Computing Realistic and Adaptive Capacity of D2D Contacts

Clément Bertier, Farid Benbadis, Vania Conan, Marcelo Dias de Amorim

Computing Realistic and Adaptive Capacity of D2D Contacts

Clément Bertier[⋆,⋄], Farid Benbadis[⋄], Vania Conan[⋄], and Marcelo Dias de Amorim[⋆]

[⋆]LIP6/CNRS – Sorbonne Université [⋄]Thales SIX GTS France

***Abstract*—Assessing the performance of opportunistic networks requires a subtle understanding of both contact and intercontact patterns. While the analysis of intercontacts has attracted significant attention from the research community, surprisingly only a few works have focused on explaining what happens during a contact. In this paper, we perform an in-depth analysis of contacts using both empirical measurements and reference models. We make several observations that allow us to better capture the adaptive nature of device-to-device (D2D) links. In particular, in the case of Wi-Fi 5, we show that a slight modification of the nominal modulation scheme is enough to achieve an accurate characterization of opportunistic contacts for some categories of propagation models. As a consequence, we confirm previous observations that the evaluation of protocols and algorithms for D2D networks based on links of fixed rate may lead to inaccurate results. We finally propose a tool that extends mobility traces with plausible values of per-link capacity.**

I. Introduction

It is no more a debate that device-to-device (D2D) and opportunistic networking open promising expectations in wireless communications [1], [2]. Designing and implementing an efficient opportunistic network is however challenging because of the highly dynamic nature of the topology. Individual nodes have their own mobility patterns, and communication opportunities appear whenever two nodes get in communication range of each other. An accurate characterization of such interactions is then a must-have step.

The literature is full of outstanding contributions that assess the underlying phenomena governing opportunistic device-to-device communications [3], [4]. In particular, we have today a good understanding of inter-contact patterns in such networks. Unfortunately, there are still critical open questions regarding the *contact* characterization. While inter-contacts depend primarily on the mobility of the nodes, characterizing a contact in detail also depends on other parameters such as the device-to-device communication technology. In this paper, we investigate issues explicitly related to the contact between two mobile nodes.

A common limitation of existing works is that they assume either that contacts have infinite capacity, allowing the transfer of any amount of data during a single encounter, or that the communication throughput during the contact follows the nominal values of the technology (e.g., Bluetooth or Wi-Fi). This leads to a sort of *fixed-rate* characterization of contacts, where either nodes are in contact (with a unique characteristic), or they are not. In this paper, we advocate that, to achieve a better characterization of opportunistic contacts, we should adopt a finer representation of the relative mobility of nodes to better capture the communication possibilities.

The central premise of our work is that the distance separating two nodes is likely to vary during a contact. Because of the inherent characteristics of the wireless medium, the quality of the signal varies with the distance between the nodes.[1] The varying node-to-node distance suggests that the link throughput between the nodes varies accordingly. This variation depends on the communication technology under consideration (in Section II, we will discuss the technology we use in our experiments), but regardless of this latter, the throughput always fluctuates as a function of the distance between the transmitter and the receiver.

Fortunately, some authors have already raised some issues with the traditional fixed-rate approach. Chowdhury et al. approximated the distance-to-throughput by first fixing a path loss equation and then observing the distances covered by each modulation in the case of IEEE 802.11n standard [5]. The authors notably propose the contact capacity as an integral of throughput capacity over time. Qayyum et al. proposed to measure the distance-to-throughput using a mobile application in Android [3]. Although sharing some goals with us, their work was restrained to short D2D links as they focused on Bluetooth links. Neto et al. proposed to estimate the contact capacity by taking into account the speed of a mobile node [6]. They transferred data, and attempt to establish a relationship between speed, RSSI, and throughput.

In this paper, we help advance the state of the art in several regards. We investigate the *impact of the varying relative distance between nodes on the capacity of opportunistic links*. By capacity, we mean the amount of data that can be transferred during the entire duration of the contact. Our ultimate goal is to establish a basis for more accurate design and evaluation of device-to-device communication strategies. We adopt an empirical methodology to achieve a realistic characterization of device-to-device links. Our experiments involve Android smartphones equipped with the Google Nearby API. We apply the observed characterization parameters to two mobility datasets, namely a vehicular one in the city of Luxembourg and a pedestrian one in Stockholm. We confront our observations

[1] Of course, other parameters come at play, but in this work we focus on the distance only and avoid including environment-dependent parameters.

978-1-7281-0270-2/19/$31.00

with the traditional fixed-rate contact characterization strategy and make some observations. We confirm the fact that considering only the duration of a contact and a fixed throughput is far from enough to capture the actual capacity of the contact.

Although the *quantitative* analysis that we show in this paper is dependent on the communication technology, on the type of devices that we considered during our data collection campaign, and on the test environment conditions, the *qualitative* observations can be extended to any setup, given that the wireless technology adopts a rate-adaptive strategy. We finally propose a transformation tool that takes mobility traces as inputs and generates realistic contact capacity traces as outputs. We are hopeful that this will entice the community to compare traditional contacts with adaptive ones. As a summary, our contributions are:

- **Data collection and analysis.** We measure the throughput of modern Android devices according to the distance to assess the maximum communication range of D2D technologies on contemporary smartphones. We use this data as a means to compute the throughput as a function of distance and set a reference model for our work.
- **Adaptive contact characterization.** We thoroughly characterize the contact capacity, meaning the amount of exchangeable data through a contact, when we consider a fixed-rate contact and an adaptive-rate contact. We show they are not equivalent, mostly due to many contacts happening at a near-maximum range.
- **Fixed parameter selection.** Even though we advocate against fixed-rate contacts as a way to estimate contact capacity, we also propose a recommended fixed-rate contact value, to yield more realistic results.
- **Dataset contact capacity tool.** We propose a transformation tool that takes mobility a dataset as an input and generates contact datasets; which include plausible contact capacity for each pairwise D2D link between nodes.

