers of user packets.

IFIT: In-situ Flow Information Telemetry is a high-level reference framework that shows how network data-plane monitoring and measurement applications can address the deployment challenges of the flow-oriented OPT techniques.

Reflective Telemetry: The reflective telemetry functions in a dynamic and closed-loop fashion. A new telemetry action is provisioned as a result of self-knowledge acquired through prior telemetry actions.

2. Architectural Concepts and Key Components of IFIT

To address the challenges mentioned in Section 1.2, a high-level framework which can help to build a workable and efficient OPT application is presented. IFIT is dedicated to OPT about user and application traffic flows. It covers OPT as the underlying techniques and works at the system level. The framework is comprised

of some key functional components (Section 2.2). By assembling these components, IFIT supports reflective telemetry that can potentially enable autonomous network operations (Section 2.3).

2.1. Reference Deployment

Figure 2 shows a reference deployment scenario of OPT.

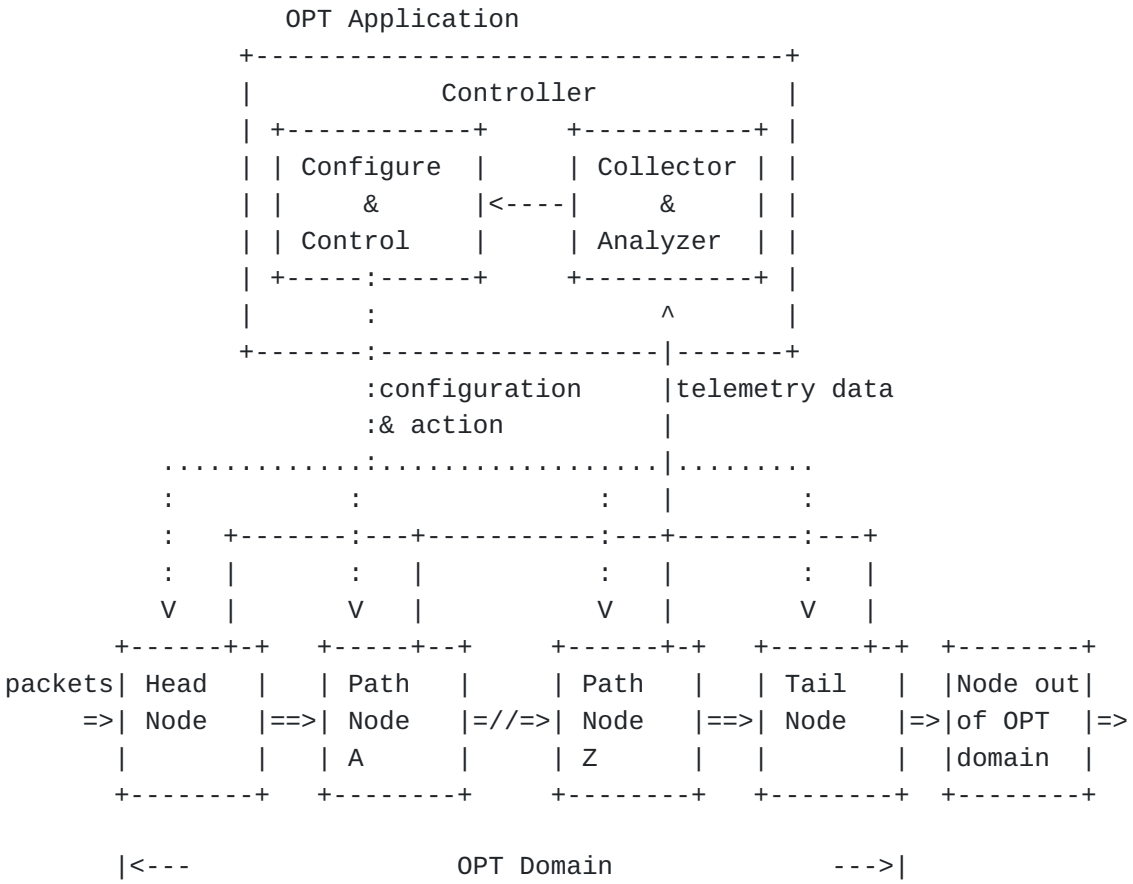

Figure 2: Deployment Scenario

An OPT application can conduct network data-plane monitoring and measurement tasks over a limited domain [RFC8799] by applying one or more underlying techniques. The application contains multiple elements, including configuring the network nodes and processing the telemetry data. The application usually uses a logically centralized controller for configuring the network nodes in the domain, and collecting and analyzing telemetry data. The configuration determines which underlying technique is used, what telemetry data are of interest, which flows and packets are concerned with, how the telemetry data are collected, etc. The process can be dynamic and interactive: after the telemetry data processing and analyzing, the application may instruct the controller to modify the configuration of the nodes, which affects the future telemetry data collection.

