Received January 9, 2020, accepted February 1, 2020, date of publication February 10, 2020, date of current version February 20, 2020.
Digital Object Identifier 10.1109/ACCESS.2020.2972938

Proximity as a Service for the Use Case of Access Enhancement Via Cellular Network-Assisted Mobile Device-to-Device

TIANHAO GUO [1], (Student Member, IEEE), JOHN SCHORMANS [1], LEXI XU [2], JINZE WU [3], AND YUE CAO [4]
[1]School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K.
[2]China Unicom Network Technology Research Institute, China Unicom, Beijing 100048, China
[3]Interdisciplinary Center of Quantum Information, State Key Laboratory of Modern Optical Instrumentation, and Zhejiang Province Key Laboratory of Quantum Technology and Device of Physics Department, Zhejiang University, Hangzhou 310027, China
[4]School of Computing and Communications, Lancaster University, Lancaster LA1 4WA, U.K.

Corresponding author: John Schormans (j.schormans@qmul.ac.uk)

This work was supported in part by the China Scholarship Council.

ABSTRACT Device-to-Device (D2D) communication is a way to treat the User Equipments (UEs) not as terminals, but as a part of the network (helpers) for service provisioning. We propose a generic framework, namely Proximity as a Service (PaaS), formulate the helper selection problem, and design and prove a heuristic helper selection policy, ContAct based Proximity (CAP), which increases the service connectivity and continuity. Design Of Experiment (DOE) is a statistical methodology that rigorously designs and conducts an experiment, and maximizes the information obtained from that experiment. We apply DOE to explore the relationship (analytic expression) between four inputs (factors) and four metrics (responses). Since different factors have different regression levels, a unified four level full factorial experiment and cubic multiple regression analysis have been carried out. Multiple regression equations are provided to estimate the different contributions and the interactions between factors. Results show that transmission range and user density are dominant and monotonically increasing, but transmission range should be restricted because of interference and energy-efficiency. After obtaining the explicit close form expressions between factors and responses, optimal values of key factors are derived. A methodology (the ϵ-constraint method) to solve the multiple-objective optimization problem has been provided and a Pareto-Optimal set of factors has been found through iteration. The fluctuation of the iterations is small and a specific solution can be chosen based on the particular scenarios (city center or countryside with different user density). The methodology of optimization informs the design rules of the operator, helping to find the optimal networking solution.

INDEX TERMS D2D, mobile computing, DOE.

I. INTRODUCTION

In recent years, the explosive growth of miscellaneous devices along with various service demands, has brought numerous challenges to current network infrastructures [1], e.g., the growth of data traffic by orders of magnitude, ubiquitous coverage and an incredible variety of service requirements.

The need to increase the network capacity and enrich services provisions has been widely recognized as key features in future networks. Several enabling technologies, e.g., massive Multiple-Input and Multiple-Output (MIMO), millimeter wave (mmWave), non-orthogonal multiple access (NOMA), femtocell and heterogeneous networks (HetNets) have been extensively studied in the literature [2] and partly applied.

Compared to other technologies, Device-to-Device (D2D) is a much cheaper solution [3], enabling direct communication between mobile devices without necessarily involving cellular links in data transmissions. Thanks to the developing features of mobile phones, User Equipment (UE) with such powerful resources would be able to play an attractive role as service suppliers in proximity by D2D, other than simply as requesters. Here we define the term Proximity as

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenyu Zhou.

a Service (PaaS) to represent the D2D collaborative framework with UEs acting as service suppliers and requesters respectively.

Up-to-now, the exploration of use cases with different proximal services is ongoing, including cellular traffic offloading [4], opportunistic crowd computing [5], mobile augmented reality [6], mobile crowdsourcing [7], computation offloading [8], [9], target marketing [10] and even block-chain applications [11].

We proposed [12] a use case of proximal service by which some (appropriately selected) UEs, named helpers [13], are eligible to opportunistically provide services for other UEs (as service requesters) as requested. Key techniques developed involve the design of the system model, appropriate knowledge of human mobility, energy-efficient D2D handling for specific services requirements (with a limited number of requesters a helper can serve simultaneously), and potential security issues (limited number of selected helpers), etc.

We further conduct experiments and data analysis via the Design Of Experiment (DOE) [14] to comprehensively investigate the fundamental impact of key factors on the D2D enabled PaaS case.

In this paper, we have completed the following:

- **Exploring human mobility, so as to identify a certain number of appropriate helpers (in terms of whom to help) for PaaS.** We formulate the "helpers selection problem" and propose the ContAct based Proximity (CAP) policy, which utilizes fruitful contact history information such as contact duration, contact frequency, and inter-meeting duration to capture human mobility for helpers selection.
- **Applying DOE to determine levels[1] of factors and conduct a full factorial design experiment.** After four separate single regression analyses for four factors and four responses, the highest order among these regression lines is seen to be cubic, which is suitable for a four-level factorial design [14]. Therefore, all factors are set with four levels (including virtual level)[2] and five repetitions are set for each combination.
- **Four polynomial multiple regression equations for four responses are obtained and comparison of regression lines and simulation results based on the equations are drawn.**
  The polynomial regression equations (analytic expressions) derived from the full factorial experiment can fit the results well, which facilitate the optimization within the domain of factors. The comparison of the influence of four factors on four responses has been made working with the different dimensions, scales, and units of the four factors. Results show that the transmission range and the user density are dominant and are both monotonically increasing factors.
- **Based on the four polynomial multiple regression equations, a non-linear multi-objective optimization has been completed.**
  The key to solve this multi-objective optimization problem is to convert it into the single-objective problem. The difficulty is that the four outputs have different dimensions, scales and units so that iterations of ϵ-constraint optimizations [15] rather than one weighted-sum single-objective optimization [15] has been completed.

[1]For every experimental design, there is a tradeoff between cost and correctness. For example, if pilot runs show that the output O of the experiment has a quadratic curve when the input I changes, at least three runs with different values of I should be conducted, which is called three-level design.

[2]In the language of DOE, virtual levels means that redundant levels are set for some factors. Therefore, all factors are unified with the largest number of levels, which can simplify the experiment.

II. RELATED WORK

The majority of previous works [16]–[18] on D2D focus on how to cognitively allocate spectrum or energy for paired UEs, in order to achieve a higher throughput with less interference.

The core of the papers [16]–[18] is the optimization of Energy Efficiency (EE). The progress in these papers is increasing the complexity of the considerations (combinations) of other constraints (or objectives). The common constraints include co-channel interference, the limited battery capacity, while the different objectives (or constraints) include spectral efficiency [16], QoS and transmission power [18].

D2D communication can use various air interfaces (in-band underlay D2D, in-band overlay D2D, network-assisted out-band D2D, and autonomous out-band D2D). In-band underlay means D2D networks and cellular networks use the same licensed spectrum [19], [20]. In-band overlay means D2D networks and cellular networks use different licensed frequencies, which means that Mobile Network Operators (MNOs) need to allocate dedicated frequencies for D2D communication. Out-band means that D2D networks work in unlicensed bands, such as Bluetooth and WiFi-Direct. If the infrastructure centrally controls the transmission, it is network-assisted out-band D2D; otherwise it is autonomous out-band D2D. The comparison is elaborated in Table 1.

TABLE 1. Comparison of different d2d air interfaces.

Characteristics	In-Band Underlay D2D	In-Band Overlay D2D	Network-Assisted Out-Band D2D	Autonomous Out-Band D2D
Dedicated D2D architecture	Yes	Yes	Yes	No

Since licensed bands are owned by MNOs, both kinds of in-band D2D air-interfaces are network-assisted. In this paper we focus on the influence of mobility on the D2D so the air-interface is transparent.[3] Therefore the interference problems in the in-band underlay D2D is not considered in

[3]In fact, this paper was entitled "via cellular network-assisted", which means autonomous out-band D2D is not included.

this paper. PaaS can be applied in the interfering scenario with only a different computational complexity.

Utilising human mobility [21] to improve network performance and exploring proximal services are also important.

In [22], stochastic geometry is used to build the system model with mobility and the "helper selection" with social trust one of the constraints. Another constraint is the transmission condition, which is then formulated as D2D success probability. Since the optimization problem is non-convex, the special unbiased optimal caching strategy is derived in a close form and a numerical searching algorithm is proposed to obtain the globally optimal solution for the general case. An iterative algorithm is proposed to get the solution that satisfies the Karush-Kuhn-Tucker conditions.