The rest of the paper is organized as follows. In Section II, we state the problem, recall the traditional definition of fixed-rate contact characterization, and present the methodology we follow in our work. Since the fine characterization of contacts depends on the technology under consideration, we detail in Section III our experimental campaign as well as the measurement-based adaptive contact characterization. We describe the evaluation scenarios and present all the analysis in Section IV. In Section V we explain how the open-source software we propose calculates the contact capacity. We postpone the related work to Section VI so that the reader has enough material to understand our contributions better and conclude the paper in Section VII.

II. Definitions and Problem Formulation

We use the common definition of *contacts* in an opportunistic network, which is the period during which two nodes have a valid wireless link and can exchange data. We note this duration as τ. Nodes are mobile, and a contact starts as soon as the two nodes are within communication range of each other. In this paper, we consider that all nodes have the same antenna characteristics leading to symmetric links. We also assume omnidirectional antennas. We note the distance between two nodes A and B as $d_{AB}(t)$.

TABLE I
Theoretical transmission rate in IEEE 802.11ac (Wi-Fi 5) using a 80MHz bandwidth and two spatial streams.

RSSI (dBm)	**Modulation**	**Rate** (Mbps)
$[-55; -]$	256-QAM 5/6	866
$[-57; -56]$	256-QAM 3/4	780
$[-58; -58]$	64-QAM 5/6	650
$[-59; -59]$	64-QAM 3/4	585
$[-63; -60]$	64-QAM 2/3	520
$[-67; -64]$	16-QAM 3/4	390
$[-70; -68]$	16-QAM 1/2	260
$[-72; -71]$	QPSK 3/4	195
$[-75; -73]$	QPSK 1/2	130
$[-; -76]$	BPSK 1/2	65

In this paper, we are interested in the *contact capacity* C_{AB}, defined as the maximum amount of data that can be transferred between A and B during a contact. The contact capacity depends on the duration of the contact and the data rate of the opportunistic link between the two nodes.

In the literature, many authors simplify the problem by considering that nodes can exchange data at a *fixed rate* δ^{fixed} when within communication range. In this case, the contact capacity is simple to obtain:

$$C_{AB}^{\text{fixed}} = \delta^{\text{fixed}} \times \tau. \tag{1}$$

Although the fixed-rate contact characterization is simple to manipulate, it falls short in capturing the actual characteristics of a contact in a real setup. Because of several physical phenomena, the throughput that we can get in a wireless link depends on several factors, including the distance separating the sender from the receiver and the propagation conditions. The consequence is that the contact capacity is seldom a linear function of the duration.

Although the wireless medium shows a continuous decreasing behavior in terms of signal delivery, existing technologies adopt a step-wise transmission data rate calculation in function of the received signal strength indication (RSSI). In Table I, we show this dependence in the case of IEEE 802.11ac (Wi-Fi 5), the one we consider in our experiments. As we can see, the ratio between the maximum and minimum achievable throughputs is higher than one order of magnitude. A contact between two nodes is likely to traverse several of these data rate levels; depending on the mobility pattern, the resulting contact capacity can be anything between $\tau \times \delta_L$ and $\tau \times \delta_H$, where τ is the contact duration and δ_L (resp. δ_H) is the lowest (resp. the highest) data rate authorized by the wireless technology. For these reasons, *considering only the contact duration may not be enough to characterize and analyze opportunistic networks*.

The rest of this paper will focus on investigating how much impact we may get from a finer characterization of contacts. The first step is to make sure that nodes do observe varying

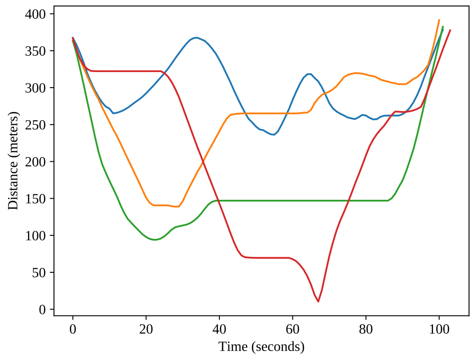


Fig. 1. Four contacts of ∼100 seconds each. Note that the distance between the nodes show the most variable patterns.

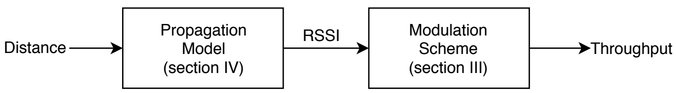


Fig. 2. From distance to throughput.

distances concerning neighbors for the same contact duration – this is precisely what we observe in practice. In Figure 1, we plot four different contacts of approximately 100 seconds in a real-world pedestrian scenario. As we can see, the four patterns show that, for the same given contact duration, the distances between the two nodes may vary unpredictably from one contact to another. We can find a case (red curve) where the nodes get very close (about 10 m) before moving away again. However, we can also find a case (blue curve) where the nodes remain very far from each other during all the duration of the contact and never approach less than 250 m.

Problem statement. The estimation of the throughput according to distance involves essentially two steps, as shown in Figure 2. The first one relates distance to RSSI based on some signal propagation model. The second step transforms RSSI into throughput, and this depends mainly on the modulation scheme. While there are common practices to determine the propagation model to fit a given scenario, finding which modulation scheme reflects best the system is more tricky. To the best of our knowledge, there is no well-defined method to transform a given RSSI into a throughput. In this paper, the main challenge will be to find models that can transform distance to throughput so that existing mobility datasets can also be extended to include information on the amount of data that can be transferred whenever two nodes meet.

A. Adaptive contact capacity

In opposition to the traditional fixed-rate contact characterization, whose contact capacity is given in Equation 1, we propose *adaptive contact characterization* as an alternative to better capture the behavior of opportunistic links.

