This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2829751, IEEE Transactions on Industrial Informatics

Joint Cloudlet Selection and Latency Minimization in Fog Networks

Mudassar Ali[1,4], Nida Riaz[1], Muhammad Ikram Ashraf [3], Saad Qaisar[1], Muhammad Naeem[2]

***Abstract*—Mobile edge or fog computing is a network architecture that brings the functionality of conventional centralized cloud to the edge nodes which are in close proximity of the end devices in internet of things (IoT) network. Fog networks have many advantages over traditional cloud networks, such as increased bandwidth utilization, enhanced security and privacy, better energy efficiency, improved performance and support for mobility. The most critical requirement of a fog architecture is to minimize the end to end latency in IoT networks, particularly in scenarios where large number of end devices (IoT nodes) and distributed computing fog nodes (cloudlets) are present. In this article, we formulated an optimization problem for joint cloudlet selection and latency minimization in fog network, subject to maximum work load and latency constraints. The problem can be epitomized as many-to-one matching game in which IoT nodes and cloudlets rank each other in order to minimize the latency. The proposed game belongs to a class of matching games with externalities. We propose an algorithm to solve this game which gives distributed and self-organizing solution. Extensive simulations have been carried out to validate the proposed algorithm.**

***Index Terms*—IoT, Fog networks, Resource management, Matching with externalities**

I. Introduction

Internet of things (IoT) is making our world smarter by integrating advance sensing, computing and communication technologies [1]. A large number of the heterogeneous sensing devices have instilled in our daily lives providing functions from monitoring home security and power consumption to monitoring our health, from enabling smart billboards to tracking our shipments, from revolutionizing agriculture to modernizing our traffic systems and many more [2]. Use of IoT for industrial informatics is a concept well established in fourth industrial revolution (industry 4.0), where sensors will be used for operating industrial robotics as well as for tracking and monitoring of industrial products. Sensors are also used for assisting in proactive maintenance at the industrial floors. It is predicted that more than 50 billion IoT nodes will be present by 2020 [3], all performing different functions, having different quality of service (QoS) requirements.

Conventional heterogeneous network architectures do not have the capabilities to handle the enormous amount of data traffic and huge computational requirements of billions of IoT nodes. Thus, to meet the service requirements of these IoT nodes, cloud computing is considered as a promising architecture, with its ability to provide flexible resources to applications on resource limited IoT nodes. However, many challenges remain unsolved, such as mobility support, location-awareness, and ultra low-latency requirements. An evolved architecture, named as the Fog network, has emerged to overcome these challenges by migrating the computing function from the cloud to the edge of the network. The concept of fog is not meant to replace cloud, but rather to complement it, and to provide additional functionalities required by IoT nodes.

[1] SEECS, National University of Sciences and Technology, Pakistan
[2]COMSATS Institute of Information Technology, Wah Campus, Wah, Pakistan
[3] University of Oulu , Finland
[4]University Engineering and Technology, Taxila 47050, Pakistan
Corresponding E-mail: muhammadnaeem@gmail.com

Fog network is three-tier architecture including cloud server, cloudlets or fog nodes and end user devices or IoT nodes as shown in Fig. 1. Cloudlets are powerful devices present at the edge, that have ample resources of storage, communication and computation. Fog networking or fog computation is a concept to use these resources at the edge to descend the concept of cloud near the edge user and provide a decentralized system [4]. Unique characteristics provided by fog networking includes location awareness, real time processing of data close to the edge, support for mobility, resource pooling and support for heterogeneity [5], [6]. These unique characteristics of fog network make it suitable for many IoT business applications [7]. It's support for mobility and localization can provide a framework for device to device (D2D) communication like an efficient and reliable vehicular ad-hoc networks (VANET) [8] [9]. Location awareness provided by fog can be used to improve QoS for many application domains like improving web user experience [10]. Network control and management can be provided close to the users rather than controlled primarily by core network gateways [11], [12]. Processing confidential information in near edge devices has leveraged fog solution as a more secure solution than cloud as presence of cloudlets near the user end not only ensure confidentiality but also makes it helpful for real time data analysis [13]. Fog networking will be an enabling technology for industry 4.0 [14].

While the deployment of fog networks provide substantial benefits over legacy cloud networks, it also presents numerous challenges in terms of network modeling, resource management and user association. Efficient distributed resource management techniques are inevitable for fog networks, which can handle dense deployment of cloudlets serving huge number of IoT nodes. Some work has already been done to provide a mathematical framework for optimizing resource allocation in fog networks, examples include search based approaches and non-linear integer formulations solved using conventional

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2829751, IEEE Transactions on Industrial Informatics

TABLE I: Summary of Literature Review: SCNs - Small Cell Networks, D2D - Device to Device Communication, UE - User Equipment, MILP - Mixed Integer Linear Programming, V2V - vehicle to vehicle, QoS - Quality of Service

