



D2D Data Offloading in Vehicular Networks with Delivery Time Selection

Loreto Pescosolido^(✉), Marco Conti, and Andrea Passarella

Institute for Informatics and Telematics (IIT),
Italian National Research Council (CNR), Pisa, Italy
{loreto.pescosolido,marco.conti,andrea.passarella}@iit.cnr.it

Abstract. Within the framework of a Device-to-Device (D2D) data offloading system for cellular networks, we propose a Content Delivery Management System (CDMS) in which the instant for transmitting a content to a requesting node, through a D2D communication, is selected to minimize the energy consumption required for transmission. The proposed system is particularly fit to highly dynamic scenarios, such as vehicular networks, where the network topology changes at a rate which is comparable with the order of magnitude of the delay tolerance. Through extensive system level simulations, we compare the energy consumed by the devices to perform D2D data offloading using the proposed scheme with the energy consumed when using a benchmark scheme (proposed in previous works) without optimal transmission instant selection. The results show that, in specific scenarios, compared to the benchmark system in which the transmission instant is not optimized, the proposed system allows a reduction of the energy consumed for D2D communications above 90%.

Keywords: D2D data offloading · Power control
Delay-tolerant applications · Radio resource management

1 Introduction

Device-to-Device (D2D) data offloading in cellular networks [1] is a powerful means to decrease congestion at the base stations, reduce the energy consumption of the overall system, and increase spectral efficiency. The idea is that, whenever a content is requested by a node, if the content is available at any of its neighbors, it must be obtained from it, rather than from the network infrastructure. We indicate the nodes that can potentially hand the desired content to a requesting node as *potential content providers*. The set of potential content providers, for a given content request, depends on scenario parameters, like the node density and the content popularity, and on the specific protocol design. For delay-tolerant applications, an interesting protocol design option is that, in case a node issuing a content request has no potential content provider in its

neighborhood at the time of request, it waits for a predefined interval, called *content timeout*, within which it is still possible to obtain the content from a new neighbor, encountered in the meantime [2,3]. Only at the expiration of the content timeout, if the content has not yet been obtained, it is transmitted by the infrastructure nodes through an Infrastructure-to-Device (I2D) transmission. This approach is particularly effective in highly dynamic scenarios, such as vehicular networks, where the network topology changes at a fast rate. The use of a content timeout allows to increase the population of potential content providers beyond the set of the requesting node's neighbors at the time of request, extending such population to the nodes that will become its neighbors in the future. In this way, the system obtains an increase of the offloading efficiency, defined as the percentage of contents delivered by using D2D communications between peer nodes (vehicles), rather than using I2D transmissions from the infrastructure nodes, bringing along the aforementioned system-level benefits.

In our prior works [4,5], we have shown that the considered type of D2D data offloading protocols for delay-tolerant applications, are very effective in reducing the overall energy consumption at system level. More in general, this is true for most D2D data offloading protocols, especially when power control is in use. However, there is still significant room for improving the performance of D2D transmissions, by taking *full* advantage of the delay tolerance of requests.

Consider two nodes, and let us define them as neighbors if and only if their distance is less than or equal to a (nominal) maximum transmission range d_{\max} . In previous works that follow the above described approach, *in the case at the time of a content request there are no potential content providers within a range d_{\max} from the requesting node* (i.e., no neighbor has the requested content in its cache), as soon as the requesting node encounters a potential content provider, the content is transmitted. In this case, it is clear that the transmission takes place at the maximum transmission range of the devices. Therefore, in a system with distance-based power control, all the requests that are not fulfilled at the time of request, inherently require the use of the maximum D2D transmit power. Furthermore, *in the opposite case, in which at the time of request there is already a potential content provider, say at distance $d < d_{\max}$* , the content delivery requires a transmit power that is higher than what would be required if the delivery was postponed to a later instant, at which the involved (or any other) content provider would be closer than d to the requesting node.

Motivated by this observation, in this work we propose the following approach. When a new request arrives, a controller, running, e.g., at the eNodeB (eNB), exploits knowledge of nodes positions and predicted movements in the near future (in the following content timeout window), to estimate which potential content providers will be in radio range of the requesting node within the content timeout. The content transmission is scheduled with the potential content provider that is predicted to be at the minimum distance from the requesting node, at the point in time when this will happen. In this way, provided that a distance-dependent transmit power control is in use, the smallest possible transmit power will be required. We will show that using this approach the energy

consumption of the considered protocol for delay-tolerant application can be dramatically reduced.