We involve the two components of the system shown in Figure 2: propagation and modulation. Let $\mathrm{RSSI}(d_{AB}(t))$ be

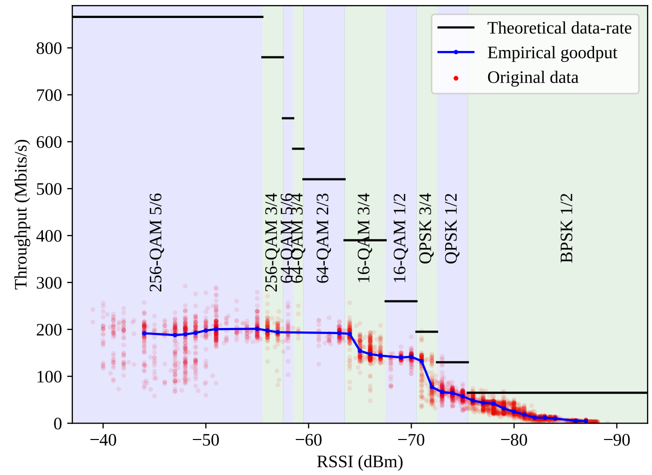


Fig. 3. RSSI-to-throughput experimental results. The step-wise black line represents the theoretical maximum values provided by the standard.

the RSSI measured on an opportunistic link when the nodes are at a distance $d_{AB}(t)$ (varying distance over time) and $\delta^{\text{adaptive}}(\mathrm{RSSI}(d_{AB}(t)))$ be the throughput.

The contact capacity for the adaptive contact characterization is given by:

$$C_{AB}^{\text{adaptive}} = \int_{T}^{T+\tau} \delta^{\text{adaptive}}(\mathrm{RSSI}(d_{AB}(t)))dt. \qquad (2)$$

where $[T, T+\tau]$ is the contact interval.

III. Empirical reference link characterization

To establish a realistic reference basis, we run an experiment to observe how the relationship between RSSI and throughput behaves in off-the-shelf devices.

A. Experimental RSSI to throughput estimation

We carried out a day-long measurement campaign to collect throughput values for different RSSIs in a quasi interference-free environment. For the communicating devices, we used two OnePlus 5T smartphones equipped with the Android 8.1 operating system and 2×2 MIMO antennas. We used the Google Nearby framework to establish device-to-device links between the nodes [7]. Google Nearby is a proprietary device-to-device library which uses Wi-Fi 5 (5 GHz band) for high-speed throughput. We initially set the devices at a one-meter distance, measure the RSSI and the throughput using Nearby, move the devices away and perform another round of measurements. We repeat the process until the two nodes lose connectivity.

Theoretical versus empirical throughput. One may believe that it suffices to use the technology's specification (e.g., table I) to obtain realistic contact capacity estimation according to distance. To verify the validity or not of this assertion, we show the results of our experiment in Figure 3. The black step-wise function represents Wi-Fi 5's theoretical data rate, and

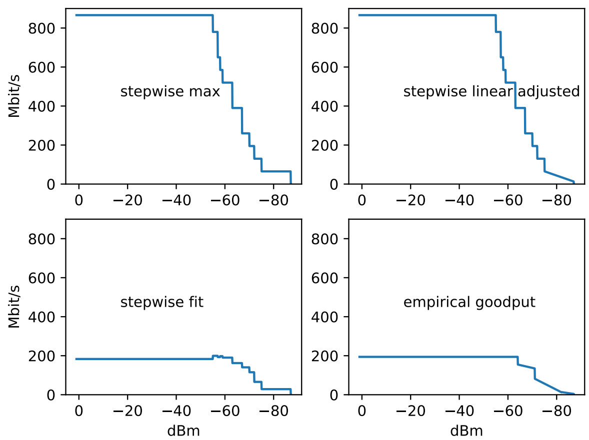


Fig. 4. Modulation schemes considered in this paper. They all have a null throughput under -87 dBm, being the minimum sensitivity of our receiver.

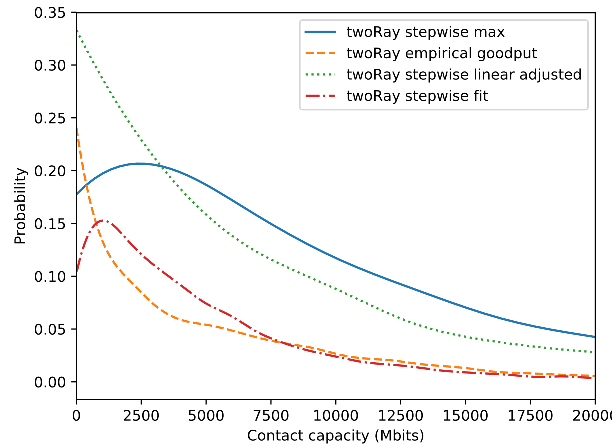


Fig. 5. Contact capacity probability density function for all modulations using a two-ray propagation model in Stockholm.

red dots show the empirical throughput we observed during our experiments. The median, shown as a blue line in the plot, represents the expected empirical throughput for a given RSSI. We refer to this median as the "empirical goodput" since it represents the observed transfer capacity of off-the-shelf devices (OnePlus 5T devices in this case).

We can make two main observations. Firstly, the theoretical data rate is too optimistic, especially for the highest RSSI values, due to hardware limitations as well as implementation choices in the Nearby framework. Secondly, for the BPSK modulation, corresponding to the lowest RSSI levels (below -75 dBm), the experimental curve goes down to 0 Mbps at -87 dBm. As we will see later on, the fact that the theoretical curve remains horizontal at this very last modulation mode has a significant impact on the behavior of the model.

