Analysis of the Failure Tolerance of Linear Access Networks

Frédéric Giroire and Juan-Carlos Maureira

Abstract— In this paper, we study the disconnection of a moving vehicle from a linear access network composed of cheap WiFi access points in the context of telecommuting in massive transportation systems. In concrete, we analyze the probability of a user experiencing a disconnection longer than a given time interval (t^*) such that all ongoing communications between the vehicle and the infrastructure network are disrupted. We provide an approximation formula considering two scenarios (intercity bus and train). We then carry out a sensitivity analysis and supply a guide for operators when choosing the parameters of the networks. Finally, we show that such systems are viable, as they attain a very low probability of long disconnections with a very low maintenance cost.

Index Terms— Access networks, failure tolerance, telecommuting, connectivity analysis, maintenance cost.

I. INTRODUCTION

NOWADAYS megacities are becoming more attractive to people for finding better jobs in order to improve their quality of life. The demographic explosion caused by this migration of people pushes megacities to expand their urban radius, causing an increase of transit times when people are commuting between home and work. This increase of transit time is fostering the proliferation of massive transportation systems, such as urban trains or inter-city buses, in which people uses to work remotely with their offices. According to the IDC (International Data Corporation) [2], the worldwide number of telecommuters increased from 919.4 million in 2008 (29% of the workforce) to 1.19 billion in 2013 (34.9% of the workforce), and projections estimates 1.76 billions of telecommuters worldwide by Luk 2020 [3].

Many efforts have been done in order to satisfy this increasing demand of telecommuting, in which 3G/4G networks have played an important role due to their penetration and availability. The first effort comes directly from commuters that uses their personal smartphones as 3G/4G modems to access
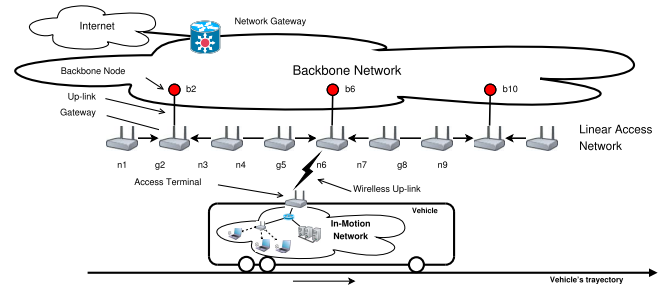
Fig. 1. Telecommuting reference scenario: commuting passengers inside a vehicle (e.g. an intercity bus or train) access to the Internet via a linear access network connected to a backbone network via gateways.

the network. This effort has shown a limited adoption due to their cost and data plans traffic quotas. A second effort comes from communication companies such as Icomera [4], [5] and Nomad Digital, that uses a bond of 3G/4G connections to offer an on-board access terminal to provide connectivity from buses or urban trains. A final effort is made by transportation system's operators, which use an on-board access terminal connected wirelessly to an infrastructure network alongside the vehicle's trajectory for providing connectivity to on-board passengers. For example, the solution [6] implemented for the train service between the Shanghai airport and downtown, is capable of delivering a continuous V2I connection up-to 16 Mbps at speeds of 500 Km/h. Motivated by such a solution, we are interested by analyzing the feasibility of providing such a connectivity to telecommuters only relying on unlicensed wireless technology. In particular, we focus on the required infrastructure network for supporting a broadband (> 16 Mbps) V2I communication only relying on cheap WiFi Access Points (APs).

Figure 1 depicts our reference scenario. At the bottom, we observe a high-speed vehicle following a given trajectory. In this vehicle, networked devices get access to the network through an on-board **access terminal**. This access terminal uses a WiFi based link to establish the connection with a set of WiFi Access Points (AP) placed along the trajectory. These APs are interconnected by an **Infrastructure Network**, which carries the traffic between trains passengers and a *Network Gateway* (depicted at the top of the figure). This *Network Gateway* provides the access to **external networks** (such as Internet).

In this paper, we study the *the feasibility of designing a trackside access network based only on cheap WiFi AP nodes*. The motivation is to reduce the construction costs for improving network resilience and use the resulting savings,

together with an appropriate defective APs replacement policy, to enable the operator to maintain network connectivity more economically, thus decreasing both initial capital investment and ongoing operational costs. So, as the number of APs may range easily from hundreds to thousands, it is important to describe how many defective APs the network can tolerate before to be unable of providing a continuous connectivity to the in-motion networks on-board the vehicle. In other words, how many failed APs are required to generate a large enough disconnection time as to produce a disruption of all on-going communications between the vehicle and the infrastructure network, event we denoted as *long disconnection*.

The contributions of this work are the following:

 (i) We give exact formulas for small networks and approximation for large networks for the probability of experiencing a long disconnection.
 (ii) We validate the analytical work by simulation.
(iii) We study two specific commuting scenarios: an *intercity train* and an *intercity bus scenarios*.
(iv) We provide a practical guide for network operator to choose the network parameters.
 (v) We exhibit a *ladder effect* showing that very small changes of parameters may impact very significantly the probability of long disconnections. It gives an opportunity for network operator to improve the resilience of their network at small cost. And
(vi) Last, we provide a cost analysis of the maintenance for a network operator and show that our solution attains a high reliability with a small maintenance cost.

The rest of the paper is structured as follows. We summarize related works in Section II. In Section 3, we present the problem. We carry out a theoretical analysis of the failure tolerance of the access network in Section 4. We then study the sensitivity to the choices of the network parameters in Section 5. Finally, we conclude the work in Section 6.

II. Related Works

Network connectivity for moving vehicles has been widely studied in the context of railway communications [7] and Vehicular Ah-Hoc Networks (VANETs) [8]. In both cases, they use track-side access nodes to wirelessly communicate the moving vehicle with an infrastructure network, which we assume it already use a suitable handover capability [9], [10] in order to seamlessly transport traffic from the in-motion network to the infrastructure network.

Connectivity of linear access networks have not received much attention in railway communication systems since 3G/4G networks [11] have dominated the scenario. Conversely in VANETs, the connectivity problem of a set of vehicles communicating each other along a highway is studied as a linear access network. Indeed, the k-connectivity analysis of one-dimensional (1D) VANET [12] networks is expressed as the probability of experiencing a disconnection after removing k-1 nodes (or vehicles). In other words, the probability of having a network partition. However, this analysis is not applicable to our reference scenario since it does not consider any infrastructure network. Nevertheless, in [13], the

analysis defines two types of nodes: the ordinary ones and the powerful ones, which makes this approach closer to the scenario depicted in Figure 1. In our linear access network, regular nodes are normal access nodes and powerful nodes are gateways. But, the problem of this approach is the same than in [14]: the node connectivity is dynamic since regular nodes are in fact moving vehicles. A unit disk connectivity rule is used to determine when they are connected or not. Therefore, the probability of being connected (or disconnected) strongly depends on the euclidean distance among nodes, which is not applicable to our case since our access nodes are fixed. Furthermore, [15] gives closed expressions for the probability of connectivity for 1D network with a variable number of nodes and with a fixed number of nodes. This latter one uses a fixed spatial distribution, which makes this approach even closer to our scenario. Nevertheless, this approach has no infrastructure network or sink nodes, so, the probability of connectivity is strictly related to node-to-node, or better said, vehicle-to-vehicle communication. In summary, as literature does not fully answer our research question, this paper aims at filling the gap by providing a connectivity analysis for one-dimensional networks in the context of massive transportation systems such as inter-city trains or buses.

