Ensuring Session Continuity for Railways Using a Stateful Programmable Data Plane

Valerio Bruschi[1], Jim Zou[2], Peter Legg[3], Aniello Cammarano[4], Luca Petrucci[1], Juan Agusti Moreno[5], Carles Terés[6], Paula Ciria[6], Daniel Camps-Mur[7], Marco Bonola[1,4], Salvatore Pontarelli[8]
[1]CNIT/University of Rome Tor Vergata, [2]ADVA Optical Networking SE, [3]Blu Wireless Technology, [4]Axbryd, [5]COMSA Corporation, [6]Ferrocarrils de la Generalitat de Catalunya, [7]i2CAT Foundation, [8]Sapienza University of Rome

***Abstract*—5G networks will provide high quality ubiquitous connectivity also in complex scenarios such as for example the railway one. Assuring reliable, efficient and scalable network connections in the railway use case requires to solve several technical issues such as for example the handover management when the network connection is realized using multiple antennas placed along the railway track. This paper presents a fast and flexible handover solution deployed entirely in the fast-path using FlowBlaze, a stateful programmable data plane. The proposed approach not only provides a low-latency solution, but also enables the deployment of additional functionalities such as traffic load balancing among the visible antennas and differentiation between user traffic and critical traffic, which will be forwarded using packet redundancy.**

I. Introduction

In recent years train operators started to offer to passengers internet access services. In order to satisfy the technical requirements needed to provide high quality services such as voice services or mobile video broadcasting, several solutions has been proposed [1], [2]. The same network can be also used to provide connectivity for the train operation control system or to increase the train safety, for example transmitting data coming from the train Closed-circuit Television (CCTV) to the ground infrastructure. Obviously these two types of traffic have very different requirements in terms of throughput, reliability, latency etc. In particular, the use of a distributed antenna system able to connect to the train using a mmWave wireless connection could provide a significant throughput able to satisfy the user requirements [3]. Furthermore, the 5G network also provides the necessary slicing needed to split the operational traffic from the passenger one [4].

However, mmWave connections provide very high throughput but have limited range (several hundred meters). Therefore, the train is going to do handovers very often, thus requiring an extremely fast network handover solution. Since the train is moving, it continuously changes the antennas to which it is connected, thus the network must be able to dynamically tracks the current active paths in order to deliver traffic toward the train. When the connected antennas change, it is required to maintain the transport session despite the change of this link. This task is termed session continuity. Without session continuity, applications need to re-establish all ongoing connections after a change of network attachment, causing unnecessary protocol overhead and extra delays. As stated in [2], fast handover is a crucial requirement for wireless data communications, since the train spends only a short period of time moving through one cell. Moreover, seamless transmission is required for delay sensitive applications such as safety-related services or other train service communications.

Even if several solutions taken from other domains could be used (e.g. mobility IP [5] or multi-path TCP [6]), the stringent requirements in terms of slicing, latency, reliability etc., suggest to develop an ad-hoc solution for this problem.

In this paper we describe a layer-2 solution which exploit programmable data planes to minimize latency, increase reliability and provide a flexible configuration of the network in order to optimize the handover management. The presented solution has been implemented in a real railway operated by Ferrocarrils de la Generalitat de Catalunya as one of the final demonstrations of the H2020 5G-PICTURE project [7].

The trial in the operational environment gave us the opportunity to gather useful insights on the feasibility of this solution and to collect real measurements of the achievable network performances. And, we demonstrate that even if the railways scenario is a very challenging vertical scenario, an efficient solution able to provide 1Gb/s to a moving train, with less than 1ms end to end latency can be deployed.

The rest of this paper is arranged as follows. Section II discusses the design problems and the developed solution, while section III presents the experimental results collected during the H2020 5G-PICTURE project final demonstration. Finally, conclusions are drawn in section IV.

II. Handover Management solution

The 5G architecture can accommodate different services with different requirements, as required in the railways scenario. To provide such flexibility, the 5G transport infrastructure has introduced programmability inside the network dataplane [8], [4]. These technologies allows developing a solution to ensure session continuity through an handover management that minimizes latency, increases reliability and provides a flexible configuration of the network.