B. *Finding an appropriate modulation scheme*

To assess the effects of the modulation scheme, we investigate the probability density function of contact capacity for the four different modulation schemes illustrated in Figure 4:

- **Step-wise maximum.** This is the theoretical step-wise function taken from the Wi-Fi 5 specification.
- **Step-wise linear adjusted.** Almost the same as the step-wise maximum, except that the slowest modulation (BPSK) has its data-rate linearly decreasing until it reaches 0 at -87 dBm.
- **Step-wise fit.** In this scheme, we move all of the steps from the Wi-Fi 5 specifications to the same levels as the experimental data.
- **Empirical goodput.** For a given RSSI, the empirical goodput is the median throughput of collected samples over the RSSI (median blue curve in Figure 3). Similar to the step-wise fit, this modulation scheme requires heavy experimentation to be correctly estimated.

To evaluate the influence of these modulation schemes on the capacity of a contact, we first need a mobility scenario. In this section, we consider the Ostermalm dataset [4], a pedestrian mobility trace in the city of Stockholm. The trace was generated with Legion Studio [8], a mobility simulator meant for architects and designers to test out pedestrian flows in large infrastructures. The trace has a duration of approximately five hours, with a total of 2,400 nodes moving within a 5,872 m^2 area. The dataset covers a period of 5 hours and shows coordinate updates every 0.6 seconds. This high frequency of position updates is necessary as we need an accurate estimation of the distance between the nodes. The population does not vary significantly throughout the trace, holding a constant value around 60 nodes.

We also need a propagation model to transform distances into RSSI (recall Figure 2 and Equation 2). We consider the two-ray propagation model as it is the one that best fits our experimental scenario (we will discuss about the propagation models in Section IV-A). The results are shown in Figure 5.

Two patterns emerge. While the step-wise maximum and the step-wise linear curves show a humped behavior, the step-wise linear adjusted is the only one that succeeds to capture the behavior of the empirical data. Even though the step-wise linear adjusted overestimates the overall capacity of the system when compared to the empirical goodput, the most important observation here is the shape of the curve, as in a real-world scenario, we expect to observe lower capacity anyway due to interference from neighboring communications.

The main takeaway from this experimental evaluation is the necessity to take the effect of the modulation scheme into account when estimating the adaptive contact capacity. Most importantly, one has to model this carefully, to take into account the asymptotic behavior of throughput according to RSSI and, by extension, throughput according to distance.

C. *Explaining the step-wise linear adjusted scheme*

As we could see in the previous subsection, the step-wise linear adjusted scheme was the only one able to capture the

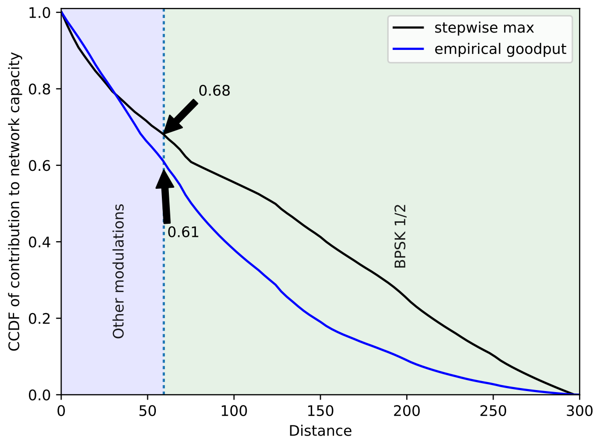


Fig. 6. CCDF of the total capacity of the Stockholm dataset according to the distance between nodes. For the adaptive contact capacity, we here chose a two-ray propagation model and plot the capacity using two modulation schemes. Regardless of the modulation scheme, the majority of the network capacity falls under BPSK 1/2.

behavior of the experimental results, albeit a small difference compared to the step-wise maximum. Our intuition is that most contacts happen at the edge of the communication range, thus using the lowest throughput modulation (in the case of Wi-Fi 5, BPSK 1/2).

We check which of short-range or long-range contacts contribute the most, by evaluating the contact capacity of a mobility trace. To this end, we consider the two-ray ground model for the propagation. For the modulation schemes, we first assess the suitability of the step-wise maximum and the empirical goodput. In Figure 6, we show the CCDF of the total capacity of the network as a function of the distance between nodes. As we can see in this figure, regardless of the idealistic (step-wise max) or realistic (empirical goodput) modulation scheme, between 68% to 61% of the total capacity of the network happens due to communications occurring at more than 60 m, which shows the importance of long-distance communications and thus of BPSK modulation. As a result, the modulation scheme used over long distances should not be neglected. The use of a realistic modulation scheme, where the throughput decreases with distance until reaching zero, is recommended.

Part of the explanation also comes from the fact that, in the case of freespace and two-ray ground reflection models, the area “covered” by the BPSK modulation is much bigger than those covered by the other modulation schemes. In Figure 7, we illustrate the differences between the areas. The bigger the area, the more neighbors are likely to communicate through longer links. Other propagation models (especially those adapted for indoors) would lead to different observations; for this reason, we will have further insights into this problem in Section IV-A.

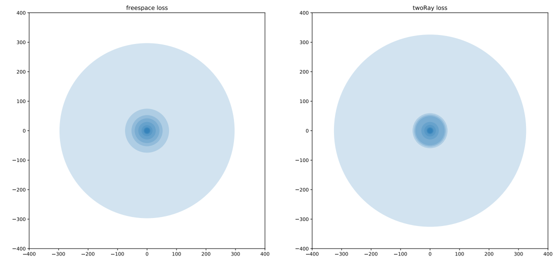


Fig. 7. Propagation and modulation on the space. The outer disk represents the coverage zone of BPSK 1/2.

