Reliability-oriented Optimization of Computation Offloading for Cooperative Vehicle-Infrastructure Systems

Jianshan Zhou, Daxin Tian, *Senior Member, IEEE,* Yunpeng Wang, Zhengguo Sheng, Xuting Duan, and Victor C.M. Leung, *Fellow, IEEE*

***Abstract*—Computation offloading is critical for mobile applications that are sensitive to computational power, while dynamic and random nature of vehicular networks makes it challenging to guarantee the reliability of vehicular computation offloading. In this letter, we propose a reliability-oriented stochastic optimization model based on dynamic programming for computation offloading in the presence of the deadline constraint on application execution. Specifically, a theoretical lower bound of the expected reliability of computation offloading is derived, and then an optimal data transmission scheduling mechanism is proposed to maximize the lower bound with consideration of randomness in vehicle-to-infrastructure (V2I) communications. Experimental results demonstrate that our mechanism can outperform the conventional scheme and benefits vehicular computation offloading in terms of reliability performance in stochastic situations.**

***Index Terms*—Cooperative vehicle infrastructure system (CVIS), vehicular communication, mobile edge computing, computation offloading, dynamic programming.**

I. Introduction

RECENT advancements in vehicular communication and networking are empowering the integration of moving vehicles' and road-side infrastructures' sensing, computing and storage capacities [1], which has spawn a new technological paradigm for vehicular telematics and infotainment applications, called *Cooperative Vehicle Infrastructure Systems* (CVIS). In CVIS, full or partial computation offloading mechanisms play a key role in boosting the cooperation between vehicles and infrastructures. Nowadays, there exist extensive research efforts such as [2]–[9] that have been made on computation offloading in a variety of mobile cloud/edge computing scenarios. Most of these current works focus on the issue of energy-aware or/and latency-aware optimization for mobile applications, since smart mobile devices (e.g., smart phones, tablets, wireless sensors, etc.) are commonly energy-hungry. Many other effective offloading mechanisms can also be found in some state-of-the-art surveys [10]–[15], which mainly concern about the resource optimization or/and load balancing in communication and computation aspects.

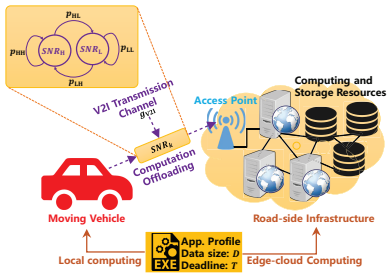

Fig. 1. Cooperative Vehicle-Infrastructure Computing Scenario.

However, little or no attention has been paid to reliability-oriented optimization of computation offloading for vehicular users, which is the main focus of this letter. Essentially different from many mobile users, vehicles are usually equipped with powerful energy, such that energy consumption issue is not the main concern in their system design. Instead, many vehicular applications, especially safety-type applications such as danger recognition and on-line diagnosis, are reliability-critical and latency-constrained. At this point, the reliability and efficiency of vehicular communications can have a great influence on the performance of offloading computation tasks to nearby service infrastructures. Moreover, due to high mobility and randomness in propagation channel fading, V2I communications may be randomly intermittent, which presents a challenge to optimization design and implementation of reliable vehicular computation offloading.

Toward this end, in this letter, we take into account the dynamic and random characteristics of V2I communications, and present a reliability-oriented stochastic optimization model for V2I-based computation offloading. We derive the successful probability of computation offloading with consideration of deadline-constrained application profile. Moreover, we obtain a lower bound of the optimal expected offloading reliability as well as an optimal data transmission scheduling mechanism by transforming the stochastic optimization into a dynamic programming paradigm. To the best of our knowledge, this work first presents reliability-oriented stochastic optimization for V2I-based computation offloading. Our mechanism can

This research was supported in part by the National Natural Science Foundation of China under Grant Nos. 61822101, 61672082 and 61711530247, the scholarship under the State Scholarship Fund (No. 201706020112), Royal Society-Newton Mobility Grant (IE160920) and The Engineering, and Physical Sciences Research Council (EPSRC) (EP/P025862/1).