Ref	Cloud server	Cloudlets	SCNs, D2D	Problem	Solutions / Techniques
[15]	✓			Joint resource allocation optimization along with minimization of carbon footprint	Proximal algorithm
[16]	✓			Optimal workload allocation to minimize the power consumption subject to QoS and bandwidth constraints	Primal decomposition / Hungarian algorithm to solve the convex problem
[17]	✓			Task scheduling and resource management to minimize task completion time is modeled as MINLP	3-stage heuristic algorithm.
[18]		✓		Maximize network utility by lowering service delay and cost by optimizing resource allocation and service rate	Many-to-many matching game and Stackelberg game
[19]		✓		Optimal UE and fog node association under the constraints of blocking probability and latency	Simplex algorithm having three different policies
[20]		✓		Optimal task distribution among peer nodes to minimize latency	Matching theory
[21]		✓		Optimal allocation of computational and radio resources to minimize power consumption, modeled as non-convex optimization problem	2-stage heuristic algorithm.
[22]		✓		Optimal access node selection under different access service requirements and front haul delay cost	Evolutionary game theory.
[23]	✓	✓		Optimizing the task distribution among fog nodes to reduce the CPU execution time and total memory usage	Bees life algorithm, an algorithm based on bees life model.
[24]	✓	✓		Joint optimization of bandwidth and computing resources in cloudlet-based mobile cloud computing environment	Stackelberg game providing an end-to-end solution.
[25]	✓	✓		Optimizing the task distribution among cloudlets (local and global) to get least power consumption and latency modeled as MILP	GNU linear programming kit.
[26]	✓	✓		Optimizing the task distribution among cloudlets to get least power consumption and latency	Centralized resources allocation using a proxy server
[27]	✓	✓		Optimizing the task distribution among cloudlets to achieve power and bandwidth efficiency	Exhaustive search, Heuristic algorithm.
[28]	✓	✓		Cloud selection optimization to minimize power consumption	Dynamic cloud selection algorithm based on exhaustive
[29]			✓	Optimal resource allocation and user association to minimize latency while considering security of communication	Dual decomposition (Prima-dual method)
[30]			✓	Optimal user association for a tradeoff between power utilization and latency	Distributed network aware heuristic algorithm
[31]			✓	Optimal resource allocation to attain energy efficiency.	Outer approximation algorithm
[32]			✓	Optimal resource allocation in V2V communication to minimize minimize latency subject to QoS constraint	Distributed algorithm based on matching game with externalities
[33]			✓	Optimal spectrum allocation in SCNs subject to cost and channel characteristics	Many-to-one matching game
[34]			✓	Optimal joint user association, channel assignment and power allocation	Matching game
[35]			✓	Optimal context aware resource allocation for SCNs	Matching theory with externalities
[36]			✓	Optimal context aware resource allocation for D2D communication	Matching theory
[37]			✓	Optimal resource allocation combining both social and physical characteristics of users in D2D communication	Matching theory (3-D matching)

algorithms like proximal algorithms and primal/dual methods. These works are discussed in details in next section.

The article is organized as follows: Detailed survey of the literature has been presented in section II which provides existing works on resource management in fog networks and their limitations. Section III presents system model and problem formulation. Section IV gives the proposed algorithm, proof of its convergence and complexity analysis . Simulation setup and numerical results are discussed in section V. Lastly, we concluded the article in section VI.

II. Literature Review

Node selection and resource management for wireless networks has been studied widely under different contexts, such as SCNs, D2D, and fog networks, summary of which is given in Table I. Literature can be found on resource allocation for SCNs, D2D communication and for our particular area of interest, i.e., fog networking. Game theory has been used for resource allocation in wireless networks because of its distributed nature, fast convergence and multi-objective optimization approach [38]. It is a mathematical framework that provides optimal resource allocation for the players in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2829751, IEEE Transactions on Industrial Informatics

a game, by maximizing the utility of all the players. It provides stable and self-organizing approaches that can be used to model the interactions among different nodes of the network having different QoS requirements [38]. Matching games has been considered as popular technique to model resource management and user associations in SCNs and D2D communications. Other approaches being used are outer approximation algorithm (OAA), primal/dual decomposition, simplex algorithm and some heuristic algorithms. Authors in [32], [36], [37] have studied resource allocation in particular domain of D2D communication and solved it using matching theory while [33–35] have applied matching theory for resource management in SCNs.

Some works focus user association and resource allocation in fog networks. As discussed earlier fog network is three-tier architecture consisting cloud sever, cloudlets and end devices/IoT nodes. Work on fog networks optimization can be categorized in two domains, resource management at cloud server and resource management at cloulets. In [15–17] authors have presented optimal resource management solutions at could server, whereas in [18–22] authors have discussed resource management at cloudlets for end devices in fog networking architecture. In [15] authors presented an optimal resource allocation scheme by using the proximal algorithm to minimize the energy consumed at the data center along with maximizing the user utility. They have used proximal algorithm instead of dual decomposition because of its fast convergence, insensitivity to step sizes and modest accuracy. In [16] authors tried to optimize the system to get balanced delay and power consumption by optimizing the workload allocation. They used primal/dual decomposition to solve the convex optimization problem in order to reach a sub-optimal solution. Similar problem has been catered in [17], where authors propose a method for task scheduling between embedded devices and computation server to minimize the task completion time. In [18] and [20] authors have solved the problem of resource allocation by using matching theory, in [18] they have considered constraints of service delay and cost while in [20] they only considered latency as a constraint. An algorithm based on evolutionary game theory has been proposed in [22] to optimally select user access mode in fog radio access network (F-RAN) scenario. The authors in [19] and [21] try to solve similar problems in F-RAN using simplex method. In [23] authors have presented an interesting solution using bee's life model to formulate task distribution among end devices, fog nodes and cloud server to optimize CPU execution time and memory usage, hence providing an end to end (E2E) solution. In [24] authors have used game theory for joint optimization of bandwidth and computing resources in distributed manner.

A. Contributions

We can conclude from literature review that matching theory has been extensively used for resource optimization in domains of D2D communication and SCN, however, it has not been explored in detail particularly for fog networks except for few recent works [18], [20], [22], which only cover constraints of latency while optimizing the resource management. We argue that matching theory can be used to model the fog network which in turn can be used to optimize node selection and resource allocation. In this article we present:

1) An optimization problem for joint cloudlet selection and latency minimization in fog network, subject to maximum work load and latency constraints.
2) A projection of optimization problem as many-to-one matching game, in which IoT nodes and cloudlets are considered as players.
3) A distributed and self-organizing algorithm to solve the matching game which results in overall improved throughput and latency.
4) Extensive simulations results to validate the proposed solution.