The paper is organized as follows. We position our work with respect to the recent research trends in this area in Sect. 2. In Sect. 3 we describe our system set-up and a possible MAC (adapted from an existing solution) for an in-band implementation of the proposed D2D Data offloading scheme. In Sect. 4 we present the proposed Content Delivery Management System (CDMS). In Sect. 5 we evaluate the performance of the proposed system through the analysis of system-level simulations results, in terms of the average energy consumption per content delivery required to satisfy a given system-wise traffic demand. Finally, Sect. 6 concludes the paper summarizing our contribution.

2 Related Work

The use of D2D communications to offload traffic from infrastructure nodes has been investigated for years by the researchers of different communities. Works like [6, 7] aim at investigating scaling laws and network throughput from a fundamental limits perspective. Works like [8, 9] (amongst many others), aim at devising radio resource allocation strategies, and/or other physical layer parameters, like coding rates and transmit power levels, assuming that D2D and/or I2D links to be accommodated are given as an input to the problem. More specific protocol-oriented works have been appearing in the last years as well. The interested reader may want to check, e.g., [1] for an extensive survey. In these works, the objective is to determine and schedule I2D and D2D offloading communications as a function of the request patterns (as opposed to the above mentioned works, in which the link to schedule are an input to the problem). In [2], the peculiarity of D2D data offloading for delay-tolerant applications was first addressed. In [3], a CDMS for contents originated from delay-tolerant applications, suited to a vehicular network scenario, was proposed. In [10], in a scenario in which content delivery mostly relies on D2D-offloading, a strategy for I2D re-injection of contents in the network is proposed to mitigate the effect of temporal content starving in a certain areas. In [11], in the framework of a content dissemination problem (i.e., when contents need to reach all the nodes, without having been explicitly requested), the authors propose a mixed I2D-multicast and D2D-relaying reinforcement-learning-based strategy, which determines which users should receive the contents through a direct I2D transmission or through D2D relaying from a neighboring device. The above mentioned works, although providing interesting insights from the perspective of offloading efficiency maximization, devote less attention to performance metrics which are closer to physical quantities, like energy consumption and spectrum efficiency. Our work is motivated by the need to take into account such metrics in the system design, and optimize it to maximize them. With the focus on energy consumption minimization, we have been working on a CDMS inspired by [3], and evaluated analytically its performance in [5]. The work in [5] investigates the effect of content popularity and vehicles speed on the D2D transmit power

by proposing and validating an analytical model for computing the offloading efficiency and the energy consumption, and then selecting the best value for the maximum D2D transmission range. The CDMS considered in [5] does not optimize the D2D transmission time, letting the nodes transmit a requested content as soon as they encounter a node requesting it. In this work, differently from the above mentioned ones, we leverage the degree of freedom entailed by delay tolerance by deferring the D2D transmission instant to the time it will require the lowest power, thus achieving impressive performance gains in terms of energy consumption. Finally, we deem it appropriate to take into account accurate channel models, since using relatively simplistic models may result in an inaccurate estimation of the performance gain of a particular design [4]. Furthermore, we consider it necessary, when dealing with the type of performance metrics discussed above, to integrate in the performance evaluation an actual radio resource management technique. Among the many available, we found it useful to implement the solution proposed in [9].

3 System Model

3.1 Network Topology, Node Mobility, and Content Requests

We consider a Region of Interest (ROI) consisting of a bidirectional street chunk which vehicle enter, traverse, and exit from both ends, as shown in Fig. 1.

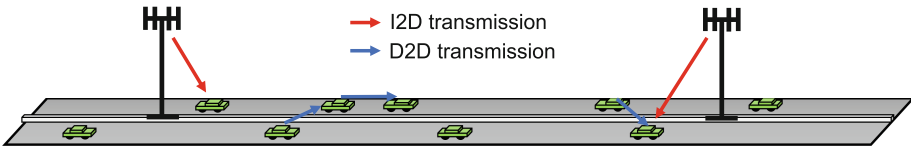


Fig. 1. Sketch of the considered scenario.