Due to the opportunistic nature of D2D communication, long/frequent disruption or delay is inevitable, especially in a multi-hop case. Therein, the end-to-end based multi-hop path from source to destination, cannot always be established due to intermittent links. Researchers in the field of Delay Tolerant Network (DTN) [23] and mobile social networks (e.g., content offloading [13] and mobile crowdsourcing [7]), have proposed a number of ways to estimate the delivery potential of the use of relays for improved message delivery.

However, in the 5G era, network facilities are almost ubiquitous wherever people live. As such, the majority of D2D communication may just rely on a one hop "network". Inevitably, due to the selfish nature of human beings and potential security issues, not all the UEs are willing to act as helpers for PaaS. From the perspective of the operator, selecting the minimum number of helpers to provide a satisfactory service is also cost-efficient. After all, the operator would prefer not to lose control. In [8], the base station selects the helpers to offload computation tasks. The process of parameter collection in [8] corresponds to the process of the generation of the utility function in this paper. In this paper, helper selection is based on the regularity of the mobility. However, in [8], the helper selection is based on the computational resource of the helper and precedence-constrained task-graph of the application.

Joint consideration of the uncertainty of mobility and resource availability is complicated. In [24], a network-assisted D2D computation offloading collaboration framework, called D2D Fogging [24], with an incentive scheme and a resource availability prediction, was formulated by Lyapunov optimization. The authors simplified the problem of mobility by introducing granularity into the time frame because nodes are static and cellular links are stable within the time frame. Network (communication) resources and computation resources are also jointly formulated. By jointly considering the execution and communication energy consumption, the proposed online algorithm achieved superior performance on energy efficient task executions.

Several PaaS systems have been proposed, mainly addressing how to achieve Quality of Service (QoS) awareness performance [6], [7], [9]–[11], through appropriate helper selection. On the contrary, we propose the CAP to provide ubiquitous (high service success ratio) and Quality of Experience (QoE) like (e.g., disruption duration and frequency) service, through extensive effort to explore human mobility and access handling run by the helpers themselves.

In the paradigm of wireless networks, there are many factors influencing the performance of the experiment and the performance evaluation also has many perspectives. For example, Vadde [25] et al. considered five factors for three measurements, which is not suitable for the use of one-factor-at-a-time experimental methodology (control variates approach).

DOE [14] is a statistical and scientific experimental methodology that can rigorously and economically conduct an experiment and extract full information of the relationship between many control factors (inputs) and many responses (outputs). In the paradigm of D2D, DOE provides a way to scientifically design the experiment, analyse the data, and optimize the performance of the system.

III. SYSTEM MODEL

In this paper, we consider a system with a certain number of Base Stations (BSs), UEs that play the role of service providers, called helpers, and UEs that request services from helpers by D2D communication, called requesters. The helpers are assumed to be willing to help other requesters access the network and act as *mobile access points* or *mobile picocells* with wireless backhaul links [26].

The whole communication process has two hops: the first hop is the communication between the requester and the helper; the second hop is the communication between the helper and the BS. Helpers are assumed to have stable connections to base stations. Nevertheless, the opportunistic contact between pairwise helper and requester inevitably results in service disruption.

The requester is assumed to generate access requests deterministically (e.g., generate access request every one minute) with the same requested access duration (T_{RAD}, e.g., half an hour). The access request needs to be served within a period of time, called the access tolerance (T_{AT}, e.g., 2 hours). T_{RAD} and T_{AT} compose the requirement of the service of the system and this requirement is the base of the four outputs of the PaaS in this paper. It is convenient to evaluate the continuity and disruption features of the PaaS with T_{RAD} and T_{AT}.

T_{RAD} and T_{AT} can also be the parameters of requirements of potential applications. If we consider the case of IoT [18], a device in D2D networks works as a machine-type gateway which can do data aggregation and a bridge between Machine Type Devices (MTDs) and the BS. There might be some applications that need to monitor the information from the MTDs (requesters) for T_{RAD} within T_{AT}. During T_{AT}, the information generated in the MTDs should immediately be sent to the base station.

A. AN EXAMPLE OF THE PAAS SYSTEM

Assume there are three helpers, one requester and one base station in the PaaS system, as shown in Fig. 1.

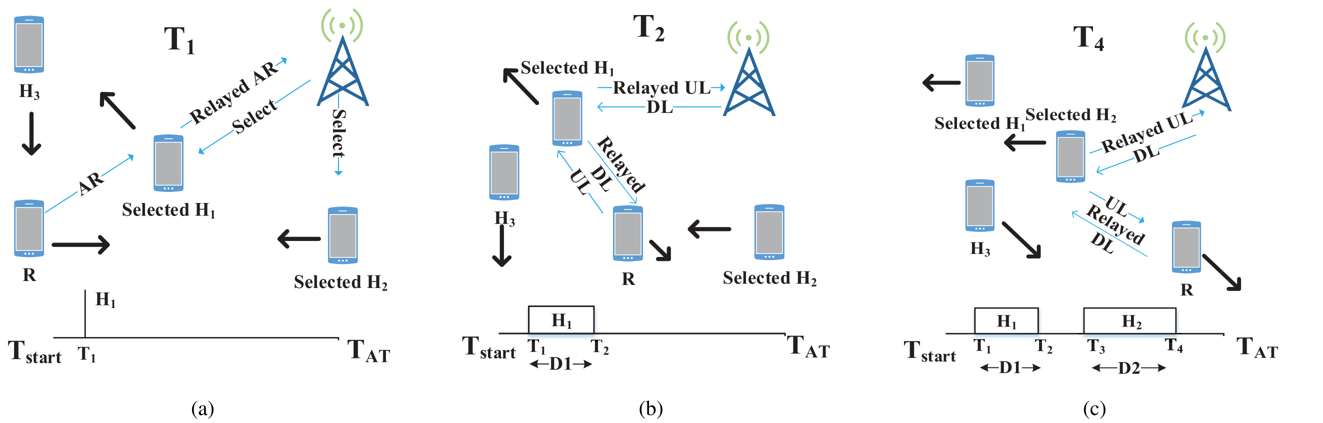


FIGURE 1. Process of D2D relaying.

The time clip of the process of the PaaS in Fig. 1 begins at the start of an access request ($T_{\text{start}} = 0$) and finishes at the end of the access tolerance (T_{AT}) of the access request. The moments T_1, T_2, and T_4 are shown to elaborate the CAP process.

- At T_1, an access request with T_{RAD} (half an hour) and T_{AT} (two hours) is generated by a requester (R). The R, running D2D air interfaces, broadcasts Access Requests (ARs) to nearby devices periodically through an associated mobile application. As illustrated in Fig. 1, when helper 1 (H_1) receives the AR and relays the AR to the base station, the base station arranges and selects appropriate helpers (H_1 and H_2) for R based on metrics (utilities) derived from their contact history.
- At T_2, all the UEs have moved to different locations, as illustrated in Fig. 1(b). H_1 has helped the requester (R) access the network for a period of duration D_1 ($D_1 < T_{\text{RAD}}$). However, H_1 is going to leave R.
- At T_4, as shown in Fig. 1(c), H_2 begins to help R at T_3 and finishes the service at T_4 ($D_1 + D_2 = T_{\text{RAD}}$) which is within T_{AT}.

The system is assumed to be based on one-to-many D2D, which means that a helper can have connections with several requesters but a requester can only communicate with one helper.

B. PROBLEM FORMULATION

Since not all helpers will be allocated with the role of helping to provide access service to a certain requester, we study how to optimally select a set (with a certain number) of helpers. An accurate mathematical description of the helper selection problem is the first step. Notations for problem formulation are given in Table 2.

At any time, the contact relationship between H_i and R_j can be in contact or not in contact:

$$H_i(t)R_j(t) = \begin{cases} 1 & \text{if in contact,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

TABLE 2. List of notations for problem formulation.