III. System Model and Problem Formulation

We consider a fog network composed of F cloudlets, geographically distributed and connected with each other as well as with distant cloud server as illustrated in Fig. 1. There are U number of IoT nodes (i.e., sensors, actuators, UEs, smart meters etc.) present, which are using applications requiring different data rate, QoS level, delay and context. IoT nodes are connected with cloudlets on a wireless interface. Cloudlets are connected with eachother via a dedicated wireless interface for information exchange and coordination, i.e., the communication among cloudlets operate at frequencies that are different than those used for communication of cloudlet with the IoT nodes. A dedicated high speed link is used to connect all cloudlets to the cloud server. We consider the *downlink* transmission of a cloudlet network, where power is uniformly divided among IoT nodes connected to a cloudlet, and all cloudlets use same band to communicate with the IoT nodes, so IoT nodes will experience interfere from near by cloudlets.

Let $\mathcal{F}$ and $\mathcal{U}$ be the sets of cloudlets and IoT nodes respectively. Let's assume an IoT node $u \in \mathcal{U}$ is in the coverage of multiple cloudlets. Each cloudlet is assigned to multiple IoT nodes, whereas one IoT node is assigned to only one cloudlet. Let $x_f^u \in \{0, 1\}$ be the assignment indicator where $x_f^u = 1$ indicates that IoT node u assigned to cloudlet f, otherwise, $x_f^u = 0$.

Each cloudlet $f \in \mathcal{F}$ can handle maximum work load of $W_f = \sum_{\forall u} w_u^f$ from the IoT nodes in the network. Workload w_u^f is function of data processing requests from IoT nodes in bits. Cloudlets can share work load of each other. The communication link between IoT node and cloudlet is considered to be one-hop. Each one hop link between IoT node and a cloudlet has different latency l_f, which depends bandwidth available, workload w_f^u of cloudlet and data rate r_f at cloudlet as follows [39]

$$l_f = x_f^u \frac{w_f^u}{r_f}, \tag{1}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2829751, IEEE Transactions on Industrial Informatics

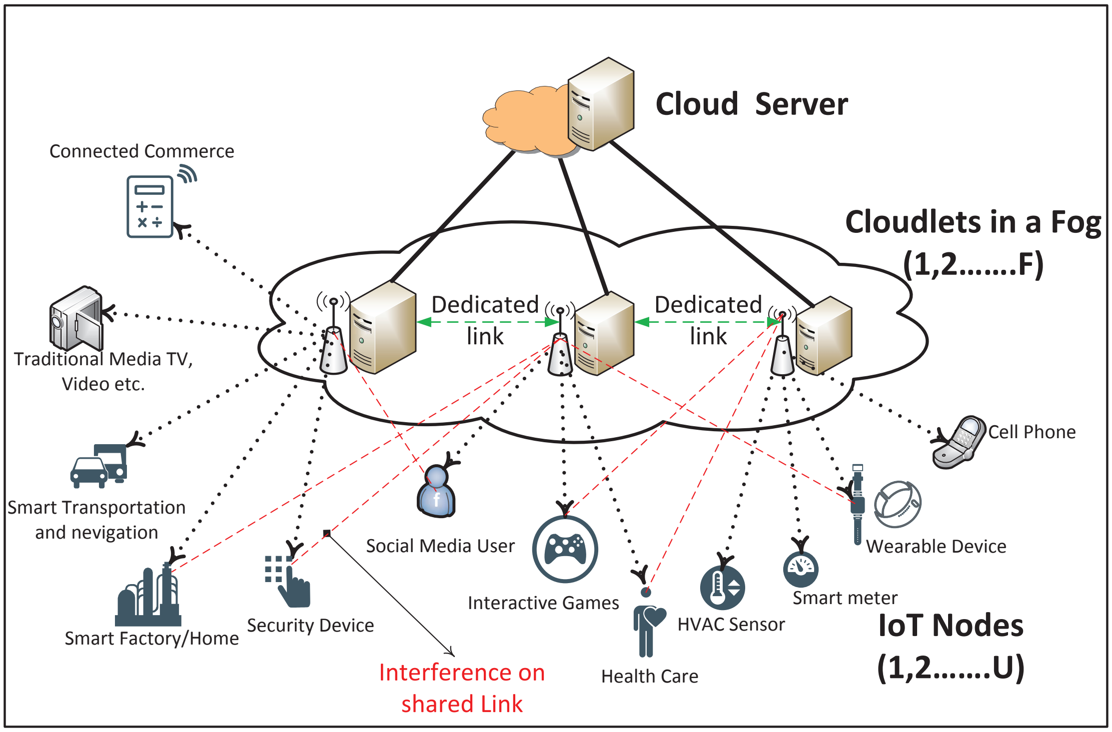

Fig. 1: Fog Network Scenario

Data rate r_f at a cloudlet is defined as the sum of data rate of all the IoT nodes connected to that cloudlet and is given by:

$$r_f = \sum_{u \in U_f} r^u, \tag{2}$$

where r^u is data rate of an IoT node, which is defined as follows:

$$r_u = \beta \ log_2 \ (1 + SINR_u), \tag{3}$$

where β is the total bandwidth assigned to IoT node u, and signal to interference plus noise ratio (SINR) at IoT node u and calculated as:

$$SINR_u = x_f^u \frac{p_f^u h_f^u}{\sum_{f' \neq f \in \mathcal{F}} p_{f'}^u h_{f'}^u + N_o}, \tag{4}$$

where p_f^u is power received by IoT node u from cloudlet f, h_f^u is channel gain between IoT node u from cloudlet f, p_f^v and h_f^v denote the power received and channel gain of interfering IoT node v respectively. The channel gain h_f^v is defined as: $h_f^u = \bar{h}_f^u \zeta G_o \left(\frac{d_o}{d}\right)^\alpha$, where G_o be the antenna gain and $\zeta 10^{\zeta/2}$ be the lognormal shadowing, ζ is the zero mean gaussian random variable with standard deviation σ [40], d is the distance between transmitter and receiver, d_o denotes antenna far field reference distance, α is path loss exponent, $\bar{h}_f^u$ is Rayleigh random variable.

