IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 14, 2018, accepted February 19, 2018, date of publication February 22, 2018, date of current version March 16, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2808598

Optimizing the Number of Fog Nodes for Cloud-Fog-Thing Networks

EREN BALEVI, (Member, IEEE), AND RICHARD D. GITLIN, (Life Fellow, IEEE)
Department of Electrical Engineering, University of South Florida, Tampa, Florida 33620 USA

Corresponding author: Eren Balevi (erenbalevi@mail.usf.edu)

ABSTRACT Fog networking has recently received considerable attention from a theoretical perspective, but in order for such networks to be practical several open areas need to be addressed. This paper determines the optimum number of nodes that should be upgraded to fog nodes with additional computing capabilities in order to maximize the average data rate and minimize the transmission delay. The optimization is performed for a given set of wireless channel conditions and a fixed total number of network nodes. It will be shown that, having more or less fog nodes than the optimum degrades the data rate. The numerical results indicate that the average data rate can increase nearly an order of magnitude for an optimized number of fog nodes in case of shadowing and fading. It is further shown that the optimum number of fog nodes does not increase in direct proportion to the increase in the total number of nodes. Furthermore, the optimum number of fog nodes decreases when channels have high path loss exponents. These findings suggest that the fog nodes must be selected among those that have the highest computation capability for densely deployed networks and high path loss exponent channels.

INDEX TERMS Fog networking, hierarchical networks, SINR, average data rate, transmission delay.

I. INTRODUCTION

A multitude of emerging applications from augmented reality, online gaming, autonomous vehicles, and smart cities envisioned for IoT/5G wireless networks are expected to produce an extraordinary increase in the amount of data. Although such a large-scale increase in data can be processed to some extent by cloud computing, the continuously growing amount of data cannot be tackled solely by cloud computing, and fog computing has emerged as a promising method to accommodate the expected demands [1]. Combining the large-scale data processing capability of cloud computing with the location aware, widely geographically distributed, low latency data processing capability, fog computing is expected to be an attractive network architecture [1]–[5]. This integration of cloud and fog networking is quite attractive in that some portion of data in the network that has stringent latency and throughput requirements may be processed by fog computing/networking, while the rest of data may be processed by cloud computing. The complementary nature of cloud and fog suggests a hierarchical network architecture dubbed a *cloud-fog-thing* network [1]–[5]. This architecture is a good compromise between fully centralized cloud networking and fully distributed fog networking.

Maximizing the average data rate of the promising *cloud-fog-thing* network depending on the signal-to-interference-plus-noise-ratio (SINR) is of paramount importance to support future 5G applications and use cases. This can also minimize the transmission delay that leads to a decrease in latency, which has a significant impact on the quality of user experience (QoE). In this regard, it is important to optimize the number of fog nodes in order to maximize the average data rate so as to minimize the transmission delay. Finding the optimum number of fog nodes will further enhance the understanding and impact of the *cloud-fog-thing* architecture. While there can be many potential nodes inside a network that could be upgraded to fog nodes, it is not clear why one does not update all the potential nodes to fog nodes to exploit all the available unused resources in the network. This paper provides answers to these questions.

A stochastic geometry analysis is used to determine the optimum number of fog nodes for a given number of nodes within the area of interest. It is important to note that the widely used Poisson Point Process (PPP) model in stochastic geometry is not applicable to this network model for two reasons. First, the PPP gives accurate models only for large-scale networks [6] whereas a fog network covers a local area, which constitutes a low-to-medium scale network. Second, and more importantly, the total number of nodes is known and finite, so that a Binomial Point Process (BPP) better represents the low-to-medium scale network whose total number of nodes is known [7].