Vehicles enter the street according to a given stationary temporal arrival process, with an average arrival rate of λ_t vehicles per second ($\lambda_t/2$ vehicles per second on each end). Each vehicle n traverses the ROI at an average speed v_n , selected as a random variable with probability density function $p_V(v)$. Particularly, we select $p_V(v)$ as a uniform random variable in a limited speed range $[v_{\min}, v_{\max}]$. Each vehicle has onboard a mobile device, which can be either a human hand-held device or part of the vehicle equipment. Along the street, a set of eNBs is regularly placed. At each instant, each device (vehicle) is under the coverage of an eNB. Each device issues content requests according to a given stationary content request process with an average content request rate of λ_Z requests per second, by sending requests messages to the eNB it is associated to at the time of request. The specific content being requested is drawn according to a content popularity distribution $p_Z(z)$. Similarly to [2, 3], we assume that the content requests can be fulfilled with some delay tolerance, i.e., they must be

served at most within a content timeout τ_c . A request may be fulfilled either by a potential content provider, through a D2D communication, or by an eNB. In this work we assume that, within the content timeout, the first option (delivery through D2D) is always privileged, and I2D content deliveries are performed only at the end of the content timeout, if it has expired before any potential content provider has been found. The rationale of this is that, in this way, we maximize the advantage of D2D transmissions in *offloading* traffic from the cellular infrastructure, which is one of the primary goals of any offloading system. A final important notion to introduce is that of the *sharing timeout*, τ_s . This is defined as the time each device keeps a received content in its cache, available for other nodes that it encounters, and after which, to avoid an indefinite increase of the cache occupation, it removes.

The above assumptions are quite general. It is clear that, for the purpose of performance evaluation, specific models need to be assumed for all the involved random processes. For the sake of readability, we leave the description of the specific assumptions used for our performance evaluation to Sect. 5.

3.2 MAC and Physical Layer Implementation

The purpose of this work, besides the presentation of the proposed CDMS, is also the evaluation of its performance with realistic medium access control and radio resource management strategies, also including physical layer aspects such as multipath frequency selective fading of the radio channels. We focus on in-band D2D offloading, in which D2D communications and I2D ones share the same spectrum. We have considered a LTE-like multi-carrier system in which the radio resources are organized as a time-frequency grid of Physical Resource Blocks (PRBs) of fixed bandwidth and duration. We have selected a very flexible Radio Resource Reuse (RRR) scheme, which allows concurrent D2D and I2D communications. Particularly, we have implemented, with some modifications, the resource-sharing oriented scheme recently proposed in [9]. However, we adapted the algorithms in [9] to use different transmit power levels across concurrent links, and included multiple eNBs and spatial frequency reuse for I2D communications (besides D2D ones) in the design. Additionally, it is worth mentioning that the solution proposed in [9] is evaluated under a flat fading channel assumption, whereas our implementation includes frequency selective channels. Further details on the considered channel model are provided in Sect. 5.

Time is organized in control intervals. In each control interval, a set of I2D and D2D links have to be scheduled for transmission. The set of I2D and D2D links to schedule in each control interval is determined by the CDMS according to the procedure described in Sect. 4. Radio Resource allocation is performed by a RRR agent residing at the eNBs. We assume that the position of each device is known to the RRR agent, and hence, it can compute the distance between any node pair.

First, the RRR agent, taking in input the distance d between transmitter and receiver of each link to be scheduled, computes the transmit power of each link. This is done by inverting an expression for the nominal per-subcarrier capacity

C , of the kind $C = \log_2(1 + P_c g(d)/\sigma_c^2)$, with respect to the per-subcarrier transmit power P_c . In the above expression, σ_c^2 is the thermal noise power on each subcarrier¹ and $g(d)$ is a *nominal* (deterministic) path loss expression², which is selected by the system according to the deployment scenario (urban, dense-urban, rural, micro-cell, macro-cell, I2D, D2D, etc.). The so obtained value of P_c is increased by a suitable link margin M , which is required to compensate the presence of random fluctuations and frequency selectivity, and of the interference that may come from concurrently transmitting links. Note that both these effects are taken into account in the implementation we used in our performance evaluation.