In this paper we formulated an optimization problem, which aims to minimize the latency of the fog network by selecting selecting a cloudlet for particular task subject to maximum work load and maximum latency constraints. Mathematically we have:

$$\begin{aligned} \min_{l,w} \quad & \sum_{u \in \mathcal{U}} \sum_{f \in \mathcal{F}} x_f^u l_f^u \\ \text{subject to} \quad & \\ & C1 : \sum_{f \in \mathcal{F}} x_f^u \leq 1, \ \forall \ u \in \mathcal{U}, \\ & C2 : \sum_{u \in \mathcal{U}} w_f^u \leq W_f, \ \forall \ f \in \mathcal{F}, \\ & C3 : \sum_{u \in \mathcal{U}} l_u^f \leq L, \ \forall \ f \in \mathcal{F}. \end{aligned} \tag{5}$$

Constraint $C1$ ensures that an IoT node will be connected to a single cloudlet at a time. Constraint $C2$ ensures that workload at a cloudlet will not exceed a threshold W_f. The maximum latency allowed at any IoT node in fog network i.e., the upper bound on latency is considered as L which depicted in constraint $C3$.

Our objective is to propose an efficient and self-organizing IoT node association scheme for fog networks. In conventional networks, end devices (IoT nodes) are attached to the base station based on the maximum signal to interference plus noise ratio (SINR) or maximum received signal strength (RSS) [41], ignoring the latency experienced by IoT nodes and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2829751, IEEE Transactions on Industrial Informatics

workload on the base stations. This motivates for investigating novel IoT node association mechanism in fog networks which considers end to end latency requirements as well as work load capacities of cloudlets.

IV. Proposed Approach

The problem formulation in equation (5) is mixed integer non-linear programming problem. Search space of (5) increases exponentially with the number of IoT nodes. Exhaustive search followed by convex optimization can be applied to reach a an optimal solution, but the complexity of exhaustive search increases exponentially with increase in number of IoT nodes. In this paper we propose matching theory based algorithm, which gives distributed and self-organizing solution to problem. The problem in equation (5) can be considered as many-to-one matching problem, where an IoT node can be matched to one cloudlet and while a cloudlet can have multiple IoT nodes. IoT nodes are accessing a fog node using combined wireless resources and association of a node will affect the utility of all other nodes. Matching with externalities is being used from the different classes of matching theory defined for wireless networks [38], as it covers the interdependency of IoT nodes while making an association. To this end, distributed and self organizing framework is proposed to optimally associate users for joint utility maximization.

A. Utilities of IoT nodes and Cloudlets

For cloudlet allocation, each IoT node needs to identify the set of cloudlets that ensure QoS requirements and reduce their latency. Thus, we define a suitable utility function for any IoT node $u \in \mathcal{U}$ over cloudlet $f \in \mathcal{F}$ for a given matching Υ as follows:

$$U_u(\Upsilon) = -l_u(\Upsilon), \tag{6}$$

where l_u is latency experienced by an IoT node in a IoT node-cloudlet pair. So the utility of a cloudlet is the sum of utilities of IoT nodes using that cloudlet for the given matching Υ :

$$U_f(\Upsilon) = \sum_{u \in \mathcal{U}} -l_u(\Upsilon). \tag{7}$$

Finally, we define a network utility for given matching Υ as $\gamma(\Upsilon) = -\sum_{u\in\mathcal{U}} \sum_{f\in\mathcal{F}} x_f^u l_f^u$. Having defined such utilities, our goal is to maximize the network utility to solve equation (5) in which each IoT node u is assigned to a cloudlet f via a matching $\Upsilon : f \leftarrow u$

Definition 1: A matching game is defined as a two-sided assignment problem between two disjoint sets of players $(\mathcal{U}, \mathcal{F})$ in which the players of each set are interested to be matched to the players of the other set, according to preference relations $(\succ_u, \succ_f)$.

A preference relation $\succ$ is defined as a complete, reflexive, and transitive binary relation between the the players of sets $\mathcal{U}$ and $\mathcal{F}$. Let's suppose $\succ_{u_1}$ be the preference relation of IoT node u_1 defined over the set $\mathcal{F}$. If IoT node u_1 prefers cloudlet f_1 over cloudlet f_2, it will be denoted by $f_1 \succ_{u_1} f_2$. Thus for any two cloudlets $f_1, f_2 \in \mathcal{F}$ and two matchings $\Upsilon, \Upsilon' \in \mathcal{U} \times \mathcal{F}$, $f_1 = \Upsilon(u_1)$, $f_2 = \Upsilon'(u_1)$

$$(f_1, \Upsilon) \succ_{u_1} (f_2, \Upsilon') \Longleftrightarrow U_{f_1,u_1}(\Upsilon) > U_{f_2,u_1}(\Upsilon'). \tag{8}$$

Similarly, we use $\succ_{f_1}$ to denote the preference relation of a cloudlet f_1 defined over the set $\mathcal{U}$. If cloudlet f_1 prefers IoT node u_1 over IoT node u_2 it will be denoted by $u_1 \succ_{f_1} u_2$. Thus for any two IoT nodes u_1, u_2 *in* $\mathcal{U}$ and two matchings $\Upsilon, \Upsilon' \in \mathcal{U} \times \mathcal{F}$, $u_1 = \Upsilon(f_1)$, $u_2 = \Upsilon'(f_1)$

$$(u_1, \Upsilon) \succ_{f_1} (u_2, \Upsilon') \Longleftrightarrow U_{f_1,u_1}(\Upsilon) > U_{f_1,u_2}(\Upsilon'). \tag{9}$$

Each IoT node and cloudlet independently rank one another based on the utilities in (6) and (7). However the selection preferences of IoT nodes (8) and cloudlets (9) depend on each other as well as affected by the existing network wide matching. Many works [42], [43], [44] consider matching games, which assumes that the preferences of a player do not depend on the other players choices. In the problem of IoT node-cloudlet association, this assumption is not valid. The external effects which effect the preferences of the players (IoT nodes and Clouldlets) as well as the performance of IoT node - Cloudlet link, are called externalities. Matching games with externalities are discussed in detail in [45]. The game we considered in this article is a matching game with externalities as the mutual interference between IoT nodes exists.