IV. Fixed vs. adaptive contact characterization

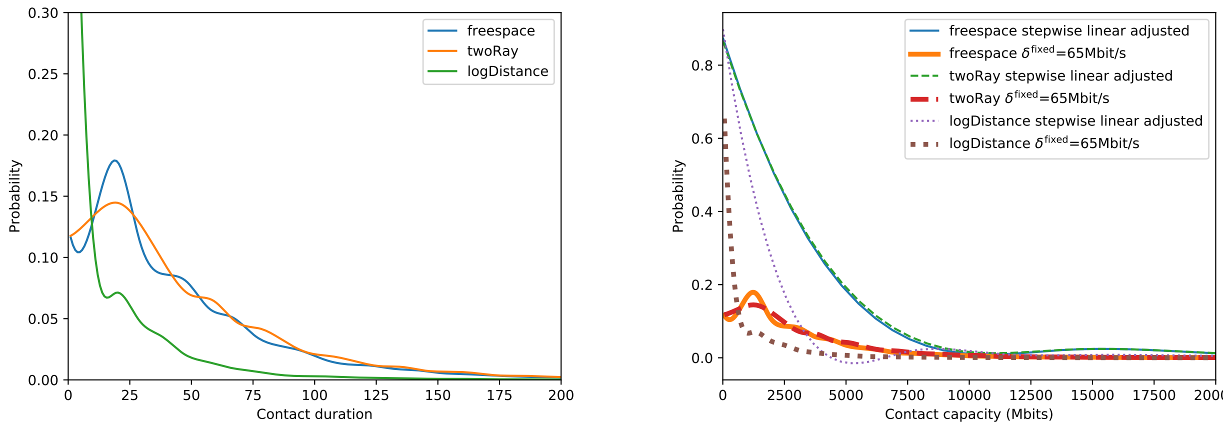
Let us now compare the capacity of contacts when they are modeled by the fixed-rate and adaptive approaches. For all contacts, we compute the capacity by applying Equations 1 and 2, respectively. To showcase the generalization of the upcoming observations, we consider two datasets, the aforementioned pedestrian Ostermalm dataset (presented in Section III-B) and the Luxembourg dataset. Luxembourg LuST [9] is a vehicular mobility trace created using SUMO [10], a state-of-the-art micro-mobility simulator. The simulation runs over 24 hours and covers a 20 km^2 area and a total of 167,000 nodes. The trace simulates a realistic daily traffic pattern, with peaks of population during the morning and evening with approximately 2,500 simultaneous nodes.

A. Influence of propagation model

We previously investigated the impact of modulation schemes on the contact capacity, and we now propose to generalize our observations using different propagation models. We consider three propagation models: freespace [11], two-ray ground [12], and log-distance [11].

As previously explained, the minimum sensitivity of our devices is -87 dBm. As the propagation model dictates the maximum communication range, we consider contact duration along with contact capacity to justify the distributions. For the sake of reference, in our work, the maximum communication range using freespace and two-ray propagation is ≈ 300 m, and for the log-distance ≈ 45 m. We set the modulation as the step-wise linear adjusted for the rest of this section. Lastly, we also plot the contact capacity for a fixed-rate contact $\delta^{\text{fixed}} = 65$ Mbps (this throughput being the same data rate as the BPSK 1/2 for Wi-Fi 5).

Let us first consider the vehicular case, shown in Figure 8. The contact duration distribution, as seen in Figure 8a, exhibits two different behaviors. The freespace and two-ray propagation models behave similarly, with a mean contact duration around 30 seconds, whereas the log-distance model leads to a higher probability for contact durations closer to zero. This is a consequence of few contacts happening at a short distance (small coverage surface); even when short-distance contacts



(a) Contact duration distribution.

(b) Contact capacity distribution.

Fig. 8. Contact duration and contact capacity distribution on a vehicular dataset (Luxembourg).

happen, the velocity of vehicles prevent the contact from lasting longer than a few seconds.

When observing the contact capacity distribution (Figure 8b), as expected, the fixed-rate curves reflect pretty well the distribution of contact duration of Figure 8a. What is eye-catching in this figure is the behavior of the adaptive capacity when using the log-distance propagation model. Contrarily to freespace and two-ray propagation models, log-distance is much more restrictive and intended to reflect indoor scenarios. This means that the coverage zone of BPSK 1/2 is much smaller compared to the other modulations. In this way, the ratio of neighbors communicating through low-throughput links decreases significantly.

In the pedestrian case (Figure 9), contact duration distribution shows a more spread behavior than Luxembourg's. This is due to the shorter relative speeds between human beings compared with the vehicular case which leads to longer contact duration. As such, this yields a more spread out distribution for the freespace and two-ray models and a completely different shape for the log-distance model. In fact, for the log-distance model, the highest contact duration probability is found between 25 to 30 seconds.

The distribution of contact capacities in the Stockholm scenario, depicted in Figure 9b, has several similarities with the plots of the Luxembourg case – the fixed-rate plots follow the same shape as the distribution of contact duration, while the adaptive plot looks like that of a decreasing exponential, except for the log-distance.

The unusual shape the log-distance with a step-wise linear adjusted modulation scheme contact capacity distribution is explained by carefully observing the behavior of nodes, which act as pedestrians. Indeed, due to the much shorter communication range in the case of log-distance, there are only two sorts of contacts. Either the contact is very short, leading to poor capacity, or the contact is long with little distance between the two nodes. These two sorts of behaviors are a direct consequence of the pedestrian mobility behavior with short communication range; either pedestrians cross each other for a brief period, or two pedestrian walk next to each other for a longer duration thus yielding higher capacity.