Second, the set of links is partitioned³ into RRR sets in order to satisfy a set of cross-interference mitigation constraints. The constraints are computed using an estimation of the interference across links obtained by computing the nominal channel gain g between any link transmitter and any link receiver among the set of links to be scheduled. A suitable amount of PRBs is assigned to each RRR set. This amount is a function of the number and size of the contents that have to be transmitted by each link in the RRR set. D2D links in the same RRR set can use the same radio resources, since their belonging to the same set stands for the fact that their cross-interference is sufficiently low not to compromise the communications. I2D links originating from the same eNB are assigned radio resources in an exclusive way, selected as a portion of the pool of PRBs assigned to the RRR set they have been included in. In its portion of PRBs, however, each I2D link is subject to the interference coming from the D2D links included in the same RRR set. Finally, I2D links originating from different eNBs, that are included in the same RRR set, can be assigned the same portion of PRBs within the pool of PRBs assigned to that RRR set. If the RRR set partitioning and consequent PRBs allocation to each RRR set, due to the cross-interference constraints and to the limited number of PRBs in a control interval, prevent to accommodate the transmission of all the data required by any of the links, the data to be transmitted are pruned until reaching a feasible amount. The pruned transmissions will be rescheduled in the next control interval. Pruning is performed giving a higher scheduling priority to content deliveries related to requests whose content timeout is closer to expire. Therefore, I2D communications have a higher priority than D2D ones, since they are by design related to content requests whose timeout has already expired. If, due to pruning, the content timeout of any content request expires, the corresponding delivery is redirected to be performed by an eNB.

¹ σ_c^2 is computed as the thermal noise power spectral density $N_0 = -174$ dBm/Hz, plus a 10 dB receiver noise figure, times the subcarrier bandwidth.

² The function $g(d)$ is one of the components of our channel model, the other component being random frequency selective fluctuations added on top of $g(d)$, see Sect. 5. In this work, we assume that $g(d)$ is computed according to Eqs. 5-4, 5-5, and 5-6 in [12]. Interestingly, the considered channel model distinguishes between I2D and D2D channels, using different path loss functions in the two cases. This is one of the main reasons to opt for this model.

³ The RRR set partitioning algorithm is similar to [9, Algorithm 1].

A more detailed description of the considered RRR scheme and physical layer model is outside the scope of this work, and will be provided in a future work.

4 Content Delivery Management System with Optimized Delivery Time

In this work, we assume that the proposed protocol is executed under the supervision of an entity that we call CDMS. This is a distributed software agent under the control of the network operator. Most of its functions, explained in the following of this section, are executed at the eNBs. Whenever a content request is generated by a user, it reaches the CDMS, which is responsible for handling it from the time it is issued by a device until its fulfillment, deciding how and when the content request will be satisfied, either through D2D or through I2D communications. The basic functions of such a CDMS have been described in detail in our previous work [5] (see also [13]). The main contribution of this paper, thus, consists in a novel algorithm used by the CDMS to estimate the optimal transmission instant for D2D communications.

The CDMS acts on a distributed database containing the up-to-date list of each node's position and an estimation of their trajectories for the next τ_c seconds. Each device may obtain a running estimation of its speed and trajectory in the next seconds, either through the use of GPS or, if it is part of the vehicle electronic equipment, directly from the speedometer, and send it periodically to the eNBs. Alternatively, the devices can send the GPS information only to the eNB, leaving the burden of trajectory estimation to the CDMS. In general, different combinations are possible, whose details are outside the scope of this work. In this way, essentially, the CDMS has a picture of how the network topology will evolve in the next seconds⁴. Each device k has an internal content cache \mathcal{C}_k populated with previously downloaded contents. We assume that, at any time, the CDMS also has an index of the contents in each node's cache, and the instants at which each content will be removed from the node's cache due to the expiration of the associated sharing timeout. Each eNB keeps the above described information for all the nodes in its coverage and all the nodes in the adjacent eNBs cells. This is obtained by a periodic exchange of information among eNBs performed through high speed fiber connections, or dedicated radio channels, forming a mesh-type backhaul network. A high level abstraction of the CDMS architecture is depicted in Fig. 2.

