OPSAWG

Internet-Draft

Intended status: Informational
Expires: December 13, 2019

H. Song, Ed.
Futurewei
F. Qin
China Mobile
P. Martinez-Julia
NICT
L. Ciavaglia
Nokia
A. Wang
China Telecom
June 11, 2019

Network Telemetry Framework draft-ietf-opsawg-ntf-01

Abstract

This document provides an architectural framework for network telemetry to address the current and future network operation challenges and requirements. As evidenced by some key characteristics and industry practices, network telemetry covers technologies and protocols beyond the conventional network Operations, Administration, and Management (OAM), so it promises better flexibility, scalability, accuracy, coverage, and performance and allows automated control loops to suit both today's and tomorrow's network operation requirements. This document clarifies the terminologies and classifies the modules and components of a network telemetry system. The framework and taxonomy help to set a common ground for the collection of related work and provide guidance for future technique and standard developments.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of $\underline{\mathsf{BCP}}$ 78 and $\underline{\mathsf{BCP}}$ 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 December 13, 2019.

Copyright Notice

Copyright (c) 2019 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to $\underline{\mathsf{BCP}}$ 78 and the IETF Trust's Legal Provisions Relating to IETF Documents

(https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

$\underline{1}$. Introduction	3
<u>1.1</u> . Requirements Language	3
<u>2</u> . Motivation	4
<u>2.1</u> . Use Cases	4
<u>2.2</u> . Challenges	6
<u>2.3</u> . Glossary	7
2.4. Network Telemetry	8
3. The Necessity of a Network Telemetry Framework	9
4. Network Telemetry Framework	<u>11</u>
4.1. Data Acquiring Mechanisms	<u>11</u>
<u>4.2</u> . Data Objects	
4.3. Function Components	14
4.4. Existing Works Mapped in the Framework	16
5. Evolution of Network Telemetry	17
	18
7. IANA Considerations	<u>19</u>
	19
9. Acknowledgments	19
	19
10.1. Normative References	19
10.2. Informative References	
Appendix A. A Survey on Existing Network Telemetry Techniques .	23
A.1. Management Plane Telemetry	23
A.1.1. Requirements and Challenges	23
A.1.2. Push Extensions for NETCONF	
A.1.3. gRPC Network Management Interface	24
A.2. Control Plane Telemetry	
A.2.1. Requirements and Challenges	
A.2.2. BGP Monitoring Protocol	
A.3. Data Plane Telemetry	
·	26

Song, et al. Expires December 13, 2019 [Page 2]

<u>A.3.2</u> .	Technique Taxonomy	<u>26</u>
<u>A.3.3</u> .	The IPFPM technology	<u>27</u>
<u>A.3.4</u> .	Dynamic Network Probe	<u>29</u>
<u>A.3.5</u> .	IP Flow Information Export (IPFIX) protocol	<u>29</u>
<u>A.3.6</u> .	In-Situ OAM	<u>29</u>
<u>A.3.7</u> .	Postcard Based Telemetry	<u>30</u>
<u>A.4</u> . Ext	ernal Data and Event Telemetry	<u>30</u>
<u>A.4.1</u> .	Requirements and Challenges	<u>30</u>
<u>A.4.2</u> .	Sources of External Events	<u>31</u>
<u>A.4.3</u> .	Connectors and Interfaces	<u>32</u>
Authors' Ad	dresses	32

1. Introduction

Network visibility is essential for network operation. Network telemetry has been considered as an ideal means to gain sufficient network visibility with better flexibility, scalability, accuracy, coverage, and performance than conventional OAM technologies. However, network telemetry is a vague term. The scope and coverage of it cause confusion and misunderstandings. It is beneficial to have an unambiguous concept and a clear architectural framework for network telemetry, so we can better align the related technology and standard work.

First, we show some key characteristics of network telemetry which set a clear distinction from the conventional network OAM and show that some conventional OAM technologies can be considered a subset of the network telemetry technologies. We then provide an architectural framework for network telemetry to meet the current and future network operation requirements. Following the framework, we classify the components of a network telemetry system so we can easily map the existing and emerging techniques and protocols into the framework. At last, we outline a roadmap for the evolution of the network telemetry system.

The purpose of the framework and taxonomy is to set a common ground for the collection of related work and provide guidance for future technique and standard developments.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP
14 [RFC2119][RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Motivation

Thanks to the advance of the computing and storage technologies, today's big data analytics gives network operators an unprecedented opportunity to gain network insights and move towards network autonomy. Some operators start to explore the application of Artificial Intelligence (AI) to make sense of network data. Software tools can use the network data to detect and react on network faults, anomalies, and policy violations, as well as predicting future events. In turn, the network policy updates for planning, intrusion prevention, optimization, and self-healing may be applied.

It is conceivable that an intent-driven autonomic network [RFC7575] is the logical next step for network evolution following Software Defined Network (SDN), aiming to reduce (or even eliminate) human labor, make the most efficient usage of network resources, and provide better services more aligned with customer requirements. Although it takes time to reach the ultimate goal, the journey has started nevertheless.

However, while the data processing capability is improved and applications are hungry for more data, the networks lag behind in extracting and translating network data into useful and actionable information. The system bottleneck is shifting from data consumption to data supply. Both the number of network nodes and the traffic bandwidth keep increasing at a fast pace. The network configuration and policy change at a much smaller time slot than ever before. More subtle events and fine-grained data through all network planes need to be captured and exported in real time. In a nutshell, it is a challenge to get enough high-quality data out of network efficiently, timely, and flexibly. Therefore, we need to examine the existing network technologies and protocols, and identify any potential technique and standard gaps based on the real network and device architectures.

In the remaining of this section, first we discuss several key use cases for today's and future network operations. Next, we show why the current network OAM techniques and protocols are insufficient for these use cases. The discussion underlines the need of new methods, techniques, and protocols which we may assign under an umbrella term - network telemetry.

2.1. Use Cases

These use cases are essential for network operations. While the list is by no means exhaustive, it is enough to highlight the requirements for data velocity, variety, volume, and veracity in networks.

