Workgroup: RAW
Published: 18 May 2020
Intended Status: Informational
Expires: 19 November 2020
Authors: P. Thubert, Ed.   G.Z. Papadopoulos   R. Buddenberg
Cisco Systems   IMT Atlantique

Reliable and Available Wireless Architecture/Framework

Abstract

Due to uncontrolled interferences, including the self-induced multipath fading, deterministic networking can only be approached on wireless links. The radio conditions may change -way- faster than a centralized routing can adapt and reprogram, in particular when the controller is distant and connectivity is slow and limited. RAW separates the routing time scale at which a complex path is recomputed from the forwarding time scale at which the forwarding decision is taken for an individual packet. RAW operates at the forwarding time scale. The RAW problem is to decide, within the redundant solutions that are proposed by the routing, which will be used for each individual packet to provide a DetNet service while minimizing the waste of resources.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Bringing determinism in a packet network means eliminating the statistical effects of multiplexing that result in probabilistic jitter and loss. This can be approached with a tight control of the physical resources to maintain the amount of traffic within a budgetted volume of data per unit of time that fits the physical capabilities of the underlying technology, and the use of time-shared resources (bandwidth and buffers) per circuit, and/or by shaping and/or scheduling the packets at every hop.

Wireless networks operate on a shared medium where uncontrolled interference, including the self-induced multipath fading, adds another dimension to the statistical effects that affect the delivery. Scheduling transmissions can alleviate those effects by leveraging diversity in the spatial, time, code, and frequency domains, and provide a Reliable and Available service while preserving energy and optimizing the use of the shared spectrum.

Deterministic Networking is an attempt to mostly eliminate packet loss for a committed bandwidth with a guaranteed worst-case end-to-end latency, even when co-existing with best-effort traffic in a shared network. This innovation is enabled by recent developments in technologies including IEEE 802.1 TSN (for Ethernet LANs) and IETF DetNet (for wired IP networks). It is getting traction in various industries including manufacturing, online gaming, professional A/V, cellular radio and others, making possible many cost and performance optimizations.

The "Deterministic Networking Architecture" [RFC8655] is composed of three planes: the Application (User) Plane, the Controller Plane, and the Network Plane. Reliable and Available Wireless (RAW) extends RAW to focus on issues that are mostly a co"ern on wireless links, and inherits the architecture and the planes. A RAW Network Plane is thus a Network Plane inherited by RAW from DetNet, composed of one or multiple hops of homogeneous or heterogeneous technologies, e.g. a Wi-Fi6 Mesh or one-hop CBRS access links federated by a 5G backhaul.

RAW networking aims at providing highly available and reliable end-to-end performances in a network with scheduled wireless segments. Uncontrolled interference and transmission obstacles may impede the transmission, and techniques such as beamforming with Multi-User MIMO can only alleviate some of those issues, so the term "deterministic" is usually not associated with short range radios, in particular in the ISM band. This uncertainty places limits to the amount of traffic that can be transmitted on a link while conforming to a RAW Service Level Agreement (SLA) that may vary rapidly.

The wireless and wired media are fundamentally different at the physical level, and while the generic "Deterministic Networking Problem Statement" [RFC8557] applies to both the wired and the wireless media, the methods to achieve RAW must extend those used to support time-sensitive networking over wires, as a RAW solution has to address less consistent transmissions, energy conservation and shared spectrum efficiency.

The development of RAW technologies has been lagging behind deterministic efforts for wired systems both at the IEEE and the IETF. But recent efforts at the IEEE and 3GPP indicate that wireless is finally catching up at the lower layer and that it is now possible for the IETF to extend DetNet for wireless segments that are capable of scheduled wireless transmissions.

The intent for RAW is to provide DetNet elements that are specialized for short range radios. From this inheritance, RAW stays agnostic to the radio layer underneath though the capability to schedule transmissions is assumed. How the PHY is programmed to do so, and whether the radio is single-hop or meshed, are unknown at the IP layer and not part of the RAW abstraction.

Still, in order to focus on real-worlds issues and assert the feasibility of the proposed capabilities, RAW will focus on selected technologies that can be scheduled at the lower layers: IEEE Std. 802.15.4 timeslotted channel hopping (TSCH), 3GPP 5G ultra-reliable low latency communications (URLLC), IEEE 802.11ax/be where 802.11be is extreme high throughput (EHT), and L-band Digital Aeronautical Communications System (LDACS). See [RAW-TECHNOS] for more.

The establishment of a path is not in-scope for RAW. It may be the product of a centralized Controller Plane as described for DetNet. As opposed to wired networks, the action of installing a path over a set of wireless links may be very slow relative to the speed at which the radio conditions vary, and it makes sense in the wireless case to provide redundant forwarding solutions along a complex path and to leave it to the Network Plane to select which of those forwarding solutions are to be used for a given packet based on the current conditions.

RAW distinguishes the longer time scale at which routes are computed from the the shorter forwarding time scale where per-packet decisions are made. RAW operates at the forwarding time scale on one DetNet flow over one path that is preestablished and installed by means outside of the scope of RAW. The scope of the RAW WG comprises Network plane protocol elements such as Operations, Administration and Maintenance (OAM) and in-band control to improve the RAW operation at the Service and at the forwarding sub-layers. RAW controls whether to use packet replication, Automatic Repeat reQuest (ARQ), Hybrid ARQ (HARQ) that includes Forward Error Correction (FEC) and coding, with a constraint to limit the use of redundancy as is really needed, e.g., when a spike of loss is observed. This is discussed in more details in Section 5.3 and the next sections.