T_{AT}	Access tolerance
T_{RAD}	Requested access duration
R_j	Requester j
H_i	Helper i
N_{H}	Total number of helpers
N_{R}	Total number of requesters
AR_{jk}	The k th access request of requester j
$T_{\text{AR}_{jk}}$	Start time of the k th access request of requester j
$SH_{\text{AR}_{jk}}$	Set of selected helper for AR_{jk}
M	limit number of the set $SH_{AR_{jk}}$
S_{AR}	Total number of all the generated access requests
SC_{AR}	Total number of the completed access requests
K	Total number of requester a helper can serve at the same time
T_{jk}^{ela}	Elapsed time since AR_{jk} requested

When BSs receive an access request (AR_{jk}), an optimal set of helpers $SH_{\text{AR}_{jk}}$ (e.g., in terms of minimized number of helpers) is selected to serve it. If H_i is in $SH_{\text{AR}_{jk}}$, the accumulated contact duration between H_i and R_j as denoted by $T(\text{H}_i\text{AR}_{jk})$, within T_{AT} of AR_{jk}, is

$$T(\text{H}_i AR_{jk}) = \int_{T_{AR_{jk}}}^{T_{AT}+T_{AR_{jk}}} H_i(t)R_j(t)\,dt. \tag{2}$$

The results of the service contributed by the selected helpers ($SH_{\text{AR}_{jk}}$) are described by two binary values 1 and 0:

$$C(\text{AR}_{jk}) = \begin{cases} 1 & \text{if } \sum\limits_{i \in SH_{\text{AR}_{jk}}} T(\text{H}_i\text{AR}_{jk}) \geqslant T_{\text{RAD}}, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The binary value 1 represents that the sum of accumulated contact duration between the selected helpers and the requester R_j is not less than T_{RAD}, while the binary value 0 means the service is not completed, i.e. the sum of accumulated contact duration between the selected helpers and the requester R_j is less than T_{RAD}

The helper selection strategy can be formulated as an optimization problem:

$$\max SC_{\mathrm{AR}} = \sum_{i=1}^{S_{\mathrm{AR}}} C(\mathrm{AR}_{jk}), \tag{4a}$$

$$\text{s.t.} \sum_{j=1}^{N_R} H_i(t)R_j(t) \leqslant K, \quad i = 1, 2, 3 \ldots, N_H,$$

$$\mathrm{Card}(SH_{\mathrm{AR}_{jk}}) \leqslant M. \quad j = 1, 2, 3 \ldots, N_R. \tag{4b}$$

It consists of a objective function (maximizing total number of the completed access requests SC_{AR}), by choosing subsets (selected helpers $SH_{\mathrm{AR}_{jk}}$) for all access requests with the total number S_{AR}, from all helpers in the system. This problem is subject to two limitations. Firstly, a helper can only serve a limited number of requesters simultaneously; secondly, the number of helpers serving a requester is limited. In other words, the cardinality of the set $SH_{\mathrm{AR}_{jk}}$, namely $\mathrm{Card}(SH_{\mathrm{AR}_{jk}})$, is limited.

C. DISCUSSION ON THE SYSTEM

The mobility of the UEs is not prior knowledge, so the information of Equation (1) is unknown before the experiment. Without this information, it is hard to estimate the results of the access requests. If AR_{j1} (the first access request) is completed, it is also hard to estimate the service completion time. A requester can only have one access request at a time, which means that the AR_{j2} can only be generated after the finishing time of AR_{j1}. In other words, the more access requests that are completed, the more access requests that will be generated. So the total number of access requests (S_{AR}) is also unknown.

Moreover, when R_j is temporarily isolated, AR_{jk} from R_j is hard to complete,

$$\sum_{i=1}^{N_H} T(\mathrm{H}_i \mathrm{AR}_{jk}) < T_{\mathrm{RAD}}. \tag{5}$$

This situation is also unpredictable.

Based on the complexity brought by the unpredictable mobility of the nodes, the heuristic policy is proposed in the next section. This new policy can learn the information of contact history and choose the best set of helpers. The helper can also raise the priority of requesters based on elapsed time T_{jk}^{ela} if the number of concurrent devices is more than K.

IV. THE POLICY OF HELPER SELECTION

The idea of the proposed policy is to learn the regularity of mobility from contact history, and then derive a utility for every helper and requester pair. The utilities are the scores to rank the appropriate helpers for a requester.

A. UTILITY FUNCTION DESIGN

The contact history between a helper and a requester is elaborated in Fig. 2. $T_{i,j}^{(C_{i,j}=1)}$ is the moment when H_i and T_j contact for the first time. The inter-meeting time is shown in Fig. 2 is $T_{i,j}^{(C_{i,j}=2)}$. For $C_{i,j} \geqslant 2$, $T_{i,j}^{(C_{i,j})} - D_{i,j}^{(C_{i,j}-1)}$ means the encounter gap, namely the disruption duration. Note that the $C_{i,j}$ can only be measured when a new contact happens.

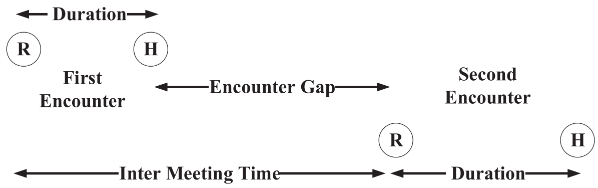


FIGURE 2. Illustration of contact history.

TABLE 3. List of notations for policy.

$D_{i,j}$	Historically contact duration between H_i and R_j
$T_{i,j}$	Historically inter-meeting time between H_i and R_j
$C_{i,j}$	Historically contact count between H_i and R_j
$U_{i,j}$	Utility value estimated of H_i and R_j
$UA_{i,j}$	Average utility for helper selection
Γ	Aging constant
$N_{i,j}$	Total number of contacts between H_i and R_j

Based on the information obtained from contact history between a pairwise helper and requester, the regularity can be found. For example, there is one requester (R) and two helpers (H_1 and H_2) in the system. H_1 contacts with R frequently, with longer contact duration and with less disruption duration. H_2 rarely contacts R, with shorter contact duration and longer disruption duration. It is easy to learn that H_1 will be a better choice for R. Moreover, $T_{\mathrm{H}_i,\mathrm{R}_j}^{(C_{i,j}=1)}$ also implies how fast R_j can get connected to the network. Based on this observation, an empirical utility function can be obtained:

$$U'_{i,j} = \frac{T_{i,j}^{(C_{i,j}=1)} + \sum_{i,j}^{N_{i,j}} (C_{i,j} \geq 2) \left(T_{i,j}^{(C_{i,j})} - D_{i,j}^{(C_{i,j}-1)} \right)}{N_{i,j}}. \tag{6}$$

Taking the Equation (6) as an example, there is one requester (R_1) and two helpers (H_1 and H_2) in the system. Assume that $T_{1,1}^{(C_{1,1}=1)} = 15$ and $D_{1,1}^{(C_{1,1}=1)} = 3$ are recorded at the first contact; $T_{1,1}^{(C_{1,1}=2)} = 10$ and $D_{1,1}^{(C_{1,1}=2)} = 6$ are recorded at the second contact; $T_{1,1}^{(C_{1,1}=3)} = 20$ is recorded at the third contact. Then $U'_{1,1}$ is calculated as:

$$U'_{1,1} = \frac{15 + (10 - 3) + (20 - 6)}{3} = 12. \tag{7}$$

Assuming $T_{2,1}^{(C_{2,1}=1)} = 20$, $D_{2,1}^{(C_{2,1}=1)} = 2$ at the first contact, while $T_{2,1}^{(C_{1,1}=2)} = 20$ is recorded at the third contact, then $U'_{2,1}$ is calculated as:

$$U'_{2,1} = \frac{20 + (20 - 2)}{2} = 19. \tag{8}$$

It is conventional that a helper with a larger value of utility has the potential to provide good quality of service, in terms of success ratio and continuity. So the $U_{i,j}$ should be inverted