A. Network Application Description

Our application is characterized by a scenario in which a train is populated by users, that want to use the internet connection deployed in the train. The objective of our application,

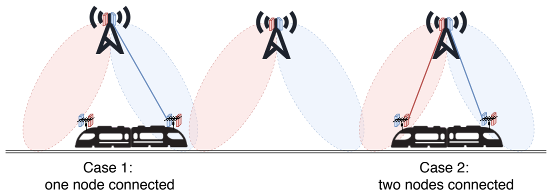


Fig. 1: Wireless connection between the train and the track segment using front and rear antenna nodes.

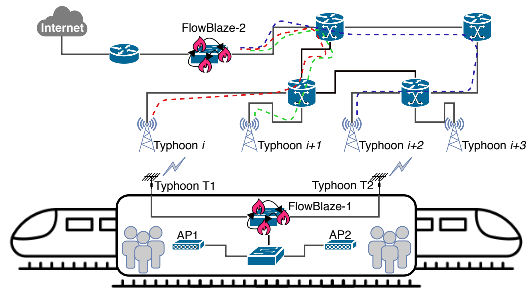


Fig. 2: Network architecture

is to guarantee the session continuity to these users, while the train is moving. The users travelling in the train, can connect to the Internet using the Access Points (APs) deployed in the train. The APs are connected to a switch, which forward traffic to the two antenna nodes deployed on the two ends of the train. These antennas are connected to some external nodes deployed close to the railway, that allow communicating with the Internet. The nodes communicate following the IEEE 802.11-2016 protocol. As in Fig. 1, every antenna placed along the railway has two facing modems in opposite directions. Therefore, the antennas on the train can be connected at the same time to the same (or to the two closest) node(s) placed along the railway.

Fig. 2 presents the network architecture: the train is equipped with two Typhoon nodes [9] (i.e. antenna nodes, T1 and T2) that forward the users traffic to the antennas deployed along the track. The network inside the train is composed by a switch and a stateful programmable data plane (i.e. the FlowBlaze node [10]). The switch aggregates traffic coming from various APs deployed in the train, while FlowBlaze-1 is responsible to manage the handover functionalities. T1 and T2 nodes are connected to some of the *Typhoon i* nodes in the track side. These nodes are connected using several switches to the Internet via the FlowBlaze-2 node. The network is configured to associate to each node along the track a different VLAN identifier (vlan_id), thus allowing to identify from which node a packet is received. Since each node belongs to a different VLAN network, in the incoming packet is pushed a specific vlan_id which identifies the current active connection. Consequently, FlowBlaze can learn the association between the train-top node and the active Typhoon i node to handle in a fast way the handover management.

B. Design problems

Since the connectivity of the antenna nodes changes over time while the train is moving, several design problems [11] come out:

- *Location tracking*. To find the correct exit-point, there should be a mechanism to detect which of the antenna nodes is able to forward traffic at that time and forward the traffic only to the connected node.
- *Flow redirection*. If the train changes its point of attachment to the network, all active flows should be redirected to the new point of attachment. The redirection process takes time, causing significant impact on handover latency. There should be a scheme that continues all active flows while minimizing the pause of data traffic in these flows.
- *Controlled flooding*. Since there are multiple antenna nodes that could be active at the same time, a packet could be forwarded to multiple paths, consequently generating multiple copies of that. Sending packets to multiple modems provides a significant flexibility, since it is possible to guarantee zero packet losses during the handover period (e.g the network can start to send duplicated packets to the next antenna just before the train arrive close to it).

As we will discuss in the next subsection, FlowBlaze is able to overcome all these limitations by implementing handover management and packet deduplication functions.

C. Handover Management

The solution that we develop for handover management work entirely on the data plane, consequently the architecture is able to react to network changes at packet level (i.e. nanoseconds scale), reducing the handover time. The programmable data plane detects which nodes are currently connected and forward the traffic only to these ones. To monitor the connection state between the train-top antennas and the nodes located along the railways, a specific binding update policy has been deployed. In the current prototype we send probe packets (i.e. ICMP ping) toward Internet to detect if the connection of a specific antenna node link is active and then we store the resulting state. Another option would be to spoofing a gratuitous ARP from the Typhoon node as soon as a new 802.11ad association come up. Even if this solution could minimize the binding update delay with respect to the use of probe packets, it is strongly dependent on the characteristics of the antenna. In fact, in the equipment used for the experiments, this feature was not available in the implementation and we decided to allocate the handover management only on the FlowBlaze nodes.