From the system-level view, it is recommended to use standardized configuration and data collection interfaces, regardless of the underlying technique. The specification of these interfaces and the implementation of the controller are out of scope for this document.

The OPT domain encompasses the head nodes and the end nodes, and may cross multiple network domains. The head nodes are responsible for enabling the OPT functions and the end nodes are responsible for terminating them. All capable nodes in this domain will be capable of executing the instructed OPT function. It is important to note that any application must, through configuration and policy, guarantee that any packet with OPT header and metadata will not leak out of the domain.

The underlying OPT techniques covered by the IFIT framework can be of any modes discussed in Section 1.1.

2.2. Key Components

The key components of IFIT to address the challenges listed in Section 1.2 are as follows. The components are described in more detail in the sections that follow.

* Flexible flow, packet, and data selection policy, addressing the challenge C1 described in Section 1;

* Flexible data export, addressing the challenge C2;

* Dynamic network probe, addressing C3;

* On-demand technique selection and integration, addressing C4.

Note that the challenges C5 and C6 are mostly standard-related, and are fundamental to IFIT. We discuss the protocol implications and guidance for solution developers in Section 3.

2.2.1. Flexible Flow, Packet, and Data Selection

In most cases, it is impractical to enable data collection for all the flows and for all the packets in a flow due to the potential performance and bandwidth impact. Therefore, a workable solution usually need to select only a subset of flows and flow packets on which to enable data collection, even though this means the loss of some information and accuracy.

In the data plane, a flow filter like those used for an Access Control List (ACL) provides an ideal means to determine the subset of flows. An application can set a sample rate or probability to a flow to allow only a subset of flow packets to be monitored, collect a different set of data for different packets, and disable or enable

data collection on any specific network node. An application can further allow any node to accept or deny the data collection process in full or partially.

Based on these flexible mechanisms, IFIT allows applications to apply flexible flow and data selection policies to suit their requirements. The applications can dynamically change the policies at any time based on the network load, processing capability, focus of interest, and any other criteria.

2.2.1.1. Block Diagram

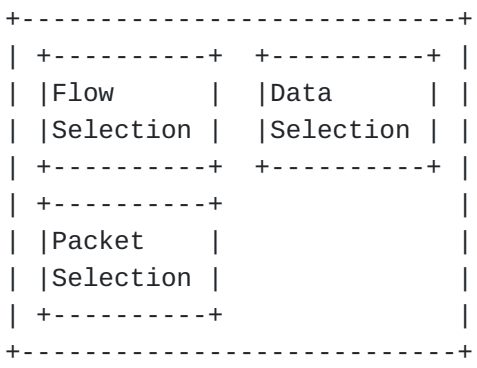

Figure 3: Flexible Flow, Packet, and Data Selection

Figure 3 shows the block diagram of this component. The flow selection block defines the policies to choose target flows for monitoring. Flow has different granularity. A basic flow is defined by 5-tuple IP header fields. Flow can also be aggregated at interface level, tunnel level, protocol level, and so on. The packet selection block defines the policies to choose packets from a target flow. The policy can be either a sampling interval, a sampling probability, or some specific packet signature. The data selection block defines the set of data to be collected. This can be changed on a per-packet or per-flow basis.

2.2.1.2. Example: Sketch-guided Elephant Flow Selection

Network operators are usually more interested in elephant flows which consume more resource and are sensitive to changes in network conditions. A CountMin Sketch [CMSketch] can be used on the data path of the head nodes, which identifies and reports the elephant flows periodically. The controller maintains a current set of elephant flows and dynamically enables OPT for only these flows.

2.2.1.3. Example: Adaptive Packet Sampling

Applying OPT on all packets of the selected flows can still be out of reach. A sample rate should be set for these flows and telemetry

should only be enabled on the sampled packets. However, the head nodes have no clue on the proper sampling rate. An overly high rate would exhaust the network resource and even cause packet drops; An overly low rate, on the contrary, would result in the loss of information and inaccuracy of measurements.

An adaptive approach can be used based on the network conditions to dynamically adjust the sampling rate. Every node gives user traffic forwarding higher priority than telemetry data export. In case of network congestion, the telemetry can sense some signals from the data collected (e.g., deep buffer size, long delay, packet drop, and data loss). The controller may use these signals to adjust the packet sampling rate. In each adjustment period (i.e., RTT of the feedback loop), the sampling rate is either decreased or increased in response of the signals. An Additive Increase/Multiplicative Decrease (AIMD) policy similar to the TCP flow control mechanism for rate adjustment can be used.

2.2.2. Flexible Data Export

The flow telemetry data can catch the dynamics of the network and the interactions between user traffic and network. Nevertheless, the data may contain redundancy. It is advisable to remove the redundancy from the data in order to reduce the data transport bandwidth and server processing load.