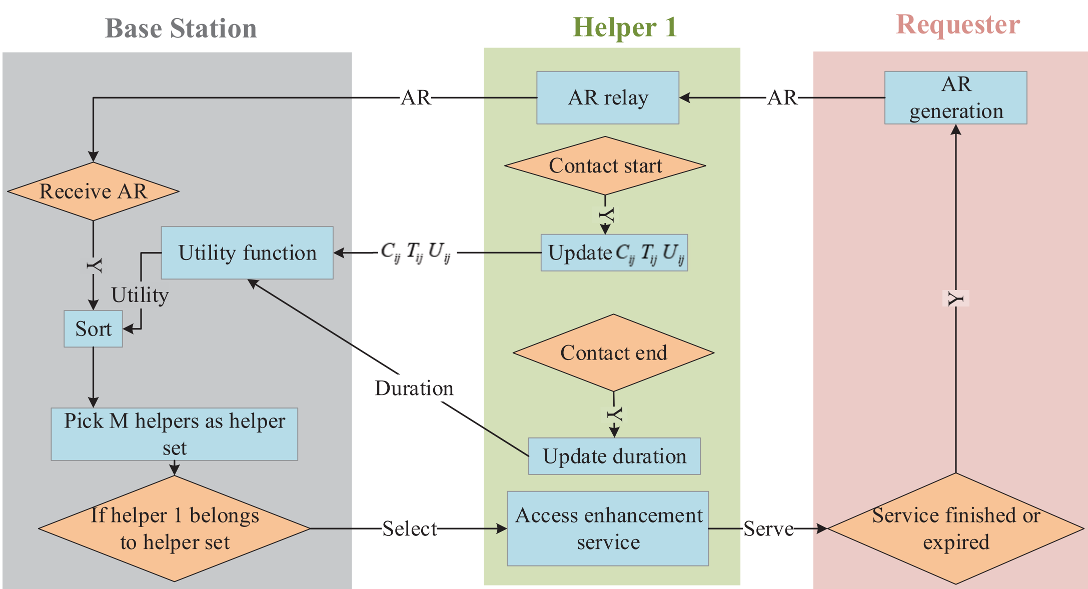


FIGURE 3. The helper selection policy.

as:

$$U_{i,j} = \frac{1}{U'_{i,j}} = \frac{N_{i,j}}{T_{i,j}^{(C_{i,j}=1)} + \sum_{i,j}^{N_{i,j}} (C_{i,j} \geq 2) \left(T_{i,j}^{(C_{i,j})} - D_{i,j}^{(C_{i,j}-1)} \right)}. \quad (9)$$

Note that the utility is updated at the beginning of the contact. Contact duration is recorded when the contact is disrupted, while inter-meeting is recorded at the beginning of the contact, similar to $C_{i,j}$, as shown in Algorithm 1.

Algorithm 1 Update Utility

```
if A contact between R_j and H_i is started then
    H_i and R_j update the number of contact C_{i,j}
    H_i and R_j update the inter-meeting time T_{i,j}
    H_i and R_j update the utility U_{i,j}
    H_i reports the information to base stations
end if
if The contact between R_j and H_i is ended then
    H_i and R_j update the contact duration
    H_i reports the information to base stations
end if
```

The goal of the helper selection policy is to find the best helpers for a requester based on the regularity of the movement model. The regularity (relationship) between a helper and a requester is quantified as utility. The regularity implies stability and the regularity that is learned from contacts happened in the time window ($UA_{i,j}$) is stabler than the one that is only learned from the latest contact ($U_{i,j}$). $UA_{i,j}$ is updated by Exponential Moving Average (EMA) by every connection, as shown below:

$$UA_{i,j}^{(\text{new})} = UA_{i,j}^{(\text{previous})} \times \Gamma + U_{i,j} \times (1 - \Gamma), \quad (10)$$

where Γ is the weighted constant to balance the influence of current and previous utility.

B. THE POLICY OF HELPER SELECTION

The helper selection is not like the algorithms previously published in [16]–[18], which keep jumping towards the optimal solution. The helper selection is sorting the helpers based on the lately updated utilities from the utility function and choose the optimal set. As shown in Fig 3, the requester generates an AR with an interval. When the requester contacts helper 1, helper 1 updates contact count ($C_{i,j}$), inter-meeting time ($T_{i,j}$), and utility value ($U_{i,j}$) between helper 1 and the requester. Helper 1 relays the AR to the base station and the base station sorts all the helpers based on the utilities derived from the utility function and the best M helpers would be selected as a set ($SH_{\text{AR}_{jk}}$). Note that a helper can only serve a limited number (K) of requesters simultaneously. If BSs arrange too many devices for this helper, the most urgent requesters would be chosen to serve based on T_{jk}^{ela} (elapsed time). Helper 1 belongs to this set, so helper 1 begins to provide the access enhancement service to the requester. When the contact ends, the updated information of duration is sent to the base station for calculating utilities. After the access tolerance (service expired) or the requested access duration is satisfied (finished), the requester begins to generate another AR after an interval.

The effectiveness of the CAP policy has been proven in the previous work [12].

V. EXPERIMENTAL SETTINGS

A. MOVEMENT MODEL

Here the evaluation is based on the medium Helsinki city scenario in Opportunistic Network Emulator (ONE) [27], as shown in Fig. 4. We deploy 4 types of points, namely Point Of Interest (POI), on this map with the consideration of the regularity of movement. For example, nodes in area-4 would randomly move to one of the 22 points by map-based shortest path movement pattern (Dijkstra's algorithm).

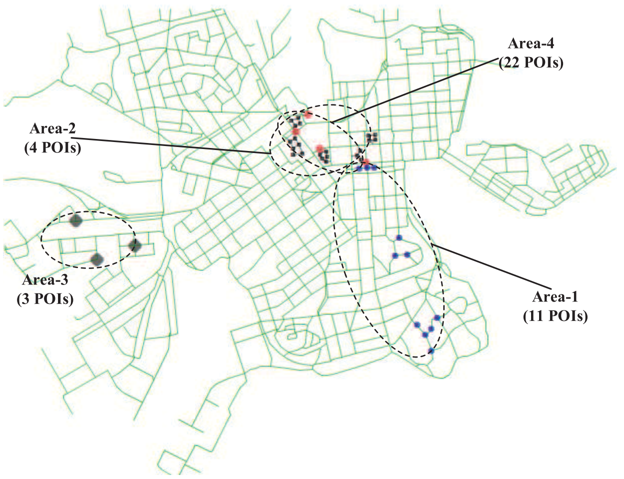


FIGURE 4. Medium Helsinki city scenario with POI.

As shown in TABLE 4, there are four groups of requesters and four groups of helpers which have different probabilities of being in four areas. For example, the requesters of the first group (Re1) spend the majority of time in Area-1 while requesters in group 2 (Re2) and group 3 (Re3) spend more time in Area-2 and Area-3, respectively. Similarly, one hundred helpers also move within 4 POIs with four mobility patterns. We assume that all the users in the system move with speed which varies over the range of 0.5∼1.5 m/s, which is the speed range for normal human walking.

TABLE 4. Probabilities of movement in four areas.

Group	Number	Area-1	Area-2	Area-3	Area-4
Re1	25	0.7	0.1	0.1	0.1
Re2	25	0.1	0.7	0.1	0.1
Re3	25	0.1	0.1	0.7	0.1
Re4	25	0.1	0.1	0.1	0.7
H1	25	0.7	0.1	0.1	0.1
H2	25	0.1	0.7	0.1	0.1
H3	25	0.1	0.1	0.7	0.1
H4	25	0.1	0.1	0.1	0.7

B. PARAMETER SETTINGS

D2D communication can use various air interfaces (in-band underlay D2D, in-band overlay D2D, network-assisted out-band D2D [28], [29], and autonomous out-band D2D). In this paper, the D2D air interface is simplified as transmission range only. The maximal transmission range of in-band D2D air interface (LTE-Direct) is designed to be 500 m [30] because a dedicated in-band D2D technique can use its own bands which are suitable for longer transmit range without interference. However, with the consideration of shadow loss, multi-path effects, and restricted battery life, the transmission range of LTE-Direct implemented in phones in urban areas cannot reach 500 m. Therefore, in this paper, the maximum D2D transmission range is set as 150 m [12].