B. Proposed Resource Allocation Algorithm

To solve the problem in (5) in a decentralized manner, we propose that cloudlets and IoT nodes have individual preferences over one another as defined in preference relations in (8) and (9). The aim of each IoT node (cloudlet) is to maximize its own utility, or equivalently, to become associated with the most preferred cloudlet (IoT node). We have to consider stable swap matching to solve matching game with externalities.

Definition 2: Given a matching Υ, a pair of IoT nodes $u1, u2 \in \mathcal{U}$ and cloudlets $f_1, f_2 \in \mathcal{F}$ with $(f_1, u_1), (f_2, u_2) \in \Upsilon$, a swap matching is defined as $\Upsilon^{u_1}_{f_1,f_2} = \{\Upsilon \backslash (f_1, u_1)\} \cup (f_2, u_1)$ Here, a matching is stable if there exist no swap matchings $\Upsilon^{u_1}_{f_1,f_2}$ such that following two proposition are satisfied:

1) $\forall y \in \{u_1, u_2, f_1, f_2\}, U_{y,\Upsilon^{u_1}_{f_1,f_2}(y)}(\Upsilon) > U_{y,\Upsilon(y)}(\Upsilon)$ and
2) $\exists y \in \{u_1, u_2, f_1, f_2\}, U_{y,\Upsilon^{u_1}_{f_1,f_2}(y)}(\Upsilon) \geq U_{y,\Upsilon(y)}(\Upsilon)$.

According to concept given in the definition, a matching Υ with pair $(f1, u1) \in \Upsilon$ is said to be stable if there does not exist any IoT node u_2 or cloudlet f_2, for which cloudlet f_1 prefers IoT node u_1 over IoT node u_2, or any IoT node u_1 which prefers cloudlet f_2 over f_1. To bring such stability in matching at network level it must be ensured that swaps are triggered only if they are beneficial for all the involved players (i.e. $\{u_1, u_2, f_1, f_2\}$).

Players are continuously changing their preferences due to externalities, resulting in formation of new IoT node-cloudlet pairs, which leads to classical solutions described in [42], [46] which are not applicable for our problem. Therefore, we propose Algorithm 1 based on stable swap matching for resource allocation which result in an increase of network-wide utility. Algorithm 1 consist of three phases. In phase 1

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2829751, IEEE Transactions on Industrial Informatics

each IoT node u_1 is initially assigned to a randomly selected cloudlet f_1, (Equivalently, the IoT node can be initially assigned to the closest cloudlet). IoT node u_1 discovers the near by cloudlet $f_2 \in \mathcal{F}$, using methods described in [41]. The SINR is calculated at each IoT node, which in turn is used to calculate utilities using equations (6) and (7). In phase 2, IoT nodes and cloudlets update their utilities and preferences based on current matching. If an IoT node u_1 is not currently connected to its preferred cloudlet (f_2), it will send a matching proposal to cloudlet f_2. When cloudlet f_2 receives the matching proposal it will update its utility and it will accept the proposal if and only if its utility $U_{f_2,u_1}(\Upsilon^{u_1}_{f_1,f_2})$ is improved by accepting the proposal. Otherwise if the matching proposal is rejected, IoT node u_1 will send a matching proposal to the next cloudlet in its preference list. Based on current matching, both IoT nodes and cloudlets update their corresponding preference lists and utilities at regular intervals. This periodic update ensures that both IoT nodes and cloudlets are connected to their corresponding best available option. Algorithm 1 arrives at a stable matching when Phase 2 is converged.

Algorithm 1 : **R**esource **A**llocation **A**lgorithm

```
Data: Each IoT node u is initially assigned to a randomly
  selected cloudlet f.
Result: Convergence to a stable matching Υ.
Phase 1: Cloudlet discovery and utility computation
Each IoT node u discovers the nearby cloudlets
At time t calculate SINR for each IoT node u using
  equation (3)
Calculate the utilities using equations (6) and (7)
Phase 2: Swap Matching
while ∄ Υ^{u1}_{f1,f2} : (f2, Υ^{u1}_{f1,f2}) ≻_{u1} (f1, Υ) and (u1, Υ^{u1}_{f1,f2}) ≻_{f2}
  (u1, Υ) do
  The utilities U_u(Υ) and U_f(Υ) are updated based on
    the current matching Υ ;
  IoT nodes and cloudlets are sorted based on ≻_{u1} and
    ≻_{f1};
  if (f2, Υ^{u1}_{f1,f2}) ≻_{u1} (f1, Υ) then
    IoT node u1 sends a proposal to cloudlet f2;
    Cloudlet f2 computes U_{f2,u1}(Υ^{u1}_{f1,f2}) for the swap
      matching Υ^{u1}_{f1,f2} ;
    if (u1, Υ^{u1}_{f1,f2}) ≻_{f2} (u1, Υ) and all three constraints in
      equation (5) are satisfied then
      L_{f2} ← L_{f2} ∪ {u1}
      Υ ← Υ^{u1}_{f1,f2}
    else
      The proposal is refused by cloudlet f2 , and IoT
        node u1 sends a proposal to the next preference.
    end if
  end if
end while
Phase 3 : Stable Matching
```

Proof of Convergence: The proof of convergence is based on following two rational assumptions.