In addition to efficient export data encoding (e.g., IPFIX [RFC7011] or protobuf), nodes have several other ways to reduce the export data by taking advantage of network device's capability and programmability. Nodes can cache the data and send the accumulated data in batches if the data is not time sensitive. Various deduplication and compression techniques can be applied on the batched data.

From the application perspective, an application may only be interested in some special events which can be derived from the telemetry data. For example, in the case that the forwarding delay of a packet exceeds a threshold, or a flow changes its forwarding path is of interest, it is unnecessary to send the original raw data to the data collecting and processing servers. Rather, IFIT takes advantage of the in-network computing capability of network devices to process the raw data and only push the event notifications to the subscribing applications.

Such events can be expressed as policies. A policy can request data export only on change, on exception, on timeout, or on threshold.

2.2.2.1. Block Diagram

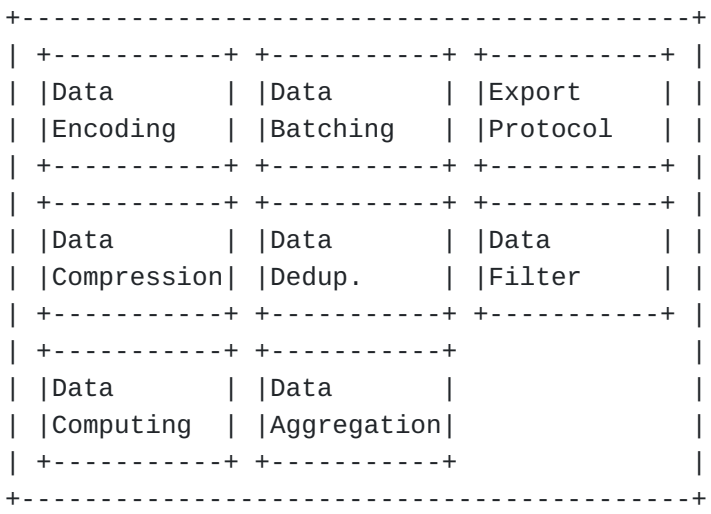

Figure 4: Flexible Data Export

Figure 4 shows the block diagram of this component. The data encoding block defines the method to encode the telemetry data. The data batching block defines the size of batch data buffered at the device side before export. The export protocol block defines the protocol used for telemetry data export. The data compression block defines the algorithm to compress the raw data. The data deduplication block defines the algorithm to remove the redundancy in the raw data. The data filter block defines the policies to filter the needed data. The data computing block defines the policies to prepocess the raw data and generate some new data. The data aggregation block defines the procedure to combine and synthesize the data.

2.2.2.2. Example: Event-based Anomaly Monitor

Network operators are interested in anomalies such as path change, network congestion, and packet drop. Such anomalies are hidden in raw telemetry data (e.g., path trace, timestamp). Such anomalies can be described as events and programmed into the device data plane. Only the triggered events are exported. For example, if a new flow appears at any node, a path change event is triggered; if the packet delay exceeds a predefined threshold in a node, the congestion event is triggered; if a packet is dropped due to buffer overflow, a packet drop event is triggered.

The export data reduction due to such optimization is substantial. For example, given a single 5-hop 10Gbps path, assume a moderate number of 1 million packets per second are monitored, and the telemetry data plus the export packet overhead consume less than 30 bytes per hop. Without such optimization, the bandwidth consumed by the telemetry data can easily exceed 1Gbps (more than 10% of the path bandwidth), When the optimization is used, the bandwidth

consumed by the telemetry data is negligible. Moreover, the pre-processed telemetry data greatly simplify the work of data analyzers.

2.2.3. Dynamic Network Probe

Due to limited data plane resource and network bandwidth, it is unlikely one can monitor all the data all the time. On the other hand, the data needed by applications may be arbitrary but ephemeral. It is critical to meet the dynamic data requirements with limited resource.

Fortunately, data plane programmability allows new data probles to be dynamically loaded. These on-demand probes are called Dynamic Network Probes (DNP). DNP is the technique to enable probes for customized data collection in different network planes. When working with an OPT technique, DNP is loaded into the data plane through incremental programming or configuration. The DNP can effectively conduct data generation, processing, and aggregation.

DNP introduces flexibility and extensibility to IFIT. It can implement the optimizations for export data reduction motioned in the previous section. It can also generate custom data as required by today's and tomorrow's applications.

2.2.3.1. Block Diagram

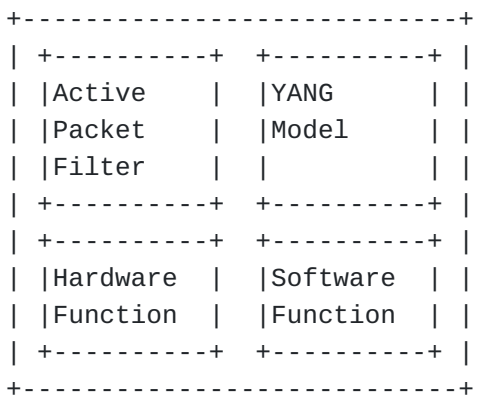

Figure 5: Dynamic Network Probes

Figure 5 shows the block diagram of this component. The active packet filter block is available in most hardware and it defines DNPs through dynamically update the packet filtering policies (including flow selection and action). YANG models can be dynamically deployed to enable different data processing and filtering functions. Some hardware allows dynamically loading hardware-based functions into the forwarding path at runtime through mechanisms such as reserved pipelines and function stubs.