The one hundred requesters generate network access requests after 10800 s warm-up time,[4] so the regularity of the mobility can be learned by base stations. Without loss of generality, the requesters are assumed busy, which means that they generate new requests after an interval of 60 s[5] after successful access service or after T_{AT} (access tolerance). All the requests are assumed to require 1800 s (half an hour) of access duration. The simulation time is set as 57600 s (16 hours).[6]

VI. DESIGN OF EXPERIMENT

A. CHOICE OF FACTORS AND RESPONSES

The four experimental responses are: Total Helper Access Duration ($R1$), Success Ratio ($R2$), Disruption Duration ($R3$), the Number of Disruption ($R4$); four factors are: Access Tolerance (A), Number of Selected Helpers (B), Helper Access Limit (C), Transmission Range (D).

1) EXPERIMENTAL RESPONSES

Success ratio [12] cannot be used as a sole means of evaluating the connectivity of the system, because the requesters being satisfied always generate more access requests afterwards, consequently bringing more access requests. The increase in access requests may result in increasing uncompleted access requests.

So in this paper, connectivity features of the system are evaluated by two responses: **Success Ratio** and **Total Helper Access Duration** (average total service time of a helper). Service time is not equal to total contact time between helpers and requesters because overloaded helpers can not provide services when contacting requesters; and requesters without an access request also do not need services when contacting helpers.

For disruptions, we are interested in two responses: 1) **The Number of Disruptions:** the average number of disruptions during each access service; 2) **Disruption Duration:** the average duration of each access request.

2) EXPERIMENTAL FACTORS

For experimental factors, the number of devices can influence the density of nodes (nodal density) because the size of the map is fixed. Therefore, the number of selected helpers determines the chance of contacts for a requester. Moreover,

[4]In this scenario, the mobility pattern usually starts to be steady after 1 hour (statistically). The warm-up time is set to be 3 hours.

[5]The interval can be constant as 60 seconds and can also be exponentially distributed with the average of 60 seconds. Preliminary experiments show that these two cases are the same. Generation pattern of the connection requests is irrelevant as there is no subsequent queueing or resource contention.

[6]Removing 3 hours of warm-up time, there are still 13 hours. The trends of all results start to be steady when simulation time reaches 12 hours. 16 hours is therefore chosen to be well in excess of the found 12 hour minimum warm up period.

TABLE 5. Levels of factors.

	$R1$	$R2$	$R3$	$R4$
A (7200 s 10800 s 14400 s 18000 s)	4	$3 \rightarrow 4$	$3 \rightarrow 4$	$3 \rightarrow 4$
B (20 25 30 35)	$2 \rightarrow 4$	$3 \rightarrow 4$	$2 \rightarrow 4$	$3 \rightarrow 4$
C (1 2 3 4)	$\rightarrow 4$	$\rightarrow 4$	$\rightarrow 4$	$\rightarrow 4$
D (50 m 80 m 100 m 150 m)	$3 \rightarrow 4$	4	$3 \rightarrow 4$	4

the number of requesters determines the loads on the helpers. If helpers are overloaded (concurrent requesters exceeding helper access limit), the chance of services are also influenced. In this work, the number of requesters is fixed and the **Number of Selected Helpers** is regarded as an experimental factor.

This paper focuses on the performance of connectivity (services, contacts, and disruptions). Therefore, bandwidth or throughput are not directly considered as input factors. The number of concurrent requesters a helper can serve (**Helper Access Limit**) is another form of bandwidth. Moreover, **Transmission Range** directly determines the connectivity between nodes.

Generally speaking, connectivity is always the goal of network researches, but the performance is always based on service requirements. The service itself can also influence the performance. For example, a network can provide services such as Transmission Control Protocol (TCP) and User Datagram Protocol (UDP). For UDP, the length of packets can influence the results of the experiment; for TCP [31], the setting of TCP slow start (for congestion control in a wired network) can influence the performance in any wireless network scenario. For the PaaS system in this paper, the contacts between helpers and requesters depends on the transient local nodal density and transmission range. The service requirement framework is designed to have two parts, access tolerance and requested access duration, where requested access duration is fixed and **Access Tolerance** is one of the factors in the experiment.

B. DETERMINATION OF THE LEVELS OF THE FOUR FACTORS

Before factorial design, levels for each factor and response pair must be determined. Extensive preliminary single factor experiments[7] were conducted for four factors and four responses by using control variates methodology. For each experiment, the factor was set with many levels (experimental points) in order to get the trend of the response and the regression equation.

Generally, the two-level factorial design is for a linear trend; the three-level factorial design is for a quadratic trend; the four-level factorial design is for a cubic trend, and so on.

As shown in Table 5, levels for different factor and response pairs (combinations of A, B, C, D and $R1$, $R2$, $R3$, $R4$) are various, which means the experiment needs a sophisticated fractional factorial design. Moreover, the levels of factor C has not been determined because C has a jump from 1 to 2, but then, the trends start to fluctuate. The possible reason might be the settings of other factors, which is a drawback of the control variates method and is eliminated by factorial design. The impact of C might be not obvious unless it interacts with other factors. Since simulations are not costly, we unify all the level of factor and response pairs as four (using virtual levels as appropriate) so that it is easy to conduct full factorial experiments.

Full factorial design needs to traverse all the combinations of factors. For one response, there are 256 (4^4) combinations because the four factors have four levels respectively. Here we set the number of repetitions for every combination as 5, so the total number of runs for each response is 1280 (5120 runs for four responses, 1280×4).

For each response, there are two regression equations: the coded regression equation and the actual regression equation. The actual regression equation can get the actual regression value (predicted response) based on a certain set of values of factors. However, the coefficients of different factors cannot indicate the importance of factors since factors represent different dimensions of quantities. Coefficients of coded regression equations can indicate the importance of different factors by comparing changes of responses along with the changes of levels.

C. DATA ANALYSIS

The goal of the experiments in this paper is to find the relationship between the inputs (factors) and the outputs (responses).[8] The advantage of DOE is that DOE can get full information about the system with a minimal number of experimental repetitions. The mathematical expressions (close form regression equations) to describe the relationship between a pairwise factor and response are the accurate and ultimate form of the results of the experiments.

After full factorial experiments, close form regression equations are obtained. Since the regression equations are too long, as shown in Equation (11), at the bottom of the next page, which is an example of the actual regression equation of R1, close form regression equations are shown in Table 6. There are regression equations of four responses. For each response, there are two regression equations: the coded regression equation and actual regression equation.

The parameters of the coded regression equation means the change of the response by one level change of the factor. Note that, levels of factors are set to fit the design space. So coded regression is to compare the importance of different factors regardless of the scale and unit. The actual regression

[7] Details of these experiments are not included in this paper

[8] Effectiveness of the CAP policy can be found in [12].

TABLE 6. Coded (actual) equations of $R1$, $R2$, $R3$ and $R4$.