Song, et al. Expires December 13, 2019 [Page 4]

- Policy and Intent Compliance: Network policies are the rules that constraint the services for network access, provide service differentiation, or enforce specific treatment on the traffic. For example, a service function chain is a policy that requires the selected flows to pass through a set of ordered network functions. An intent is a high-level abstract policy which requires a complex translation and mapping process before being applied on networks. While a policy is enforced, the compliance needs to be verified and monitored continuously.
- SLA Compliance: A Service-Level Agreement (SLA) defines the level of service a user expects from a network operator, which include the metrics for the service measurement and remedy/penalty procedures when the service level misses the agreement. Users need to check if they get the service as promised and network operators need to evaluate how they can deliver the services that can meet the SLA.
- Root Cause Analysis: Any network failure can be the cause or effect of a sequence of chained events. Troubleshooting and recovery require quick identification of the root cause of any observable issues. However, the root cause is not always straightforward to identify, especially when the failure is sporadic and the related and unrelated events are overwhelming. While machine learning technologies can be used for root cause analysis, it up to the network to sense and provide all the relevant data.
- Network Optimization: This covers all short-term and long-term network optimization techniques, including load balancing, Traffic Engineering (TE), and network planning. Network operators are motivated to optimize their network utilization and differentiate services for better ROI or lower CAPEX. The first step is to know the real-time network conditions before applying policies for traffic manipulation. In some cases, micro-bursts need to be detected in a very short time-frame so that fine-grained traffic control can be applied to avoid network congestion. The long-term network capacity planning and topology augmentation also rely on the accumulated data of the network operations.
- Event Tracking and Prediction: The visibility of user traffic path and performance is critical for healthy network operation.

 Numerous related network events are of interest to network operators. For example, Network operators always want to learn where and why packets are dropped for an application flow. They also want to be warned of issues in advance so proactive actions can be taken to avoid catastrophic consequences.

2.2. Challenges

For a long time, network operators have relied upon SNMP [RFC3416], Command-Line Interface (CLI), or Syslog to monitor the network. Some other OAM techniques as described in [RFC7276] are also used to facilitate network troubleshooting. These conventional techniques are not sufficient to support the above use cases for the following reasons:

- o Most use cases need to continuously monitor the network and dynamically refine the data collection in real-time and interactively. The poll-based low-frequency data collection is ill-suited for these applications. Subscription-based streaming data directly pushed from the data source (e.g., the forwarding chip) is preferred to provide enough data quantity and precision at scale.
- o Comprehensive data is needed from packet processing engine to traffic manager, from line cards to main control board, from user flows to control protocol packets, from device configurations to operations, and from physical layer to application layer. Conventional OAM only covers a narrow range of data (e.g., SNMP only handles data from the Management Information Base (MIB)). Traditional network devices cannot provide all the necessary probes. An open and programmable network device is therefore needed.
- o Many application scenarios need to correlate network-wide data from multiple sources (i.e., from distributed network devices, different components of a network device, or different network planes). A piecemeal solution is often lacking the capability to consolidate the data from multiple sources. The composition of a complete solution, as partly proposed by Autonomic Resource Control Architecture(ARCA)

 [I-D.pedro-nmrg-anticipated-adaptation], will be empowered and guided by a comprehensive framework.
- o Some of the conventional OAM techniques (e.g., CLI and Syslog) are lack of formal data model. The unstructured data hinder the tool automation and application extensibility. Standardized data models are essential to support the programmable networks.
- o Although some conventional OAM techniques support data push (e.g., SNMP Trap [RFC2981][RFC3877], Syslog, and sFlow), the pushed data are limited to only predefined management plane warnings (e.g., SNMP Trap) or sampled user packets (e.g., sFlow). We require the data with arbitrary source, granularity, and precision which are beyond the capability of the existing techniques.

Song, et al. Expires December 13, 2019 [Page 6]

o The conventional passive measurement techniques can either consume too much network resources and render too much redundant data, or lead to inaccurate results; the conventional active measurement techniques can interfere with the user traffic and their results are indirect. We need techniques that can collect direct and ondemand data from user traffic.

2.3. Glossary

Before further discussion, we list some key terminology and acronyms used in this documents. We make an intended distinction between network telemetry and network OAM.

AI: Artificial Intelligence. Use machine-learning based technologies to automate network operation.

BMP: BGP Monitoring Protocol

DNP: Dynamic Network Probe

DPI: Deep Packet Inspection

gNMI: gRPC Network Management Interface

gRPC: gRPC Remote Procedure Call

IPFIX: IP Flow Information Export Protocol

IPFPM: IP Flow Performance Measurement

IOAM: In-situ OAM

NETCONF: Network Configuration Protocol

Network Telemetry: Acquiring network data remotely for network monitoring and operation. A general term for a large set of network visibility techniques and protocols, with the characteristics defined in this document. Network telemetry addresses the current network operation issues and enables smooth evolution toward intent-driven autonomous networks.

NMS: Network Management System

OAM: Operations, Administration, and Maintenance. A group of network management functions that provide network fault indication, fault localization, performance information, and data and diagnosis functions. Most conventional network monitoring techniques and protocols belong to network OAM.

Song, et al. Expires December 13, 2019 [Page 7]

PBT: Postcard-Based Telemetry

SNMP: Simple Network Management Protocol

YANG: A data modeling language for NETCONF

YANG FSM: A YANG model to define device side finite state machine

YANG PUSH: A method to subscribe pushed data from remote YANG datastore

2.4. Network Telemetry

Network telemetry has emerged as a mainstream technical term to refer to the newer data collection and consumption techniques, distinguishing itself from the convention techniques for network OAM. The representative techniques and protocols include IPFIX [RFC7011] and gPRC [I-D.kumar-rtgwg-grpc-protocol]. Network telemetry allows separate entities to acquire data from network devices so that data can be visualized and analyzed to support network monitoring and operation. Network telemetry overlaps with the conventional network OAM and has a wider scope than it. It is expected that network telemetry can provide the necessary network insight for autonomous networks and address the shortcomings of conventional OAM techniques.

One difference between the network telemetry and the network OAM is that the network telemetry assumes machines as data consumer rather than human operators. Hence, the network telemetry can directly trigger the automated network operation, while the conventional OAM tools usually help human operators to monitor and diagnose the networks and guide manual network operations. The difference leads to very different techniques.