J. Zhou, D. Tian, Y. Wang and X. Duan are with Beijing Advanced Innovation Center for Big Data and Brain Computing, Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems & Safety Control, School of Transportation Science and Engineering, Beihang University, Beijing 100191, China. E-mail: jianshanzhou@foxmail.com, dtian@buaa.edu.cn

Z. Sheng is with Department of Engineering and Design, the University of Sussex, Richmond 3A09, UK.(E-mail: z.sheng@sussex.ac.uk)

V. Leung is with Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, B.C., V6T 1Z4 Canada.(E-mail: vleung@ece.ubc.ca)

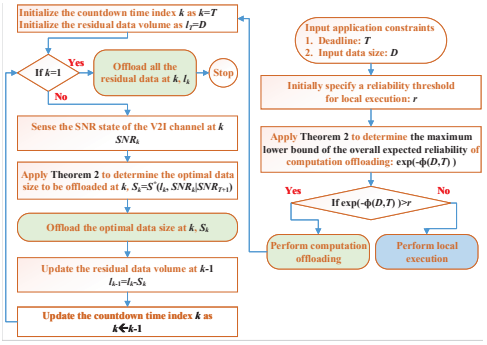

Fig. 2. Reliability-oriented Vehicular Computation Offloading.

be used to design a threshold-based decision-making policy for computation offloading in stochastic situations, which may facilitate a paradigm shift from the local vehicular computing to cooperative vehicle-infrastructure computing in CVIS.

II. System Model

A. Mobile Application Model

Following much existing literature such as [2], [16], we can use two key parameters to characterize the profile of a mobile application: i) The input data size D that is the total number of data bits as the application input. These D-bit data can be partitioned and offloaded to a roadside infrastructure as the cloud edge for remote execution. ii) The application completion deadline T that denotes the maximum number of successive time slots before which the mobile application must be completed. In addition, we use k to represent the index of a time slot, and these D-bit data can also be partitioned into a series of smaller pieces $s_k \in [0, D]$, where s_k denotes the number of data bits that should be transmitted to the infrastructure in the k-th time slot.

B. Vehicle-to-Infrastructure Transmission Channel Model

We take into account dynamic nature and randomness of the V2I communications from the physical-layer perspective. Since vehicle mobility and many environmental factors, such as trees, buildings and nearby moving vehicles as the obstacles in the signal propagation path, can lead to stochastic variation of signal-to-noise ratio (SNR) condition in V2I communications, we present stochastic modeling on the V2I channel. In order to capture the randomness, we adopt a two-state Markov chain model to characterize the SNR. It is also worth pointing out that the Markov modeling approach has been widely used to characterize a stochastic channel in many other works, such as Gilbert-Elliott channels [2] and Rayleigh channels [17].

Specifically, we consider that there are two traversable states of the V2I transmission channel, which include a high-SNR state and a low-SNR state. The SNRs corresponding to these two states are denoted by SNR_{H} and SNR_{L}, respectively. Let SNR_k be the SNR level of the transmission channel in time slot k. Thus, $SNR_k \in \{SNR_{\mathrm{H}}, SNR_{\mathrm{L}}\}$. We denote by p_{HH} the state transition probability that the next channel state has a high SNR given that its current state has a high SNR and by p_{LL} the probability that the next channel state has a low SNR given that its current state has a low SNR. Additionally, the transition probability that the channel changes from the current state with a high SNR to that with a low SNR can be calculated by $p_{\mathrm{HL}} = 1 - p_{\mathrm{HH}}$, and the transition probability from a low-SNR state to a high-SNR state can be $p_{\mathrm{LH}} = 1 - p_{\mathrm{LL}}$. The proposed two-state Markov model is represented by the diagram in Figure 1. Based on the two-state Markov model, the mean durations (represented by the number of time slots) of being in the high-SNR state and in the low-SNR state can be estimated by $T_{\mathrm{H}} = \frac{1}{1-p_{\mathrm{HH}}}$ and $T_{\mathrm{L}} = \frac{1}{1-p_{\mathrm{LL}}}$, respectively. Given the fading coefficient of the V2I channel as g_{V2I} and the corresponding fading variance as σ^2, we can express the channel capacity between the vehicle and the wireless access point of the infrastructure by

$$I_{\mathrm{V2I}} = \log_2\left(1 + SNR_k \left|g_{\mathrm{V2I}}\right|^2\right) \tag{1}$$