Fog networking is clearly outlined with its benefits in [1] and [2] and, the hierarchical *cloud-fog-thing* network is justified with different use cases in [3]–[5]. Further, fog computing based radio access networks (RANs) are discussed in [8] and [9]. One of the primary ideas common to all these papers is to upgrade some number of nodes into a fog node. However, the optimum number of nodes that will be upgraded to fog nodes as well as the incentive of not upgrading all nodes to a fog node are not stated. This study aims to fulfill this gap in the literature of the *cloud-fog-thing* network architecture. The optimum number of fog nodes will be found by assuming that each node elects itself as a fog node with some probability. Then, the number of fog nodes becomes np, if there are n nodes within the area of interest, each of which can be a fog node with probability p. The same approach is used to determine the cluster-heads or leaders of each cluster in wireless sensor networks [10]–[14], however, all those papers assume that the probability of being a cluster-head is given as *a priori* information instead of determining this by analysis. Reference [15] determines the optimum cluster-head probability using a PPP model to optimize energy efficiency for wireless sensor networks, which has some different notions than fog networking and is quite different than the situation addressed in this paper where the probability of being a fog node is found using a BPP model.

As stated in a recent survey paper, determining the optimum number of fog nodes is an open research problem, and affects the overall network efficiency [16]. Based on this motivation, the optimum number of fog nodes is determined for channels with different path loss exponents using a BPP model. Interestingly, our analysis indicates that too large or too small number of fog nodes decreases the average data rate. In addition, how the fog nodes scales with the incremental total number of nodes for different channels is quantified. Additionally, the optimum number of nodes that can be controlled by a fixed number of fog nodes will also be found in this paper. This analysis might be useful in the design of the efficient virtual machines in the cloud, in the determination of the value of K in K-means clustering algorithm, which may be used to find the optimum locations of fog nodes, and in enhancing caching efficiency.

The paper is organized as follows. The network model and the problem statement are presented in Section II. In Section III, the problem is formulated to find the optimum number of fog nodes. Section IV introduces a stochastic geometry analysis for a BPP model. The derived closed-form derivations are validated in Section V and the benefits and planned future research are given in Section VI. The paper ends with the concluding remarks in Section VII.

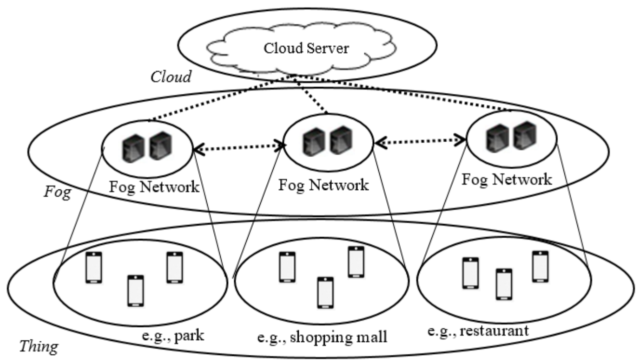


FIGURE 1. ***Cloud-fog-thing* type hierarchical network model.**

II. NETWORK MODEL AND PROBLEM STATEMENT

As noted above, it is expected that various applications supported by 5G wireless networks, and beyond, will require an interplay between cloud and fog networks. Accordingly, some portion of data is processed at the fog networks and the remaining portion of data is conveyed to the cloud. In this model, the inherent features of the fog layer such as widely deployed geographical distribution and location awareness is linked with the large-scale data management capability of the cloud layer. A system of smart traffic light is one example that illustrates the interplay between cloud and fog networks so that distributed traffic lights are connected to each other as well as vehicles, and pedestrians and bikes intelligently control the traffic such that local decisions are made by fog nodes, and long-term statistics are gathered at the cloud [1]. Another example is a smart pipeline monitoring system in which the combination of fog and cloud networks sequentially process the data coming from the massive number of sensors [3]. The same network model is highlighted in [2] and [4] as well. In these papers, the network model is composed of the hierarchical combination of fog and cloud termed as *cloud-fog-thing* network as depicted in Fig. 1. Typically, the fog layer may be composed of many local fog networks located at parks, shopping malls, restaurants to name a few, and the thing layer involves the end devices that may be various type of sensors, IoT devices, or mobile phones.