Although the network telemetry techniques are just emerging and subject to continuous evolution, several characteristics of network telemetry have been well accepted (Note that network telemetry is intended to be an umbrella term covering a wide spectrum of techniques, so the following characteristics are not expected to be held by every specific technique):

- o Push and Streaming: Instead of polling data from network devices, the telemetry collector subscribes to the streaming data pushed from data sources in network devices.
- o Volume and Velocity: The telemetry data is intended to be consumed by machines rather than by human being. Therefore, the data volume is huge and the processing is often in realtime.

- o Normalization and Unification: Telemetry aims to address the overall network automation needs. The piecemeal solutions offered by the conventional OAM approach are no longer suitable. Efforts need to be made to normalize the data representation and unify the protocols.
- o Model-based: The telemetry data is modeled in advance which allows applications to configure and consume data with ease.
- o Data Fusion: The data for a single application can come from multiple data sources (e.g., cross-domain, cross-device, and cross-layer) and needs to be correlated to take effect.
- o Dynamic and Interactive: Since the network telemetry means to be used in a closed control loop for network automation, it needs to run continuously and adapt to the dynamic and interactive queries from the network operation controller.

Note that a technique does not need to have all the above characteristics to be qualified as telemetry. An ideal network telemetry solution may also have the following features or properties:

- o In-Network Customization: The data can be customized in network at run-time to cater to the specific need of applications. This needs the support of a programmable data plane which allows probes to be deployed at flexible locations.
- o Direct Data Plane Export: The data originated from data plane can be directly exported to the data consumer for efficiency, especially when the data bandwidth is large and the real-time processing is required.
- o In-band Data Collection: In addition to the passive and active data collection approaches, the new hybrid approach allows to directly collect data for any target flow on its entire forwarding path.
- o Non-intrusive: The telemetry system should avoid the pitfall of the "observer effect". That is, it should not change the network behavior and affect the forwarding performance.

3. The Necessity of a Network Telemetry Framework

Big data analytics and machine-learning based AI technologies are applied for network operation automation, relying on abundant data from networks. The single-sourced and static data acquisition cannot meet the data requirements. It is desirable to have a framework that

integrates multiple telemetry approaches from different layers. This allows flexible combinations for different applications. The framework would benefit application development for the following reasons:

- o The future autonomous networks will require a holistic view on network visibility. All the use cases and applications need to be supported uniformly and coherently under a single intelligent agent. Therefore, the protocols and mechanisms should be consolidated into a minimum yet comprehensive set. A telemetry framework can help to normalize the technique developments.
- o Network visibility presents multiple viewpoints. For example, the device viewpoint takes the network infrastructure as the monitoring object from which the network topology and device status can be acquired; the traffic viewpoint takes the flows or packets as the monitoring object from which the traffic quality and path can be acquired. An application may need to switch its viewpoint during operation. It may also need to correlate a service and impact on network experience to acquire the comprehensive information.
- o Applications require network telemetry to be elastic in order to efficiently use the network resource and reduce the performance impact. Routine network monitoring covers the entire network with low data sampling rate. When issues arise or trends emerge, the telemetry data source can be modified and the data rate can be boosted.
- o Efficient data fusion is critical for applications to reduce the overall quantity of data and improve the accuracy of analysis.

So far, some telemetry related work has been done within IETF. However, the work is fragmented and scattered in different working groups. The lack of coherence makes it difficult to assemble a comprehensive network telemetry system and causes repetitive and redundant work.

A formal network telemetry framework is needed for constructing a working system. The framework should cover the concepts and components from the standardization perspective. This document clarifies the layers on which the telemetry is exerted and decomposes the telemetry system into a set of distinct components that the existing and future work can easily map to.

4. Network Telemetry Framework

Network telemetry techniques can be classified from multiple dimensions. In this document, we provide three unique perspectives: data acquiring mechanisms, data objects, and function components.

4.1. Data Acquiring Mechanisms

Broadly speaking, network data can be acquired through subscription (push) and query (poll). A subscriber may request data when it is ready. It follows a Publish-Subscription (Pub-Sub) mode or a Subscription-Publish (Sub-Pub) mode. In the Pub-Sub mode, predefined data are published and multiple qualified subscribers can subscribe the data. In the Sub-Pub mode, a subscriber designates what data are of interest and demands the network devices to deliver the data when they are available.

In contrast, a querier expects immediate feedback from network devices. It is usually used in a more interactive environment. The queried data may be directly extracted from some specific data source, or synthesized and processed from raw data.

There are four types of data from network devices:

Simple Data: The data that are steadily available from some data store or static probes in network devices. such data can be specified by YANG model.

Complex Data: The data need to be synthesized or processed from raw data from one or more network devices. The data processing function can be statically or dynamically loaded into network devices.

Event-triggered Data: The data are conditionally acquired based on the occurrence of some event. An event can be modeled as a Finite State Machine (FSM).

Streaming Data: The data are continuously or periodically generated. It can be time series or the dump of databases. The streaming data reflect realtime network states and metrics and require large bandwidth and processing power.

The above data types are not mutual exclusive. For example, event-triggered data can be simple or complex, and streaming data can be event triggered. The relationships of these data types are illustrated in Figure 1

Song, et al. Expires December 13, 2019 [Page 11]

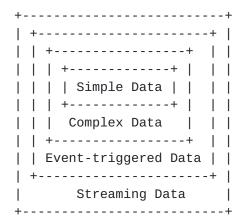


Figure 1: Data Type Relationship

Subscription usually deals with event-triggered data and streaming data, and query usually deals with simple data and complex data. It is easy to see that conventional OAM techniques are mostly about querying simple data only. While these techniques are still useful, advanced network telemetry techniques pay more attention on the other three data types, and prefer event/streaming data subscription and complex data query over simple data query.

4.2. Data Objects

Telemetry can be applied on the forwarding plane, the control plane, and the management plane in a network, as well as other sources out of the network, as shown in Figure 2. Therefore, we categorize the network telemetry into four distinct modules.