Generally, there is usually no dominant line-of-sight (LoS) propagation path between a mobile communication terminal and an access infrastructure (e.g., a basestation) in urban environments that are heavily built-up. In such situations, Rayleigh fading model has been widely used to represent the characteristics (e.g., path loss) of radio signal propagation. Hence, we also employ Rayleigh fading model for V2I communications, in which $\left|g_{\mathrm{V2I}}\right|^2$ is a random variable following an exponential distribution with parameter σ^{-2}. Accordingly, the successful probability of transmitting s_k-bit data in a time slot k through the V2I channel can be calculated by

$$p\left(s_k, SNR_k\right) = \mathrm{Prob}\left\{I_{\mathrm{V2I}} > s_k\right\} = \exp\left(-\frac{2^{s_k}-1}{\sigma^2 SNR_k}\right) \tag{2}$$

III. Stochastic Optimization Model for V2I Data Transmission Scheduling

We consider the optimization of transmission scheduling of D-bit application data in T time slots. For simplicity, we arrange the time slot index, k, in descending order, i.e., $k = T, T-1, \ldots, 1$. That is, k is used as a countdown timer, and the initial time slot is indexed by $k = T+1$. We also denote by $\boldsymbol{s} = \left[s_T, s_{T-1}, \ldots, s_1\right]^{\mathrm{T}}$ a feasible scheduling solution and by $\mathbb{S}$ the corresponding feasible region. Then, since SNR_k are random variables, we propose the optimization model for V2I transmission scheduling based on (2) with the goal to maximize the expected successful probability of computation offloading:

$$\begin{aligned} \max_{\boldsymbol{s}\in\mathbb{S}} :& \mathbb{E}\left[\prod_{k=1}^{T} p\left(s_k, SNR_k\right)\right] = \mathbb{E}\left[\exp\left(-\sum_{k=1}^{T}\frac{2^{s_k}-1}{\sigma^2 SNR_k}\right)\right] \\ s.t.\ & \begin{cases} \sum_{k=1}^{T} s_k = D; \\ \forall s_k \geq 0. \end{cases} \end{aligned} \tag{3}$$

In fact, it is difficult or even impractical to solve (3). Thus, we consider to bound the expected overall successful probability of computation offloading. The result is given in Theorem 1.

***Theorem** 1:* For the expected overall successful probability of computation offloading in (3), it always holds that

$$\mathbb{E}\left[\exp\left(-\sum_{k=1}^{T}\frac{2^{s_k}-1}{\sigma^2 SNR_k}\right)\right] \geq \exp\left(-\mathbb{E}\left[\sum_{k=1}^{T}\frac{2^{s_k}-1}{\sigma^2 SNR_k}\right]\right). \tag{4}$$

Proof: It is noted that $\exp(-x)$ is a convex function with respect to x. Thus, applying Jensen's inequality to the convex function $\exp(-x)$ can immediately derive Theorem 1. ■

Accordingly, to maximize the objective function in (3) as much as possible, we can turn to solve the following model rather than the original model (3):

$$\min_{\boldsymbol{s}\in\mathbb{S}} :\mathbb{E}\left[\sum_{k=1}^{T}\frac{2^{s_k}-1}{\sigma^2 SNR_k}\right] \quad s.t.\begin{cases}\sum_{k=1}^{T} s_k = D;\\ \forall s_k \geq 0.\end{cases} \tag{5}$$

From (5), the optimal solution depends on the SNR of the channel in the initial time slot $k = T+1$, i.e., $SNR_{T+1} = SNR_{\mathrm{H}}$ or $SNR_{T+1} = SNR_{\mathrm{L}}$. To solve (5), we denote the number of data bits that are remained to be offloaded at the beginning of the time slot k by l_k. Thus, we can have $l_{k-1} = l_k - s_k$ for $k = T, T-1, \ldots, 2$, and $l_T = D$. We also denote the optimal number of data bits to be scheduled in time slot k under the condition $SNR_{T+1} = SNR_{\mathrm{H}}$ by $s_k^*(l_k, SNR_k|SNR_{\mathrm{H}})$ and that under $SNR_{T+1} = SNR_{\mathrm{L}}$ by $s_k^*(l_k, SNR_k|SNR_{\mathrm{L}})$. Let $\phi(D,T|SNR_H)$ and $\phi(D,T|SNR_L)$ be the optimal value of the objective functions in (5) under these two conditions, respectively. Now, we derive the optimal data bits to be transmitted in each time slot by using dynamic programming and stochastic analysis as in Theorem 2.