III. Infrastructure Network: Terminology, Topology, and Scenarios

In this section, we introduce the terminology used in our study. We also formally describe the topology of a linear access network and enunciate the two scenarios we used to validate our results.

A. Network Topology

The users (for example, the passengers of a train or of a bus) connect to external networks (such as Internet) through an *infrastructure network*. This infrastructure networks is composed of two distinct networks:

- a *linear access network* made of a large number n of cheap WiFi access points, which we name *access nodes* from now on. The access nodes are placed along the trajectory of the vehicle (hence the name linear) every c meters. They have a *coverage area of radius r*, with $r > c$, so that two neighboring access nodes can exchange data. The size of the access network might vary from hundreds to thousands of nodes.
- a *backbone network* made of a smaller number of more reliable backbone nodes. The root of the backbone network is connected to an ISP network, so the traffic from the access network flows across the backbone nodes and links until reaching the root node, and from there, external networks. As our focus is in the access network, a discussion about possible topologies for the backbone network is out of the scope of this paper. Nevertheless, see [16] for a discussion about some proposed backbone topologies for linear access networks.

A subset of the access nodes (m of them) serve as *gateways* between both networks, see Figure 1.

When a user wants to send data, he first connects to the closest access nodes to reach the linear network. The message

TABLE I

SUMMARY OF NOTATIONS USED IN THE PAPER AND DEFAULT VALUES CORRESPONDING TO TWO SCENARIOS: A RAILROAD LINE
OF 100 km FOR AN INTERCITY TRAIN AND AN INTERCITY BUS LINE OF 30 km

Variable	Description	Train scenario	Bus scenario
v	speed of the vehicle	100 km/h	30 km/h
n	number of access nodes	2000	600
m	number of gateways	665	199
$\mathcal{D}$	number of access nodes between two gateways	3	3
r	radius of the coverage area of an access node	115 m	115 m
c	distance between two access nodes	50 m	50 m
p	probability of failure of an access node (1/MTTF)	0.001 (MTTF: 3 years)	0.001 (MTTF: 3 years)
t^*	maximum acceptable disconnection time	3 sec	3 sec
l^*	minimum length of a defective segment leading to a long disconnection	6	5

is then sent to the closest gateway and finally to the root of the backbone topology through the backbone network.

A summary of notations is provided in Table I along with two sets of default values, corresponding to two scenarios of reference detailed below. These default values serve to the reader as an example of possible values taken by the parameters.

1) Intercity Train Scenario: We consider an intercity train linking two cities at a distance of 100 km. The average speed of the train is 100 km/h. To provide connectivity to the train passengers, the infrastructure of the linear access network is made of $n = 2000$ access nodes, separated by a distance of $c = 50$ m. Among these access nodes, one over three of them is a gateway.

2) Intercity Bus Scenario: The second scenario we consider is the one of a bus linking two cities 30 km away. The average speed of the bus is 30 km/h and the number of access nodes required to cover the bus path is $n = 600$.

B. Network Failures

The linear access network uses cheap hardware and is thus prone to failures of access nodes. In this study, we consider that the backbone topology is reliable, or in other words, that a failure in the backbone topology is a lot less likely to happen – backbone nodes are a lot fewer and thus, network operator may use hardware with larger Mean Time To Failure (MTTF).

1) Node Failures at the Linear Access Network: An access node may not be able to deliver a service to the user because of two types of **node failure**:

1) - *Defective Node*: A node is said defective, when it is experiencing hardware, power supply, or any other problem which makes it not operate as it should.
2) - *Isolated Node*: A node is said isolated, if it has no way to reach the root of the backbone network because of a set of defective nodes.

The first type of error requires an external intervention to be recovered. The second type of error is automatically recovered when the defective nodes (causing the failure) are recovered. Figure 2 presents different cases of isolated nodes when the network has two defective nodes. Generally, x defective nodes might produce i isolated nodes, resulting in $l = x + i$ failed nodes.

2) Disconnection Time and Long Disconnection: When the vehicle (containing the in-motion network) is passing
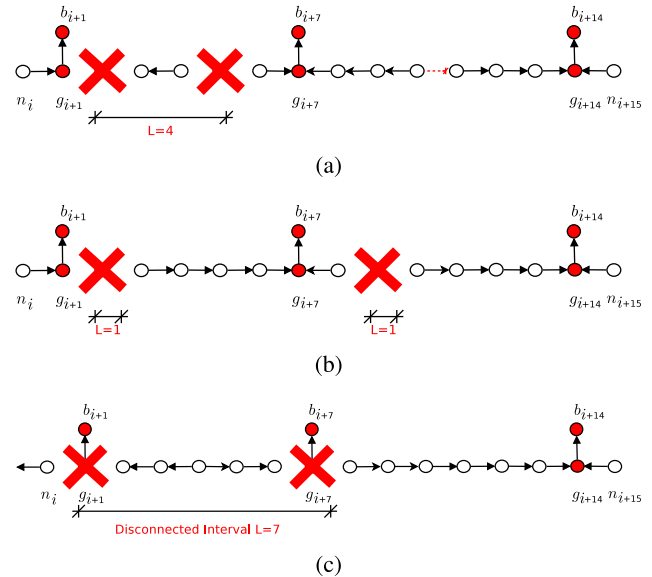

Fig. 2. Different cases of isolated nodes produced by two defective nodes. (a) Two isolated nodes. (b) No isolated node. (c) Worst case: two consecutive gateways are defective.

by a segment of several failed access nodes, the in-motion network is disconnected from the infrastructure network. The disconnection is perceptible to the user when its duration is larger than t^* seconds.

The network operator thus has to define a threshold value t^* for the maximum disconnection time allowed for passengers. This threshold depends on the type of applications and of the Quality of Service (QoS) the network operator wants to offer. For example, an on-going data stream may be broken, if it suffers a disruption of 3 seconds or more (it would depend on the buffer size of the user). Note that the duration of this disconnection depends on the speed of the vehicle as discussed in the next subsection. If a user experiences a disconnection, whose duration is larger than t^*, we say that it experience a *Long Disconnection.*

C. Speed Effects on the Disconnection Time

In this section, we discuss the effect of the speed on the disconnection time of the in-motion network produced by contiguous failed access nodes.