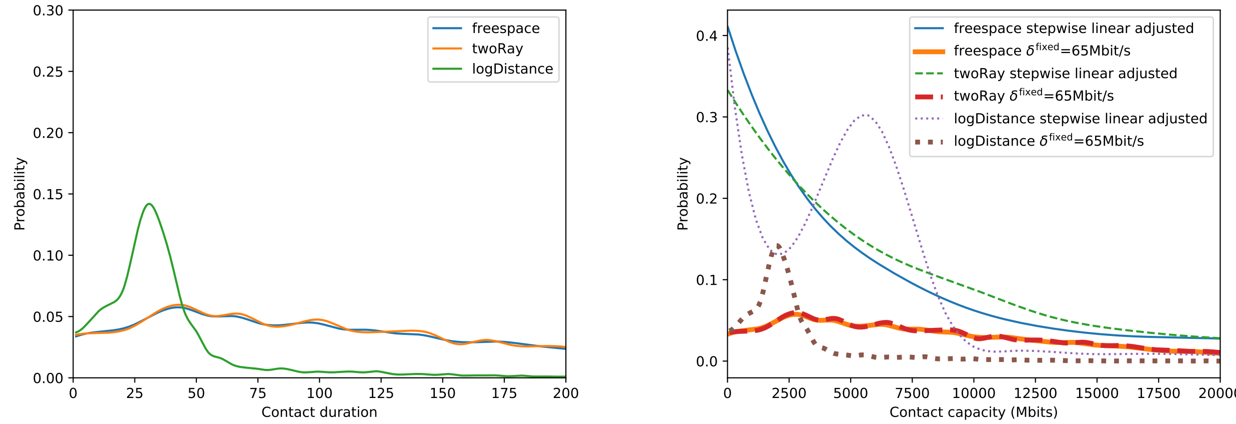
We also confirm once again that the fixed-rate approach exhibits an unrealistic behavior, and we conclude from our results that designers willing to set up an opportunistic network based on device-to-device communications should account for the type of modulation the underlying technology relies on. The influence of the environment, represented by the propagation model, is limited compared to the modulation scheme. By adopting this strategy, they may be able to obtain much more accurate results compared to what they would obtain with the fixed-rate characterization.

B. Picking a better fixed-rate value

If for some reason, one still prefers to use a fixed-rate characterization, she ought to better select δ^{fixed}. To select a better fixed-rate contact throughput value, we compute the vertical least squares fitting curve based on a simple linear equation for each of the distributions as shown in Figure 10. We remind the reader that in the case of contact capacity as a function of contact duration, a fixed-rate contact follows the linear equation $\delta^{\text{fixed}} = ax + b = ax$, since b is always equal to 0; a contact of 0 second yielding 0 contact capacity.

The boxplots in Figure 10 represent the distribution of contact capacity for a given contact duration, considering an adaptive contact. With each box having a height corresponding to 50% of the data, and the bottom (resp. top) whiskers corresponding to the 5th (resp. 95th) percentile. The curve with triangle symbols holds the arithmetic mean of all values found inside the bin.

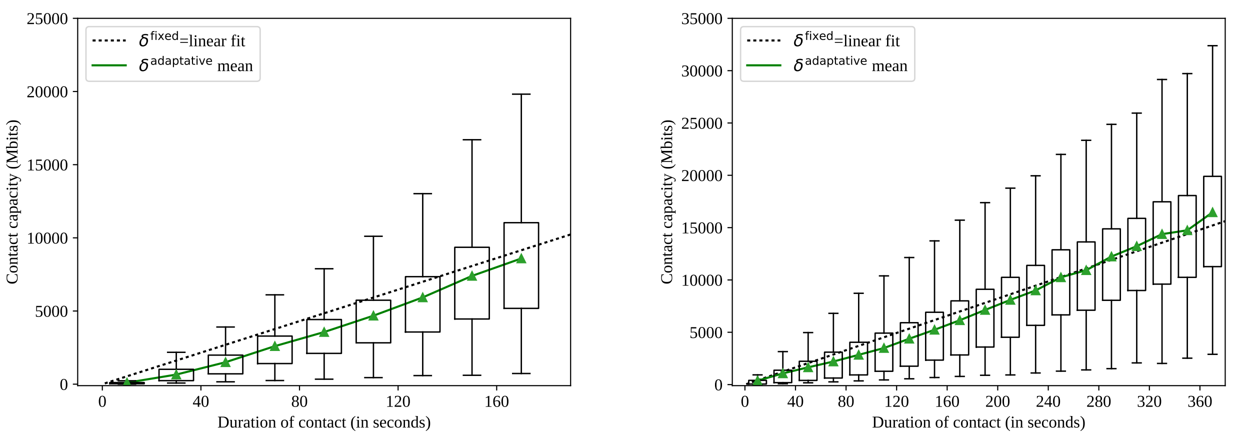
When focusing on the vehicular case Figure 10a, we can see that the linear fit for the fixed-rate contact does not seem like a good match. For contacts of less than 40 seconds



(a) Contact duration distribution.

(b) Contact capacity distribution.

Fig. 9. Contact duration and contact capacity distribution on a pedestrian dataset (Stockholm).



(a) Luxembourg

(b) Stockholm

Fig. 10. Contact capacity as a function of contact duration.

long, the spread around the empirical mean is quite small, but the contact capacity growth between one bin to the next is less than linear. This shows that most contacts under 40 seconds happen on the edge of the node communication range, where the data rates are the slowest. Knowing from the distribution of contact times in section IV that the majority of contacts in Luxembourg happen under 40 seconds, these edge-communications concern most contacts. With this in mind, the closest recommendation for the value of a fixed-rate contact for Luxembourg we can make is 53.84 Mbits/s.

For the Stockholm case, as represented on Figure 10b, the linear fitted function seems to be a better match. Since the surface in which nodes evolve inside the Stockholm dataset is significantly inferior to the Luxembourg one, making the average contact about 80 seconds long. Our recommendation for the value of a fixed-rate contact for Stockholm is 41.10 Mbits/s.

Overall, having a fixed-rate contact is more realistic in the pedestrian case than the vehicular case. We, however, remind the reader that the significant spread around the means shows fixed-rate contacts are far from being the optimal way to portray realistic contact capacity.