***Theorem** 2:* For the optimization model of V2I transmission scheduling (5), the optimal transmission scheduling is

$$s_k^*(l_k, SNR_k|SNR_{\mathrm{H}}) = \begin{cases} l_1, & k=1;\\ \frac{l_k}{k} + \log_2\left(\prod_{t=1}^{k-1} H_t^{\frac{t}{k}}\right) + \frac{k-1}{k}\log_2\left(SNR_k\right), & k\geq 2\end{cases} \tag{6}$$

under the condition $SNR_{T+1} = SNR_{\mathrm{H}}$, where H_t is

$$H_t = \frac{p_{\mathrm{HH}}}{SNR_{\mathrm{H}}^{\frac{1}{t}}} + \frac{p_{\mathrm{HL}}}{SNR_{\mathrm{L}}^{\frac{1}{t}}}; \tag{7}$$

and

$$s_k^*(l_k, SNR_k|SNR_{\mathrm{L}}) = \begin{cases} l_1, & k=1;\\ \frac{l_k}{k} + \log_2\left(\prod_{t=1}^{k-1} L_t^{\frac{t}{k}}\right) + \frac{k-1}{k}\log_2\left(SNR_k\right), & k\geq 2\end{cases} \tag{8}$$

under $SNR_{T+1} = SNR_{\mathrm{L}}$, where L_t is

$$L_t = \frac{p_{\mathrm{LL}}}{SNR_{\mathrm{L}}^{\frac{1}{t}}} + \frac{p_{\mathrm{LH}}}{SNR_{\mathrm{H}}^{\frac{1}{t}}}. \tag{9}$$

The optimal values of the objective function in (5) under $SNR_{T+1} = SNR_{\mathrm{H}}$ and $SNR_{T+1} = SNR_{\mathrm{L}}$ are

$$\phi(D,T|SNR_H) = \frac{T2^{\frac{l_T}{T}}\left(\prod_{t=1}^{T} H_t^{\frac{t}{T}}\right)}{\sigma^2} - \frac{TH_1}{\sigma^2} \tag{10}$$

$$\phi(D,T|SNR_L) = \frac{T2^{\frac{l_T}{T}}\left(\prod_{t=1}^{T} L_t^{\frac{t}{T}}\right)}{\sigma^2} - \frac{TL_1}{\sigma^2} \tag{11}$$

respectively, where $l_T = D$. The minimum expected objective function in (5), denoted by $\Phi(D,T)$, is

$$\Phi(D,T) = \frac{T_{\mathrm{H}}}{T_{\mathrm{H}}+T_{\mathrm{L}}}\phi(D,T|SNR_{\mathrm{H}}) + \frac{T_{\mathrm{L}}}{T_{\mathrm{H}}+T_{\mathrm{L}}}\phi(D,T|SNR_{\mathrm{L}}). \tag{12}$$

Proof: Recall $l_{k-1} = l_k - s_k$ for $k \geq 2$. Applying the dynamic programming principle to (5), we can rearrange the optimization model into a series of recursive equations, i.e., the Bellman equation [18], as follows

$$f_k(l_k, SNR_k) = \begin{cases} \frac{2^{l_1}-1}{\sigma^2}\frac{1}{SNR_1}, & k=1;\\ \min\limits_{s_k\in[0,l_k]}\left\{\begin{array}{l}\frac{2^{s_k}-1}{\sigma^2}\frac{1}{SNR_k}\\ +\mathbb{E}\left[f_{k-1}(l_k-s_k, SNR_{k-1})\right]\end{array}\right\}, & k\geq 2.\end{cases} \tag{13}$$

First, we prove the theorem under the condition $SNR_{T+1} = SNR_{\mathrm{H}}$ by adopting mathematical induction as follows.