2. Terminology

RAW reuses terminology defined for DetNet in the "Deterministic Networking Architecture" [RFC8655], e.g., PREOF for Packet Replication, Elimination and Ordering Functions.

RAW also reuses terminology defined for 6TiSCH in [6TiSCH-ARCH] such as the term Track. 6TiSCH defined a Track as a complex path with associated PAREO operations.

RAW uses the term OAM as defined in [RFC6291].

RAW defines the following terms:

PAREO: Packet (hybrid) ARQ, Replication, Elimination and Ordering. PAREO is a superset Of DetNet's PREOF that includes radio-specific techniques such as short range broadcast, MUMIMO, constructive interference and overhearing, which can be leveraged separately or combined to increase the reliability.

Flapping: In the context of RAW, a link flaps when the wireless connectivity is interrupted for short transient times, typically of a subsecond duration.

In the context of the RAW work, Reliability and Availability are defined as follows:

Reliability: Reliability is a measure of the probability that an item will perform its intended function for a specified interval under stated conditions. For RAW, the service that is expected is delivery within a bounded latency and a failure is when the packet is either lost or delivered too late. RAW expresses reliability in terms of Mean Time Between Failure (MTBF) and Maximum Consecutive Failures (MCF). More in [NASA].

Availability: Availability is a measure of the relative amount of time where a path operates in stated condition, in other words (uptime)/(uptime+downtime). Because a serial wireless path may not be good enough to provide the required availability, and even 2 parallel paths may not be over a longer period of time, the RAW availability implies a path that is a lot more complex than what DetNet typically envisages (a Track).

3. Related Work at The IETF

RAW intersects with protocols or practices in development at the IETF as follows:

- The Dynamic Link Exchange Protocol (DLEP) [RFC8175] from [MANET] can be leveraged at each hop to derive generic radio metrics (e.g., based on LQI, RSSI, queueing delays and ETX) on individual hops.

- OAM work at [detnet] such as [DetNet-IP-OAM] for the case of the IP Data Plane observes the state of DetNet paths, typically MPLS and IPv6 pseudowires [DetNet-DP-FW], in the direction of the traffic. RAW needs feedback that flows on the reverse path and gathers instantaneous values from the radio receivers at each hop to inform back the source and replicating relays so they can make optimized forwarding decisions. The work named ICAN may be related as well.

- [BFD] detect faults in the path between an ingress and an egress forwarding engines, but is unaware of the complexity of a path with replication, and expects bidirectionality. BFD considers delivery as success whereas with RAW the bounded latency can be as important as the delivery itself.

- [SPRING] and [BIER] define in-band signaling that influences the routing when decided at the head-end on the path. There's already one RAW-related draft at BIER [BIER-PREF] more may follow. RAW will need new in-band signaling when the decision is distributed, e.g., required chances of reliable delivery to destination within latency. This signaling enables relays to tune retries and replication to meet the required SLA.

- *[CCAMP] defines protocol-independent metrics and parameters (measurement attributes) for describing links and paths that are required for routing and signaling in technology-specific networks. RAW would be a source of requirements for CCAMP to define metrics that are significant to the focus radios.

4. Use Cases and Requirements Served

[RFC8578] presents a number of wireless use cases including Wireless for Industrial Applications, Pro-Audio and SmartGrid. [RAW-USE-CASES] adds a number of use cases that demonstrate the need for RAW capabilities for new applications such as Pro-Gaming and drones. The use cases can be abstracted in two families, Loose Tracks, e.g., for first op Radio Access Protection and Strict Tracks, e.g., for End-to-End Protection in a wireless mesh.

4.1. Radio Access Protection

To maintain the committed reliability at all times, a wireless host may use more than one Radio Access Network (RAN) in parallel.

```
                                             ***   **
                          RAN 1  -----  ***     **  ***
                       /               *    **        ****
             +----+  /              *          **      ****
             |    |-                *                   *****
             |Host|--zzz- RAN 2 -- *      Internet       *****
             |    |-                 *                  *****
             +----+  $$ù              *              *******
                       \               ***   ***    *****
                          RAN n  --------  ***  *****

          zzz = flapping now  $$$ expensive
```

Figure 1: Radio Access Protection

The RANs may be heterogeneous, e.g., 5G [I-D.farkas-raw-5g] and Wi-Fi [RAW-TECHNOS] for high-speed communication, in which case a Layer-3 abstraction becomes useful to select which of the RANs are used at a particular point of time, and the amount of traffic that is distributed over each RAN.

The idea is that the rest of the path to the destination(s) is protected separately (e.g., uses non-congruent paths) and/or is a lot more reliable, e.g., wired. In that case, RAW observes reliability of the path through each of the RANs but only operates on the first hop.

4.2. End-to-End Protection in a Wireless Mesh

In radio technologies that support mesh networking (e.g., Wi-Fi and TSCH), a Track is a complex path with distributed PAREO capabilities. In that case, RAW operates through the multipath and makes decisions either at the Ingress or at every hop (more in Section 6.2).

```
         A-------B-------C-----D
        /  \   /       /        \
 Ingress ----M-------N--zzzzz--- Egress
        \      \   /            /
         P--zzz--Q-------------R

zzz = flapping now
```

Figure 2: End-to-End Protection

The Protection may be imposed by the source based on end-to-end OAM, or performed hop-by-hop, in which case the OAM must enables the intermediate Nodes to estimate the quality of the rest of the feasible paths in the sub-Track to the destination.