Coded (Actual) $R1$		Coded (Actual) $R2$		Coded (Actual) $R3$		Coded (Actual) $R4$		=
8.973E+05	(4.651E+05)	1.030E+00	(-7.141E+00)	2.112E+03	(2.692E+03)	1.176E+01	(-1.220E+01)	
-1.456E+04	(1.948E+02)	-2.440E-02	(6.760E-04)	-1.666E+02	(2.352E+00)	-3.050E-02	(4.398E-03)	A
2.099E+05	(-1.134E+05)	-2.500E-03	(1.718E-01)	-6.885E+02	(5.049E+01)	-1.070E-02	(1.490E+00)	B
2.553E+04	(1.460E+05)	-7.200E-03	(3.526E-01)	-3.180E+01	(-2.786E+02)	7.800E-03	(4.645E-01)	C
4.703E+05	(-3.295E+04)	-2.870E-02	(8.352E-02)	-1.630E+03	(-1.483E+02)	-5.260E+00	(-7.322E-02)	D
-5.278E+03	(-1.228E+00)	-4.580E-02	(-8.826E-06)	-1.570E+02	(-4.428E-02)	-3.282E-01	(-6.300E-05)	AB
–	–	-1.080E-02	(-1.500E-05)	-2.591E+01	(-2.452E-02)	-7.180E-02	(-3.400E-05)	AC
-1.454E+04	(-4.529E-01)	-1.182E-01	(-3.977E-06)	-3.313E+02	(-1.389E-02)	-8.922E-01	(-3.700E-05)	AD
-9.913E+02	(2.933E+03)	-4.200E-03	(-2.482E-03)	–	–	–	–	BC
-2.347E+04	(1.641E+03)	-5.440E-02	(-1.184E-03)	3.935E+02	(1.040E+00)	-5.521E-01	(-1.384E-02)	BD
-2.590E+03	(2.579E+03)	-9.600E-03	(-1.164E-03)	6.270E+00	(-9.405E-01)	-2.560E-02	(-3.476E-03)	CD
-2.225E+04	(-1.142E-02)	-6.670E-02	(-2.169E-08)	-1.647E+02	(-6.800E-05)	-4.336E-01	(-9.418E-08)	A^2
-2.825E+04	(2.955E+03)	-1.340E-02	(-1.192E-03)	1.485E+02	(-5.797E-01)	-1.139E-01	(-8.196E-03)	B^2
-4.330E+04	(-9.887E+04)	-9.300E-03	(-5.518E-02)	8.943E+01	(1.502E+02)	–	–	C^2
-1.331E+05	(2.598E+02)	-7.300E-02	(-3.350E-04)	8.808E+02	(1.480E+00)	3.100E+00	(3.119E-03)	D^2
–	–	3.400E-03	(5.596E-08)	–	–	–	–	ABC
–	–	5.010E-02	(2.476E-08)	2.368E+02	(1.170E-04)	5.073E-01	(2.505E-07)	ABD
–	–	9.800E-03	(2.428E-08)	3.292E+01	(8.100E-05)	1.008E-01	(2.488E-07)	ACD
-1.699E+04	(-3.021E+01)	–	–	–	–	–	–	BCD
9.528E+03	(4.356E-05)	2.680E-02	(1.227E-10)	1.272E+02	(5.817E-07)	2.552E-01	(1.167E-09)	A^2B
–	–	8.000E-03	1.831E-10	–	–	–	–	A^2C
2.309E+04	(1.583E-05)	7.990E-02	(5.479E-11)	2.467E+02	(1.692E-07)	6.886E-01	(4.723E-10)	A^2D
–	–	1.100E-02	(3.612E-08)	7.761E+01	(2.560E-04)	–	–	AB^2
–	–	1.330E-02	(1.091E-06)	3.206E+01	(2.639E-03)	–	–	AC^2
–	–	9.570E-02	(7.086E-09)	3.359E+02	(2.500E-05)	9.504E-01	(7.040E-08)	AD^2
-2.923E+04	(-1.039E+01)	1.400E-02	(4.976E-06)	–	–	1.736E-01	(6.200E-05)	B^2D
–	–	4.700E-03	(2.810E-04)	–	–	–	–	BC^2
-9.908E+04	(-5.284E+00)	4.250E-02	(2.269E-06)	-1.372E+02	(-7.319E-03)	5.450E-01	(2.900E-05)	BD^2
–	–	1.050E-02	(9.300E-05)	–	–	–	–	C^2D
-3.343E+04	(-8.916E+00)	5.000E-03	(1.326E-06)	–	–	–	–	CD^2
3.280E+04	(2.083E-07)	4.200E-02	(2.670E-13)	1.235E+02	(7.845E-10)	–	–	A^3
-1.237E+04	(-2.931E+01)	–	–	–	–	–	–	B^3
3.583E+04	(1.062E+04)	9.100E-03	(2.700E-03)	-6.467E+01	(-1.916E+01)	–	–	C^3
-6.060E+04	(-4.847E-01)	6.270E-02	(5.016E-07)	-5.167E+02	(-4.134E-03)	-1.490E+00	(-1.200E-05)	D^3
–	(9.935E-01)	–	(9.803E-01)	–	(9.934E-01)	–	(9.944E-01)	R^2
–	(9.934E-01)	–	(9.798E-01)	–	(9.933E-01)	–	(9.943E-01)	AdjR^2
–	(9.933E-01)	–	(9.793E-01)	–	(9.931E-01)	–	(9.942E-01)	PredR^2

equation is ‘‘actual’’, which can be used to draw regression lines and can be used to do optimization.

From the coefficients of coded regression equations, we can draw the conclusion that the factor D is more important than the other factors for $R1$, $R3$ and $R4$ by at least one order of magnitude. For $R2$ (success ratio), A (access tolerance) is directly associated with service completion since more tolerance time means more chance to satisfy the requested duration. For R4, D is the only factor that matters (by two orders of magnitudes). For R1 and R3, B is the second important factor.

Based on the actual regression equations, regression lines for four factors and four responses are plotted and compared with the results of the experiments in Fig. 5. Although the actual equations are derived from multiple regression analysis, the regression lines can only be plotted by changing one factor with other three factors being fixed. The default fixed values of factors of Fig. 5 are shown in Table 7.

$R2$ is success ratio, the values over 1 are meaningless but it is inevitable that regression lines will produces such results. The dominance and monotonicity of B and D for all responses are proved by regression lines in Fig. 5 and partial derivation.

Overall, the R-Square [14] (R^2) of all responses are high (close to 1). It means that regression lines fit the results of experiments well, which is then verified by the closeness between results and the regression lines. The good performance of the fitting of regression lines is not because of the number of predictors (factors) and over-fit since the Adjusted R-Square [14] and Predicted R-Square [14] are basically the same with R^2. Moreover, the trends of regression lines drawn by actual equations and the coefficients of the coded equations are mutually verified.

$$\begin{aligned} R1 = {} & 1.860E+16+(7.792E+12)A+(-4.536E+15)B+(5.839E+15)C+(-1.318E+15)D+(-4.912E+10)AB \\ & +(-1.811E+10)AD+(1.173E+14)BC+(6.565E+13)BD+(1.032E+14)CD+(-4.567E+08)A^2 \\ & +(1.182E+14)B^2+(-3.955E+15)C^2+(1.039E+13)D^2+(-1.208E+12)BCD+(1.742E+06)A^2B \\ & +(6.334E+05)A^2D+(-4.156E+11)B^2D+(-2.114E+11)BD^2+(-3.566E+11)CD^2+(8.332E+03)A^3 \\ & +(-1.172E+12)B^3+(4.247E+14)C^3+(6.991\times 10^3)D^3 \end{aligned} \tag{11}$$

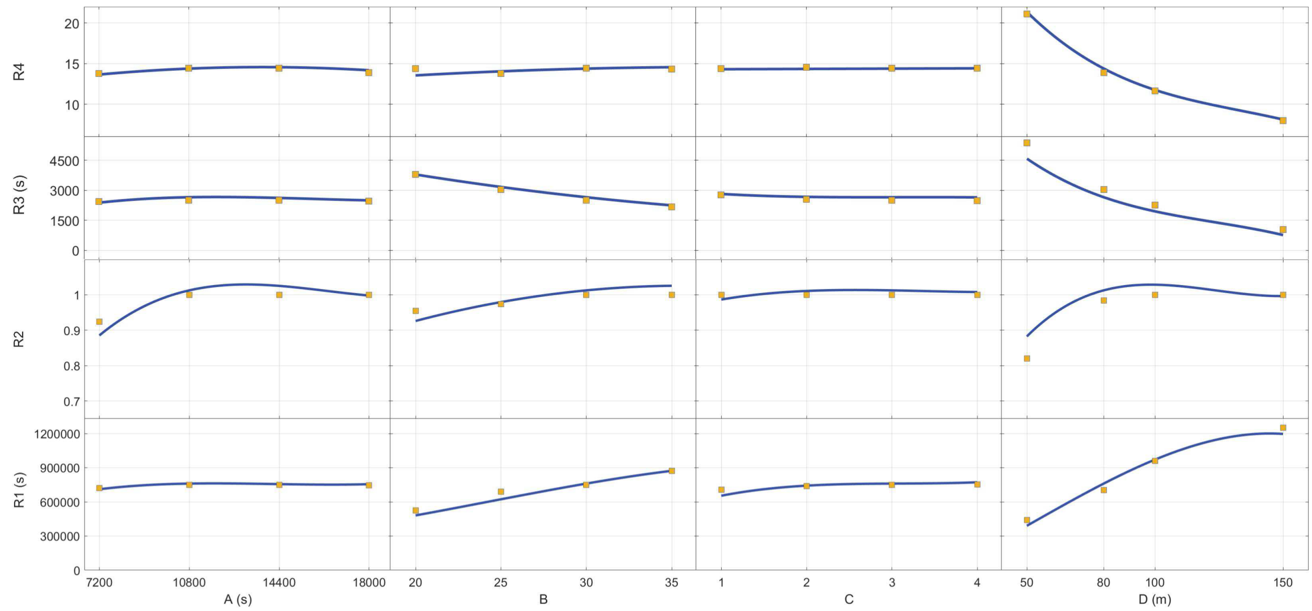


FIGURE 5. Regression lines of four responses and four factors.