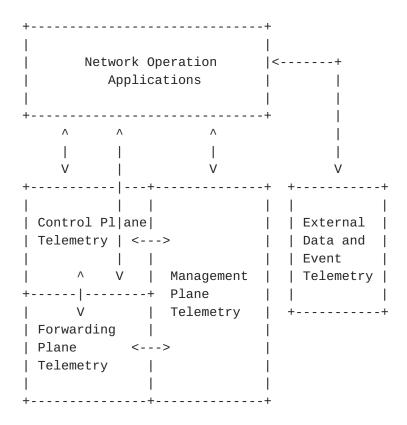


Figure 2: Layer Category of the Network Telemetry Framework

The rationale of this partition lies in the different telemetry data objects which result in different data source and export locations. Such differences have profound implications on in-network data programming and processing capability, data encoding and transport protocol, and data bandwidth and latency.

We summarize the major differences of the four modules in the following table. Some representative techniques are shown in some table blocks to highlight the technical diversity of these modules.

++	+			
Module 	Control Plane	Management Plane	Forwarding Plane	External Data
Object 	control protocol & signaling, RIB, ACL	config. & operation state, MIB	flow & packet QoS, traffic stat., buffer & queue stat.	social & environ-
Export Location 	main control CPU, linecard CPU or fwding chip	main control CPU 	fwding chip or linecard CPU; main control CPU unlikely	various
Model	YANG, custom	MIB, syslog, YANG, custom	template, YANG, custom	YANG
Encoding 	GPB, JSON, XML, plain	GPB, JSON, XML	plain	GPB, JSON XML, plain
Protocol 	gRPC,NETCONF, IPFIX,mirror	gPRC, NETCONF,	IPFIX, mirror 	gRPC
Transport 	HTTP, TCP, UDP	HTTP, TCP	UDP	HTTP,TCP UDP

Figure 3: Layer Category of the Network Telemetry Framework

Note that the interaction with the network operation applications can be indirect. For example, in the management plane telemetry, the management plane may need to acquire data from the data plane. Some of the operational states can only be derived from the data plane such as the interface status and statistics. For another example, the control plane telemetry may need to access the FIB in data plane. On the other hand, an application may involve more than one plane simultaneously. For example, an SLA compliance application may require both the data plane telemetry and the control plane telemetry.

4.3. Function Components

At each plane, the telemetry can be further partitioned into five distinct components:

Song, et al. Expires December 13, 2019 [Page 14]

- Data Query, Analysis, and Storage: This component works at the application layer. On the one hand, it is responsible for issuing data queries. The queries can be for modeled data through configuration or custom data through programming. The queries can be one shot or subscriptions for events or streaming data. On the other hand, it receives, stores, and processes the returned data from network devices. Data analysis can be interactive to initiate further data queries.
- Data Configuration and Subscription: This component deploys data queries on devices. It determines the protocol and channel for applications to acquire desired data. This component is also responsible for configuring the desired data that might not be directly available form data sources. The subscription data can be described by models, templates, or programs.
- Data Encoding and Export: This component determines how telemetry data are delivered to the data analysis and storage component. The data encoding and the transport protocol may vary due to the data exporting location.
- Data Generation and Processing: The requested data needs to be captured, processed, and formatted in network devices from raw data sources. This may involve in-network computing and processing on either the fast path or the slow path in network devices.
- Data Object and Source: This component determines the monitoring object and original data source. The data source usually just provides raw data which needs further processing. A data source can be considered a probe. A probe can be statically installed or dynamically installed.

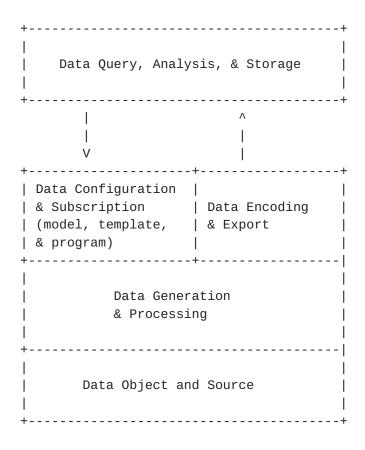


Figure 4: Components in the Network Telemetry Framework

Since most existing standard-related work belongs to the first four components, in the remainder of the document, we focus on these components only.

4.4. Existing Works Mapped in the Framework

The following two tables provide a non-exhaustive list of existing works (mainly published in IETF and with the emphasis on the latest new technologies) and shows their positions in the framework. The details about the mentioned work can be found in Appendix A.

	Query	Subscription
Simple Data	SNMP, NETCONF, YANG, BMP, IOAM, PBT,gPRC	ĺ
Custom Data	DNP, YANG FSM gRPC, NETCONF	
Event-triggered Data 	 	gRPC, NETCONF, YANG PUSH, DNP IOAM, PBT, YANG FSM
Streaming Data 		gRPC, NETCONF, IOAM, PBT, DNP IPFIX, IPFPM

Figure 5: Existing Work Mapping I

+			+
	Management Plane	Plane	Forwarding Plane
·	gRPC, NETCONF,	•	NETCONF/YANG, YANG FSM
data gen. & processing	DNP, YANG	DNP, YANG	In-situ OAM, PBT, IPFPM, DNP
data	gRPC, NETCONF YANG PUSH	BMP, NETCONF	IPFIX

Figure 6: Existing Work Mapping II

5. Evolution of Network Telemetry

As the network is evolving towards the automated operation, network telemetry also undergoes several levels of evolution.

- Level 0 Static Telemetry: The telemetry data is determined at design time. The network operator can only configure how to use it with limited flexibility.
- Level 1 Dynamic Telemetry: The telemetry data can be dynamically programmed or configured at runtime, allowing a tradeoff among resource, performance, flexibility, and coverage. DNP is an effort towards this direction.
- Level 2 Interactive Telemetry: The network operator can continuously customize the telemetry data in real time to reflect the network operation's visibility requirements. At this level, some tasks can be automated, although ultimately human operators will still need to sit in the middle to make decisions.
- Level 3 Closed-loop Telemetry: Human operators are completely excluded from the control loop. The intelligent network operation engine automatically issues the telemetry data request, analyzes the data, and updates the network operations in closed control loops.

While most of the existing technologies belong to level 0 and level 1, with the help of a clearly defined network telemetry framework, we can assemble the technologies to support level 2 and make solid steps towards level 3.