V. Contact Network Capacity Computation Tool

As a contribution to the community, we propose an open-source software to quantify the contact capacity of a mobility trace.[2] The software, implemented as a Python library, proposes several preset models for the propagation model and modulation scheme to meet the requirements of the user. As an incentive for users to fine-tune the software to their specific

[2]https://github.com/Bertier/OpportunistiKapacity

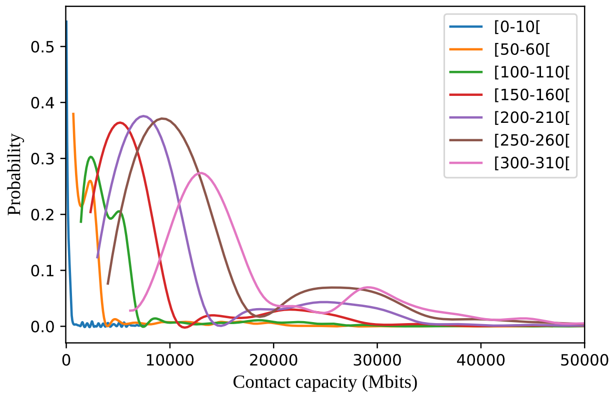


Fig. 11. Contact capacity probability according to contact duration in the Stockholm scenario. Here, the two-ray propagation and step-wise linear adjusted modulation scheme is used.

needs, we provide access to lower-level functions in case one wishes to implement customized models. The tool can handle both mobility and contact traces.

A. Mobility traces

For this type of trace, the computation is straightforward. The software takes a mobility trace as input (i.e., a sequence of GPS coordinates) and calculates the distance between nodes at each snapshot to determine if there is a contact or not. A contact exists if the distance between the nodes is below the threshold set by the model. If a contact is found, it computes the capacity using the throughput estimation based on the distance between the two nodes, until they lose connectivity.

B. Duration-based contact capacity estimation

Contact traces offer the duration of a contact but not the geographical coordinates of the nodes. Thus, we cannot handle these traces as we do with mobility traces. Recall that, in Section II, we presented Figure 1 that clearly showed that contacts of the same duration are likely to hide different mobility patterns. To circumvent this issue, we propose to rely on probability density functions that relate duration to capacity.

Let us illustrate this idea with an example. Let us assume the user chooses the two-ray propagation model along with the step-wise linear adjusted modulation scheme. Firstly, we calculate the contact capacity for each contact duration, as seen in Figure 11. This figure comes from the measured data we showed in Section III, and we expect to extend our reference dataset with other measurement campaigns in the future. The software computes then a contact capacity based on the duration of the contact directly from the appropriate probability density function.

VI. Related work

In the literature, fixing a data exchange rate for a contact has been an on-going issue for more than a decade. While it depends on the wireless technology used, as well as the mobility scenario and the environment, most often these issues are entirely overlooked.

Delay tolerant networks (DTN), or more specifically DTN routing/diffusion algorithms, is a field where contacts and inter-contact times are used as a way to quantify exchange opportunities [13]–[16]. Some algorithms focus on the fact that nodes (i.e., transmitting or receiving devices) have a finite amount of memory that limits their receiving/forwarding capabilities and forces fine buffer management [14], [16]. While it is a common practice to take in consideration hardware limitation due to memory, we are surprised to see that transfer speed is ignored. In other words, it is commonly assumed that the packet transfer between two nodes always succeeds if two nodes are within communication range (though, these algorithms assume small packet size).

Let us pick the example of the Opportunistic Network Environment simulator (The ONE), a state-of-the-art simulator aiming to simplify the evaluation of DTN algorithms in a realistic fashion [17]. For the throughput of mobile devices, they chose a conservative and simplistic estimate by setting a fixed-rate throughput for contacts. For instance, they assume that nodes using Bluetooth can exchange data at a 2 Mbps throughput within a 10 m radius and that nodes using Wi-Fi have a 4.5 Mbps throughput within a 30 m radius [17]. A recent example of such a practice is proposed by Zhu *et al.* [18], where a contact happens with any vehicle inside a 200 m range with an exchange rate of 20 Mbps. Our study shows that such a characterization is not very precise and that distance between communicating nodes should be taken into account to model these interactions accurately.

Fortunately, some authors shed light onto this issue. The closest work to our study has investigated the idea of tuning throughput according to the distance. For instance, Chowdhury *et al.* [5] observed that the Wi-Fi protocol adapted its rate according to the RSSI, and therefore proposed to calculate the amount of transferable data between a Wi-Fi access point and a mobile node according to the distance between them. Neto *et al.* [6] take this idea and empirically try to verify this model by measuring the throughput between a mobile node (car) and a base station. Last, Qayyum *et al.* [3] take an empirical measurement of the throughput according to distance using a mobile application.

Contrary to the approaches mentioned above, in this paper, we consider several new aspects. First, instead of assuming only one of the two nodes is mobile, which is unrealistic in a pedestrian or vehicular environment, we suppose that both nodes in contact are mobile. Moreover, where the other studies simulate random positions of nodes or real traces with a small number of nodes, we use realistic traces with hundreds, even thousands, of mobile nodes to understand how the adaptive rate affects the contact capacity in a large network.

Another essential difference between our study and the previously mentioned works is the method used to estimate the throughput between two devices according to the distance separating them. Chowdhury *et al.* [5] took the theoretical link speed as specified in the Wi-Fi 4 (IEEE 802.11n) standard, and tried to fit a piece-wise and polynomial function to obtain the link speed from the RSSI. Then, they estimate the RSSI according to distance by using a log-distance propagation model. Qayyum *et al.* [3] first empirically measure the distance to throughput by implementing a Bluetooth data transfer scheme in a mobile application, and then propose a distance to throughput model using cubic spline interpolation. Compared to our model, they report a notably lower throughput, and in their case, shorter range ($\approx$10 m). While detailed in a companion paper, our model takes the best of these two approaches, by using a mobile application to measure the RSSI to throughput on contemporary hardware (notably, using IEEE 802.11ac instead of IEEE 802.11n) and by using an empirically verified model for RSSI estimation.