i) When $k = 1$, the whole remaining data bits, l_1, must be transmitted in the last time slot to meet the deadline imposed on the computation offloading. Thus, the optimal number of data bits scheduled in time slot $k = 1$ is $s_1^*(l_1, SNR_1|SNR_{\mathrm{H}}) = l_1$. Besides, given that the SNR level of the V2I transmission channel is high in time slot $k = 2$, i.e., $SNR_2 = SNR_{\mathrm{H}}$, we can calculate the expected optimal objective function in time slot $k = 1$ by

$$\begin{aligned}\mathbb{E}\left[f_1(l_1, SNR_1|SNR_{\mathrm{H}})\right] &= \mathbb{E}\left[\frac{2^{l_1}-1}{\sigma^2}\frac{1}{SNR_1}\right]\\ &= \frac{2^{l_1}-1}{\sigma^2}\left(\frac{p_{\mathrm{HH}}}{SNR_{\mathrm{H}}} + \frac{p_{\mathrm{HL}}}{SNR_{\mathrm{L}}}\right) = \frac{2^{l_1}-1}{\sigma^2}H_1,\end{aligned} \tag{14}$$

which is in accordance with (10) by setting $T = 1$ under $SNR_{T+1} = SNR_{\mathrm{H}}$.

ii) Suppose that (6), (7) and (10) hold for $k-1$ and $k \geq 3$. Based on the induction hypothesis, we can present (13) as

$$f_k(l_k, SNR_k|SNR_{\mathrm{H}}) = \min_{s_k\in[0,l_k]}\left\{\begin{array}{l}\frac{2^{s_k}-1}{\sigma^2}\frac{1}{SNR_k}\\ +\frac{(k-1)2^{\frac{l_k-s_k}{k-1}}\left(\prod_{t=1}^{k-1} H_t^{\frac{t}{k-1}}\right)}{\sigma^2} - \frac{(k-1)H_1}{\sigma^2}\end{array}\right\}. \tag{15}$$

iii) Let $g(s_k)$ denote the objective function in (15). Now, to obtain the optimal solution for (15), we solve the equation $\frac{\mathrm{d}g(s_k)}{\mathrm{d}s_k} = 0$, which is equivalent to solving

$$\frac{2^{s_k}\ln 2}{\sigma^2 SNR_k} - \frac{2^{\frac{l_k-s_k}{k-1}}\ln 2\left(\prod_{t=1}^{k-1} H_t^{\frac{t}{k-1}}\right)}{\sigma^2} = 0. \tag{16}$$

Solving (16) can directly get $s_k^*(l_k, SNR_k|SNR_{\mathrm{H}})$ as in (6). By substituting $s_k^*(l_k, SNR_k|SNR_{\mathrm{H}})$ into (15), we can derive

$$f_k(l_k, SNR_k|SNR_{\mathrm{H}}) = \frac{k2^{\frac{l_k}{k}}\left(\prod_{t=1}^{k-1} H_t^{\frac{t}{k}}\right)}{\sigma^2}\frac{1}{SNR^{\frac{1}{k}}} - \frac{kH_1}{\sigma^2}. \tag{17}$$

Noting that $\mathbb{E}[\frac{1}{SNR_k^{\frac{1}{k}}}|SNR_{k-1} = SNR_{\mathrm{H}}] = H_k$ as given in (7), we can derive the mathematical expectation

$$\mathbb{E}\left[f_k(l_k, SNR_k|SNR_{\mathrm{H}})\right] = \frac{k2^{\frac{l_k}{k}}\left(\prod_{t=1}^{k} H_t^{\frac{t}{k}}\right)}{\sigma^2} - \frac{kH_1}{\sigma^2}, \tag{18}$$

which can lead to (10) by setting $k = T$. Therefore, we can prove (6), (7) and (10) under the condition $SNR_{T+1} = SNR_{\mathrm{H}}$.

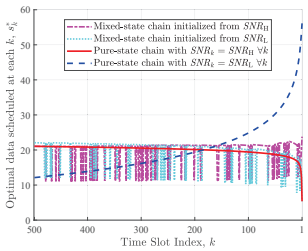

Fig. 3. Optimal V2I transmission scheduling with $D = 10^4$ bits, $T = 500$, $p_{\mathrm{HH}} = 0.9$ and $p_{\mathrm{LL}} = 0.2$.

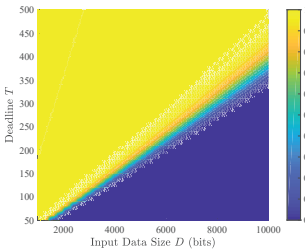

Fig. 4. The lower bound of the expected reliability, $\exp(-\Phi(D,T))$, with $p_{\mathrm{HH}} = 0.9$ and $p_{\mathrm{LL}} = 0.2$.