- Each IoT node can be connected to limited number of nearby cloudlets because it has limited transmit power so there are finite number of possible swaps. Moreover, a swap will occur if and only if it results in improved utility of IoT node or cloudlet.
- Once all the probable swaps have been assessed, Phase 2 ends and each IoT node will remain connected to the most preferred cloudlet, and vice versa. Further improvement can not be achieved by swaps among neighboring IoT nodes and cloudlets. Therefore, Algorithm 1 converges to a stable matching after a finite number of iterations.

Complexity Analysis: In order to calculate the complexity of the proposed algorithm, we consider a simple case in which IoT nodes have strict preference ordering. We calculate the overhead and complexity for the matching Υ. We assume that U_f is the maximum number of IoT nodes matched to cloudlet $f \in \mathcal{F}$ and U_m is the total number of IoT nodes matched $U_m = \sum_{f \in \mathcal{F}} U_f$ such that $U_f = U_{f'}$ and $f \neq f'$. The value of U_f depends on the available bandwidth. Let the number of satisfied matched IoT nodes are represented by $U_s (\leq U_m)$, which is based on the preference order. For simplicity, let's assume U_s is constant during all the iterations.To calculate the complexity, two worst case scenarios are considered, when all IoT nodes $u \in \mathcal{U}$ in network are matched to a single cloudlet:

1) When the number of IoT nodes are less then the total number of matched IoT nodes i.e., $U < U_m, |\mathcal{U}| = U$.
2) When the number of IoT node is greater than the total number of matched IoT nodes i.e., $U > U_m$.

Our objective is to find the maximum number of iterations required for convergence and the maximum number of proposals sent from IoT nodes to cloudlets (overhead) for both cases. In each iteration i , IoT nodes send proposal to their most preferred cloudlet, and the cloudlet accepts or rejects the received proposal based on its preference order and available capacity. Therefore, the number of matched but unsatisfied IoT nodes at each iteration is less or equal than the available capacity.

Considering the first case, all IoT nodes are matched to a single cloudlet, when the algorithm is converged. The worst case scenario happens, only if all the IoT nodes have the same preference order. Thus, at the end of each iteration i we have $U - U_s i$ unsatisfied IoT nodes. All IoT nodes are matched and satisfied when the maximum number of iterations i_{max} is obtained. i.e., $U - U_s t_{max} = 0$. Hence, the complexity is of order $O(U)$. Moreover, we have $U - U_s i$ proposal messages at each iteration i, and the total overhead for sending such messages is given by:

$$\eta = \sum_{i=1}^{i_{max}} (U - U_s i + U_s) = \frac{U(U + U_s)}{2U_s} \tag{10}$$

Considering the second case where $U > U_m$, there are $U - U_m$ unallocated IoT nodes when algorithm approaches convergence. The worst case happens, only if all IoT nodes have same preference order. Hence at i_{max} iteration we have $U - U_s i_{max}$ IoT nodes unallocated. The complexity of the order $O(U_m)$, and the overhead is given by:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2829751, IEEE Transactions on Industrial Informatics

$$\eta = \sum_{i=1}^{U_m} U - U_s i \tag{11}$$

V. Simulation Setup and Results

We consider a fog network in which a number of IoT nodes and cloudlets are uniformly distributed. The transmit power of each cloudlet is 30 dBm. The transmit power of each IoT node is 13 dBm. Transmissions are affected by distance dependent path loss and shadowing according to 3GPP specifications [47]. We assume that there is no power control, and thus the power is uniformly divided among IoT nodes. It is also assumed that the bandwidth is divided equally among the IoT nodes. The minimum SINR required by each IoT node is 9.5 dB , and the noise power spectral density is -174 dBm/Hz. The transmission radius of a cloudlet is 50 m. Minimum number of IoT nodes allowed are 20, where as maximum number of IoT nodes allowed are 100 with an increment of 20. We use a full-buffer traffic model for all IoT nodes in our simulations. To compare our results, we consider a random allocation of IoT node to the cloudlets which represents a baseline solutions for the cloudlet selection problem.

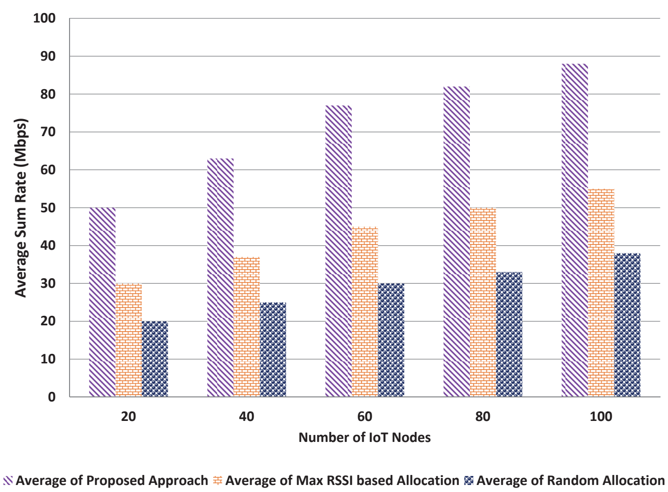

Fig. 2: Average sum rate as a function of number of IoT nodes in fog network

Fig. 2 depicts the average sum rate as a function of number of IoT nodes in fog network, it is clear from the Fig. 2 , that by increasing number of IoT node average sum rate increases. Average sum rate of proposed method is better than the base line solution, which allocates IoT nodes to cloudlets randomly. The proposed method is also compared with maximum RSSI based association method, in which an IoT node is assigned to a cloudlet from which it receives the maximum RSSI. The proposed method outperforms maximum RSSI based association as well. As depicted in Fig. 3 there is approximately 65% improvement in network throughput when we compare proposed scheme with maximum RSSI based association, and there is approximately 145% improvement in network throughput when we compare proposed scheme with random association.