5. RAW Considerations

5.1. Reliability and Availability

5.1.1. High Availability Engineering Principles

The reliability criteria of a critical system pervade through its elements, and if the system comprises a data network then the data network is also subject to the inherited reliability and availability criteria. It is only natural to consider the art of high availability engineering and apply it to wireless communicaitons in the context of RAW.

There are three principles [pillars] of high availability engineering:

1. elimination of single points of failure
2. reliable crossover
3. prompt detection of failures as they occur.

These principles are common to all high availability systems, not just ones with Internet technology at the center. Examples of both non-Internet and Internet are included.

5.1.1.1. Elimination of Single Points of Failure

Physical and logical components in a system happen to fail, either as the effect of wear and tear, when used beyond acceptable limits, or due to a software bug. It is necessary to decouple component failure from system failure to avoid the latter. This allows failed components to be restored while the rest of the system continues to function.

A non-Internet example is a standby generator available to power the system on failure of grid power. An Internet example is more than one communication several non-congruent link/path between Nodes in a routable network.

There is a rather open-ended issue over alternate routes -- for example, when links are cabled through the same conduit, they form a shared risk link group (SRLG), and will share the same fate if the bundle is cut. Just how distributed the infrastructure is a matter of discussion; there is no single right answer. It should be noted that intermediate Nodes such as routers, switches, and the air medium itself can become single points of failure; this must be avoided, using link- and Node-disjoint paths, and, for RAW, a high degree of diversity in the transmissions over the air.

From an economics standpoint, executing this principle properly generally increases capitalization expense because of the redundant equipment. In a constrained network where the waste of energy and bandwidth should be minimized, an excessive use of redundant links must be avoided; for RAW this means that the extra bandwidth must only be used as a replacement of that lost due to a failure.

5.1.1.2. Reliable Crossover

Having a backup equipment has a limited value unless it can be reliably switched into use within the down-time parameters.

Using the backup generator example: one that does not automatically sense grid power failure, start itself, and place itself on line does not represent reliable crossover.

Routers and IGPs execute reliable crossover continuously because the routers will use any alternate routes that are available [RFC0791]. This is due to the stateless nature of IP datagrams and the dissociation of the datagrams from the forwarding routes they take. The "IP Fast Reroute Framework" [FRR] analyzes mechanisms for fast failure detection and path repair for IP Fast-Reroute, and discusses the case of multiple failures and SRLG. Examples of FRR techniques include Remote Loop-Free Alternate [RLFA-FRR] and backup label-switched path (LSP) tunnels for the local repair of LSP tunnels using RSVP-TE [RFC4090].

The DetNet PREOF leverages 1+1 redundancy whereby a packet is sent twice, over non-congruent paths. This avoids the gap during the fast reroute operation, but doubles the traffic in the network. In the case of RAW, the expectation is that multiple transient faults may happen in overlapping time windows, in which case the 1+1 redundancy with delayed reestablishment of the second path will not provide the required guarantees. The Data Plane must be configured with a sufficient degree of redundancy to select an alternate redundat path immediately upon a fault, without the need for a slow intervention from the controller plane.

5.1.1.3. Prompt Notification of Failures

The execution of the two above principles is likely to render a system where the user will rarely see a failure. But someone needs to in order to direct maintenance.

There are many reasons for system monitoring (FCAPS for fault, configuration, accounting, performance, security is a handy mental checklist) but fault monitoring is sufficient reason [STD 62] describes how to use SNMP to observe and correct long-term faults. "Overview and Principles of Internet Traffic Engineering" [TE] discusses the importance of measurement for network protection, and provides abstract an method for network survivability with the analysis of a traffic matrix as observed by SNMP, probing techniques, FTP, IGP link state advertisements, and more.

Using the art of SNMP, the above described backup generator would include an SNMP agent that can report the status of the generator (get messages) on demand, and report changes in status (e.g. startup, amount of fuel in the tank) (trap messages).

Those measurements are needed in the context of RAW to inform the controller and make the long term reactive decision to rebuild a complex path. But RAW itself operates in the Network Plane at a faster time scale. To act on the Data Plane, RAW needs live information from the Operational Plane , e.g., using Bidirectional Forwarding Detection [BFD] and its variants (bidirectional and remote BFD) to protect a link, and OAM techniques to protect a path.

5.1.2. Applying Reliability Concepts to Networking

The terms Reliaility and Availability are defined for use in RAW in Section 2 and the reader is invited to read [NASA] for more details on the general definition of Reliability. Practically speaking a number of nines is often used to indicate the reliability of a data link, e.g., 5 nines indicate a Packet Delivery Ratio (PDR) of 99.999%.

This number is typical in a wired environment where the loss is due to a random event such as a solar particle that affects the transmission of a particular frame, but does not affect the previous or next frame, nor frames transmitted on other links. Note that the QoS requirements in RAW may include a bounded latency, and a packet that arrives too late is a fault and not considered as delivered.

For a periodic pattern such as an automation control loop, this number is proportional to the Mean Time Between Failures (MTBF). If a single fault can have dramatic consequences, then the MTBF is the expression of the chances that an unwanted event occurs. In data networks, this is rarely the case. Packet loss cannot never be fully avoided and the systems are built to resist to one loss, e.g., using redundancy with Retries (HARQ) or Packet Replication and Elimination (PRE), or, in a typical control loop, by linear interpolation from the previous measuremnents.

But the linear interpolation method can not resist to multiple consecutive losses, and a high MTBF is desired as a guarantee that this will not happen, IOW that the losses-in-a-row can be bounded. In that case, what's really desired is a Maximum Consecutive Failures (MCF). If the number of losses in a row passes the MCF, the control loop has to abort. Engineers that build automated processes may use the network reliability expressed in nines or as an MTBF to provide an MCF, e.g., as described in section 7.4 of [RFC8578].