On a coarse timescale, with respect to a given content request, the requesting node and the proposed CDMS act as follows. Upon receiving a request of content z from a node j , the eNB performs the following operations:

1. It compares the estimated trajectory of the requesting node j , for the next τ_c seconds, with those of all the nodes that have content z in their caches. In

⁴ In this work, we assume a perfect prediction of the vehicles' trajectory for an amount of time equal to the content timeout, leaving the evaluation of the robustness of the system with respect to trajectory prediction errors to a future work.

- practice, the set of potential content providers is limited by a maximum speed parameter v_{\max} , the content timeout τ_c , and the maximum D2D transmission range d_{\max} . These parameters are system parameters known to the CDMS and which determine the set of nodes that the requesting node j is supposed to encounter before the content timeout for the request of content z expires.
2. It truncates the estimated trajectory of each potential content provider with the instant its *sharing* timeout for content z will expire, if such instant comes before the expiration of the considered request *content* timeout, obtaining what we call “filtered” trajectories.
 3. On the basis of the filtered trajectories of all the potential content providers, it computes which potential content provider, say device k , will reach the shortest distance from device j , the value of such distance, say $d_{(j,k)}^{(z)}$, and the instant $t_{(j,k)}^{(z)}$ at which the two nodes are going to find themselves that close to each other.
 4. It schedules the transmission of content z from the potential provider k to the requesting node j at the time instant $t_{(j,k)}^{(z)}$.
 5. During the content timeout, the CDMS, with respect to the considered content request, keeps track of any device other than j and k which (i) is not included in the initial set of potential content providers and (ii) is supposed to encounter node j within the content timeout. If any such node receives the same content z during the content timeout, the closest distance it will reach from node j is computed, and if it is found to be shorter than $d_{(j,k)}^{(z)}$, the content delivery is rescheduled to be performed by the newly found potential provider, say k' , at the time it will supposedly be at the newly found shortest distance.

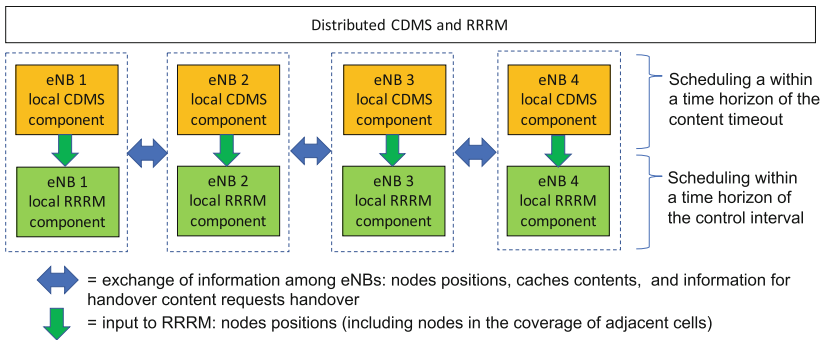


Fig. 2. High level abstraction of the distributed CDMS.

The requesting node, upon issuing a request, just waits for the content to be delivered to it. At the end of the content timeout, if it has not yet received the content, it will anyway receive it from the network infrastructure through an I2D transmission.

On a finer time scale, the operations described above are executed, in practice, in discrete-time, according to control intervals of duration typically much lower than the content timeout. For instance, the content timeout can be in the order of one minute, and the control interval duration is in the order of 1 s. The scheduled content delivery instants $t_{(j,k)}^{(z)}$ are hence computed in terms of number of control intervals, and mapped to the future control intervals, contributing, at the prescribed time, to the input to the RRR and allocation scheme described in Sect. 3.2.

5 Performance Evaluation

To evaluate the performance of our system, we used a custom simulator written in Matlab⁵. Our simulator implements a state of the art frequency selective channel model selected from [12, 14], which includes random Lognormal shadowing and Rician small scale multipath fading⁶.