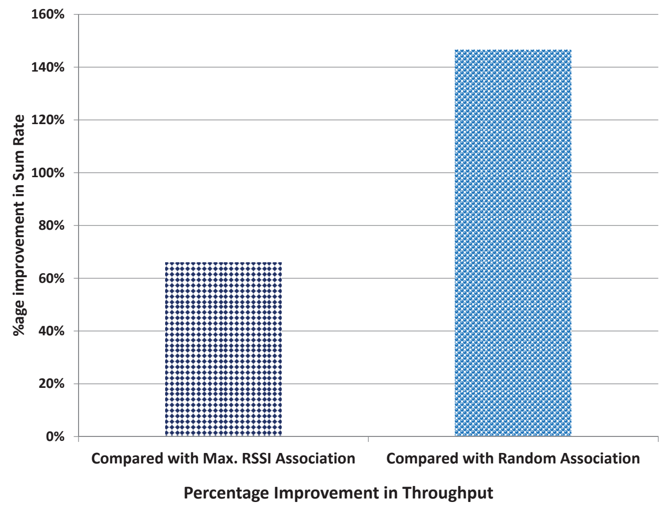

Fig. 3: Percentage improvement in throughput: Comparing proposed method with random association and Max. RSSI association

VI. Conclusion

In this article, we formulated an optimization problem for joint cloudlet selection and latency minimization in fog network, subject to maximum work load and latency constraints. We cast the optimization problem as many-to-one matching game, in which IoT nodes and cloudlets are players who want to maximize their own utility. We propose is a distributed and self-organizing method to solve the matching game which results in overall improved sum rate of network. The proposed method is compared with random association and maximum RSSI based association. The proposed method outperforms both random association and maximum RSSI based association.

References

[1] R. Zhu, X. Zhang, X. Liu, W. Shu, T. Mao, and B. Jalaian, "Erdt: Energy-efficient reliable decision transmission for intelligent cooperative spectrum sensing in industrial iot," *IEEE Access*, vol. 3, pp. 2366–2378, 2015.

[2] S. Nigam, S. Asthana, and P. Gupta, "Iot based intelligent billboard using data mining," in *Innovation and Challenges in Cyber Security (ICICCS-INBUSH), 2016 International Conference on*. IEEE, 2016, pp. 107–110.

[3] D. Evans, "The internet of things: How the next evolution of the internet is changing everything," *CISCO white paper*, vol. 1, no. 2011, pp. 1–11, 2011.

[4] F. Computing, "the internet of things: Extend the cloud to where the things are," 2016.

[5] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*. ACM, 2012, pp. 13–16.

[6] S. Yi, C. Li, and Q. Li, "A survey of fog computing: concepts, applications and issues," in *Proceedings of the 2015 Workshop on Mobile Big Data*. ACM, 2015, pp. 37–42.

[7] M. Chiang. (2016) Fog networking: An overview on research opportunities. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1601/1601.00835.pdf

[8] A. Hussein, I. H. Elhajj, A. Chehab, and A. Kayssi, "Sdn vanets in 5g: An architecture for resilient security services," in *Software Defined Systems (SDS), 2017 Fourth International Conference on*. IEEE, 2017, pp. 67–74.

[9] N. B. Truong, G. M. Lee, and Y. Ghamri-Doudane, "Software defined networking-based vehicular adhoc network with fog computing," in *Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on*. IEEE, 2015, pp. 1202–1207.

[10] J. Zhu, D. S. Chan, M. S. Prabhu, P. Natarajan, H. Hu, and F. Bonomi, "Improving web sites performance using edge servers in fog computing architecture," in *Service Oriented System Engineering (SOSE), 2013 IEEE 7th International Symposium on*. IEEE, 2013, pp. 320–323.

[11] P. S. Khodashenas, C. Ruiz, M. S. Siddiqui, A. Betzler, and J. F. Riera, "The role of edge computing in future 5g mobile networks: concept and challenges," *Cloud and Fog Computing in 5G Mobile Networks*, p. 349, 2017.

[12] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5g network edge architecture & orchestration," *IEEE Communications Surveys & Tutorials*, 2017.

[13] J. K. Zao, T. T. Gan, C. K. You, S. J. R. Méndez, C. E. Chung, Y. Te Wang, T. Mullen, and T. P. Jung, "Augmented brain computer interaction based on fog computing and linked data," in *Intelligent Environments (IE), 2014 International Conference on*. IEEE, 2014, pp. 374–377.

[14] G. Peralta, M. Iglesias-Urkia, M. Barcelo, R. Gomez, A. Moran, and J. Bilbao, "Fog computing based efficient iot scheme for the industry 4.0," in *Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM), 2017 IEEE International Workshop of*. IEEE, 2017, pp. 1–6.

[15] C. T. Do, N. H. Tran, C. Pham, M. G. R. Alam, J. H. Son, and C. S. Hong, "A proximal algorithm for joint resource allocation and minimizing carbon footprint in geo-distributed fog computing," in *Information Networking (ICOIN), 2015 International Conference on*. IEEE, 2015, pp. 324–329.

[16] R. Deng, R. Lu, C. Lai, T. H. Luan, and H. Liang, "Optimal workload allocation in fog-cloud computing toward balanced delay and power consumption," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1171–1181, 2016.

[17] D. Zeng, L. Gu, S. Guo, Z. Cheng, and S. Yu, "Joint optimization of task scheduling and image placement in fog computing supported software-defined embedded system," *IEEE Transactions on Computers*, vol. 65, no. 12, pp. 3702–3712, 2016.

[18] H. Zhang, Y. Xiao, S. Bu, D. Niyato, F. R. Yu, and Z. Han, "Computing resource allocation in three-tier iot fog networks: a joint optimization approach combining stackelberg game and matching," *IEEE Internet of Things Journal*, 2017.

[19] K. Intharawijitr, K. Iida, and H. Koga, "Analysis of fog model considering computing and communication latency in 5g cellular networks," in *Pervasive Computing and Communication Workshops (PerCom Workshops), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–4.

[20] M. S. Elbamby, M. Bennis, and W. Saad, "Proactive edge computing in latency-constrained fog networks," *arXiv preprint arXiv:1704.06749*, 2017.