As each access node (AN) represents a *Wireless Access Point* device, they are characterized by an area of radio
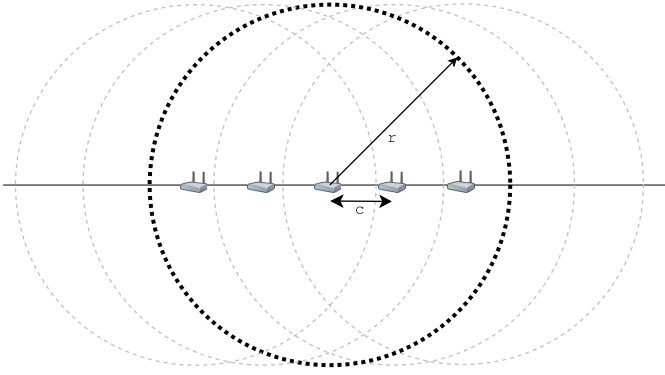
Fig. 3. Radio coverage of access nodes (ANs) in linear access networks. ANs are placed every c meters and have a radio coverage of r meters.
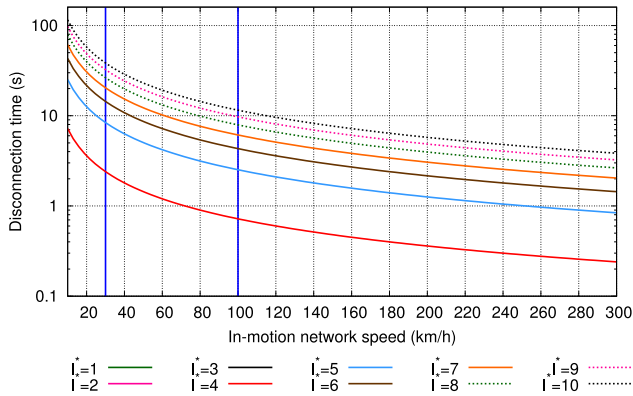

Fig. 4. Disconnection time (log-scale) for different speeds and different number of failed nodes.

coverage being a disk of radius r meters. The access nodes are placed at c meters apart, with $c < r$. The radio coverages of two contiguous ANs thus they overlap, as depicted in Figure 3. As our objective is to evaluate disruptions in the communication caused by the event of long disconnections only, we assume a perfect communication within the AN range (r) in order to avoid nuisance factors in the computation of the event's probability. Nevertheless, other types of events causing disconnections can be determined by simulating the wireless communication over the train trajectory as discussed in [17]–[19]. For l contiguous failed access nodes, the radio coverage is down over a distance of $\max((l-1)c + 2(c-r), 0)$. The corresponding disconnection time when the vehicle is moving at v m/s is thus described by the following expression:

$$t_{disc}(l, v) = \max\left(\frac{c \cdot (l+1) - 2r}{v}, 0\right). \tag{1}$$

Figure 4 shows the disconnection time for different numbers of failed nodes l and in terms of the speed of the in-motion network v. For example, at a speed of 30 km/h, a user experiences a disconnection of 2 seconds if 4 consecutive access nodes are failing and of 8 seconds if 5 consecutive access nodes are failing. At higher speed, the corresponding disconnection times are of course smaller, as the vehicle passes faster along the segment of failing ANs. For example, at 100 km/h, the disconnection times are 0.7 seconds and
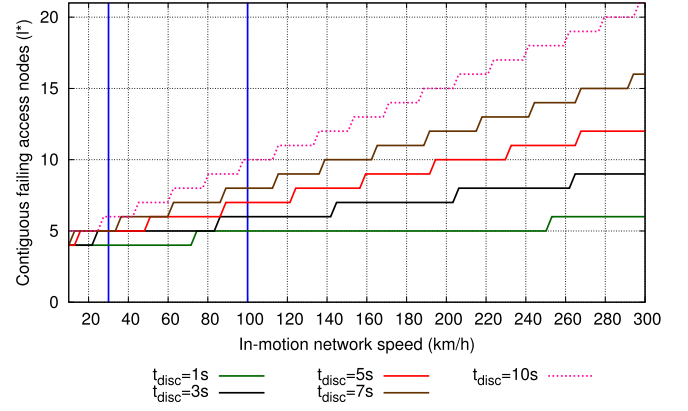

Fig. 5. Minimum number of contiguous failing access nodes leading to a long disconnection, l^*, as a function of the vehicle speed for different values of t^*, the maximum disconnection time tolerated.

1.5 seconds for $l = 4$ and $l = 5$ Equation (1) yields

$$l^*(t^*, v) = \left\lceil \frac{vt^* + 2r}{c} - 1 \right\rceil \tag{2}$$

Using Equation (2), a network operator can determine l^* *the minimum number of contiguous failing access nodes leading to a Long Disconnection* corresponding to the maximum disconnection time t^* for a given speed.

Figure 5 presents the value of l^* as a function of the speed of the vehicle v for different values of t^*. Note first that, for any value of t^* and any speed, the network can support three consecutive failing nodes as the radio coverage has a radius $r = 115$ m while consecutive ANs are at distance $c = 50$ m. Thus, two ANs at distance 4 cover the 3 ANs between them (see Figure 3). Letting $t^* = 3$ seconds (the default value), the network can support three failed nodes for any speed, four failed nodes for speeds greater than 22 km/h; five failed nodes for $v > 35$ km/h, and so on. The access network of the intercity bus scenario thus experiences a long disconnection when $l^* = 5$ consecutive nodes are failing and, the one of the intercity train scenario, for $l^* = 6$ consecutive failing nodes.

IV. NETWORK FAILURE ANALYSIS

In this section, we study the probability of experiencing a network failure producing a long disconnection from the infrastructure network, that is the probability of observing a segment of l^* contiguous access nodes in a failure state, which causes a disconnection time larger than t^* within an observation period of time τ.

For starting, we show that the probability of observing l^* consecutive failed nodes can be computed exactly for small topology sizes ($n \leq 100$). However, as the number of terms to be computed is exponential with n, the exact computation becomes untractable for larger networks. To overcome this difficulty, we provide an approximation formula which can be computed for any network size. The approximation is valid when p is small, that is when $np^x \ll 1$. The approximation formula provides the additional advantage of extracting more clearly the influence of each system parameter on the probability of failure.

A. Probability Analysis (Exact Formula)

1) Model: In this section, we describe the probability of having at least l^* consecutive nodes in failure state somewhere in the access network. We denote this probability by $\mathbb{P}[S]$. The event of having l^* failed nodes is triggered by x defective nodes, which may or may not lead to a larger number of failed nodes by creating isolated nodes. Let p be the probability for a node to become defective within an observation period of duration τ. The probability p is obtained from the mean-time-to-failure (MTTF) of an access node, when observing a sample of devices working under certain operational conditions. Considering a MTTF of one year and an observation period τ of one day, the probability of observing a failed node is $p = \frac{1}{365} \approx 0.0027$. As far as we know, there are no published studies on the failure probability of an Access Point device. Therefore, we assume for this study values of p of 0.01, 0.001 and 0.0001, which are reasonable when considering the depicted MTTF and the observation period for the two scenarios considered. We note F the number of defective nodes during the time of observation τ. We consider regular topologies with a constant number of access nodes between contiguous gateways, let us say $\mathcal{D} = \frac{n-1}{m-1}$, and there is always a gateway at the beginning of the linear access network.

2) Probability of Network Failure: We aim at estimating $\mathbb{P}[S]$, so by the law of total probability we have:

$$\mathbb{P}[S] = \sum_{x=0}^{n} \mathbb{P}[S|F=x]\mathbb{P}[F=x]. \tag{3}$$

The distribution of the number of failures during the observation window is given by:

$$\mathbb{P}[F=x] = \binom{n}{x} p^x (1-p)^{n-x}.$$

3) Exact Computation for Small Networks: $\mathbb{P}[S]$ can be computed exactly for small networks using Equation 3. We need to compute $\mathbb{P}[S|F=x]$. We can compute it by testing all cases of possible failures. The size of the combinatorial space to explore is exponential with the topology size. The computation, thus, becomes quickly untractable. However, this is possible for small topology sizes ($n \leq 100$). We generate the $\binom{n}{x}$ different cases of having x defective access nodes among n access nodes. We then determine which nodes are isolated. When it is done, we check if there is a segment of length $\geq l^*$ of consecutive defective or isolated nodes.