Dynamically loadable software functions can be implemented in the control processors in capable nodes.

2.2.3.2. Examples

Following are some possible DNPs that can be dynamically deployed to support applications.

On-demand Flow Sketch: A flow sketch is a compact online data structure (usually a variation of multi-hashing table) for approximate estimation of multiple flow properties. It can be used to facilitate flow selection. The aforementioned CountMin Sketch [CMSketch] is such an example. Since a sketch consumes data plane resources, it should only be deployed when actually needed.

Smart Flow Filter: The policies that choose flows and packet sampling rate can change during the lifetime of an application.

Smart Statistics: An application may need to count flows based on different flow granularity or maintain hit counters for selected flow table entries.

Smart Data Reduction: DNP can be used to program the events that conditionally trigger data export.

2.2.4. On-demand Technique Selection and Integration

With multiple underlying data collection and export techniques at its disposal, IFIT can flexibly adapt to different network conditions and different application requirements.

For example, depending on the types of data that are of interest, IFIT may choose either passport or postcard mode to collect the data; if an application needs to track down where the packets are lost, switching from passport mode to postcard mode should be supported.

IFIT can further integrate multiple data plane monitoring and measurement techniques together and present a comprehensive data plane telemetry solution.

Based on the application requirements and the real-time telemetry data analysis results, new configurations and actions can be deployed.

2.2.4.1. Block Diagram

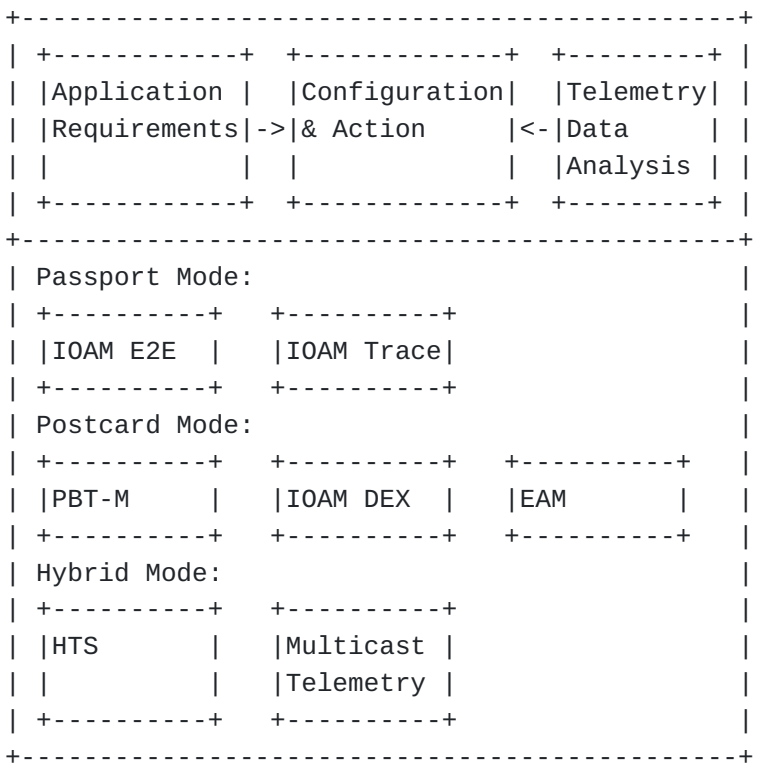

Figure 6: Technique Selection and Integration

Figure 6 shows the block diagram of this component, which lists the candidate OPT techniques.

Located in the logically centralized controller, this component makes all the control and configuration dynamically to the capable nodes in the domain which will affect the future telemetry data. The configuration and action decisions are based on the inputs from the application requirements and the realtime telemetry data analysis results. Note that here the telemetry data source is not limited to the data plane. The data can come form all the sources mentioned in [RFC9232], including external data sources.

2.3. IFIT for Reflective Telemetry

The components described in Section 2.2 can work together to support reflective telemetry, as shown in Figure 7.