As described before, the network is configured to associate to each node along the track a different VLAN identifier (vlan_id), thus allowing to identify from which node a packet is received. Since each node belongs to a different VLAN network, in the incoming packet is pushed a specific vlan_id which identifies the current active connection. The FlowBlaze-2 node (see Fig. 2) deployed in the ground segment can learn the association between the MAC address of the train antenna and the vlan_id corresponding to the active Typhoon i node, thus identifying the active binding. The packets coming from the ground toward one of the trains in the railway can be directed to the right path simply pushing the last vlan_id associated to the destination MAC.

When the train will be connected to a different node, a new vlan_id will be automatically pushed into the flow, thus it is easy to recognize that the connection has changed the path. When such switching occurs, the packets that was already tagged with the previous vlan_id can be lost.

The number of lost packets during a connection change is the sum of the following elements:

1) n_{flight}, the number of packets that are already travelling into the network toward the previous antenna when the connection changes;
2) n_{delay}, the number of packets that will be send toward the previous antenna while the binding update connection signal (ICMP or gratuitous ARP) is received;
3) n_{update}, the number of packets that will be send toward the previous antenna while FlowBlaze updates the binding in its table.

The use of gratuitous ARPs minimizes n_{delay}, while FlowBlaze allows a fast data plane binding update, thus minimizing n_{update}. Unfortunately these solutions have no impact in reducing n_{flight}. However, the programmability of FlowBlaze allows developing controlled flooding policies able to completely avoid packet losses. An easy option to avoid packet loss is to send packets both to the i-th active node and also to the $(i+1)$-th node. In this way a copy of the packet will arrive to the train also if the train moves to the next antenna during the packet transmission delay. The FlowBlaze-2 node is in charge of providing this functionality.

Moreover, the proposed method has the ability also to perform traffic load balancing. When only one link between the train and the antenna is active, the switch sends all the traffic to this link. If both the links are active, the traffic can be split between the active connections. Finally, it is possible to duplicate packets coming from the train for specific traffic classes. For example, train service communications using the same infrastructure should provide the smallest possible packet rate loss. Sending redundantly this traffic to both the active antenna can increase the transmission reliability.

D. Packet De-Duplication

As mentioned before, we can duplicate some packets to minimize packet loss. Consequently, multiple copies of a packet can be present in the network and we need to detect and discard all duplicates. To do that, we need to memorize an accurate unique representation of a packet. In our case we decided to save the 5-tuple and the CRC code for each packet encountered by the FlowBlaze nodes. Therefore, when a new packet with the same 5-tuple and CRC of one memorized is detected, this packet should be discarded. The 5-tuple and the CRC code guarantee a good unique representation. In fact, according to the birthday paradox, the collision probability of two packets will be around 0.01% even if we suppose that there are in flight 1 thousand of packets of the same flow. According to [12], we can estimate that the system should be able to store few thousand of packets against which compare an incoming packet to avoid duplication. Considering that we store 17 bytes for each packet, this correspond to a memory requirement of 170KB if we set the windows to 10'000 packets. This requirement is quite manageable with the current technology solutions, since the FlowBlaze implementation [10] provides a context table of 256KB.

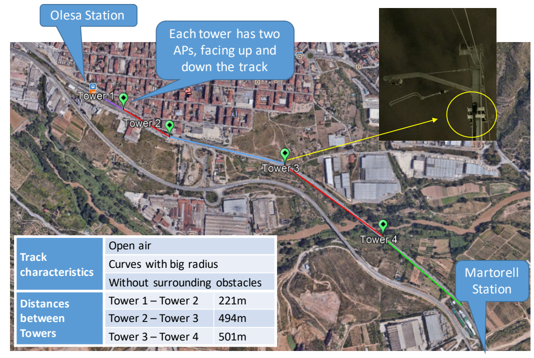


Fig. 3: track segment used in experiments.