Figure 6 provides the execution time to compute $\mathbb{P}[S]$ for networks of sizes between $n = 10$ and $n = 90$ access nodes. We confirm that the computation time is exponential with n. It takes less than 10 minutes to compute $\mathbb{P}[S]$ for a network of size 65, but more than 1 hour for a network of size 83. For networks with $n \geq 100$, it is impractical to use this method of computation. We thus need approximation formulas to compute the probability for larger networks.

B. Analytical Approximations

When the probability p of a node to become defective is small ($np^x \ll 1$), Expression 3 can be approximated,
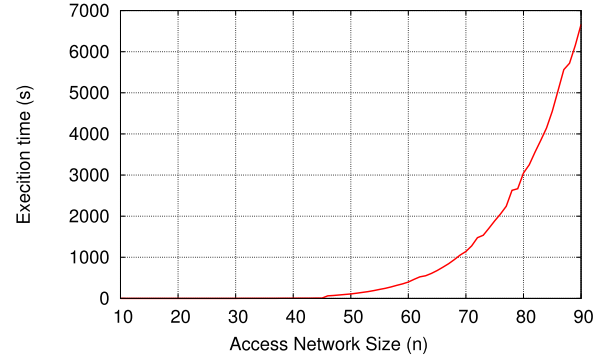

Fig. 6. Computation time of the formula of the probability to experience a long disconnection, Equation 3.

see Theorem 1. $\mathbb{P}[S]$ is the sum of $\mathbb{P}[S|F=x]\mathbb{P}[F=x]$ over all possible values of x, the number of defective ANs in the linear access network. In fact, it is enough to compute the expression for only a single value of x. Indeed, it can be shown that the dominant term of the sum is the one for the minimum number of failures causing a disconnected interval of length l^*, $x(l^*)$ defined below.

Definition 1 (Minimum Number of Failures): We note $x(l^*)$ the minimum number of failures necessary to have a disconnected interval of size l^*, i.e., $x(l^*) = 2 + \lfloor \frac{l^*-2}{\mathcal{D}} \rfloor$.

This minimum number corresponds to cases in which failures happen to a maximum number of gateways.

The next step is to determine and count the number of possible positions for the $x(l^*)$ defective nodes leading to a long disconnection. Doing so, we get the following result:

Theorem 1: When $np^{l^*} \ll 1$, the probability to experience at least a long disconnection in the linear access network is well estimated by

$$\mathbb{P}[S] \approx \lfloor \frac{n-1-l^*}{\mathcal{D}} + \mathbb{1}_{\mu \neq n-1} \rfloor \frac{1}{2} (\mathcal{D}-\mu)(\mathcal{D}-\mu+1) p^{x(l^*)}, \tag{4}$$

where $\mu = ((l^*-2) \mod \mathcal{D})$. Recall that $x(l^*)$ is the minimum number of failures necessary to have a disconnected interval of size l^*, i.e., $x(l^*) = 2 + \lfloor \frac{l^*-2}{\mathcal{D}} \rfloor$.

Moreover, when $\mathcal{D} \ll n$ and $l^* \ll n$ (this is the case for practical scenarios and, in particular, for the two scenarios we considered), it gives:

$$\mathbb{P}[S] \approx \frac{n}{2\mathcal{D}} (\mathcal{D}-\mu)(\mathcal{D}-\mu+1) p^{x(l^*)}. \tag{5}$$

1) Formula and Parameter Impacts: The approximation formula exhibits the influence of the system parameters. Note the importance of the factor $p^{x(l^*)}$. $\mathbb{P}[S]$ thus depends exponentially on $x(l^*)$, showing the strong impact of l^* (and thus indirectly of the speed of the vehicle v) and of the number of ANs between gateways, $\mathcal{D}$. However, $\mathbb{P}[S]$ depends only linearly on n. Let us now prove the theorem.

Proof: Let us prove the result. Recall that a *long disconnection* is a disconnection of the users during a time $t \geq t^*$. This disconnection is caused by a segment of contiguous failing (defective or isolated) nodes of length $l \geq l^*$. We call such a segment a *syndrome*. We define the event S^i as "There exists a syndrome (of length $l \geq l^*$) starting at access
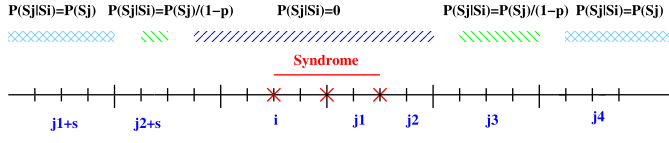
Fig. 7. Syndrome dependence: $\mathbb{P}[S^j|S^i] \leq \mathbb{P}[S^j]/(1-p)$.

Fig. 8. Sketch of the proof for a syndrome.

node i". We note $\mathcal{S}$ the event "There exists **at least** one syndrome in the network". We have $\mathcal{S} = \cup_{i \in N} S^i$.

It gives

$$\mathbb{P}[\mathcal{S}] = \sum_{i \in N} \mathbb{P}[S^i] - \sum_{i,j \in N} \mathbb{P}[S^i \cap S^j] + \dots$$

By definition, we have $\mathbb{P}[S^i \cap S^j] = \mathbb{P}[S^i|S^j]\mathbb{P}[S^j]$. Note now that it is impossible to have two overlapping syndromes, since the occurrence of such case implies to observe a larger syndrome of length $s_i + s_j - \delta$, where δ is the number of nodes they overlap of (see Figure 7) and s_i is the length of the syndrome S_i. Therefore, we have:

$\mathbb{P}[S^i|S^j] = 0$ if j and i "overlaps"
$\mathbb{P}[S^i|S^j] \leq \mathbb{P}[S^i]/(1-p)$ if $j < i$ or if j and i "close"
$\mathbb{P}[S^i|S^j] = \mathbb{P}[S^i]$ otherwise

where overlaps means $i - (i \mod \mathcal{D}) - 1 \leq j \leq (i+s_i - (i+s_i \mod \mathcal{D}) + \mathcal{D} + 1$. Hence,

$$\mathbb{P}[S^i \cap S^j] \leq \frac{\mathbb{P}[S^i]\mathbb{P}[S^j]}{1-p}.$$

Similarly, when intersecting all possible syndromes in the topology, we have

$$\mathbb{P}[\cap_{l*} S_l^{i*}] \leq \frac{\prod_{l*} \mathbb{P}[S_l^{i*}]}{(1-p)^{l*-1}}.$$

Hence, $\mathbb{P}[\cap_z S^{i_z}] \leq N^z \max_i P[S^i]^z/(1-p)^z$. As $\forall i \in N$ $\mathbb{P}[S^i]/(1-p) \ll 1$ (anticipating from Equation 8, we get

$$\mathbb{P}[\mathcal{S}] \approx \sum_{i \in N} \mathbb{P}[S^i].$$

Therefore, we compute the probability to have a syndrome of length l starting at node i, denoted by $\mathbb{P}[S_l^i]$. We define $\mu = l - 2 \mod \mathcal{D}$.