5.1.3. Reliability in the Context of RAW

In contrast with wired networks, errors in transmission are the predominent source of packet loss in wireless networks. The root cause may be of multiple origins:

Multipath Fading: A destructive interference by a reflection of the original signal.

A radio signal may be received directly (line-of-sight) and/or as a reflection on a physical structure (echo). The reflections take a longer path and are delayed by the extra distance divided by the speed of light in the medium. Depending on the frequency, the echo lands with a different phase which may add up to (constructive interference) or destroy the signal (destructive interference).

The affected frequencies depend on the relative position of the sender, the receiver, and all the reflecting objects in the environment. A given hop will suffer from multipath fading for multiple packets in a row till the something moves that changes the reflection patterns.

Co-channel Interference: Energy in the spectrum used for the transmission confuses the receiver.

The wireless medium itself is a Shared Risk Link Group (SRLG) for nearby users of the same spectrum, as an interference may affect multiple co-channel transmissions between different peers within the interference domain of the interferer, possibly even when they use different technologies.

Obstacle in Fresnel Zone: The optimal transmission happens when the Fresnel Zone between the sender and the receiver is free of obstacles.

As long as a physical object (e.g., a metallic trolley between peers) that affects the transmission is not removed, the quality of the link is affected.

In an environment that is rich of metallic structures and mobile objects, a single radio link will provide a fuzzy service, meaning that it cannot be trusted to transport the traffic reliably over a long period of time.

Transmission errors are typically not independent, and their nature and duration are unpredictable; as long as a physical object (e.g., a metallic trolley between peers) that affects the transmission is not removed, or as long as the interferer (e.g., a radar) keeps transmitting, a continuous stream of packets will be affected.

The key word to combat losses is diversity. A single packet may be sent at different times over different paths that rely on different radio frequencies and different PHY technologies, e.g., narrowband vs. spread spectrum. It is typically retried a number of times in case of a loss, and if possible the retries should again vary all possible parameters. Each form of diversity combats a particular cause of loss and use of diversity must be maximised to optimize the PDR.

5.2. RAW Prerequisites

A prerequisite to the RAW work is that an end-to-end routing function computes a complex sub-topology along which forwarding can happen between a source and one or more destinations. For 6TiSCH, this is a Track. The concept of Track is specified in the 6TiSCH Architecture [6TiSCH-ARCH]. Tracks provide a high degree of redundancy and diversity and enable RAW PREOF, end-to-end network coding, and possibly radio-specific abstracted techniques such as

ARQ, overhearing, frequency diversity, time slotting, and possibly others.

How the routing operation computes the Track is out of scope for RAW. The scope of the RAW operation is one Track, and the goal of the RAW operation is to optimize the use of the Track at the forwarding timescale to maintain the expected service while optimizing the usage of constrained resources such as energy and spectrum.

Another prerequisite is that an IP link can be established over the radio with some guarantees in terms of service reliability, e.g., it can be relied upon to transmit a packet within a bounded latency and provides a guaranteed BER/PDR outside rare but existing transient outage windows that can last from split seconds to minutes. The radio layer can be programmed with abstract parameters, and can return an abstract view of the state of the Link to help forwarding decision (think DLEP from MANET). In the layered approach, how the radio manages its PHY layer is out of control and out of scope. Whether it is single hop or meshed is also unknown and out of scope.

5.3. Routing Time Scale vs. Forwarding Time Scale

With DetNet, the end-to-end routing can be centralized and can reside outside the network. In wireless, and in particular in a wireless mesh, the path to the controller that performs the route computation and maintenance expensive in terms of critical resources such as air time and energy.

Reaching to the routing computation can also be slow in regards to the speed of events that affect the forwarding operation at the radio layer. Due to the cost and latency to perform a route computation, the controller plane is not expected to be sensitive/ reactive to transient changes. The abstraction of a link at the routing level is expected to use statistical operational metrics that aggregate the behavior of a link over long periods of time, and represent its availability as shades of gray as opposed to either up or down.

```
                 +----------------+
                 |  Controller    |
                 |    (PCE)       |
                 |  [Routing ]    |
                 |  [Function]    |
                 +----------------+
                         ^
                         |
                       Slow
                         |
     _-._-._-._-._-._-.  |  ._-._-._-._-._-._-._-._-._-._-._-
   _-._-._-._-._-._-._-. | _-._-._-._-._-._-._-._-._-._-._-
                         |
                     Expensive
                  ....   |  .......
              ....    .  | .       .....
           ....          v             ...
         ..   A-------B-------C---D     ..
      ...    /  \   /       /      \     ..
     .      I ----M-------N--zzz-- E  ..
     ..      \      \   /         /     .
       ..     P--zzz--Q----------R   ..
         ..                         ..
           .......               ...
                  ...............
    zzz = flapping now
```

Figure 3: Time Scales

In the case of wireless, the changes that affect the forwarding decision can happen frequently and often for short durations, e.g., a mobile object moves between a transmitter and a receiver, and will cancel the line of sight transmission for a few seconds, or a radar measures the depth of a pool and interferes on a particular channel for a split second.

There is thus a desire to separate the long term computation of the route and the short term forwarding decision. In such a model, the routing operation computes a complex Track that enables multiple Non-Equal Cost Multi-Path (N-ECMP) forwarding solutions, and leaves it to the Data Plane to make the per-packet decision of which of these possibilities should be used.