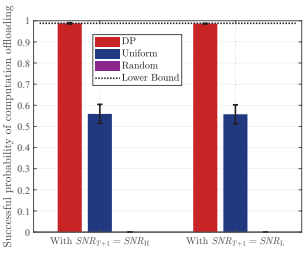

Fig. 5. Reliability performance comparison with $D = 10^4$ bits, $T = 500$, $p_{\mathrm{HH}} = 0.9$ and $p_{\mathrm{LL}} = 0.2$.

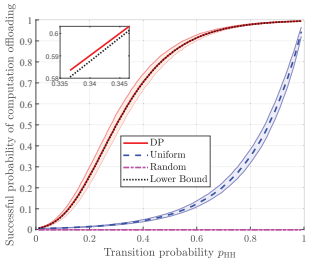

Fig. 6. Reliability performance comparison with $D = 10^4$ bits, $T = 500$ and $SNR_{T+1} = SNR_{\mathrm{H}}$.

The proof of (8), (9) and (11) under $SNR_{T+1} = SNR_{\mathrm{L}}$ can also be achieved by following the same induction logic, which is omitted here for the sake of brevity.

In the steady state, the probability that the channel has a high SNR or a low SNR is $\frac{T_{\mathrm{H}}}{T_{\mathrm{H}}+T_{\mathrm{L}}}$ and $\frac{T_{\mathrm{L}}}{T_{\mathrm{H}}+T_{\mathrm{L}}}$, respectively. Thus, we can derive (12) based on (10) and (11). ■

***Proposition** 1:* Given the application profile (D,T), the expected overall successful probability of computation offloading in (3) is not less than $\exp(-\Phi(D,T))$.

Proof: The proposition is proven by combining Theorems 1 and 2. ■

In addition, we further present an implementation framework of the reliability-oriented vehicular computation offloading based on the dynamic programming approach as in Figure 2. Under the framework, a vehicular node is able to dynamically schedule its data transmissions within the imposed deadline of the application, such that the overall expected reliability of the computation offloading can be well guaranteed in the stochastic V2I communication situation.

IV. Numerical Results

We conduct different simulation experiments where we set $SNR_{\mathrm{H}} = 50\,\mathrm{dB}$ and $SNR_{\mathrm{L}} = 20\,\mathrm{dB}$ to simulate the good and the bad SNR conditions of the V2I channel, respectively, and let $\sigma = 10^3$. In Figure 3, we illustrate the optimal data transmission scheduling s_k^* in different specified cases of the channel state variation. From two extreme cases (See the red solid and the blue dashed lines), we can see that the number of data bits to be transmitted in each k decreases as time proceeds when the SNR is always high, while it will increase when the SNR always stays low. The main reason is that since the expected SNR in the next time slot is lower given that the current SNR is high, the data bits to be transmitted in the next time slot should be reduced in order to guarantee the offloading reliability. In contrast, given the current channel is at a low-SNR level, the expected SNR in the next time slot is higher, such that the data bits to be transmitted in the next time slot should be larger. This proposition can also be confirmed in other two specific cases where the transition chain consists of mixed states. It is worth pointing out that since an optimal scheduling solution for the stochastic optimization model (5) is obtained from the optimal expectation perspective, it may not be the best for a specific and deterministic case. For instance, in Figure 3, in the extreme case where the SNR is always low, the corresponding scheduling solution is not the best solution for that case, and may even perform worse than a simple solution that schedules equal data bits in each time slot. Nonetheless, in actual implementation, we can decide to performance local execution instead of computation offloading when the lower bound of the optimal expected successful probability of computation offloading associated with an optimal scheduling solution, $\exp(-\Phi(D,T))$, is lower than a given threshold. Moreover, we show the profile of $\exp(-\Phi(D,T))$ under different application profiles in Figure 4. A finite region with high reliability of computation offloading (where $\exp(-\Phi(D,T))$ is more than 0.9) does exist.

Next, we compare our mechanism based on dynamic programming (marked as 'DP') with two other offloading mechanisms, one of which (marked as 'Uniform') schedules equal data bits in each time slot while the other mechanism (marked as 'Random') randomly creates the data partitions by following the uniform distribution. Extensive Monte Carlo simulations have been carried out with 1000 replications per initial state condition. The numerical results are given in Figure 5, where the average level of the successful probability of computation offloading is illustrated with the corresponding standard deviation. Besides, in Figure 6, we set $p_{\mathrm{LL}} = 1 - p_{\mathrm{HH}}$ and vary p_{HH} from 0.01 to 0.99 to simulate different randomness. Figure 6 also gives the average result as well as the standard deviation interval. From these two figures, it can be seen that the proposed mechanism can achieve the highest performance under different initial channel states and different channel randomness. This confirms the advantage of our proposed method in dynamic and stochastic communication situations.