Proposition 1: There exist two classes of nodes.

- *First case ($0 \leq i \mod n_1 < n_1 - \mu$): $x(l^*)$ failures are necessary to disconnect the interval. The probability of a syndrome is*

$$\mathbb{P}_1[S] = p^{x(l^*)}(1-p)^{(i(l^*)+1)\mathcal{D}-l+1}.$$

We have v_1 such cases with

$$v_1 = (\mathcal{D} - \mu)\lfloor \frac{n-1-l^*}{\mathcal{D}} + \mathbb{1}_{\mu \neq n-1}\rfloor.$$

- *Second case: $x(l^*) + 1$ failures are necessary to disconnect the interval. The probability of a syndrome is*

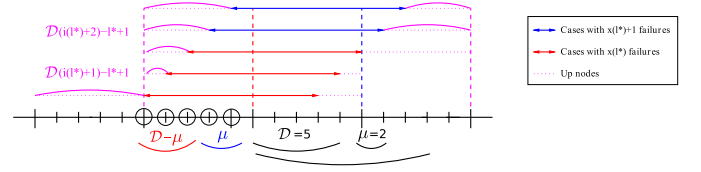$$\mathbb{P}_2[S] = p^{x(l^*)+1}(1-p)^{(i(l^*)+2)\mathcal{D}-l^*+1}.$$

We have v_2 such cases with

$$v_2 = \mu\lfloor \frac{n-1-l^*}{\mathcal{D}} \rfloor + 1 + \mathbb{1}_{\mu \neq n-1}.$$

We verify that we have $v_1 + v_2 = n - l + 1$ possible syndromes.

Proof: Given a syndrome, i.e., a disconnected interval length l^*, the first and last nodes of the interval, plus the intermediate gateways have to fail to obtain a syndrome. It happens with probability $p^{x(l^*)}$ (first case) and $p^{x(l^*)+1}$ (second case). The nodes inside the interval can have any state. The nodes outside the interval between the two[1] neighboring gateways have to be up and running. There are $(i(l^*)+1)\mathcal{D} - l^* + 1$ of them (first case) and $(i(l^*)+2)\mathcal{D} - l^* + 1$ of them (second case), giving the probabilities of failure. To obtain v_1 and v_2, consider the nodes between two gateways and the first gateway (the $\mathcal{D}$ circled nodes in Figure 8). If the first node of the disconnected interval is one of the $\mathcal{D} - \mu$ first such nodes, only $x(l^*)$ failures are necessary (first case). For the next μ, we are in the second case. There are $\lfloor \frac{n-1-l^*}{\mathcal{D}} \rfloor$ such intervals. The remaining terms of the formulas deal with the border of the network (first node and last non complete interval). ∎

Therefore, when p is small,

$$\mathbb{P}_1[S] = p^{x(l^*)}(1-p)^{(i(l^*)+1)\mathcal{D}-l+1} \approx p^{x(l^*)}, \qquad (6)$$
$$\mathbb{P}_2[S] = p^{x(l^*)+1}(1-p)^{(i(l^*)+2)\mathcal{D}-l+1} \approx p^{x(l^*)+1}. \qquad (7)$$

Hence $\mathbb{P}_1[S] \gg \mathbb{P}_2[S]$.

2) Approximation of $\mathbb{P}[\mathcal{S}]$ When p Is Small: We have seen that

$$\mathbb{P}[\mathcal{S}] \approx \sum_{i \in N} \mathbb{P}[S^i] \quad \text{and} \quad \mathbb{P}[S^i] = \sum_{z \geq l} P[S_z^i],$$

giving

$$\mathbb{P}[\mathcal{S}] \approx \sum_{i \in N} \sum_{z \geq l} P[S_z^i].$$

Equations (6) and (7) say that $P[S_z^i] \approx p^{x(z)}$ or $P[S_z^i] \approx p^{x(z)+1}$. As $x(z)$ is a non decreasing function, only the z such that $x(z) = x(l^*)$ are contributing to the double sum. Furthermore, only the nodes of the first class are contributing to $\mathbb{P}[\mathcal{S}]$ (basically, only the terms in $p^{x(l^*)}$ contribute).

We thus only express $\mathbb{P}[S^i]$ for the nodes of the first class. In an interval (syndromes starting in $\mathcal{D} \leq i + j$ with $0 \leq j < \mathcal{D}$), we have $\mathcal{D} - \mu$ nodes of class 1 (nodes of index $0 \leq i \mod \mathcal{D} < \mathcal{D} - \mu$), see Figure 9. For the last node, only a syndrome of size l^* contributes. For the one before last node, two syndromes of sizes l^* and l^*+1 contribute. More in

---

[1] There are two special cases on the border of the network that we not consider for this analysis.
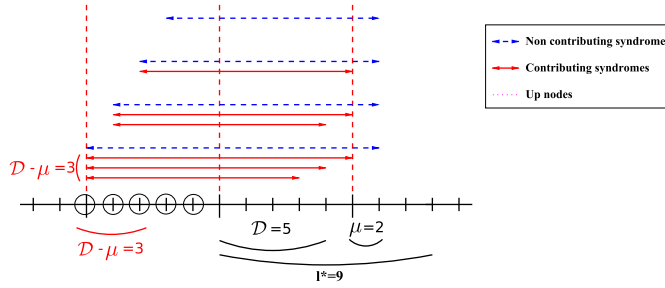
Fig. 9.   Length of the contributing syndromes starting in an interval.

general, $\mathcal{D} - \mu - j$ syndromes of size $l^* \leq j \leq l^* + \mathcal{D} - \mu - 1$ contribute. We get

$$\mathbb{P}[S^i] \underset{p \approx 0}{=} (\mathcal{D} - \mu - 1 - (i \mod \mathcal{D})) p^{x(l^*)}. \tag{8}$$

It gives

$$\mathbb{P}[\mathcal{S}] \approx \lfloor \frac{n - 1 - l^*}{\mathcal{D}} + \mathbb{1}_{\mu \neq n-1} \rfloor \sum_{k=1}^{\mathcal{D}-\mu} k p^{x(l^*)},$$

Finally, $\sum_{k=1}^{\mathcal{D}-\mu} k = \frac{1}{2}(\mathcal{D} - \mu)(\mathcal{D} - \mu + 1)$, yielding Equation 4 of Theorem 1.

Moreover, in practical settings, $\mathcal{D}$ and l^* can be considered as small constants compared to n ($\mathcal{D}$ is 3 and l^* is 6 and 5 respectively in the train and bus scenarios and would not be larger than few tens, when n is in the order of hundreds or thousands). Thus, we get:

$$\lfloor \frac{n - 1 - l^*}{\mathcal{D}} + \mathbb{1}_{\mu \neq n-1} \rfloor = \lfloor \frac{n}{\mathcal{D}}(1 - \frac{1}{n} - \frac{l^*}{n} + \frac{\mathbb{1}_{\mu \neq n-1}\mathcal{D}}{n}) \rfloor$$
$$\approx \frac{n}{\mathcal{D}},$$

yielding Equation 5 of Theorem 1.