6. Security Considerations

Given that this document has proposed a framework for network telemetry and the telemetry mechanisms discussed are distinct (in both message frequency and traffic amount) from the conventional network OAM concepts, we must also reflect that various new security considerations may also arise. A number of techniques already exist for securing the data plane, control plane, and the management plane in a network, but the it is important to consider if any new threat vectors are now being enabled via the use of network telemetry procedures and mechanisms.

Security considerations for networks that use telemetry methods may include:

- o Telemetry framework trust and policy model;
- o Role management and access control for enabling and disabling telemetry capabilities;
- o Protocol transport used telemetry data and inherent security capabilities;

- o Telemetry data stores, storage encryption and methods of access;
- o Tracking telemetry events and any abnormalities that might identify malicious attacks using telemetry interfaces.

Some of the security considerations highlighted above may be minimized or negated with policy management of network telemetry. In a network telemetry deployment it would be advantageous to separate telemetry capabilities into different classes of policies, i.e., Role Based Access Control and Event-Condition-Action policies. Also, potential conflicts between network telemetry mechanisms must be detected accurately and resolved quickly to avoid unnecessary network telemetry traffic propagation escalating into an unintended or intended denial of service attack.

Further discussion and development of this section will be required, and it is expected that this security section, and subsequent policy section will be developed further.

7. IANA Considerations

This document includes no request to IANA.

8. Contributors

The other major contributors of this document are listed as follows.

- o Tianran Zhou
- o Zhenbin Li
- o Daniel King

Acknowledgments

We would like to thank Adrian Farrel, Randy Presuhn, Joe Clarke, Victor Liu, James Guichard, Uri Blumenthal, Giuseppe Fioccola, Yunan Gu, Parviz Yegani, Young Lee, Alexander Clemm, Qin Wu, and many others who have provided helpful comments and suggestions to improve this document.

10. References

10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
 Requirement Levels", BCP 14, RFC 2119,
 DOI 10.17487/RFC2119, March 1997,
 https://www.rfc-editor.org/info/rfc2119.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC
 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174,
 May 2017, https://www.rfc-editor.org/info/rfc8174>.

10.2. Informative References

[I-D.brockners-inband-oam-requirements]

Brockners, F., Bhandari, S., Dara, S., Pignataro, C., Gredler, H., Leddy, J., Youell, S., Mozes, D., Mizrahi, T., Lapukhov, P., and r. remy@barefootnetworks.com, "Requirements for In-situ OAM", draft-brockners-inband-oam-requirements-03 (work in progress), March 2017.

[I-D.fioccola-ippm-multipoint-alt-mark]

Fioccola, G., Cociglio, M., Sapio, A., and R. Sisto, "Multipoint Alternate Marking method for passive and hybrid performance monitoring", draft-fioccola-ippm-multipoint-alt-mark-04 (work in progress), June 2018.

[I-D.ietf-grow-bmp-adj-rib-out]

Evens, T., Bayraktar, S., Lucente, P., Mi, K., and S. Zhuang, "Support for Adj-RIB-Out in BGP Monitoring Protocol (BMP)", draft-ietf-grow-bmp-adj-rib-out-05 (work in progress), June 2019.

[I-D.ietf-grow-bmp-local-rib]

Evens, T., Bayraktar, S., Bhardwaj, M., and P. Lucente, "Support for Local RIB in BGP Monitoring Protocol (BMP)", draft-ietf-grow-bmp-local-rib-04 (work in progress), June 2019.

[I-D.ietf-netconf-udp-pub-channel]

Zheng, G., Zhou, T., and A. Clemm, "UDP based Publication Channel for Streaming Telemetry", <u>draft-ietf-netconf-udp-pub-channel-05</u> (work in progress), March 2019.

[I-D.ietf-netconf-yang-push]

Clemm, A. and E. Voit, "Subscription to YANG Datastores", draft-ietf-netconf-yang-push-25 (work in progress), May 2019.

- [I-D.kumar-rtgwg-grpc-protocol]
 Kumar, A., Kolhe, J., Ghemawat, S., and L. Ryan, "gRPC
 Protocol", draft-kumar-rtgwg-grpc-protocol-00 (work in
 progress), July 2016.
- [I-D.openconfig-rtgwg-gnmi-spec]
 Shakir, R., Shaikh, A., Borman, P., Hines, M., Lebsack,
 C., and C. Morrow, "gRPC Network Management Interface
 (gNMI)", <u>draft-openconfig-rtgwg-gnmi-spec-01</u> (work in progress), March 2018.
- [I-D.pedro-nmrg-anticipated-adaptation]

 Martinez-Julia, P., "Exploiting External Event Detectors to Anticipate Resource Requirements for the Elastic Adaptation of SDN/NFV Systems", draft-pedro-nmrg-anticipated-adaptation-02 (work in progress), June 2018.
- [I-D.song-ippm-postcard-based-telemetry]
 Song, H., Zhou, T., Li, Z., and J. Shin, "Postcard-based
 On-Path Flow Data Telemetry", draft-song-ippm-postcard-based-telemetry-03 (work in progress), April 2019.
- [I-D.song-opsawg-dnp4iq]
 Song, H. and J. Gong, "Requirements for Interactive Query with Dynamic Network Probes", draft-song-opsawg-dnp4iq-01 (work in progress), June 2017.
- [I-D.zhou-netconf-multi-stream-originators]
 Zhou, T., Zheng, G., Voit, E., Clemm, A., and A. Bierman,
 "Subscription to Multiple Stream Originators", draft-zhounetconf-multi-stream-originators-04 (work in progress),
 March 2019.
- [RFC1157] Case, J., Fedor, M., Schoffstall, M., and J. Davin,
 "Simple Network Management Protocol (SNMP)", RFC 1157,
 DOI 10.17487/RFC1157, May 1990,
 https://www.rfc-editor.org/info/rfc1157>.
- [RFC3416] Presuhn, R., Ed., "Version 2 of the Protocol Operations
 for the Simple Network Management Protocol (SNMP)",
 STD 62, RFC 3416, DOI 10.17487/RFC3416, December 2002,
 https://www.rfc-editor.org/info/rfc3416>.

- [RFC3877] Chisholm, S. and D. Romascanu, "Alarm Management Information Base (MIB)", <u>RFC 3877</u>, DOI 10.17487/RFC3877, September 2004, https://www.rfc-editor.org/info/rfc3877>.