In the case of wires, the concept is known in traffic engineering where an alternate path can be used upon the detection of a failure in the main path, e.g., using OAM in MPLS-TP or BFD over a collection of SD-WAN tunnels. RAW formalizes a forwarding time scale that is an order(s) of magnitude shorter than the controler plane routing time scale, and separates the protocols and metrics that are used at both scales. Routing can operate on long term statistics such as delivery ratio over minutes to hours, but as a first approximation can ignore flapping. On the other hand, the RAW forwarding decision is made at packet speed, and uses information that must be pertinent at the present time for the current transmission.

6. RAW Architecture Elements

6.1. PAREO Functions

In a nutshell, PRE establishes several paths in a network to provide redundancy and parallel transmissions to bound the end-to-end delay to traverse the network. Optionally, promiscuous listening between paths is possible, such that the Nodes on one path may overhear transmissions along the other path. Considering the scenario shown in Figure 4, many different paths are possible for S to reach R. A simple way to benefit from this topology could be to use the two independent paths via Nodes A, C, E and via B, D, F. But more complex paths are possible by interleaving transmissions from the lower level of the path to the upper level.

PRE may also take advantage of the shared properties of the wireless medium to compensate for the potential loss that is incurred with radio transmissions. For instance, when the source sends to A, B may listen also and get a second chance to receive the frame without an additional transmission. Note that B would not have to listen if it already received that particular frame at an earlier timeslot in a dedicated transmission towards B.

```
                    (A)   (C)   (E)

     source (S)                        (R) (root)

                    (B)   (D)   (F)
```

Figure 4: A Typical Ladder Shape with Two Parallel Paths Toward the Destination

The PRE model can be implemented in both centralized and distributed scheduling approaches. In the centralized approach, a Path Computation Element (PCE) scheduler calculates the routes and schedules the communication among the Nodes along a circuit such as a Label switched path. In the distributed approach, each Node selects its route to the destination, typically using a source routing header. In both cases, at each Node in the paths, a default parent and alternative parent(s) should be selected to set up complex tracks.

In the following Subsections, all the required operations defined by PRE, namely, Alternative Path Selection, Packet Replication, Packet Elimination and Promiscuous Overhearing, are described.

6.1.1. Packet Replication

The objective of PRE is to provide deterministic networking properties: high reliability and bounded latency. To achieve this goal, determinism in every hop of the forwarding paths MUST be guaranteed. By employing a Packet Replication procedure, each Node forwards a copy of each data packet to multiple parents: its Default Parent (DP) and multiple Alternative Parents (APs). To do so, each Node (i.e., source and intermediate Node) transmits the data packet multiple times in unicast to each parent. For instance, in Figure 5, the source Node S is transmitting the packet to both parents, Nodes A and B, at two different times. An example schedule is shown in Table 1. Thus, the packet can use non-congruent paths to the destination.

```
            ===> (A) => (C) => (E) ===
          //         \\//   \\//       \\
source (S)           //\\   //\\         (R) (root)
          \\       //  \\ //  \\       //
            ===> (B) => (D) => (F) ===
```

Figure 5: Packet Replication: S transmits twice the same data packet, to its DP (A) and to its AP (B).

Channel	0	1	2	3	4	5	6
0	S->A	S->B	B->C	B->D	C->F	E->R	F->R
1		A->C	A->D	C->E	D->E	D->F	

Table 1: Packet Replication: Sample schedule

6.1.2. Packet Elimination

The replication operation increases the traffic load in the network, due to packet duplications. Thus, a Packet Elimination operation SHOULD be applied at each RPL DODAG level to reduce the unnecessary traffic. To this aim, once a Node receives the first copy of a data packet, it discards the subsequent copies. Because the first copy that reaches a Node is the one that matters, it is the only copy that will be forwarded upward. Then, once a Node performs the Packet Elimination operation, it will proceed with the Packet Replication operation to forward the packet toward the RPL DODAG Root.

6.1.3. Promiscuous Overhearing

Considering that the wireless medium is broadcast by nature, any neighbor of a transmitter may overhear a transmission. By employing the Promiscuous Overhearing operation, a DP and some AP(s) eventually have more chances to receive the data packets. In Figure 6, when Node A is transmitting to its DP (Node C), the AP (Node D) and its sibling (Node B) may decode this data packet as well. As a result, by employing corellated paths, a Node may have multiple opportunities to receive a given data packet. This feature not only enhances the end-to-end reliability but also it reduces the end-to-end delay and increases energy efficiency.

```
                  ===> (A) ====> (C) ====> (E) ====
                //     ^ | \\                      \\
      source (S)       | |   \\                      (R) (root)
                \\     | v     \\                  //
                  ===> (B) ====> (D) ====> (F) ====
```

Figure 6: Unicast to DP with Overhearing: by employing Promiscuous Overhearing, DP, AP and the sibling Nodes have more opportunities to receive the same data packet.

6.1.4. Constructive Interference

Constructive Interference can be seen as the reverse of Promiscuous Overhearing, and refers to the case where two senders transmit the exact same signal in a fashion that the emitted symbols add up at the receiver and permit a reception that would not be possible with a single sender at the same PHY mode and the same power level.

Constructive Interference was proposed on 5G, Wi-Fi7 and even tested on IEEE 802.14.5. The hard piece is to synchronize the senders to the point that the signals are emitted at slightly different time to offset the difference of propagation delay that corresponds to the difference of distance of the transmitters to the receiver at the speed of light to the point that the symbols are superposed long enough to be recognizable.