III. Experimental Results

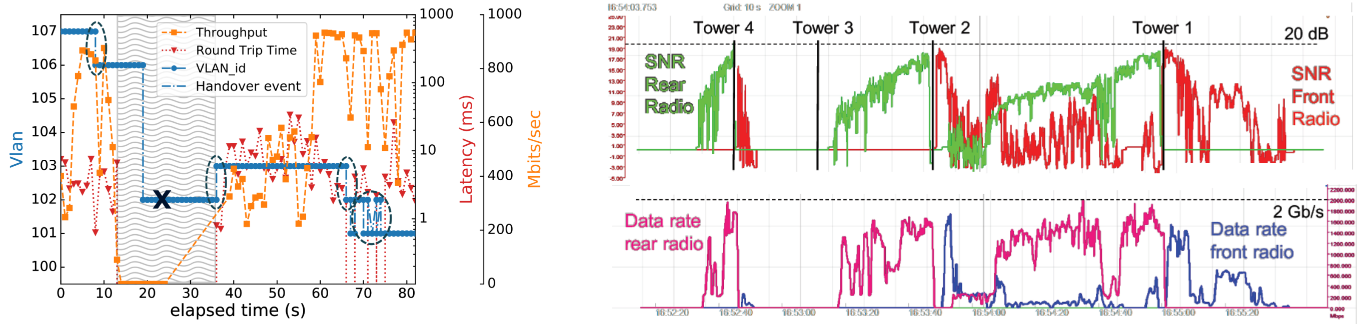
A. Location description

The presented solution has been implemented in a real railway operated by Ferrocarrils de la Generalitat de Catalunya as one of the final demonstrations of the H2020 5G-PICTURE project [7]. The demo site comprises approximately 1,5 Km of track as shown in Fig. 3. Deployment in track was composed of 4 towers from Olesa station to the south. Four stanchions equipped each with 2 mmWave nodes interconnected with fibre forms a telecommunications infrastructure that connects a set of on-board devices mounted in a specific train that passes by this section with Martorell station, which represents a very simplified OCC (Operations Control Centre). Along this track segment, commercial trains speed is over 60 Km/h, up to 90 Km/h[1]. The results were measured over a time windows of two minutes in movement and one minute in stationary.

B. Network components

The network architecture has been deployed following the scheme reported in Fig. 2. The mmWave antenna over the train are composed of dual antenna, dual baseband unit for IEEE 802.11-2012 DMG / WiGig at 60GHz. The antennas are interconnected through two 10G Ethernet switches that implement an Ethernet ring inside the train. Along the rail, the mmWave nodes are distributed in some stanchions of the track. The nodes are interconnected by a passive WDM link, with optical add/drop filter nodes distributed along the track and into Martorell and Olesa stations. The traffic multiplexed from the passive WDM link is further aggregated to a 100GbE backhaul network at the central office in Martorell station. Finally, the FlowBlaze nodes are based on the NetFPGA SUME SmartNIC, an x8 Gen3 PCIe card containing a Xilinx Virtex-7 690T FPGA providing four 10G Ethernet links.

[1]It is worth notice that the mentioned commercial train speed is not a technological limit of the proposed approach but just a deployment limit.



(a) End-to-end performance test

(b) Antenna-to-antenna performance test

Fig. 4: Experimental results of the complete track segment: from Martorell to Olesa stations. (a) the VLAN id, the measured RTT and the throughput are shown. It is possible to note an uncovered region (the gray area) due to the stanchion 3 malfunctioning, and several handover events (the dashed circles). (b) SNR and data rate between the towers and the Front and Rear modems of a single antenna unit of the train. It can be seen the uncovered region (around tower 3) and the overlapping zone between towers 2 and 1, where the train is connected with two antennas.