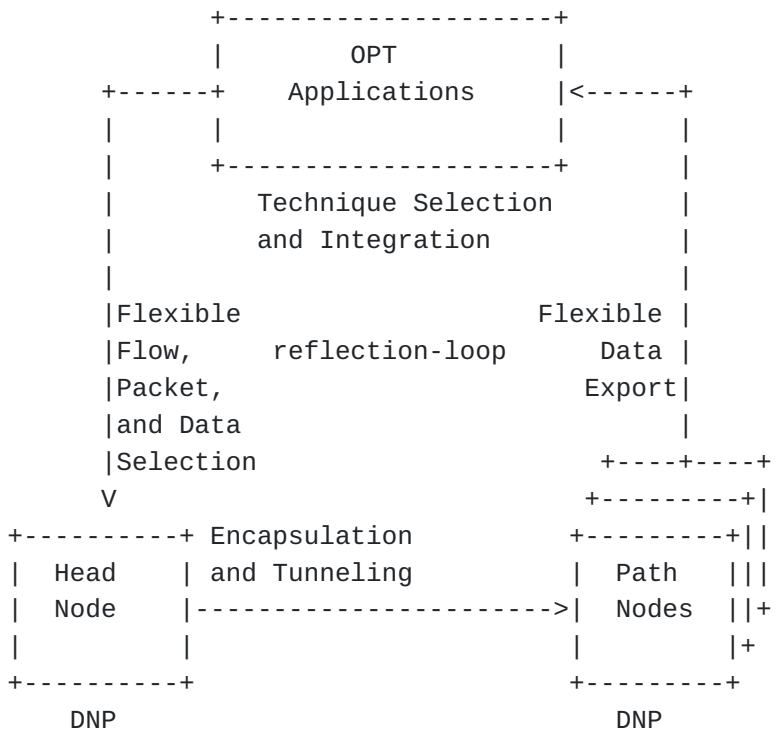

Figure 7: IFIT-based Reflective Telemetry

An application may pick a suite of telemetry techniques based on its requirements and apply an initial technique to the data plane. It then configures the head nodes to decide the initial target flows/packets and telemetry data set, the encapsulation and tunneling scheme based on the underlying network architecture, and the IFIT-capable nodes to decide the initial telemetry data export policy. Based on the network condition and the analysis results of the telemetry data, the application can change the telemetry technique, the flow/data selection policy, and the data export approach in real time without breaking the normal network operation. Many of such dynamic changes can be done through loading and unloading DNPs.

The reflective telemetry enabled by IFIT allows numerous new applications. Two examples are provided below.

2.3.1. Intelligent Multipoint Performance Monitoring

[RFC8889] describes an intelligent performance management based on the network condition. The idea is to split the monitoring network into clusters. The cluster partition that can be applied to every type of network graph and the possibility to combine clusters at different levels enable the so-called Network Zooming. It allows a controller to calibrate the network telemetry, so that it can start without examining in depth and monitor the network as a whole. In case of necessity (packet loss or too high delay), an immediate

detailed analysis can be reconfigured. In particular, the controller, that is aware of the network topology, can set up the most suitable cluster partition by changing the traffic filter or activate new measurement points and the problem can be localized with a step-by-step process.

An application on top of the controllers can manage such mechanism, whose dynamic and reflective operations are supported by the IFIT framework.

2.3.2. Intent-based Network Monitoring

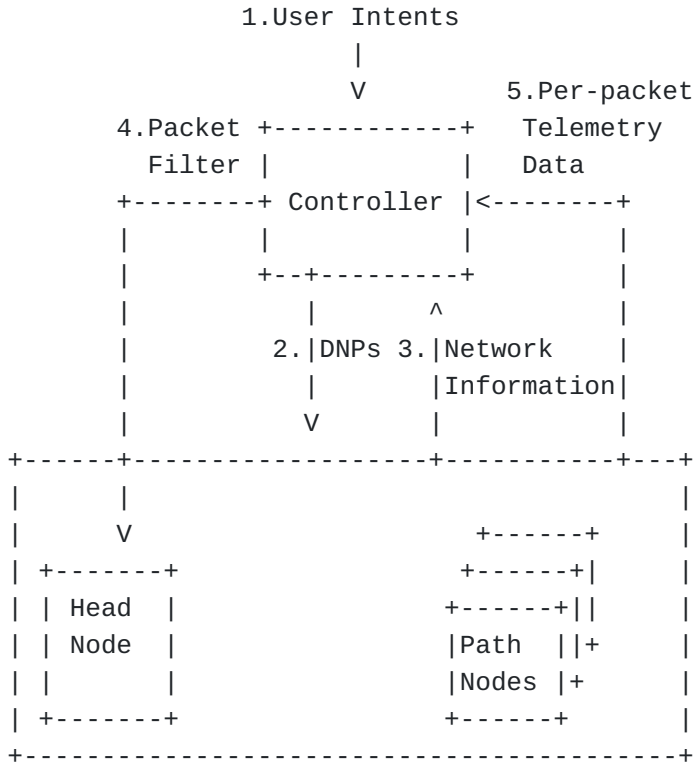

Figure 8: Intent-based Monitoring

In this example, a user can express high level intents for network monitoring. The controller translates an intent and configures the corresponding DNPs in capable nodes which collect necessary network information. Based on the real-time information feedback, the controller runs a local algorithm to determine the suspicious flows. It then deploys specific packet filters to the head node to initiate the high precision per-packet OPT for these flows.