6.2. Wireless Tracks

The "6TiSCH Architecture" [6TiSCH-ARCH] introduces the concept of Track a a possibly complex path with the PAREO functions operated within.

A simple track is composed of a direct sequence of reserved hops to ensure the transmission of a single packet from a source Node to a destination Node across a multihop path.

A Complex Track is designed as a directed acyclic graph from a source Node towards a destination Node to support multi-path forwarding, as introduced in "6TiSCH Architecture" [6TiSCH-ARCH]. By employing PRE functions [RFC8655], several paths may be computed, and these paths may be more or less independent. For example, a complex Track may branch off and rejoin over non-congruent paths (branches).

Some more details for Deterministic Network PRE techniques are presented in the following Section.

7. RAW Architecture

RAW inherits the conceptual model described in section 4 of the DetNet Architecture [RFC8655].

A Controller Plane Function (CPF) called the Path Computation Element(PCE) [RFC4655] interacts with RAW Nodes over a Southbound API. The RAW Nodes are DetNet relays that are capable of additional diversity mechanisms and measurement functions related to the radio interface, in particular the PAREO redundancy mechanisms.

The PCE defines a complex path between an Ingress End System and an Egress End System, and indicates to the RAW Nodes where the PAREO operations may be actioned in the Network Plane. The path may be loosely expressed in order to traverse a non-RAW subnetwork. In that case, the expectation is that the non-RAW subnetwork can be neglected in the RAW computation, that is, considered infinitely fast, reliable and/or available in comparison with the links between RAW nodes.

```
                 CPF         CPF              CPF              CPF

       -+-+-+-+-+-+-+ Southbound -+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-

                     RAW  --z   RAW  --z   RAW  --z   RAW
                 z-- Node  z--  Node  z--  Node  z--  Node --z
      Ingress --z    /          /                     /     z-- Egress
      End           Z          Z                     Z          End
      Node   ---z   /          /                     /      z-- Node
               z-- RAW  --z   RAW   ( non-RAW ) --- RAW  ---z
                   Node  z--  Node --- ( Nodes  )   Node

         --z   radio                      wired
          z--  link                   --- link
```

Figure 7: RAW Nodes

The Link-Layer metrics are reported to the PCE in a time-aggregated, e.g., statistical fashion. Example Link-Layer metrics include typical Link bandwidth (the medium speed depends dynamically on the PHY mode and the number of users sharing the spectrum) and average availability and reliability figures.

Based on those metrics, the PCE installs a complex path with enough redundant forwarding solutions to ensure that the Network Plane can reliably deliver the packets within a System Level Agreement (SLA) associated to the flow. The SLA defines end-to-end reliability and availability figures, where reliability may be expressed a successful delivery within a bounded delay. One a path is established, end-to-end subpath and overall reliability and availability metrics are also reported to the PCE to assure that the SLA is continuously served and recompute the path if not.

Depending on the SLA, the path or a leg of the path may include non-RAW Nodes, either interleaved inside the path, or more typically till the Egress End Node. RAW observes the Lower-Layer Links between RAW nodes (typically, radio links) and the end-to-end Network Layer subpath to decide at all times which of the PAREO redundancy is actioned by which RAW Nodes.

7.1. PCE vs. PSE

Section 5.3 shows that the time scale at which RAW needs to operate is not that of the Controller Plane that needs to deal with a possibly large whole network and make global optimization across multiple flows that may contend for limited resources.

RAW separates the path computation time scale at which a complex path is recomputed from the path selection time scale at which the forwarding decision is taken for one or a few packets. RAW operates at the path selection time scale. The RAW problem is to decide, within the redundant solutions that are proposed by the PCE, which will be used for each packet to provide a Reliable and Available service while minimizing the waste of resources.

To that effect, RAW defines the Path Selection Engine (PSE) that is the counter-part of the PCE to perform rapid local adjustments of the forwarding tables to avoid excessive use of the resource diversity that the PCE selects. The PSE enables to exploit the richer forwarding capabilities with PAREO and scheduled transmissions at a faster time scale over the smaller domain that is the Track, either Loose or Strict.

	PCE (Not in Scope)	**PSE (In Scope)**
Operation	Centralized	Source-Routed or Distributed
Communication	Slow, expensive	Fast, local
Time Scale	Long (hours, days)	Short (seconds, sub-second)
Network Size	Large, many Tracks to optimize globally	Small, within one Track
Considered Metrics	Averaged, Statistical, Shade of grey	Instant values / boolean condition

Table 2: PCE vs. PSE

7.2. RAW OAM

The RAW OAM operation in the Network Plane observes a subset of the links along that redundant path and the RAW PSE makes the decision on which PAREO function in actioned at which RAW Node, for a packet or a small collection of packets.

In the case of a End-to-End Protection in a Wireless Mesh, the Track is strict and congruent with the path so all links are observed. Conversely, in the case of Radio Access Protection, the Track is Loose and in that case only the first hop is observed; the rest of the path is abstracted and considered infinitely reliable, meaning that the loss of a packet that was sent over one of the possible first hops is attributed to that first hop, even what a particular loss effectively happens farther down the path.

```
                                        ***   **
                     RAN 1  -----  ***     **  ***
                  /              *    **         ****
     +-------+  /              *          **      ****    +------+
     |Ingress|-                *                    *****  |Egress|
     |  End  |------ RAN 2 -- *      Internet       ****---| End  |
     |System |-                 *                  *****   |System|
     +-------+  \              *              *******      +------+
                  \               ***   ***    *****
                     RAN n  --------  ***  *****

             <------------------> <---------------------->
                Observed by OAM       Opaque to OAM
```

Figure 8: Observed Links in Radio Access Protection

The Links that are not observed by OAM are opaque to it, meaning that the OAM information is carried and possibly echoed as data. In the example above, the Internet is opaque and not controlled by RAW, but RAW measures the end-to-end latency and delivery ratio for packets sent over each if RAN 1, RAN 2 and RAN 3, and determines whether a packet should be sent over either or a collection of those access links.