3) Remark on the Approximation: The terms of error in Equation 5 is basically $\frac{1}{n} - \frac{l^*}{n} + \frac{\mathbb{1}_{\mu \neq n-1}\mathcal{D}}{n}$. If l^* and $\mathcal{D}$ are less than x% of n, the error of the approximation is less than x%. As an example, for the train scenario, we get an approximation within 0.2% and within 0.5% for the bus scenario. ∎

V. RESULTS

In this section, we first validate the analytical approximation provided in the previous section, by comparing it with the exact computations for small networks. We then study the probability of failure (that is of experiencing a long disconnection) of larger linear access networks using the approximation formula. We make a sensitivity analysis of the probability of failure to the different parameters of the network topology, in particular for the two scenarios we presented (intercity bus and train). Doing so, we provide a guide to network operators for choosing the right values of the parameters as a function of the maximum probability of failure they want their users to experience and as a function of the cost of maintenance of the network.

A. Validation of the Analysis

To validate the approximation formula provided in Theorem 1, we compare its results with the ones of the exact computations for small access networks of sizes between $n = 10$ and $n = 50$ access nodes. We provide in Figure 10 the comparison for different values of p, as the precision of the approximation depends on p. We chose values of p which may be encountered in practice: $p = 0.0001$, $p = 0.001$, $p = 0.01$, respectively for the Left, Middle, and Right plots. The values of the other parameters correspond to the bus scenario. We see that the approximation formula matches almost perfectly the exact computations for the three values of p considered. Thus, we use it in the following.

B. Sensitivity Analysis to Network Parameters

We carry out an analysis of the impact of each parameter on the probability to experience a long disconnection, $\mathbb{P}[S]$. We consider the two scenarios presented above, the intercity train and bus. We set the parameters at their default values given in Table I. We then vary the values of the parameters one by one. The parameters considered are: n, the topology size, $\mathcal{D}$, the number of access nodes between gateways, v, the speed of the vehicle, t^*, the maximum disconnection time a user may experience, and c, the distance between two access nodes. Results are presented for the two scenarios in Figure 11. The vertical bars in the plots correspond to the default values of the studied parameter.

1) Impact of the Topology Size: We first study the impact of the topology size on $\mathbb{P}[S]$, i.e. of its number of access nodes (ANs), n. We have seen in the analysis of the previous section that, when n is large enough, $\mathbb{P}[S]$ depends linearly on n (see Equation 5 of Theorem 1). We confirm this in Figure 11 (Left). We consider topologies with sizes from 10 to 5000 ANs. Note the two vertical bars corresponding to the default sizes of 600 ANs for the bus scenario and of 2000 ANs for the train scenario. We observe that the probability of failure is low in both scenarios: 1.9×10^{-6} for the train scenario with 2000 ANs and 1.1×10^{-6} for the bus scenario with 600 ANs. The probability stays low even if we double for example the network size: 3.2×10^{-6} for a train network with 4000 ANs and 2.5×10^{-6} for a bus network with 1200 ANs.

2) Effect of the Number of Nodes Between Gateways: The impact of $\mathcal{D}$ is more complex (see Equation 5). $\mathbb{P}[S]$ is very sensitive to small changes for small values of $\mathcal{D}$. As an example, the probability of failure is 2×10^{-6} for the default values in the train scenario, that is when $\mathcal{D} = 3$. But, it is only 3×10^{-9} for $\mathcal{D} = 2$ (see Figure 11b). Conversely, for the bus scenario, the probability of failure is a lot higher when $\mathcal{D} = 4$ than for the default value: 1×10^{-4} to be compared with 1×10^{-6}. A single unit of change of the value of $\mathcal{D}$ corresponds to a change of several order of magnitude for $\mathbb{P}[S]$. However, when $\mathcal{D}$ is larger (here ≥ 8), its impact decreases and the relation with $\mathbb{P}[S]$ becomes close to linear.

3) Impact of t^*: We provide in Figure 11 (Middle) the probability of to experience a long disconnection as a function of the corresponding maximum allowed disconnection time, t^*. We see that $\mathbb{P}[S]$ is a step function. The explanation is that $\mathbb{P}[S]$ depends on the maximum size of a disconnected interval l^* and that we have one value of $\mathbb{P}[S]$ for one value of l^*, as shown in the figure (dotted lines). Note that, for the
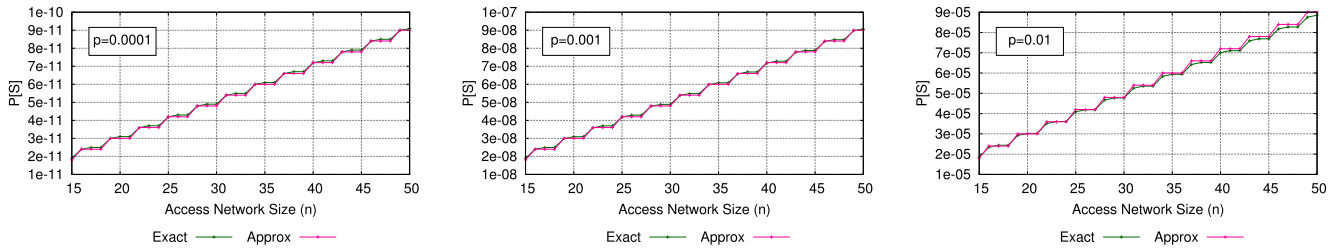
Fig. 10. Validation of the approximation formula for different values of p, the probability of failure of an access node.

bus and train scenario, t^* has a default value of 3 seconds. But, we can ensure for free (meaning with the same probability of failure), that we will not experience a disconnection larger than 2.5 seconds, the previous threshold value between two steps.

4) Effect of the Vehicle Speed v: We now study the impact of speed. We also observe in Figure 11 (Right) that $\mathbb{P}[S]$ is a step function of the speed. Indeed, the value of l^* is discrete and depends on v as expressed in Equation 2. An bus driving at a speed between 22.5 km/h and 75 km/h would experience the same probability of failure (1.1×10^{-6}) than at its default value of 30 km/h. Driving more slowly than 22.5 km/h would increase a lot the probability of failure to $2e - 4$). The train scenario experience less variations around the default value for the probability of failure. A train driving between 75 km/h and 142 km/h experiences a probability of 1.9×10^{-6}, of 4×10^{-6} between 22.5 km/h and 75 km/h, and of 2×10^{-7} between 142 km/h and 200 km/h.

5) Effect of the Distance Between Access Nodes c: The last parameter we study is c the distance between access nodes. It has a very strong influence on $\mathbb{P}[S]$. Indeed, Equation 2 shows, that c greatly impacts the value of l^*. As an example, for the train scenario, $\mathbb{P}[S]$ is 10^{-12} for $c = 25$, 10^{-6} for $c = 25$, and 10^{-3} for $c = 75$. The variation is of several order of magnitude. Reducing the distance between gateways greatly improves the reliability of the system.

C. Discussion About Train and Bus Scenarios, and Take-Aways for Operators

We have estimated the impact of the network parameters, n (linear impact), t^*, v, c, and $\mathcal{D}$ (very important impact). We further discuss the importance for a network operator to *choose their values in the right range*. Then, we carry out an *analysis of the maintenance periodicity* (and thus of the *maintenance cost for an operator*) on the mean time to experience a long disconnection.