TABLE 7. Default levels of factors of Fig. 5.

Factors	Values
Access Tolerance ($R1$)	10800 s
Helper Access Limit ($R2$)	3
The Number of Selected Helpers ($R3$)	30
Transmission Range ($R4$)	80 m

D. OPTIMIZATION OF THE FACTORS

In this section, we address the optimization of 4-tuple responses $\{R_1, R_2, R_3, R_4\}$, which is formulated as:

$$\begin{aligned} &\max R_1(\mathbf{x}) \\ &\max R_2(\mathbf{x}) \\ &\min R_3(\mathbf{x}) \\ &\min R_4(\mathbf{x}) \\ &\text{s.t. } \mathbf{x} = (A, B, C, D)^{\mathrm{T}} \in \mathbf{R}^4, \\ &\quad 7200 \leq A \leq 18000, \\ &\quad 20 \leq B \leq 35, \\ &\quad 1 \leq C \leq 4, \\ &\quad 50 \leq D \leq 150. \end{aligned} \tag{12}$$

It is a non-linear Multi-Objective Optimization Problem (MOOP) [15]. A common methodology is to convert it into single-objective optimization problems. The importance of four objectives of responses in this system is difficult to weight because different responses have different dimensions, scales, and units, which means the method of *weighted sum* [15] is not feasible. Since the four responses are intrinsically consistent, we chose the ϵ-constraint method [15] which is formulated as:

$$\begin{aligned} &\max/\min R_j(\mathbf{x}) \\ &\text{s.t. } \mathbf{x} = (A, B, C, D)^{\mathrm{T}} \in \mathbf{R}^4, \\ &\quad R_i(\mathbf{x}) \geq \epsilon_i \text{ or } R_i(\mathbf{x}) \leq \epsilon_i, \\ &\quad \forall\, i \in \{1, 2, 3, 4\} \backslash \{j\}, \\ &\quad 7200 \leq A \leq 18000, \\ &\quad 20 \leq B \leq 35, \\ &\quad 1 \leq C \leq 4, \\ &\quad 50 \leq D \leq 150, \end{aligned} \tag{13}$$

where ϵ_i is the lower or upper bound in the last optimization procedure and is one of the non-linear constraints. In each step of iterations, there is only one optimization problem and three constraints which have been optimized in previous steps.

For the ϵ-constraint method, there might be no iteration result that is better than the other iterations. In other words, the solution set is a non-dominated set [15], which is referred as a Pareto-optimal set [15].

By using the platform proposed in [32] based on genetic algorithms [33], it is convenient to produce a Pareto-optimal set, which is listed in Table 8. The differences between iterations are small and a specific iteration can be chosen

TABLE 8. Producing a Pareto-optimal set of the factors.

Iteration	A (s)	B	C	D (m)	$R1$ (s)	$R2$	$R3$ (s)	$R4$
1	12137	32	2	86	8.6822E+05	1.034	2227.767	13.52
2	12825	35	3	150	1.1838E+06	0.972	560.251	8.174
3	12140	33	2	131	1.2116E+06	0.987	1046.72	9.248
4	12314	32	2	94	9.5645E+05	1.031	1927.994	12.378
5	12096	32	3	92	9.4273E+05	1.032	1992.645	12.673
6	12143	32	2	86	8.6477E+05	1.034	2240.447	13.499
7	12195	33	3	120	1.1684E+06	0.998	1267.368	9.939
8	12188	33	2	108	1.0855E+06	1.016	1539.796	10.928
9	12302	32	3	98	9.8728E+05	1.027	1840.091	11.998
10	12060	34	2	102	1.0579E+06	1.021	1626.752	11.53
11	12179	33	2	126	1.1929E+06	0.992	1157.145	9.566
12	12137	32	2	87	8.8695E+05	1.034	2172.969	13.358
13	12174	32	3	105	1.0558E+06	1.018	1641.25	11.219
14	12362	33	2	99	1.0234E+06	1.023	1716.856	11.783
15	12143	32	2	90	9.1054E+05	1.034	2081.453	12.928
16	11876	35	2	136	1.2283E+06	0.98	902.137	8.958
17	12210	30	4	150	1.1744E+06	0.986	721.616	8.102

based on particular needs. A confirmation experiment for the second iteration was conducted and the results verify the table (RA1 to RA4: 1.2012E+06 s, 1, 612.323 s, and 8.012).

VII. CONCLUSION

This paper has reported how we built the PaaS system and the CAP policy, and has shown how we conducted DOE and further levels of investigation revealing the relationship between inputs (factors) and outputs (responses) of the system. The results further reveal the influence of the four factors, e.g., service tolerance, the number of helpers allocated, the number of concurrent devices supported by each helper, and the transmit range, on the PaaS through an opportunistic D2D relay. This work contributes to the introduction of DOE to the field of mobile computing and D2D. With the precise close form multiple regression equations, it is easy to predict the influence of any change of the four factors within a certain domain.

With the multiple close form regression equations, we found that two factors (transmission range and the selected helper number) have a dramatic influence on performance. Since service requirements and the ability of devices are not controlled by the service operators, transmission range and user density are the most significant metrics. There is always a particular value of transmission range which fits a particular user density. For example, in the city center, a small range is enough and a long transmission range may disturb proximal cells. In the rural area, the transmission range needs to be long and can be long.

Our future work will focus on testing a large number of devices in practical scenarios with the concern of energy efficiency for service discovery.

REFERENCES

[1] Z. Zhou, C. Sun, R. Shi, Z. Chang, S. Zhou, and Y. Li, "Robust energy scheduling in vehicle-to-grid networks," *IEEE Netw.*, vol. 31, no. 2, pp. 30–37, Mar. 2017.

[2] M. A. Habibi, M. Nasimi, B. Han, and H. D. Schotten, "A comprehensive survey of RAN architectures toward 5G mobile communication system," *IEEE Access*, vol. 7, pp. 70371–70421, 2019.

[3] M. F. Hossain, A. U. Mahin, T. Debnath, F. B. Mosharrof, and K. Z. Islam, "Recent research in cloud radio access network (C-RAN) for 5G cellular systems—A survey," *J. Netw. Comput. Appl.*, vol. 139, pp. 31–48, Aug. 2019.

[4] P. Duan, Y. Jia, L. Liang, J. Rodriguez, K. M. S. Huq, and G. Li, "Space-reserved cooperative caching in 5G heterogeneous networks for industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 14, no. 6, pp. 2715–2724, Jun. 2018.

[5] D. Chatzopoulos, M. Ahmadi, S. Kosta, and P. Hui, "OPENRP: A reputation middleware for opportunistic crowd computing," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 115–121, Jul. 2016.

[6] D. Chatzopoulos and P. Hui, "Readme: A real-time recommendation system for mobile augmented reality ecosystems," in *Proc. ACM Multimedia Conf. (MM)*, New York, NY, USA, 2016, pp. 312–316.

[7] C. Bermejo, D. Chatzopoulos, and P. Hui, "How sustainable is social based mobile crowdsensing? An experimental study," in *Proc. IEEE 24th Int. Conf. Netw. Protocols (ICNP)*, Nov. 2016, pp. 1–6.

[8] J. Xie, Y. Jia, Z. Chen, Z. Nan, and L. Liang, "D2D computation offloading optimization for precedence-constrained tasks in information-centric IoT," *IEEE Access*, vol. 7, pp. 94888–94898, 2019.

[9] K. Sucipto, D. Chatzopoulos, S. Kosta, and P. Hui, "Keep your nice friends close, but your rich friends closer—Computation offloading using NFC," 2016, *arXiv:1612.03000*. [Online]. Available: https://arxiv.org/abs/1612.03000

[10] N. Allurwar, B. Nawale, and S. Patel, "Beacon for proximity target marketing," *Int. J. Eng. Comput. Sci*, vol. 15, pp. 16359–16364, 2016.