7.3. Source-Routed vs. Distributed Forwarding Decision

Within a large routed topology, the route-over mesh operation builds a particular complex Track with one source and one or more destinations; within the Track, packets may follow different paths and may be subject to RAW forwarding operations that include replication, elimination, retries, overhearing and reordering.

The RAW forwarding decisions include the selection of points of replication and elimination, how many retries can take place, and a limit of validity for the packet beyond which the packet should be destroyed rather than forwarded uselessly further down the Track.

The decision to apply the RAW techniques must be done quickly, and depends on a very recent and precise knowledge of the forwarding conditions within the complex Track. There is a need for an observation method to provide the RAW Data Plane with the specific knowledge of the state of the Track for the type of flow of interest (e.g., for a QoS level of interest). To observe the whole Track in quasi real time, RAW will consider existing tools such as L2-triggers, DLEP, BFD and in-band and out-of-band OAM.

One possible way of making the RAW forwarding decisions is to make them all at the ingress and express them in-band in the packet, which requires new loose or strict Hop-by-hop signaling. To control the RAW forwarding operation along a Track for the individual packets, RAW may leverage and extend known techniques such as DetNet tagging, Segment Routing (SRv6) or BIER-TE such as done with [BIER-PREF].

An alternate way is to enable each forwarding Node to make the RAW forwarding decisions for a packet on its own, based on its knowledge of the expectation (timeliness and reliability) for that packet and a recent observation of the rest of the way across the possible paths within the Track. Information about the service should be placed in the packet and matched with the forwarding Node's capabilities and policies.

In either case, a per-flow state is installed in all intermediate Nodes to recognize the flow and determine the forwarding policy to be applied.

7.4. Flow Identification

Section 4.7 of the DetNet Architecture [RFC8655] ties the app-flow identification which is an appliation layer concept with the network path identification that depends on the networking technology by "exporting of flow identification", e.g., to a MPLS label.

With RAW, this exporting operation is injective but not bijective. e.g., a flow is fully placed within one RAW Track, but not all packets along that Track are necessarily part of the same flow. For instance, out-of-band OAM packets must circulate in the exact same fashion as the flows that they observe. It results that the flow identification that maps to to app-flow at the network layer must be separate from the path identification that is used to forward a packet.

```
        Flow 1 (6-tuple) ----+
                             |
   Flow 2 (6-tuple)  ---+    |
                        |    |
OAM     -----------+    |    |
                   |    |    |
                   |    |    |
              |    |    |    |    |
              |    v    v    v    |
              |                   |
              +---------+---------+
                        |
                        |
                        +------------> Track 1
                               (IP address, instanceId)
```

Figure 9: Flow Injection

Section 3.4 of the DetNet data-plane framework [DetNet-DP-FW] indicates that for a DetNet IP Data Plane, a flow is identified by an IPv6 6-tuple. With RAW, that 6-tuple is not what indicates the Track, in other words, the flow ID is not the Track ID.

For instance, the 6TiSCH Architecture [6TiSCH-ARCH] uses a combination of the address of the Ingress End System and an instance identifier in a Hop-by-hop option to indicate a Track. Packets that are tagged with the same (address, instance ID) tuple will experience the same forwarding behavior regardless of the IPv6 6-tuple, and regardless of whether they transport application flows or OAM.

8. Security Considerations

9. IANA Considerations

This document has no IANA actions.

10. Contributors

Xavi Vilajosana: Wireless Networks Research Lab, Universitat Oberta de Catalunya

Rex Buddenberg:

Remous-Aris Koutsiamanis: IMT Atlantique

Nicolas Montavont:
IMT Atlantique

11. Acknowledgments

TBD

12. References

12.1. Normative References

[6TiSCH-ARCH] Thubert, P., "An Architecture for IPv6 over the TSCH mode of IEEE 802.15.4", Work in Progress, Internet-Draft, draft-ietf-6tisch-architecture-28, 29 October 2019, <https://tools.ietf.org/html/draft-ietf-6tisch-architecture-28>.

[RAW-TECHNOS] Thubert, P., Cavalcanti, D., Vilajosana, X., and C. Schmitt, "Reliable and Available Wireless Technologies", Work in Progress, Internet-Draft, draft-thubert-raw-technologies-04, 6 January 2020, <https://tools.ietf.org/html/draft-thubert-raw-technologies-04>.

[RAW-USE-CASES] Papadopoulos, G., Thubert, P., Theoleyre, F., and C. Bernardos, "RAW use cases", Work in Progress, Internet-Draft, draft-bernardos-raw-use-cases-03, 8 March 2020, <https://tools.ietf.org/html/draft-bernardos-raw-use-cases-03>.

[RFC4655] Farrel, A., Vasseur, J.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <https://www.rfc-editor.org/info/rfc4655>.

[BFD] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010, <https://www.rfc-editor.org/info/rfc5880>.

[RFC6291] Andersson, L., van Helvoort, H., Bonica, R., Romascanu, D., and S. Mansfield, "Guidelines for the Use of the "OAM" Acronym in the IETF", BCP 161, RFC 6291, DOI 10.17487/RFC6291, June 2011, <https://www.rfc-editor.org/info/rfc6291>.