1) Ladder Effect: As we have seen, for a large number of parameters of linear access networks, the probability to experience a long disconnection follows a step function of the parameter. We talk of a *ladder effect.* The explanation is double: the size of the disconnected segment is discrete when the speed is continuous. Moreover, more importantly, an crucial parameter is $x(l^*)$, the minimum number of access node failures to obtain a disconnected interval of size l^*. $x(l^*)$ is also discrete and follows a step function of l^*. A change of 1 unit of the value of $x(l^*)$ changes the order of the value of $\mathbb{P}[S]$ as shown in Equation 5. It is important for a network operator to be aware of this ladder effect, as a

small change of the value of a parameter can have a large impact on the probability of experiencing a long disconnection. It is important to know the threshold values and, if possible, to choose the parameters of the system so that the system is just before a step. As an example, if we consider a high speed train scenario, a change of speed from 200 km/h to 205 km/h makes $\mathbb{P}[S]$ drops from 1×10^{-6} to 3×10^{-9}. A very small relative change of speed (which may or may not be possible to the operator of the transportation system) has a huge impact of several order of magnitude.

D. Network Maintenance Cost for an Operator

In this section, we consider the maintenance cost of a network operator. As ANs are cheap and not very reliable, defective ANs have to be changed. However, it would be too expensive to change an AN as soon as it becomes defective. The best way is to plan a periodic maintenance (e.g. every month or every two weeks), so that an employee replaces all the ANs that have failed during the period. We study here how to choose the maintenance periodicity and if this solution can be put into reality with a reasonable cost for the bus and train scenarios. We plot in Figure 11f the mean time (in month) to experience a long disconnection as a function of the periodicity of the maintenance (in days). We discuss below both scenarios.

1) Bus Scenario: The probability of experiencing a long disconnection during one day is $P = 1.1 \times 10^{-6}$. If an operator wants to have low maintenance costs, the wifi access nodes may be replaced only once every two weeks. We consider that this period is reasonable because of the short length of the bus line considered (30 km) and of the fact that all access nodes are accessible along a road in this scenario and not far from the city center. During two weeks, in average, 8.4 access nodes fail. However, the probability to experience a long disconnection over this period is only of 0.3%.[2] This means that regular passengers of the bus experience a disconnection only around every 27 years and 9 months (see Figure 11f). This is a very large number! To carry out such a maintenance, a single employee just has to travel along the 30 km line and replace the failed access nodes. This can easily be done in less than half a day.

2) Train Scenario: In the train scenario, the probability to experience a long disconnection during one day is

---

[2]The probability to experience a long disconnection over a period of τ days is computed by setting the prob. of an access node to fail to $1 - (1 - p)^\tau$, where p is the prob. to fail during a single day.
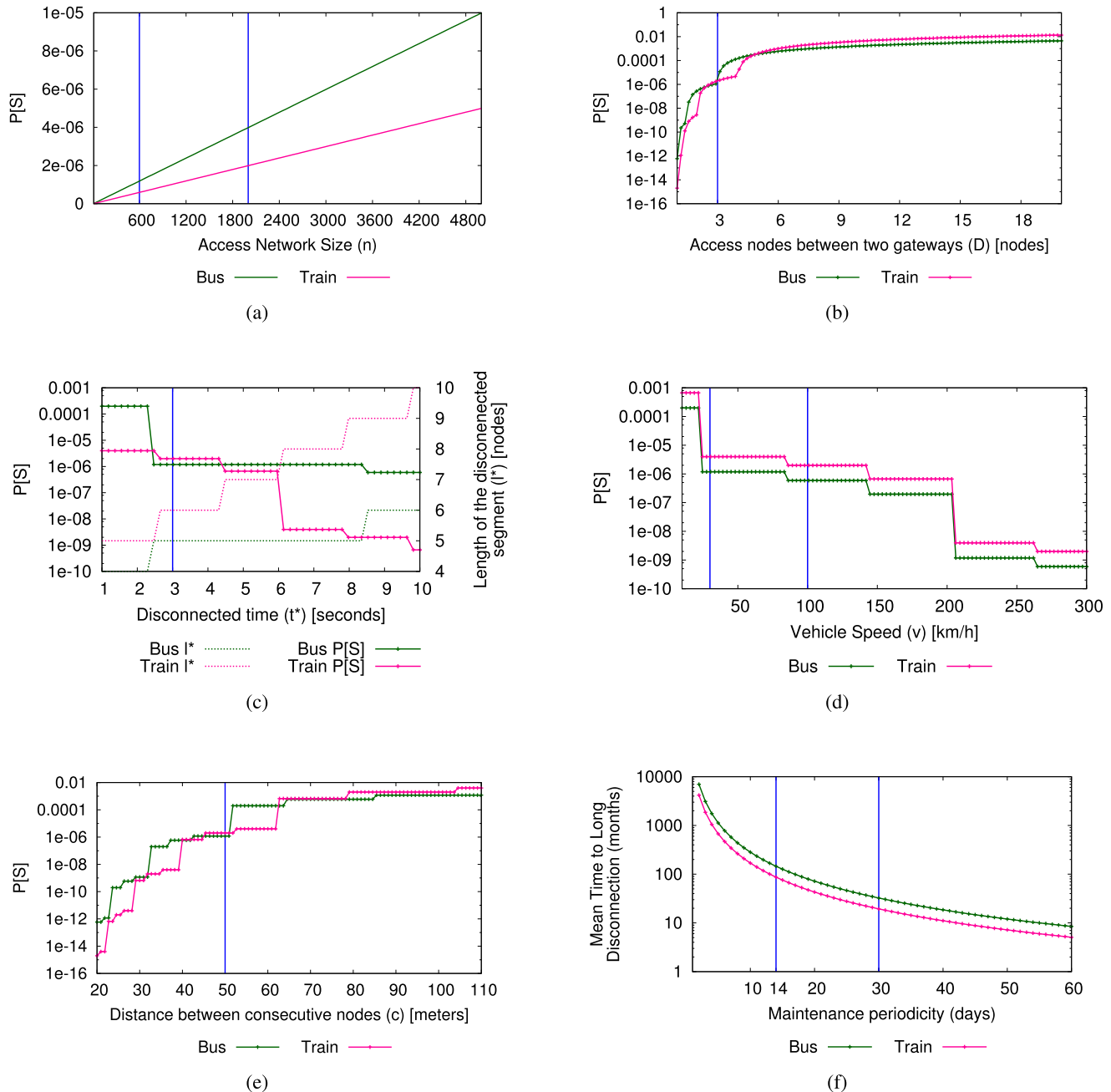
Fig. 11. Sensitivity analysis to the network parameters of the probability to experience a long disconnection for the intercity bus and train scenarios. Vertical lines correspond to the default values of both scenarios. (a) Sensitivity to the number of access nodes of the network topology. (b) Sensitivity to the number of access nodes between two gateways. (c) Sensitivity to the maximum disconnection time acceptable for a user. (d) Sensitivity to the vehicle speed. (e) Sensitivity to the distance between two access nodes. (f) Mean time to a long disconnection for the linear access network as a function of the periodicity of the network maintenance.