3. Guidance for Solution Developers

Having a high-level framework covering a class of related techniques promotes a holistic approach for standard development and helps to

avoid duplicated efforts and piecemeal solutions that only focus on a specific technique while omitting the compatibility and extensibility issues, which is important to a healthy ecosystem for network telemetry.

A complete IFIT-based solution needs standard interfaces for configuration and data extraction, and standard encapsulation on various transport protocols. It may also need standard API and primitives for application programming and deployment. [I-D.ietf-ippm-ioam-deployment] summarizes some techniques for encapsulation and data export for IOAM. Solution developers need to consider the aspects set out in the following subsections.

3.1. Encapsulation in Transport Protocols

Since the introduction of IOAM, the IOAM option header encapsulation schemes in various network protocols have been defined (e.g., [I-D.ietf-ippm-ioam-ipv6-options]). Similar encapsulation schemes are needed to cover the other OPT techniques. Meanwhile, the OPT header/data encapsulation schemes in some popular protocols, such as MPLS and SRv6, are also needed. PBT-M [I-D.song-ippm-postcard-based-telemetry] does not introduce new headers to the packets so the trouble of encapsulation for a new header is avoided. While there are some proposals which allow new header encapsulation in MPLS packets (e.g., [I-D.song-mpls-extension-header]) or in SRv6 packets (e.g., [I-D.song-spring-siam]), they are still in their infancy stage and require further work. Before standards are available, in a confined domain, pre-standard encapsulation approaches may be applied.

3.2. Tunneling Support

In carrier networks, it is common for user traffic to traverse various tunnels for QoS, traffic engineering, or security. Both the uniform mode and the pipe mode for tunnel support are required and described in [I-D.song-ippm-ioam-tunnel-mode]. The uniform mode treats the nodes in a tunnel uniformly as the nodes outside of the tunnel on a path. In contrast, the pipe mode abstracts all the nodes between the tunnel ingress and egress as a circuit so no nodes in the tunnel is visible to the nodes outside of the tunnel. With such flexibility, the operator can either gain a true end-to-end visibility or apply a hierarchical approach which isolates the monitoring domain between customer and provider.

3.3. Deployment Automation

Standard approaches that automate the function configuration, and capability query and advertisement, could either be deployed in a centralized fashion or a distributed fashion. The draft

[I-D.ietf-ippm-ioam-yang] provides a YANG model for IOAM configuration. Similar models needs to be defined for other techniques. It is also helpful to provide standards-based approaches for configuration in various network environments. For example, in Segment Routing (SR) networks, extensions to BGP or Path Computation Element Communication Protocol (PCEP) can be defined to distribute SR policies carrying OPT information, so that telemetry behavior can be enabled automatically when the SR policy is applied. [I-D.chen-pce-sr-policy-ifit] defines extensions to PCEP to configure SR policies for OPT. [I-D.ietf-idr-sr-policy-ifit] defines extensions to BGP for the same purpose. Additional capability discovery and dissemination will be needed for other types of networks.

To realize the potential of OPT, programming and deploying DNPs are important. ForCES [RFC5810] is a standard protocol for network device programming, which can be used for DNP deployment. Currently some related works such as [I-D.wwx-netmod-event-yang] and [I-D.bwd-netmod-eca-framework] have proposed to use YANG models to define the smart policies which can be used to implement DNPs. In the future, other approaches for hardware and software-based functions can be development to enhance the programmability and flexibility.

4. Security Considerations

In addition to the specific security issues discussed in each individual document on OPT, this document considers the overall security issues at the system level. This should serve as a guide to the OPT application developers and users. General security and privacy considerations for any network telemetry system are also discussed in [RFC9232].

Since the OPT techniques work on the network forwarding plane, the IFIT framework poses some security risks. The important and sensitive information about a network could be exposed to an attacker. Further, the OPT data might swamp various parts of the network, leading to a possible DoS attack.

Fortunately, security measures can be enforced on various parts of the framework to mitigate such threats. For example, the configuration can filter and rate limit the monitored traffic; encryption and authentication can be applied on the exported telemetry data; different underlying techniques can be chosen to adapt to the different network conditions.

5. IANA Considerations

This document includes no request to IANA.

6. Contributors

Other major contributors of this document include Giuseppe Fioccola, Daniel King, Zhenqiang Li, Zhenbin Li, Tianran Zhou, and James Guichard.

7. Acknowledgments

We thank Diego Lopez, Shwetha Bhandari, Joe Clarke, Adrian Farrel, Frank Brockners, Al Morton, Alex Clemm, Alan DeKok, Benoit Claise, and Warren Kumari for their constructive suggestions for improving this document.