[11] D. Chatzopoulos, S. Gujar, B. Faltings, and P. Hui, "LocalCoin: An ad-hoc payment scheme for areas with high connectivity: Poster," in *Proc. 17th ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)*, New York, NY, USA, 2016, pp. 365–366.

[12] T. Guo, G. Huang, J. Schormans, T. Wang, and Y. Cao, "CAP: A ContAct based proximity service via opportunistic device-to-device relay," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2017, pp. 337–342.

[13] B. Han, P. Hui, V. Kumar, M. V. Marathe, G. Pei, and A. Srinivasan, "Cellular traffic offloading through opportunistic communications: A case study," in *Proc. 5th ACM Workshop Challenged Netw.*, 2010, pp. 31–38.

[14] D. C. Montgomery, *Design and Analysis of Experiments*. Hoboken, NJ, USA: Wiley, 2017.

[15] K. Miettinen, *Nonlinear Multiobjective Optimization* (International Series in Operations Research & Management Science). New York, NY, USA: Springer, 1999.

[16] Z. Zhou, M. Dong, K. Ota, J. Wu, and T. Sato, "Energy efficiency and spectral efficiency tradeoff in device-to-device (D2D) communications," *IEEE Wireless Commun. Lett.*, vol. 3, no. 5, pp. 485–488, Oct. 2014.

[17] Z. Zhou, K. Ota, M. Dong, and C. Xu, "Energy-efficient matching for resource allocation in D2D enabled cellular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5256–5268, Jun. 2017.

[18] Z. Zhou, M. Dong, K. Ota, G. Wang, and L. T. Yang, "Energy-efficient resource allocation for D2D communications underlaying cloud-RAN-based LTE—A networks," *IEEE Internet Things J.*, vol. 3, no. 3, pp. 428–438, Jun. 2016.

[19] S. Wang, W. Guo, Z. Zhou, Y. Wu, and X. Chu, "Outage probability for multi-hop D2D communications with shortest path routing," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 1997–2000, Nov. 2015.

[20] Z. Zhou, G. Ma, C. Xu, Z. Chang, and T. Ristaniemi, "Energy-efficient resource allocation in cognitive D2D communications: A game-theoretical and matching approach," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[21] M. Waqas, Y. Niu, Y. Li, M. Ahmed, D. Jin, S. Chen, and Z. Han, "Mobility-aware device-to-device communications: Principles, practice and challenges," *IEEE Commun. Surveys Tuts.*, to be published.

[22] Y. Pan, C. Pan, Z. Yang, M. Chen, and J. Wang, "A caching strategy towards maximal D2D assisted offloading gain," *IEEE Trans. Mobile Comput.*, to be published.

[23] Y. Cao and Z. Sun, "Routing in delay/disruption tolerant networks: A taxonomy, survey and challenges," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 654–677, 2nd Quart., 2013.

[24] L. Pu, X. Chen, J. Xu, and X. Fu, "D2D fogging: An energy-efficient and incentive-aware task offloading framework via network-assisted D2D collaboration," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3887–3901, Dec. 2016.

[25] K. Vadde and V. Syrotiuk, "Factor interaction on service delivery in mobile ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 7, pp. 1335–1346, Sep. 2004.

[26] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless video content delivery through distributed caching helpers," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1107–1115.

[27] A. Keränen, J. Ott, and T. Kärkkäinen, "The one simulator for DTN protocol evaluation," in *Proc. 2nd Int. Conf. Simulation Tools Techn.*, 2009, p. 55.

[28] X. Lin, J. G. Andrews, and A. Ghosh, "A comprehensive framework for device-to-device communications in cellular networks," 2013, *arXiv:1305.4219*. [Online]. Available: https://arxiv.org/abs/1305.4219

[29] A. Asadi and V. Mancuso, "WiFi direct and LTE D2D in action," in *Proc. IFIP Wireless Days (WD)*, Nov. 2013, pp. 1–8.

[30] *LTE Direct Always-on Device-to Device Proximal Discovery*, Qualcomm Technol., Chennai, Indian, 2014.

[31] Z. Jian, G. Tianhao, L. Jilong, and Y. Mingwu, "Adaptive flow control proxy for enhancing downlink TCP throughput in infrastructure WLANs," in *Proc. 14th Int. Symp. Commun. Inf. Technol. (ISCIT)*, Sep. 2014, pp. 56–60.

[32] Y. Tian, R. Cheng, X. Zhang, and Y. Jin, "PlatEMO: A MATLAB platform for evolutionary multi-objective optimization [educational forum]," *IEEE Comput. Intell. Mag.*, vol. 12, no. 4, pp. 73–87, Nov. 2017.

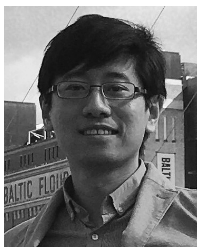
[33] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, vol. 16. Hoboken, NJ, USA: Wiley, 2001.

TIANHAO GUO received the M.Phil. degree from the State Key Laboratory of Integrated Services Network, Xidian University, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, U.K. From 2016 to 2017, he was working as a Visiting Ph.D. Student with the Department of Computer and Information Sciences, Northumbria University, U.K. His research interests focus on D2D and mobile computing.

JOHN SCHORMANS received the B.Sc. degree in electrical and electronic engineering, in 1984, having been a sponsored as an undergraduate with the avionics and communications company GEC Marconi, and the Ph.D. degree in 1990. In 1987, he was further sponsored by GEC to join the Telecommunications Research Group, Queen Mary, University of London, under Prof. Laurie Cuthbert. He returned to GEC to work on future airborne telecommunications systems, and then joined Anglia Polytechnic University, in 1992 as a lecturer. He returned to the Department of Electronic Engineering, Queen Mary University of London as a Lecturer in 1995, where he was promoted to a Senior Lecturer in 2000. He has been a BT Short Term Research Fellow working on the evaluation of packet switched telecommunications systems at BT research Labs, where he is now working with the London based start-up company on research aimed at measuring QoS metrics and evaluating user-experience in mobile networks, and as a member of their Advisory Board. He is currently the Deputy Director of education with the School of Electronic Engineering and Computer Science, Queen Mary University of London. His researches focus on the analysis, simulation and measurement of network performance metrics and user quality-of-experience in packet networks, wired, and wireless. He is a member of the Academic Committee of the QMUL-BUPT Joint Programme, with pathway responsibility for telecommunications as a Chair of Telecoms Programme with the School of Electronic Engineering and Computer Science.

LEXI XU received the Ph.D. degree from the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, U.K., in 2013. He is currently a Senior Engineer with the China Unicom Network Technology Research Institute and a Senior Researcher with National Engineering Laboratory for Next Generation Internet Broadband Service Application. He is the Industrial Tutor of BUPT, NHU, and SAU. He has been actively involved in 30 big data research and network optimization projects. He has published more than 80 technical articles and applied more than 20 patents, and one book and two international conferences proceedings. His research interests include big data, self-organizing networks, smart mobility, and radio resource management in LTE-A/5G. He is also a China Unicom Delegate in ITU, ETSI, and CCSA. He served as a Workshop Chair/TPC Co-Chair of ICSINC, IEEE ISCIT, 5GWN, and IEEE IUCC.

JINZE WU received the Ph.D. degree from the State Key Laboratory of Quantum Optics and Quantum Optics Devices, Institute of Opto-Electronics, Shanxi University, China, in 2019. He currently holds the postdoctoral position with the Interdisciplinary Center of Quantum Information, State Key Laboratory of Modern Optical Instrumentation, and also with the Zhejiang Province Key Laboratory of Quantum Technology and Device of Physics Department, Zhejiang University, China. His research interest focuses on quantum optics and its applications in precise metrology.



YUE CAO received the Ph.D. degree from the Institute for Communication Systems (ICS), University of Surrey, U.K., in 2013. He was the Research Fellow with ICS, University of Surrey, U.K., and also a Lecturer and Senior Lecturer with the Department of Computer and Information Sciences, Northumbria University, U.K. He has been the International Lecturer with the School of Computing and Communications, Lancaster University, U.K. His research interest focuses on intelligent transport systems. He is an Associate Editor of IEEE Access and *KSII Transactions on Internet and Information Systems*.