V. Conclusion And Future Work

In this letter we explore the problem of reliability-oriented optimization of computation offloading in dynamic and stochastic V2I communications. We have proposed an optimal V2I transmission scheduling mechanism based on dynamic programming, the goal of which is to improve the reliability in computation offloading by maximizing the lower bound of the expected successful probability of data transmissions. Numerical results verify the effectiveness and advantage of our method in terms of guaranteeing the reliability performance. In the future, we will extend our method to a highly reliable cooperative computing framework where vehicles and infrastructures are coordinated to process distributed applications.

REFERENCES

[1] J. A. Guerrero-ibanez, S. Zeadally, and J. Contreras-Castillo, "Integration challenges of intelligent transportation systems with connected vehicle, cloud computing, and internet of things technologies," *IEEE Wireless Communications*, vol. 22, no. 6, pp. 122–128, 2015.

[2] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4569–4581, September 2013.

[3] O. Muoz, A. Pascual-Iserte, and J. Vidal, "Optimization of radio and computational resources for energy efficiency in latency-constrained application offloading," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4738–4755, Oct 2015.

[4] S. Lagn, A. Pascual-Iserte, O. Muoz, and J. Vidal, "Energy efficiency in latency-constrained application offloading from mobile clients to multiple virtual machines," *IEEE Transactions on Signal Processing*, vol. 66, no. 4, pp. 1065–1079, Feb 2018.

[5] E. Baccarelli, N. Cordeschi, A. Mei, M. Panella, M. Shojafar, and J. Stefa, "Energy-efficient dynamic traffic offloading and reconfiguration of networked data centers for big data stream mobile computing: review, challenges, and a case study," *IEEE Network*, vol. 30, no. 2, pp. 54–61, March 2016.

[6] C. You, K. Huang, H. Chae, and B. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1397–1411, March 2017.

[7] G. Li, J. He, S. Peng, W. Jia, C. Wang, J. Niu, and S. Yu, "Energy efficient data collection in large-scale internet of things via computation offloading," *IEEE Internet of Things Journal*, pp. 1–1, 2018.

[8] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2795–2808, October 2016.

[9] K. Zhang, Y. Mao, S. Leng, Q. Zhao, L. Li, X. Peng, L. Pan, S. Maharjan, and Y. Zhang, "Energy-efficient offloading for mobile edge computing in 5g heterogeneous networks," *IEEE Access*, vol. 4, pp. 5896–5907, 2016.

[10] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1628–1656, thirdquarter 2017.

[11] W. Yu, F. Liang, X. He, W. G. Hatcher, C. Lu, J. Lin, and X. Yang, "A Survey on the Edge Computing for the Internet of Things," *IEEE Access*, vol. 6, pp. 6900–6919, 2018.

[12] D. Xu, Y. Li, X. Chen, J. Li, P. Hui, S. Chen, and J. Crowcroft, "A survey of opportunistic offloading," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2198–2236, thirdquarter 2018.

[13] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A survey on mobile edge networks: Convergence of computing, caching and communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.

[14] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2322–2358, Fourthquarter 2017.

[15] R. Mahmud, R. Kotagiri, and R. Buyya, *Fog Computing: A Taxonomy, Survey and Future Directions*. Singapore: Springer Singapore, 2018, pp. 103–130. [Online]. Available: https://doi.org/10.1007/978-981-10-5861-5_5

[16] Z. Sheng, C. Mahapatra, V. C. M. Leung, M. Chen, and P. K. Sahu, "Energy efficient cooperative computing in mobile wireless sensor networks," *IEEE Transactions on Cloud Computing*, vol. 6, no. 1, pp. 114–126, Jan 2018.

[17] Q. Zhang and S. A. Kassam, "Finite-state markov model for rayleigh fading channels," *IEEE Transactions on Communications*, vol. 47, no. 11, pp. 1688–1692, Nov 1999.

[18] M. Sniedovich, *Dynamic programming: foundations and principles*. Boca Raton: CRC Press, 2010.