8. References

8.1. Normative References

[RFC7799] Morton, A., "Active and Passive Metrics and Methods (with Hybrid Types In-Between)", RFC 7799, DOI 10.17487/RFC7799, May 2016, <https://www.rfc-editor.org/info/rfc7799>.

[RFC8799] Carpenter, B. and B. Liu, "Limited Domains and Internet Protocols", RFC 8799, DOI 10.17487/RFC8799, July 2020, <https://www.rfc-editor.org/info/rfc8799>.

8.2. Informative References

[CMSketch] Cormode, G. and S. Muthukrishnan, "An improved data stream summary: the count-min sketch and its applications", 2005, <http://dx.doi.org/10.1016/j.jalgor.2003.12.001>.

[I-D.bwd-netmod-eca-framework] Boucadair, M., Wu, Q., Wang, M., King, D., and C. Xie, "Framework for Use of ECA (Event Condition Action) in Network Self Management", Work in Progress, Internet-Draft, draft-bwd-netmod-eca-framework-00, 3 November 2019, <https://www.ietf.org/archive/id/draft-bwd-netmod-eca-framework-00.txt>.

[I-D.chen-pce-sr-policy-ifit] Chen, H., Yuan, H., Zhou, T., Li, W., Fioccola, G., and Y. Wang, "PCEP SR Policy Extensions to Enable IFIT", Work in Progress, Internet-Draft, draft-chen-pce-sr-policy-ifit-02, 10 July 2020, <https://www.ietf.org/archive/id/draft-chen-pce-sr-policy-ifit-02.txt>.

[I-D.ietf-idr-sr-policy-ifit] Qin, F., Yuan, H., Yang, S., Zhou, T., and G. Fioccola, "BGP SR Policy Extensions to Enable IFIT", Work in Progress, Internet-Draft, draft-ietf-idr-

sr-policy-ifit-05, 24 October 2022, <https://datatracker.ietf.org/api/v1/doc/document/draft-ietf-idr-sr-policy-ifit/>.

[I-D.ietf-ippm-ioam-deployment] Brockners, F., Bhandari, S., Bernier, D., and T. Mizrahi, "In-situ OAM Deployment", Work in Progress, Internet-Draft, draft-ietf-ippm-ioam-deployment-02, 12 October 2022, <https://www.ietf.org/archive/id/draft-ietf-ippm-ioam-deployment-02.txt>.

[I-D.ietf-ippm-ioam-direct-export] Song, H., Gafni, B., Brockners, F., Bhandari, S., and T. Mizrahi, "In-situ OAM Direct Exporting", Work in Progress, Internet-Draft, draft-ietf-ippm-ioam-direct-export-11, 23 September 2022, <https://www.ietf.org/archive/id/draft-ietf-ippm-ioam-direct-export-11.txt>.

[I-D.ietf-ippm-ioam-ipv6-options] Bhandari, S. and F. Brockners, "In-situ OAM IPv6 Options", Work in Progress, Internet-Draft, draft-ietf-ippm-ioam-ipv6-options-09, 11 October 2022, <https://www.ietf.org/archive/id/draft-ietf-ippm-ioam-ipv6-options-09.txt>.

[I-D.ietf-ippm-ioam-yang] Zhou, T., Guichard, J., Brockners, F., and S. Raghavan, "A YANG Data Model for In-Situ OAM", Work in Progress, Internet-Draft, draft-ietf-ippm-ioam-yang-04, 7 July 2022, <https://www.ietf.org/archive/id/draft-ietf-ippm-ioam-yang-04.txt>.

[I-D.ietf-mboned-multicast-telemetry] Song, H., McBride, M., Mirsky, G., Mishra, G., Asaeda, H., and T. Zhou, "Multicast On-path Telemetry using IOAM", Work in Progress, Internet-Draft, draft-ietf-mboned-multicast-telemetry-04, 11 August 2022, <https://www.ietf.org/archive/id/draft-ietf-mboned-multicast-telemetry-04.txt>.

[I-D.li-apn-framework] Li, Z., Peng, S., Voyer, D., Li, C., Liu, P., Cao, C., and G. S. Mishra, "Application-aware Networking (APN) Framework", Work in Progress, Internet-Draft, draft-li-apn-framework-06, 30 September 2022, <https://www.ietf.org/archive/id/draft-li-apn-framework-06.txt>.

[I-D.mirsky-ippm-hybrid-two-step] Mirsky, G., Lingqiang, W., Zhui, G., Song, H., and P. Thubert, "Hybrid Two-Step Performance Measurement Method", Work in Progress, Internet-Draft, draft-mirsky-ippm-hybrid-two-step-13, 25 April 2022, <https://www.ietf.org/archive/id/draft-mirsky-ippm-hybrid-two-step-13.txt>.

[I-D.song-ippm-ioam-tunnel-mode]
Song, H., Li, Z., Zhou, T., and Z. Wang, "In-situ OAM Processing in Tunnels", Work in Progress, Internet-Draft, draft-song-ippm-ioam-tunnel-mode-00, 27 June 2018, <https://www.ietf.org/archive/id/draft-song-ippm-ioam-tunnel-mode-00.txt>.