We considered a two-lane street chunk of length 1.8 Km and width 20 m. The two lanes correspond to opposite marching directions. Four eNBs are placed at the horizontal coordinates of 0, 600, 1200, and 1800 m, respectively, at the center of the street (see Fig. 1), with height 10 m. The distance between the center of the two lanes is 10 meters. This is also the closest distance a vehicle can get to any vehicle marching in the opposite direction. We modeled the vehicles arrival as a temporal Poisson Point Process (PPP). In all the simulations whose results are presented in the following, the vehicle arrival rate was kept fixed at $\lambda_t = 1/3$ vehicles per second. Similarly, we used a PPP for modeling the request arrival process of each node, and kept the content request rate per device fixed at $\lambda_z = 1/6$ requests per second (10 content requests per minute). The content requests processes of different devices were set to be statistically independent. The selected content popularity distribution was a Zipf distribution with parameter $\alpha = 1.1$, i.e., $p_Z(z) \sim \frac{1}{z^{(\alpha)}} z^{-\alpha}$, truncated to a library size of 10^4 contents. The sharing timeout was also fixed and equal to $\tau_s = 600$ s. The content size was fixed and equal to a payload of 500 KBytes, which we assumed to

⁵ The reason to use a custom simulator, as opposed to classic network simulators like ns-3 or OMNET++, is to obtain a fine grain control on the physical layer aspects, retaining an acceptable level of scalability, using a state of the art channel model.

⁶ We implemented the Geometry-based Stochastic Channel Model (GSCM) described in [12, 14], up to the detail of generating the spatially correlated large scale parameters (LSPs): delay spread, shadow fading standard deviation, and Rician K-factor mean and variance, according to the procedure described in [12]. These parameters are then used to generate the random component of the set of spatially correlated frequency selective channels between any two points in a square grid with spatial step-size of 5 m, representing the region of interest. The frequency response experienced by each transmission is obtained by adding the so generated random component on top of the deterministic path loss component $g(d)$ described in Sect. 3.2, taking the random component between the grid points closest to the transmitter and receiver, respectively.

be encoded in a packet of 625 KBytes using a FEC coding rate equal to 0.8. The MAC parameters we used (see Subsect. 3.2) are as follows: each control interval lasts one second, and is divided in time slots of duration 1 ms. Each PRB lasts for 1 time slot and has width 200 KHz. In each PRB bandwidth, there are 12 subcarriers, the overall system bandwidth is 10 MHz, and in each control interval, a maximum of 50000 PRBs could be scheduled to concurrent I2D and D2D transmissions (possibly spatially reusing the same PRBs across non-interfering links).

The following set of results shows the performance obtained by letting a system parameter vary while keeping the rest of the parameters fixed. We focus on three parameters: the speed range $[v_{\min}, v_{\max}]$ in which each vehicle's speed falls, the content timeout τ_c , and the maximum D2D transmission distance. For each value of the varying system parameter, we run 10 independent i.i.d simulations, each lasting 1 h, reinitializing the random number generator seed with the same state at the beginning of each batch of 10 simulations. Each simulation is initialized with a random number of vehicles, positions, speeds, and cache content of each node according to the results in our previous work [5], in which we computed the steady state average number of vehicles and cache content distribution. In each simulation, we used a different independent realization of the whole set of random component of the channels between any two points in the grid, and between any eNB and any point in the grid. The next figures show the average value of the offloading efficiency and of the energy, E_{D2D} , consumed by D2D transmissions, measured in dB Re 1mJ⁷. Each value is displayed along with the related 95% confidence interval. Dark bars refer to a benchmark scheme, in which offloading is performed as soon as there is a potential content provider within range. Light bars refer to the proposed scheme.

Figure 3 shows the system performance with different selections of the speed range (measured in m/s). The increase in offloading efficiency with increasing vehicles speed, although a higher speed entails a lower vehicles' density (if the vehicle arrival rate is the same), is due to the fact, with higher speeds, the rate at which vehicles encounter each other increases, and hence the number of potential content providers met during the content timeout increases. From the energy consumption point of view, it can be seen that the proposed system allows for a more than 10 dB reduction of the energy consumed by the devices, on average, to deliver contents to their neighbors. This reduction is considerable, since it represents a more than 90% energy saving, and is consistent across the different selected values of the speed range.