- [RFC7276] Mizrahi, T., Sprecher, N., Bellagamba, E., and Y.
 Weingarten, "An Overview of Operations, Administration,
 and Maintenance (OAM) Tools", RFC 7276,
 DOI 10.17487/RFC7276, June 2014,
 https://www.rfc-editor.org/info/rfc7276>.

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

- [RFC8321] Fioccola, G., Ed., Capello, A., Cociglio, M., Castaldelli,
 L., Chen, M., Zheng, L., Mirsky, G., and T. Mizrahi,
 "Alternate-Marking Method for Passive and Hybrid
 Performance Monitoring", RFC 8321, DOI 10.17487/RFC8321,
 January 2018, https://www.rfc-editor.org/info/rfc8321.

Appendix A. A Survey on Existing Network Telemetry Techniques

We provide an overview of the challenges and existing solutions for each network telemetry module.

A.1. Management Plane Telemetry

A.1.1. Requirements and Challenges

The management plane of the network element interacts with the Network Management System (NMS), and provides information such as performance data, network logging data, network warning and defects data, and network statistics and state data. Some legacy protocols are widely used for the management plane, such as SNMP and Syslog. However, these protocols are insufficient to meet the requirements of the automatic network operation applications.

New management plane telemetry protocols should consider the following requirements:

- Convenient Data Subscription: An application should have the freedom to choose the data export means such as the data types and the export frequency.
- Structured Data: For automatic network operation, machines will replace human for network data comprehension. The schema languages such as YANG can efficiently describe structured data and normalize data encoding and transformation.
- High Speed Data Transport: In order to retain the information, a server needs to send a large amount of data at high frequency. Compact encoding formats are needed to compress the data and improve the data transport efficiency. The push mode, by replacing the poll mode, can also reduce the interactions between clients and servers, which help to improve the server's efficiency.

A.1.2. Push Extensions for NETCONF

NETCONF [RFC6241] is one popular network management protocol, which is also recommended by IETF. Although it can be used for data collection, NETCONF is good at configurations. YANG Push [I-D.ietf-netconf-yang-push] extends NETCONF and enables subscriber applications to request a continuous, customized stream of updates from a YANG datastore. Providing such visibility into changes made upon YANG configuration and operational objects enables new capabilities based on the remote mirroring of configuration and operational state. Moreover, distributed data collection mechanism [I-D.zhou-netconf-multi-stream-originators] via UDP based publication channel [I-D.ietf-netconf-udp-pub-channel] provides enhanced efficiency for the NETCONF based telemetry.

A.1.3. gRPC Network Management Interface

gRPC Network Management Interface (gNMI)

[I-D.openconfig-rtgwg-gnmi-spec] is a network management protocol based on the gRPC [I-D.kumar-rtgwg-grpc-protocol] RPC (Remote Procedure Call) framework. With a single gRPC service definition, both configuration and telemetry can be covered. gRPC is an HTTP/2 [RFC7540] based open source micro service communication framework. It provides a number of capabilities which are well-suited for network telemetry, including:

- o Full-duplex streaming transport model combined with a binary encoding mechanism provided further improved telemetry efficiency.
- o gRPC provides higher-level features consistency across platforms that common HTTP/2 libraries typically do not. This characteristic is especially valuable for the fact that telemetry data collectors normally reside on a large variety of platforms.
- o The built-in load-balancing and failover mechanism.

A.2. Control Plane Telemetry

A.2.1. Requirements and Challenges

The control plane telemetry refers to the health condition monitoring of different network protocols, which covers Layer 2 to Layer 7. Keeping track of the running status of these protocols is beneficial for detecting, localizing, and even predicting various network issues, as well as network optimization, in real-time and in fine granularity.

One of the most challenging problems for the control plane telemetry is how to correlate the E2E Key Performance Indicators (KPI) to a specific layer's KPIs. For example, an IPTV user may describe his User Experience (UE) by the video fluency and definition. Then in case of an unusually poor UE KPI or a service disconnection, it is non-trivial work to delimit and localize the issue to the responsible protocol layer (e.g., the Transport Layer or the Network Layer), the responsible protocol (e.g., ISIS or BGP at the Network Layer), and finally the responsible device(s) with specific reasons.

Traditional OAM-based approaches for control plane KPI measurement include PING (L3), Tracert (L3), Y.1731 (L2) and so on. One common issue behind these methods is that they only measure the KPIs instead of reflecting the actual running status of these protocols, making them less effective or efficient for control plane troubleshooting and network optimization. An example of the control plane telemetry is the BGP monitoring protocol (BMP), it is currently used to monitoring the BGP routes and enables rich applications, such as BGP peer analysis, AS analysis, prefix analysis, security analysis, and so on. However, the monitoring of other layers, protocols and the cross-layer, cross-protocol KPI correlations are still in their infancy (e.g., the IGP monitoring is missing), which require substantial further research.

A.2.2. BGP Monitoring Protocol

BGP Monitoring Protocol (BMP) [RFC7854] is used to monitor BGP sessions and intended to provide a convenient interface for obtaining route views.

The BGP routing information is collected from the monitored device(s) to the BMP monitoring station by setting up the BMP TCP session. The BGP peers are monitored by the BMP Peer Up and Peer Down Notifications. The BGP routes (including Adjacency_RIB_In [RFC7854], Adjacency_RIB_out [I-D.ietf-grow-bmp-adj-rib-out], and Local_Rib [I-D.ietf-grow-bmp-local-rib] are encapsulated in the BMP Route Monitoring Message and the BMP Route Mirroring Message, in the form of both initial table dump and real-time route update. In addition, BGP statistics are reported through the BMP Stats Report Message, which could be either timer triggered or event-driven. More BMP extensions can be explored to enrich the applications of BGP monitoring.

A.3. Data Plane Telemetry

A.3.1. Requirements and Challenges

An effective data plane telemetry system relies on the data that the network device can expose. The data's quality, quantity, and timeliness must meet some stringent requirements. This raises some challenges to the network data plane devices where the first hand data originate.