C. Results

In Fig. 4a, we present the end-to-end (e2e) performance test of the complete track segment in which the train starts from Martorell station and traverses all the towers deployed in the railway track. The picture shows latency (as the Round Trip Time), throughput and active VLAN during the train trip of about 80 seconds, corresponding to the 2 Km path covered by the antenna. It can be noted that the system performs several handovers, (i.e. at seconds 10, 20, 35, 65 70) and both throughput and latency experienced wide variations. As a first consideration, we would like to stress how was challenging the vertical scenario in which our solution has been deployed as the full picture reports. In fact, there is a time interval between 15th and 35th second in which the connection has stopped. We were aware of a malfunction in the stanchion corresponding to these values, so the connection loss was not unexpected.

Excluding the area of not coverage the e2e latency is always below 10 ms and the throughput is around 900Mb/s or 400Mb/s depending on the channel conditions. We reported in Fig. 4b the results of a TCP iperf session between modems on the track and the train (i.e. not an end to end analysis). This test uses a single antenna node (with front and rear-facing modems). The figure shows two areas: (i) the region of no coverage and (ii) an overlapping region covered by two antennas (i.e. the area between tower 2 and tower 1). This overlapping enables the opportunity to duplicate packets, consequently avoiding packet losses. Finally, the antenna telemetry shows (in blue/pink) that the single mmWave modems are able to reach a data rate of up to 2Gbit/s, that is not visible in the previous experiment because the Iperf3 client/server nodes are bounded by the 1GbE interface that are equipped with. The figure also shows the SNR measured by each train-top modem (around 19 dB when close to the towers).

IV. Conclusions

This paper presented a novel layer-2 handover management functionality. The solution works entirely in the data plane, minimizing the binding update latency and reducing the traffic toward the network controller. The proposed solution has been deployed in a real railway operated by Ferrocarrils de la Generalitat de Catalunya as one of the final demonstrations of the H2020 5G-PICTURE project [7] and is one of the first deployments of 5G technologies in the railway sector. We demonstrated that, even in a very challenging vertical scenario, an efficient solution able to provide 1Gb/s to a moving train with less than 1ms end to end latency can be deployed.

Acknowledgment

This work is partially supported by the EU Commission in the frame of the H2020 projects 5G-PICTURE (grant #762057) and 5G-MED (grant #951947).

References

[1] B. Ai *et al.*, "Challenges toward wireless communications for high-speed railway," *IEEE transactions on intelligent transportation systems*, vol. 15, no. 5, pp. 2143–2158, 2014.

[2] S. Xu *et al.*, "A survey on high-speed railway communications: A radio resource management perspective," *Computer Communications*, vol. 86, pp. 12–28, 2016.

[3] F. Boccardi *et al.*, "Five disruptive technology directions for 5g," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.

[4] P. Öhlén *et al.*, "Data plane and control architectures for 5g transport networks," *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1501–1508, March 2016.

[5] C. Perkins, "IP mobility support for IPv4, revised," Tech. Rep., 2010.

[6] H. Han *et al.*, "Multi-path TCP: a joint congestion control and routing scheme to exploit path diversity in the internet," *IEEE/ACM Transactions on Networking (TON)*, vol. 14, no. 6, pp. 1260–1271, 2006.

[7] H2020 5G-PICTURE. [Online]. Available: https://www.5g-picture-project.eu/

[8] R. Bifulco and G. Rétvári, "A survey on the programmable data plane: Abstractions, architectures, and open problems," in *IEEE 19th Int. Conf. on High Performance Switching and Routing (HPSR)*, 2018, pp. 1–7.

[9] (2019) Typhoon Family. [Online]. Available: https://www.bluwireless.co.uk/products/evaluation-modules/

[10] S. Pontarelli *et al.*, "Flowblaze: Stateful packet processing in hardware," in *USENIX Symposium on Networked Systems Design and Implementation (NSDI 19)*, 2019.

[11] C. Chen *et al.*, "Mobility management for low-latency handover in SDN-based enterprise networks," in *Wireless Communications and Networking Conference (WCNC)*. IEEE, 2016, pp. 1–6.

[12] M. Ruiz *et al.*, "An FPGA-based approach for packet deduplication in 100 gigabit-per-second networks," in *International Conference on ReConFigurable Computing and FPGAs*. IEEE, 2017, pp. 1–6.