Figure 4 shows the behavior of the two systems with different values of the maximum range d_{\max} within which potential content providers, for each content request, are searched by the CDMS. The offloading efficiency increases accordingly to the increase of d_{\max} , but the energy consumption increases as well. This happens because the additional D2D offloading opportunities obtained by widening the potential content providers search space, come at the price of a higher

⁷ $E_{D2D}[\text{dB Re 1mJ}] = 10 \log_{10}(E_{D2D}[\text{mJ}])$, e.g, a value of 10 dB Re 1mJ means that $E_{D2D}[\text{mJ}] = 10 \text{ mJ}$. A value of -10 dB Re 1mJ, means that $E_{D2D}[\text{mJ}] = 0.1 \text{ mJ}$.

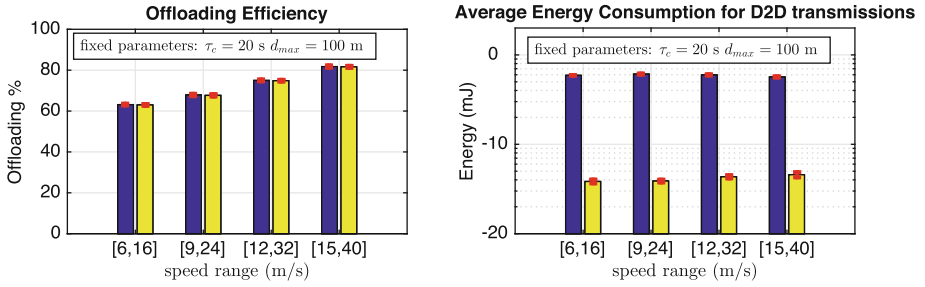


Fig. 3. Offloading efficiency and energy consumption with varying speed range. Benchmark offloading scheme (dark bars) and proposed scheme (light bars).

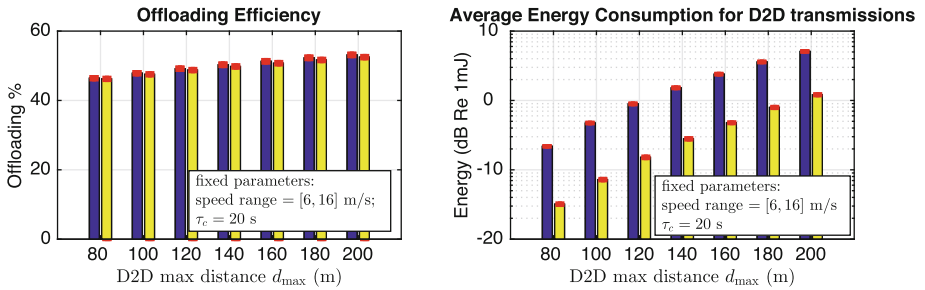


Fig. 4. Offloading efficiency and energy consumption with varying maximum D2D transmission distance d_{max} . Benchmark offloading scheme (dark bars) and proposed scheme (light bars).

required transmit power (for those additional opportunities), since the related involved transmission range is inevitably higher. Keep in mind that, due to the finite content timeout, the requesting node cannot wait indefinitely for those additional, farthest, potential content providers, to come close to it. The relative gain of the proposed D2D offloading transmission instant selection scheme, however, is consistently in the order of magnitude of 10 dB, with worst case 7 dB, which represents an 80% energy saving.

Finally, Fig. 5 displays the performance trend with respect to an increase of the content timeout. As expected, for both the benchmark and the proposed system, the offloading efficiency increases, since possibility of encountering potential content providers, within a larger time window, obviously increases. What is even more interesting, however, is the increase of the relative gain, in terms of energy consumption, of the proposed system with respect to the benchmark, with larger content timeouts. This happens because, with larger a time window, the probability that a potential content provider comes *very* close to the requesting device increases, thus entailing a lower transmit power during the content transmission. With a content timeout of 120 s, the energy saving for D2D offloading transmissions achieves 97%.

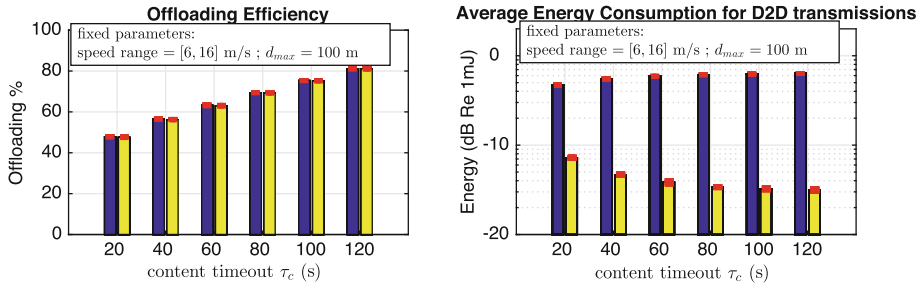


Fig. 5. Offloading efficiency and energy consumption with varying content timeout. Benchmark offloading scheme (dark bars) and proposed scheme (light bars).