$P = 1.9 \times 10^{-6}$. As the length of the train line (100 km) is longer than the one of the bus line, an operation to replace the access nodes is more complex. We estimate that it would take a single employee a whole day to travel along a 100 km line and replace the failed access nodes. We thus consider in this case a longer maintenance periodicity of one month. During one month, in average, 59 access nodes fail. The probability to experience a long disconnection over this period is 5%. This means that a single long disconnection will happen in average every 20 months (see Figure 11f). This is a still a large number. Recall, that we are discussing an connection

interruption of only 3 seconds. We considered here only very low maintenance cost, done by a single employee in one day. An operator may want to decrease the maintenance period to two weeks even for a train line, doubling the cost. In this case, the mean time to a long disconnection would be 16 years and 8 months.

VI. CONCLUSION

In this work, we have analyzed the tolerance to failures of a linear access network offering connectivity to passengers commuting in an intercity transportation system. We provided

an approximation formula to estimate precisely the probability for a passenger to experience a long disconnection. We studied two practical scenarios corresponding respectively to an intercity train and to an intercity bus. We showed that, for reasonable values of the parameters, the probability for the networks to experience a long disconnection is very low. If an employee of the operator of the linear access networks replace the cheap defective access nodes once two weeks (for a linear network of tens of kilometers), a long disconnection would happen only around every 15 years. This shows that such solutions are viable.

Furthermore, we provide a sensibility analysis of the probability to network parameters allowing a network operator to smartly choose the parameter of the systems, such as distance between nodes, distance between gateways, speed of the vehicles. We exhibit a ladder effect and show that a very small changes of the value of some parameters may significantly increase the resilience of the system. This is of great value for the network operators which may improve the reliability of the system with very small changes.

For the next steps, it would be interesting to compare the analytical results and the simulations provided in the paper with measurement based data.

REFERENCES

[1] F. Giroire and J.-C. Maureira, "Analysis of the failure tolerance of linear access networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, USA, Dec. 2016, pp. 1–7.

[2] S. D. Drake, R. Boggs, M. Sandler, R. Giusto, S. K. Sudan, and S. Ryan, "Worldwide mobile worker population 2007–2011 forecast EXCERPT," IDC, Framingham, MA, USA, Tech. Rep. 256194, Mar. 2008.

[3] G. Luk, "Global mobile workforce forecast update 2016–2022," Strategy Analytics, Boston, MA, USA, Tech. Rep., Nov. 2015. [Online]. Available: https://www4.strategyanalytics.com/default.aspx?mod=offices

[4] M. Agervald, K. Axelsson, M. Bergek, and M. Karlsson, "A system for data transmission via several communication routes," U.S. Patent WO2000 067 435 A1, Jan. 30, 2002. [Online]. Available: http://www.google.com/patents/EP1175757A1?cl=ru

[5] M. Karlsson, "Wireless train communication system," U.S. Patent WO2016 013 968 A1, Jan. 28, 2016. [Online]. Available: https://encrypted.google.com/patents/WO2016013968A1?cl=un

[6] M. Emmelmann, "System design and proof-of-concept implementation of seamless handover support for communication-based train control," in *Vehicular Networking: Automotive Applications and Beyond* (Intelligent Transport Systems). New York, NY, USA: Wiley, Jun. 2010.

[7] D. T. Fokum and V. S. Frost, "A survey on methods for broadband Internet access on trains," *IEEE Commun. Surveys Tuts.*, vol. 12, no. 2, pp. 171–185, 2nd Quart., 2010.

[8] S. Al-Sultan, M. M. Al-Doori, A. H. Al-Bayatti, and H. Zedan, "A comprehensive survey on vehicular ad hoc network," *J. Netw. Comput. Appl.*, vol. 37, no. 1, pp. 380–392, Jan. 2014. [Online]. Available: http://dx.doi.org/10.1016/j.jnca.2013.02.036

[9] Y. Zhou and B. Ai, "Handover schemes and algorithms of high-speed mobile environment: A survey," *Comput. Commun.*, vol. 47, pp. 1–15, Jul. 2014.

[10] J.-C. Maureira, "Spiderman handover for railway communications," *IEEE Intell. Transp. Syst. Mag.*, vol. 8, no. 4, pp. 54–67, Oct. 2016.

[11] L. Verstrepen *et al.*, "Making a well-founded choice of the wireless technology for train-to-wayside data services," in *Proc. 9th Conf. Telecommun. Internet Media Techno Econ. (CTTE)*, Jun. 2010, pp. 1–7.

[12] Z. Yan, H. Jiang, Z. Shen, Y. Chang, and L. Huang, "*K*-connectivity analysis of one-dimensional linear VANETs," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 426–433, Jan. 2012.

[13] S. C. Ng, G. Mao, and B. D. O. Anderson, "On the properties of one-dimensional infrastructure-based wireless multi-hop networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2606–2615, Jul. 2012.

[14] W. Zhang *et al.*, "Multi-hop connectivity probability in infrastructure-based vehicular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 740–747, May 2012.

[15] A. Sharif-Nassab and F. Ashtiani, "Connectivity analysis of one-dimensional ad hoc networks with arbitrary spatial distribution for variable and fixed number of nodes," *IEEE Trans. Mobile Comput.*, vol. 11, no. 10, pp. 1425–1435, Oct. 2012.

[16] J.-C. Maureira, "Internet on rails," Ph.D. dissertation, Dept. I3S, Univ. Nice Sophia Antipolis, Nice, France, 2011.

[17] J.-C. Maureira, P. Uribe, O. Dalle, T. Asahi, and J. Amaya, "Component based approach using OMNeT++ for train communication modeling," in *Proc. 9th Int. Conf. Intell. Transp. Syst. Telecommun. (ITST)*, Oct. 2009, pp. 441–446.

[18] S. Banerjee, M. Hempel, and H. Sharif, "A survey of wireless communication technologies & their performance for high speed railways," *J. Transp. Technol.*, vol. 6, no. 1, 2016, Art. no. 62604.

[19] T. Koch, "Is the assumption of perfect channel-state information in fading channels a good assumption?" in *Proc. 2nd Int. Symp. Appl. Sci. Biomed. Commun. Technol.*, Nov. 2009, pp. 1–6.

Frédéric Giroire received the Ph.D. degree from Pierre and Marie Curie University in 2006. He was with Sprint research laboratories, CA, USA, in 2002, for six months, and with Intel Research laboratories, Berkeley, in 2007, for one year. He has been a Research Scientist with CNRS since 2008, inside the Joint Team Coati between I3S Laboratory, CNRS, University of Nice-Sophia Antipolis and Inria. He holds three patents. His research interests include algorithmic graph theory and combinatorial optimization for network design and management issues.

Juan-Carlos Maureira received the degree in computer sciences from Northern Catholic University, Chile, in 2004 and the Ph.D. degree in discrete simulation and network optimization from Inria Sophia-Antipolis (Mascotte Team) under the supervision of J.-C. Bermond and O. Dalle. His Ph.D. thesis was on Internet on Rails, approved with a Trs Honorable mention, in 2011. From 2002 to 2003, he authored his engineering thesis with the Telematics Institute, Karlsruhe University, Germany (DAAD Scholarship). Then, he joined the Center for Mathematical Modeling, University of Chile, as System Engineer in 2004, and then an HPC Engineer in 2006. Then, he rejoined the Center for Mathematical Modeling, University of Chile, as a Scientist in HPC and astroinformatics research areas.