[RFC8578] Grossman, E., Ed., "Deterministic Networking Use Cases", RFC 8578, DOI 10.17487/RFC8578, May 2019, <https://www.rfc-editor.org/info/rfc8578>.

[RFC8175] Ratliff, S., Jury, S., Satterwhite, D., Taylor, R., and B. Berry, "Dynamic Link Exchange Protocol (DLEP)", RFC

8175, DOI 10.17487/RFC8175, June 2017, <https://www.rfc-editor.org/info/rfc8175>.

[RFC8557] Finn, N. and P. Thubert, "Deterministic Networking Problem Statement", RFC 8557, DOI 10.17487/RFC8557, May 2019, <https://www.rfc-editor.org/info/rfc8557>.

[RFC8655] Finn, N., Thubert, P., Varga, B., and J. Farkas, "Deterministic Networking Architecture", RFC 8655, DOI 10.17487/RFC8655, October 2019, <https://www.rfc-editor.org/info/rfc8655>.

12.2. Informative References

[RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791, DOI 10.17487/RFC0791, September 1981, <https://www.rfc-editor.org/info/rfc791>.

[TE] Awduche, D., Chiu, A., Elwalid, A., Widjaja, I., and X. Xiao, "Overview and Principles of Internet Traffic Engineering", RFC 3272, DOI 10.17487/RFC3272, May 2002, <https://www.rfc-editor.org/info/rfc3272>.

[STD 62] Harrington, D., Presuhn, R., and B. Wijnen, "An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks", STD 62, RFC 3411, DOI 10.17487/RFC3411, December 2002, <https://www.rfc-editor.org/info/rfc3411>.

[RFC4090] Pan, P., Ed., Swallow, G., Ed., and A. Atlas, Ed., "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, DOI 10.17487/RFC4090, May 2005, <https://www.rfc-editor.org/info/rfc4090>.

[FRR] Shand, M. and S. Bryant, "IP Fast Reroute Framework", RFC 5714, DOI 10.17487/RFC5714, January 2010, <https://www.rfc-editor.org/info/rfc5714>.

[RLFA-FRR] Bryant, S., Filsfils, C., Previdi, S., Shand, M., and N. So, "Remote Loop-Free Alternate (LFA) Fast Reroute (FRR)", RFC 7490, DOI 10.17487/RFC7490, April 2015, <https://www.rfc-editor.org/info/rfc7490>.

[BIER-PREF] Thubert, P., Eckert, T., Brodard, Z., and H. Jiang, "BIER-TE extensions for Packet Replication and Elimination Function (PREF) and OAM", Work in Progress, Internet-Draft, draft-thubert-bier-replication-elimination-03, 3 March 2018, <https://tools.ietf.org/html/draft-thubert-bier-replication-elimination-03>.

[DetNet-IP-OAM] Mirsky, G., Chen, M., and D. Black, "Operations, Administration and Maintenance (OAM) for Deterministic Networks (DetNet) with IP Data Plane", Work in Progress, Internet-Draft, draft-mirsky-detnet-ip-oam-02, 23 March 2020, <https://tools.ietf.org/html/draft-mirsky-detnet-ip-oam-02>.

[DetNet-DP-FW] Varga, B., Farkas, J., Berger, L., Malis, A., and S. Bryant, "DetNet Data Plane Framework", Work in Progress, Internet-Draft, draft-ietf-detnet-data-plane-framework-06, 6 May 2020, <https://tools.ietf.org/html/draft-ietf-detnet-data-plane-framework-06>.

[I-D.farkas-raw-5g] Farkas, J., Dudda, T., Shapin, A., and S. Sandberg, "5G - Ultra-Reliable Wireless Technology with Low Latency", Work in Progress, Internet-Draft, draft-farkas-raw-5g-00, 1 April 2020, <https://tools.ietf.org/html/draft-farkas-raw-5g-00>.

[NASA] Adams, T., "RELIABILITY: Definition & Quantitative Illustration", , <https://kscddms.ksc.nasa.gov/Reliability/Documents/150814-3bWhatIsReliability.pdf>.

[MANET] IETF, "Mobile Ad hoc Networking", , <https://dataTracker.ietf.org/doc/charter-ietf-manet/>.

[detnet] IETF, "Deterministic Networking", , <https://dataTracker.ietf.org/doc/charter-ietf-detnet/>.

[SPRING] IETF, "Source Packet Routing in Networking", , <https://dataTracker.ietf.org/doc/charter-ietf-spring/>.

[BIER] IETF, "Bit Indexed Explicit Replication", , <https://dataTracker.ietf.org/doc/charter-ietf-bier/>.

[BFD] IETF, "Bidirectional Forwarding Detection", , <https://dataTracker.ietf.org/doc/charter-ietf-bfd/>.

[CCAMP] IETF, "Common Control and Measurement Plane", , <https://dataTracker.ietf.org/doc/charter-ietf-ccamp/>.

Authors' Addresses

Pascal Thubert (editor)
Cisco Systems, Inc
Building D
45 Allee des Ormes - BP1200
06254 MOUGINS - Sophia Antipolis
France

Phone: +33 497 23 26 34
Email: pthubert@cisco.com

Georgios Z. Papadopoulos
IMT Atlantique
Office B00 - 114A
2 Rue de la Chataigneraie
35510 Cesson-Sevigne - Rennes
France

Phone: +33 299 12 70 04
Email: georgios.papadopoulos@imt-atlantique.fr

Rex Buddenberg
CA
United States of America

Email: buddenbergr@gmail.com