- o A data plane device's main function is user traffic processing and forwarding. While supporting network visibility is important, the telemetry is just an auxiliary function, and it should not impede normal traffic processing and forwarding (i.e., the performance is not lowered and the behavior is not altered due to the telemetry functions).
- o The network operation applications requires end-to-end visibility from various sources, which results in a huge volume of data. However, the sheer data quantity should not stress the network bandwidth, regardless of the data delivery approach (i.e., through in-band or out-of-band channels).
- o The data plane devices must provide timely data with the minimum possible delay. Long processing, transport, storage, and analysis delay can impact the effectiveness of the control loop and even render the data useless.
- o The data should be structured and labeled, and easy for applications to parse and consume. At the same time, the data types needed by applications can vary significantly. The data plane devices need to provide enough flexibility and programmability to support the precise data provision for applications.
- o The data plane telemetry should support incremental deployment and work even though some devices are unaware of the system. This challenge is highly relevant to the standards and legacy networks.

The industry has agreed that the data plane programmability is essential to support network telemetry. Newer data plane chips are all equipped with advanced telemetry features and provide flexibility to support customized telemetry functions.

A.3.2. Technique Taxonomy

There can be multiple possible dimensions to classify the data plane telemetry techniques.

Active and Passive: The active and passive methods (as well as the hybrid types) are well documented in [RFC7799]. The passive methods include TCPDUMP, IPFIX [RFC7011], sflow, and traffic mirror. These methods usually have low data coverage. The bandwidth cost is very high in order to improve the data coverage. On the other hand, the active methods include Ping, Traceroute, OWAMP [RFC4656], and TWAMP [RFC5357]. These methods are intrusive and only provide indirect network measurement results. The hybrid methods, including in-situ OAM

[I-D.brockners-inband-oam-requirements], IPFPM [RFC8321], and Multipoint Alternate Marking

[I-D.fioccola-ippm-multipoint-alt-mark], provide a well-balanced and more flexible approach. However, these methods are also more complex to implement.

In-Band and Out-of-Band: The telemetry data, before being exported
to some collector, can be carried in user packets. Such methods
are considered in-band (e.g., in-situ OAM
 [I-D.brockners-inband-oam-requirements]). If the telemetry data
is directly exported to some collector without modifying the user
packets, Such methods are considered out-of-band (e.g., postcard-based INT). It is possible to have hybrid methods. For example,
only the telemetry instruction or partial data is carried by user
packets (e.g., IPFPM [RFC8321]).

E2E and In-Network: Some E2E methods start from and end at the network end hosts (e.g., Ping). The other methods work in networks and are transparent to end hosts. However, if needed, the in-network methods can be easily extended into end hosts.

Flow, Path, and Node: Depending on the telemetry objective, the methods can be flow-based (e.g., in-situ OAM

[I-D.brockners-inband-oam-requirements]), path-based (e.g., Traceroute), and node-based (e.g., IPFIX [RFC7011]).

A.3.3. The IPFPM technology

The Alternate Marking method is efficient to perform packet loss, delay, and jitter measurements both in an IP and Overlay Networks, as presented in IPFPM [RFC8321] and [I-D.fioccola-ippm-multipoint-alt-mark].

This technique can be applied to point-to-point and multipoint-to-multipoint flows. Alternate Marking creates batches of packets by alternating the value of 1 bit (or a label) of the packet header. These batches of packets are unambiguously recognized over the network and the comparison of packet counters for each batch allows the packet loss calculation. The same idea can be applied to delay

Song, et al. Expires December 13, 2019 [Page 27]

measurement by selecting ad hoc packets with a marking bit dedicated for delay measurements.

Alternate Marking method needs two counters each marking period for each flow under monitor. For instance, by considering n measurement points and m monitored flows, the order of magnitude of the packet counters for each time interval is n*m*2 (1 per color).

Since networks offer rich sets of network performance measurement data (e.g packet counters), traditional approaches run into limitations. One reason is the fact that the bottleneck is the generation and export of the data and the amount of data that can be reasonably collected from the network. In addition, management tasks related to determining and configuring which data to generate lead to significant deployment challenges.

Multipoint Alternate Marking approach, described in [I-D.fioccola-ippm-multipoint-alt-mark], aims to resolve this issue and makes the performance monitoring more flexible in case a detailed analysis is not needed.

An application orchestrates network performance measurements tasks across the network to allow an optimized monitoring and it can calibrate how deep can be obtained monitoring data from the network by configuring measurement points roughly or meticulously.

Using Alternate Marking, it is possible to monitor a Multipoint Network without examining in depth by using the Network Clustering (subnetworks that are portions of the entire network that preserve the same property of the entire network, called clusters). So in case there is packet loss or the delay is too high the filtering criteria could be specified more in order to perform a detailed analysis by using a different combination of clusters up to a perflow measurement as described in IPFPM [RFC8321].

In summary, an application can configure end-to-end network monitoring. If the network does not experiment issues, this approximate monitoring is good enough and is very cheap in terms of network resources. However, in case of problems, the application becomes aware of the issues from this approximate monitoring and, in order to localize the portion of the network that has issues, configures the measurement points more exhaustively. So a new detailed monitoring is performed. After the detection and resolution of the problem the initial approximate monitoring can be used again.

A.3.4. Dynamic Network Probe

Hardware-based Dynamic Network Probe (DNP) [I-D.song-opsawg-dnp4iq] provides a programmable means to customize the data that an application collects from the data plane. A direct benefit of DNP is the reduction of the exported data. A full DNP solution covers several components including data source, data subscription, and data generation. The data subscription needs to define the complex data which can be composed and derived from the raw data sources. The data generation takes advantage of the moderate in-network computing to produce the desired data.

While DNP can introduce unforeseeable flexibility to the data plane telemetry, it also faces some challenges. It requires a flexible data plane that can be dynamically reprogrammed at run-time. The programming API is yet to be defined.

A.3.5. IP Flow Information Export (IPFIX) protocol

Traffic on a network can be seen as a set of flows passing through network elements. IP Flow Information Export (IPFIX) [RFC7011] provides a means of transmitting traffic flow information for administrative or other purposes. A typical IPFIX enabled system includes a pool of Metering Processes collects data packets at one or more Observation Points, optionally filters them and aggregates information about these packets. An Exporter then gathers each of the Observation Points together into an Observation Domain and sends this information via the IPFIX protocol to a Collector.