6 Conclusion

We have proposed a content delivery management system for D2D data offloading in cellular networks tailored to scenarios, such as vehicular networks, where the topology varies at a fast rate, and to delay-tolerant applications. We evaluated the system level performance using an accurate system level simulator which includes a radio resource reuse scheme for allocating resources over a time-frequency radio resource grid, and incorporates a quite detailed channel model including small scale frequency selective fading. The proposed system, in which the D2D transmission instant is selected to minimize the transmission range, allows energy savings of around 90% in most of the considered settings, peaking at 97% for a delay tolerance of 2 min. The proposed system exploits the availability of nodes mobility predictions at the CDMS. As a future research direction, we plan to investigate the effect of trajectory prediction mismatches.

Acknowledgement. This work was partially funded by the EC under the H2020 REPLICATE (691735), SoBigData (654024) and AUTOWARE (723909) projects.

References

1. Rebecchi, F., et al.: Data offloading techniques in cellular networks: a survey. *IEEE Commun. Surv. Tutor.* **17**(2), 580–603 (2015)
2. Whitbeck, J., Lopez, Y., Leguay, J., Conan, V., de Amorim, M.D.: Push-and-track: saving infrastructure bandwidth through opportunistic forwarding. *Pervasive Mob. Comput.* **8**(5), 682–697 (2012)
3. Bruno, R., Masaracchia, A., Passarella, A.: Offloading through opportunistic networks with dynamic content requests. In: *Proceedings of the IEEE MASS 2014*, October 2014
4. Pescosolido, L., Conti, M., Passarella, A.: Performance evaluation of an energy efficient traffic offloading protocol for vehicular networks. In: *Proceedings of the 1st International Balkan Conference on Communications and Networking (BalkanCom 2017)*, Tirana, AL, 30 May–2 June 2017 (2017)

5. Pescosolido, L., Conti, M., Passarella, A.: Performance analysis of a device-to-device offloading scheme for vehicular networks. In: Proceedings of the IEEE WoW-MoM 2018, Chania, Greece, 12–15 June 2018 (2018)
6. Ji, M., Caire, G., Molisch, A.F.: Wireless device-to-device caching networks: basic principles and system performance. *IEEE J. Sel. Areas Commun.* **34**(1), 176–189 (2016)
7. Ji, M., Caire, G., Molisch, A.F.: Fundamental limits of caching in wireless D2D networks. *IEEE Trans. Inf. Theory* **62**(2), 849–869 (2016)
8. Lin, X., Andrews, J.G., Ghosh, A.: Spectrum sharing for device-to-device communication in cellular networks. *IEEE Trans. Wirel. Commun.* **13**(12), 6727–6740 (2014)
9. Yang, Y., Liu, T., Ma, X., Jiang, H., Liu, J.: FRESH: push the limit of D2D communication underlying cellular networks. *IEEE Tran. Mob. Comput.* **16**(6), 1630–1643 (2017)
10. Rebecchi, F., de Amorim, M.D., Conan, V.: Circumventing plateaux in cellular data offloading using adaptive content reinjection. *Comput. Netw.* **106**, 49–63 (2016)
11. Rebecchi, F., Valerio, L., Bruno, R., Conan, V., De Amorim, M.D., Passarella, A.: A joint multicast/D2D learning-based approach to LTE traffic offloading. *Comput. Commun.* **72**, 26–37 (2015)
12. ICT METIS Project Deliverable 1.4: METIS Channel Models. Technical Report (2015)
13. Pescosolido, L., Conti, M., Passarella, A.: Performance analysis of a device-to-device offloading scheme in a vehicular network environment (2018). <https://arxiv.org/pdf/1801.09082>
14. 3GPP: Study on 3D channel model for LTE (Release 12). 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, T.R. 36.873, V12.4.0, March 2017