[21] J. Oueis, E. C. Strinati, and S. Barbarossa, "The fog balancing: Load distribution for small cell cloud computing," in *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*. IEEE, 2015, pp. 1–6.

[22] S. Yan, M. Peng, M. A. Abana, and W. Wang, "An evolutionary game for user access mode selection in fog radio access networks," *IEEE Access*, vol. 5, pp. 2200–2210, 2017.

[23] S. Bitam, S. Zeadally, and A. Mellouk, "Fog computing job scheduling optimization based on bees swarm," *Enterprise Information Systems*, pp. 1–25, 2017.

[24] S. Meng, Y. Wang, Z. Miao, and K. Sun, "Joint optimization of wireless bandwidth and computing resource in cloudlet-based mobile cloud computing environment," *Peer-to-Peer Networking and Applications*, pp. 1–11, 2017.

[25] Y. Jararweh, M. Al-Ayyoub, M. Al-Quraan, A. T. Loai, and E. Benkhelifa, "Delay-aware power optimization model for mobile edge computing systems," *Personal and Ubiquitous Computing*, pp. 1–11, 2017.

[26] A. Mukherjee, D. De, and D. G. Roy, "A power and latency aware cloudlet selection strategy for multi-cloudlet environment," *IEEE Transactions on Cloud Computing*, 2016.

[27] A. Akbar and P. R. Lewis, "Towards the optimization of power and bandwidth consumption in mobile-cloud hybrid applications," in *Fog and Mobile Edge Computing (FMEC), 2017 Second International Conference on*. IEEE, 2017, pp. 213–218.

[28] K. Gai, M. Qiu, H. Zhao, L. Tao, and Z. Zong, "Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing," *Journal of Network and Computer Applications*, vol. 59, pp. 46–54, 2016.

[29] E. Paraskevas, T. Jiang, and J. S. Baras, "Trust-aware network utility optimization in multihop wireless networks with delay constraints," in *Control and Automation (MED), 2016 24th Mediterranean Conference on*. IEEE, 2016, pp. 593–598.

[30] T. Han and N. Ansari, "Network utility aware traffic load balancing in backhaul-constrained cache-enabled small cell networks with hybrid power supplies," *IEEE Transactions on Mobile Computing*, 2017.

[31] M. Ali, S. Qaisar, M. Naeem, and S. Mumtaz, "Energy efficient resource allocation in d2d-assisted heterogeneous networks with relays," *IEEE Access*, vol. 4, pp. 4902–4911, 2016.

[32] M. I. Ashraf, M. Bennis, C. Perfecto, and W. Saad, "Dynamic proximity-aware resource allocation in vehicle-to-vehicle (v2v) communications," in *Globecom Workshops (GC Wkshps), 2016 IEEE*. IEEE, 2016, pp. 1–6.

[33] O. Semiari, W. Saad, Z. Daw, and M. Bennis, "Matching theory for backhaul management in small cell networks with mmwave capabilities," in *Communications (ICC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3460–3465.

[34] T. LeAnh, N. H. Tran, W. Saad, L. Le, D. Niyato, T. Ho, and C. S. Hong, "Matching theory for distributed user association and resource allocation in cognitive femtocell network," *IEEE Transactions on Vehicular Technology*, 2017.

[35] N. Namvar, W. Saad, B. Maham, and S. Valentin, "A context-aware matching game for user association in wireless small cell networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 439–443.

[36] J. Zhao, Y. Liu, K. K. Chai, M. Elkashlan, and Y. Chen, "Matching with peer effects for context-aware resource allocation in d2d communications," *IEEE Communications Letters*, vol. 21, no. 4, pp. 837–840, 2017.

[37] C. Xu, C. Gao, Z. Zhou, Z. Chang, and Y. Jia, "Social network-based content delivery in device-to-device underlay cellular networks using matching theory," *IEEE Access*, vol. 5, pp. 924–937, 2017.

[38] Y. Gu, W. Saad, M. Bennis, M. Debbah, and Z. Han, "Matching theory for future wireless networks: fundamentals and applications," *IEEE Communications Magazine*, vol. 53, no. 5, pp. 52–59, 2015.

[39] T. C. Chiu, W. H. Chung, A. C. Pang, Y. J. Yu, and P. H. Yen, "Ultra-low latency service provision in 5g fog-radio access networks," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2016, pp. 1–6.

[40] M. Hanif and P. Smith, "On the statistics of cognitive radio capacity in shadowing and fast fading environments," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 2, pp. 844–852, February 2010.

[41] T. Q. Quek, *Small cell networks: Deployment, PHY techniques, and resource management*. Cambridge University Press, 2013.

[42] E. A. Jorswieck, "Stable matchings for resource allocation in wireless networks," in *2011 17th International Conference on Digital Signal Processing (DSP)*, July 2011, pp. 1–8.

[43] S. Bayat, R. H. Y. Louie, Z. Han, Y. Li, and B. Vucetic, "Multiple operator and multiple femtocell networks: Distributed stable matching," in *2012 IEEE International Conference on Communications (ICC)*, June 2012, pp. 5140–5145.

[44] L. Jiang, S. Parekh, and J. Walrand, "Base station association game in multi-cell wireless networks (special paper)," in *2008 IEEE Wireless Communications and Networking Conference*, March 2008, pp. 1616–1621.

[45] A. E. Roth and M. Sotomayor, "Two-sided matching," *Handbook of game theory with economic applications*, vol. 1, pp. 485–541, 1992.

[46] Y. Wu, H. Viswanathan, T. Klein, M. Haner, and R. Calderbank, "Capacity optimization in networks with heterogeneous radio access technologies," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, Dec 2011, pp. 1–5.

[47] S. Sesia, M. Baker, and I. Toufik, *LTE-the UMTS long term evolution: from theory to practice*. John Wiley & Sons, 2011.