A.3.6. In-Situ OAM

Traditional passive and active monitoring and measurement techniques are either inaccurate or resource-consuming. It is preferable to directly acquire data associated with a flow's packets when the packets pass through a network. In-situ OAM (iOAM) [I-D.brockners-inband-oam-requirements], a data generation technique, embeds a new instruction header to user packets and the instruction directs the network nodes to add the requested data to the packets. Thus, at the path end, the packet's experience gained on the entire forwarding path can be collected. Such firsthand data is invaluable to many network OAM applications.

However, iOAM also faces some challenges. The issues on performance impact, security, scalability and overhead limits, encapsulation difficulties in some protocols, and cross-domain deployment need to be addressed.

A.3.7. Postcard Based Telemetry

PBT [I-D.song-ippm-postcard-based-telemetry] is an alternative to IOAM. PBT directly exports data at each node through an independent packet. PBT solves several issues of IOAM. It can also help to identify packet drop location in case a packet is dropped on its forwarding path.

A.4. External Data and Event Telemetry

Events that occur outside the boundaries of the network system are another important source of telemetry information. Correlating both internal telemetry data and external events with the requirements of network systems, as presented in Exploiting External Event Detectors to Anticipate Resource Requirements for the Elastic Adaptation of SDN/NFV Systems [I-D.pedro-nmrg-anticipated-adaptation], provides a strategic and functional advantage to management operations.

A.4.1. Requirements and Challenges

As with other sources of telemetry information, the data and events must meet strict requirements, especially in terms of timeliness, which is essential to properly incorporate external event information to management cycles. Thus, the specific challenges are described as follows:

- o The role of external event detector can be played by multiple elements, including hardware (e.g. physical sensors, such as seismometers) and software (e.g. Big Data sources that analyze streams of information, such as Twitter messages). Thus, the transmitted data must support different shapes but, at the same time, follow a common but extensible ontology.
- o Since the main function of the external event detectors is to perform the notifications, their timeliness is assumed. However, once messages have been dispatched, they must be quickly collected and inserted into the control plane with variable priority, which will be high for important sources and/or important events and low for secondary ones.
- o The ontology used by external detectors must be easily adopted by current and future devices and applications. Therefore, it must be easily mapped to current information models, such as in terms of YANG.

Organizing together both internal and external telemetry information will be key for the general exploitation of the management possibilities of current and future network systems, as reflected in

the incorporation of cognitive capabilities to new hardware and software (virtual) elements.

A.4.2. Sources of External Events

To ensure that the information provided by external event detectors and used by the network management solutions is meaningful for the management purposes, the network telemetry framework must ensure that such detectors (sources) are easily connected to the management solutions (sinks). This requires the specification of a simple taxonomy of detectors and match it to the connectors and/or interfaces required to connect them.

Once detectors are classified in such taxonomy, their definitions are enlarged with the qualities and other aspects used to handle them and represented in the ontology and information model (e.g. YANG). Therefore, differentiating several types of detectors as potential sources of external events is essential for the integrity of the management framework. We thus differentiate the following source types of external events:

- o Smart objects and sensors. With the consolidation of the Internet of Things~(IoT) any network system will have many smart objects attached to its physical surroundings and logical operation environments. Most of these objects will be essentially based on sensors of many kinds (e.g. temperature, humidity, presence) and the information they provide can be very useful for the management of the network, even when they are not specifically deployed for such purpose. Elements of this source type will usually provide a specific protocol for interaction, especially one of those protocols related to IoT, such as the Constrained Application Protocol (CoAP). It will be used by the telemetry framework to interact with the relevant objects.
- o Online news reporters. Several online news services have the ability to provide enormous quantity of information about different events occurring in the world. Some of those events can impact on the network system managed by a specific framework and, therefore, it will be interested on getting such information. For instance, diverse security reports, such as the Common Vulnerabilities and Exposures (CVE), can be issued by the corresponding authority and used by the management solution to update the managed system if needed. Instead of a specific protocol and data format, the sources of this kind of information usually follow a relaxed but structured format. This format will be part of both the ontology and information model of the telemetry framework.

Song, et al. Expires December 13, 2019 [Page 31]

o Global event analyzers. The advance of Big Data analyzers provides a huge amount of information and, more interestingly, the identification of events detected by analyzing many data streams from different origins. In contrast with the other types of sources, which are focused in specific events, the detectors of this source type will detect very generic events. For example, a sports event takes place and some unexpected movement makes it highly interesting and many people connects to sites that are covering such event. The systems supporting the services that cover the event can be affected by such situation so their management solutions should be aware of it. In contrast with the other source types, a new information model, format, and reporting protocol is required to integrate the detectors of this type with the management solution.

Additional types of detector types can be added to the system but they will be generally the result of composing the properties offered by these main classes. In any case, future revisions of the network telemetry framework will include the required types that cover new circumstances and that cannot be obtained by composition.

A.4.3. Connectors and Interfaces

For allowing external event detectors to be properly integrated with other management solutions, both elements must expose interfaces and protocols that are subject to their particular objective. Since external event detectors will be focused on providing their information to their main consumers, which generally will not be limited to the network management solutions, the framework must include the definition of the required connectors for ensuring the interconnection between detectors (sources) and their consumers within the management systems (sinks) are effective.

In some situations, the interconnection between the external event detectors and the management system is via the management plane. For those situations there will be a special connector that provides the typical interfaces found in most other elements connected to the management plane. For instance, the interfaces will accomplish with a specific information model (YANG) and specific telemetry protocol, such as NETCONF, SNMP, or gRPC.

Authors' Addresses

Haoyu Song (editor) Futurewei 2330 Central Expressway Santa Clara USA

Email: hsong@futurewei.com

Fengwei Qin China Mobile No. 32 Xuanwumenxi Ave., Xicheng District Beijing, 100032 P.R. China

Email: qinfengwei@chinamobile.com

Pedro Martinez-Julia NICT 4-2-1, Nukui-Kitamachi Koganei, Tokyo 184-8795 Japan

Email: pedro@nict.go.jp

Laurent Ciavaglia Nokia Villarceaux 91460 France

Email: laurent.ciavaglia@nokia.com

Aijun Wang China Telecom Beiqijia Town, Changping District Beijing, 102209 P.R. China

Email: wangaj.bri@chinatelecom.cn