VII. Conclusion

In this paper, we investigated contact characterization based on the throughput between the nodes, namely fixed and adaptive. Our adaptive contact characterization relies on the principle that one should take the distance between two nodes into account to properly estimate the throughput between them. The adaptive contact characterization adopts a two-step calculation strategy: firstly, the distance is turned into a received signal strength (propagation model), then the signal strength is turned into a throughput (modulation scheme).

We applied our strategy in two mobility datasets. We observed that the most important factor when capturing the behavior of a contact is that the modulation scheme's throughput must decrease with distance until reaching zero. Also, although distant contacts have the poorest data-rate, they contribute the most in terms of global network capacity regardless of optimistic or realistic modulation schemes. By comparing the fixed and adaptive contact capacity distributions, we noticed that their distribution shapes are entirely different, thus showcasing that fixed-rate contacts are not enough to capture the essence of realistic contact capacities.

We additionally provide a piece of software that computes the contact capacity of a mobility trace, either based on the distance between the nodes or, if the latter is not available, on the contact duration with greater accuracy than fixed-rate contacts. With this tool as a means to simplify the contact capacity computation, we hope that the community will get further insights into how the varying contact throughputs impact protocols and algorithms for D2D networks.

In our future work, we intend to provide a mathematical framework to calculate the contact capacity based on the duration of contact. Additionally, we also intend to conduct more experimental campaigns so that our tool encompasses more scenarios (e.g., indoors).

References

[1] R. Ansari, C. Chrysostomou, S. Hassan, M. Guizani, S. Mumtaz, J. Rodriguez, and J. Rodrigues, "5G D2D networks: Techniques, challenges, and future prospects," *IEEE Systems Journal*, vol. 12, no. 4, pp. 3970 – 3984, Dec. 2018.

[2] F. Jameel, Z. Hamid, F. Jabeen, S. Zeadally, and M. A. Javed, "A survey of device-to-device communications: Research issues and challenges," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2133–2168, thirdquarter 2018.

[3] S. Qayyum, M. Shahriar, M. Kumar, and S. K. Das, "PCV: Predicting contact volume for reliable and efficient data transfers in opportunistic networks," *Proceedings - Conference on Local Computer Networks, LCN*, pp. 801–809, 2013.

[4] Ó. Helgason, S. T. Kouyoumdjieva, and G. Karlsson, "Opportunistic communication and human mobility," *IEEE Transactions on Mobile Computing*, vol. 13, no. 7, pp. 1597–1610, 2014.

[5] H. Chowdhury, J. Lehtomäki, J.-P. Mäkelä, and S. Kota, "Data Downloading on the Sparse Coverage-Based Wireless Networks," *Journal of Electrical and Computer Engineering*, vol. 2010, pp. 1–7, 2010. [Online]. Available: http://www.hindawi.com/journals/jece/2010/843272/

[6] J. B. P. Neto, E. Mota, P. Almeida, and R. Rojas, "A model for contact volume prediction in DTNs," in *2012 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, jul 2012, pp. 000 199–000 202. [Online]. Available: http://ieeexplore.ieee.org/document/6249293/

[7] Google, "Nearby - Connections API Overview." [Online]. Available: https://developers.google.com/nearby/connections/overview

[8] Bentley Systems Incorporated, "Legion Software," 2018. [Online]. Available: http://www.legion.com/legion-software

[9] L. Codeca, R. Frank, and T. Engel, "Luxembourg SUMO Traffic (LuST) Scenario: 24 hours of mobility for vehicular networking research," in *2015 IEEE Vehicular Networking Conference (VNC)*, vol. 2016-Janua. IEEE, dec 2015, pp. 1–8. [Online]. Available: http://ieeexplore.ieee.org/document/7385539/

[10] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "SUMO Simulation of Urban MObility," *Iaria*, no. c, pp. 55–60, 2011.

[11] K. Benkic, M. Malajner, P. Planinsic, and Z. Cucej, "Using RSSI value for distance estimation in wireless sensor networks based on ZigBee," *Proceedings of the 15th International Conference on Systems, Signals and Image Processing*, pp. 303–306, 2008. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4604427

[12] C. Sommer, S. Joerer, and F. Dressler, "On the applicability of Two-Ray path loss models for vehicular network simulation," in *2012 IEEE Vehicular Networking Conference (VNC)*. IEEE, nov 2012, pp. 64–69. [Online]. Available: http://ieeexplore.ieee.org/document/6407446/

[13] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 7, no. 3, p. 19, jul 2003. [Online]. Available: http://doi.acm.org/10.1145/961268.961272http://portal.acm.org/citation.cfm?doid=961268.961272

[14] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine, "MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks," in *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, vol. 00, no. c. IEEE, 2006, pp. 1–11. [Online]. Available: http://ieeexplore.ieee.org/document/4146881/

[15] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant MANETs," in *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing - MobiHoc '07*. New York, New York, USA: ACM Press, 2007, p. 32. [Online]. Available: http://dl.acm.org/citation.cfm?id=1288107.1288113http://portal.acm.org/citation.cfm?doid=1288107.1288113

[16] A. Balasubramanian, B. Levine, and A. Venkataramani, "DTN routing as a resource allocation problem," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 4, p. 373, oct 2007. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1282427.1282422

[17] A. Keränen and J. Ott, "Increasing reality for DTN protocol simulations," *Helsinki University of Technology, Tech. Rep*, 2007. [Online]. Available: http://www.netlab.tkk.fi/{~}jo/papers/2007-ONE-DTN-mobility-simulator.pdf

[18] X. Zhu, Y. Li, D. Jin, and J. Lu, "Contact-Aware Optimal Resource Allocation for Mobile Data Offloading in Opportunistic Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7384–7399, aug 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7852453/