[I-D.song-ippm-postcard-based-telemetry]
Song, H., Mirsky, G., Filsfils, C., Abdelsalam, A., Zhou, T., Li, Z., Graf, T., Mishra, G., Shin, J., and K. Lee, "Marking-based Direct Export for On-path Telemetry", Work in Progress, Internet-Draft, draft-song-ippm-postcard-based-telemetry-14, 7 September 2022, <https://www.ietf.org/archive/id/draft-song-ippm-postcard-based-telemetry-14.txt>.

[I-D.song-mpls-extension-header] Song, H., Zhou, T., Andersson, L., Zhang, Z. J., and R. Gandhi, "MPLS Network Actions using Post-Stack Extension Headers", Work in Progress, Internet-Draft, draft-song-mpls-extension-header-11, 15 October 2022, <https://www.ietf.org/archive/id/draft-song-mpls-extension-header-11.txt>.

[I-D.song-spring-siam] Song, H., Mishra, G., and T. Pan, "SRv6 In-situ Active Measurement", Work in Progress, Internet-Draft, draft-song-spring-siam-04, 6 September 2022, <https://www.ietf.org/archive/id/draft-song-spring-siam-04.txt>.

[I-D.wwx-netmod-event-yang] Wu, Q., Bryskin, I., Birkholz, H., Liu, X., and B. Claise, "A YANG Data model for ECA Policy Management", Work in Progress, Internet-Draft, draft-wwx-netmod-event-yang-10, 1 November 2020, <https://www.ietf.org/archive/id/draft-wwx-netmod-event-yang-10.txt>.

[I-D.zhou-ippm-enhanced-alternate-marking]
Zhou, T., Fioccola, G., Liu, Y., Cociglio, M., Lee, S., and W. Li, "Enhanced Alternate Marking Method", Work in Progress, Internet-Draft, draft-zhou-ippm-enhanced-alternate-marking-11, 29 August 2022, <https://www.ietf.org/archive/id/draft-zhou-ippm-enhanced-alternate-marking-11.txt>.

[passport-postcard] Handigol, N., Heller, B., Jeyakumar, V., Mazieres, D., and N. McKeown, "Where is the debugger for

my software-defined network?", 2012, <https://doi.org/10.1145/2342441.2342453>.

[RFC5810] Doria, A., Ed., Hadi Salim, J., Ed., Haas, R., Ed., Khosravi, H., Ed., Wang, W., Ed., Dong, L., Gopal, R., and J. Halpern, "Forwarding and Control Element Separation (ForCES) Protocol Specification", RFC 5810, DOI 10.17487/RFC5810, March 2010, <https://www.rfc-editor.org/info/rfc5810>.

[RFC7011] Claise, B., Ed., Trammell, B., Ed., and P. Aitken, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information", STD 77, RFC 7011, DOI 10.17487/RFC7011, September 2013, <https://www.rfc-editor.org/info/rfc7011>.

[RFC8889] Fioccola, G., Ed., Cociglio, M., Sapio, A., and R. Sisto, "Multipoint Alternate-Marking Method for Passive and Hybrid Performance Monitoring", RFC 8889, DOI 10.17487/RFC8889, August 2020, <https://www.rfc-editor.org/info/rfc8889>.

[RFC8993] Behringer, M., Ed., Carpenter, B., Eckert, T., Ciavaglia, L., and J. Nobre, "A Reference Model for Autonomic Networking", RFC 8993, DOI 10.17487/RFC8993, May 2021, <https://www.rfc-editor.org/info/rfc8993>.

[RFC9197] Brockners, F., Ed., Bhandari, S., Ed., and T. Mizrahi, Ed., "Data Fields for In Situ Operations, Administration, and Maintenance (IOAM)", RFC 9197, DOI 10.17487/RFC9197, May 2022, <https://www.rfc-editor.org/info/rfc9197>.

[RFC9232] Song, H., Qin, F., Martinez-Julia, P., Ciavaglia, L., and A. Wang, "Network Telemetry Framework", RFC 9232, DOI 10.17487/RFC9232, May 2022, <https://www.rfc-editor.org/info/rfc9232>.

Authors' Addresses

Haoyu Song
Futurewei
2330 Central Expressway
Santa Clara,
United States of America

Email: haoyu.song@futurewei.com

Fengwei Qin
China Mobile
No. 32 Xuanwumenxi Ave., Xicheng District

Beijing, 100032
China

Email: qinfengwei@chinamobile.com

Huanan Chen
China Telecom
China

Email: chenhuan6@chinatelecom.cn

Jaehwan Jin
LG U+
South Korea

Email: daenamu1@lguplus.co.kr

Jongyoon Shin
SK Telecom
South Korea

Email: jongyoon.shin@sk.com