Bearing in mind the overall network structure and the general notion of fog networking is that any existing node in the network can be upgraded to a fog node[1] raise the question of the optimum number of fog nodes that should be upgraded from the existing nodes in the network, which is addressed in this paper. Assume that there is a square planar with one side being is $2a$, i.e., from $-a$ to a and the cloud is located at the center, and the n nodes are randomly and uniformly distributed around the cloud. At the beginning, these nodes are assumed to be ordinary, and then, some of them are specialized as fog nodes that constitute the fog layer and the rest of them will stay ordinary that constitute the

[1]We make the assumption that all nodes have fog node capabilities that may be turned on (or off) using over the air activation (or deactivation).

thing layer. To find the number of fog nodes, it is assumed that each node becomes a fog node with probability p, and this yields n_0 and n_1 number of ordinary nodes and fog nodes respectively as $n_0 = n(1-p)$, and $n_1 = np$. In this way, we transform the problem of explicitly optimizing n_1 into determining the optimum value of p. Note that this paper solely focuses on finding the optimum number of fog nodes deferring the question of which nodes should be updated as fog nodes to [24] and [25].

Clearly there has to be a criterion to determine the optimum probability of being a fog node p, and thus the optimum number of fog nodes. In this analysis, the criterion to find the optimum number of fog nodes is to maximize the data rate depending on SINR so as to minimize the transmission delay. Hence, the probability of being a fog node p is optimized, and the optimum values of n_0 and n_1 will be found accordingly. In general, this paper provides a mathematical framework to specify the optimum number of fog nodes under one fog network so that one can find the optimum number of fog nodes dynamically even if the total number of nodes changes. Using this framework, one can determine the maximum possible nodes that can be controlled by a fixed number of fog nodes as well.

III. PROBLEM FORMULATION

A stochastic geometry analysis is performed to determine the optimum number of fog nodes when the end devices send their packets to the fog nodes, which forward the data to the cloud after processing some part of the data. In this model, fog nodes and end devices are considered as points in 2-dimensional Euclidean space. Throughout our analysis, it is assumed that the total number of points residing in the area of interest is known, though the number may dynamically change. Additionally, the number of nodes in the fog layer and in the thing layer may change. By this is meant that some nodes in the fog layer may be downgraded to the nodes in the thing layer or vice versa depending on the change in the network geometry due to mobility, or arrival or departure of the nodes in the network. A widely used PPP model to accurately model the large-scale networks for random number of nodes in stochastic geometry [17] cannot be applied to this problem, because the total number of nodes is known and finite. Indeed, this knowledge means that a BPP is the appropriate model [7]. Furthermore, the sub-regions covered by fog networks are not large-scale, i.e., they may be classified as low-to-medium scale network. Relying on these factors, it is more appropriate to model the underlying network model as a BPP.

The *cloud-fog-thing* network architecture can be simplified as a hierarchical tree based topology for one fog network as depicted in Fig. 2.[2] Here, nodes in the thing layer are termed as end devices that constitute Tier-0, which are controlled by the fog nodes located at Tier-1 and the cloud server is situated at the top layer that is able to control a square planar region with a side of $2a$. Note that fog nodes are connected to each

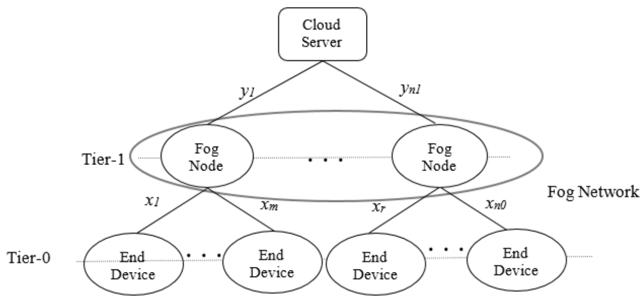


FIGURE 2. A simplified tree based hierarchical network model.

other in a circular, fully connected mesh topology, and form the fog network, which is a generic and an appropriate model consistent with the definition of a fog network [1], [2].

There is an interplay between the number of nodes at Tier-1 and Tier-0 so that the number of fog nodes will be dynamically determined according to the number of end devices. More specifically, suppose that there are $n_0 = n(1-p)$ and $n_1 = np$ end devices and fog nodes, respectively, and $n = n_0 + n_1$. To find the relation among n, n_0, and n_1, the optimum probability of being fog node p is found.

Assume that the packet size is M bits and the packet is partially processed in the fog node, e.g., K bits of the packet are processed, and the rest, i.e., $M-K$ bits are relayed to the cloud. Then, the transmission delay becomes

$$\tau_{trans} = \frac{M}{R_{fog}} + \frac{M-K}{R_{cloud}} \tag{1}$$

where R_{fog} and R_{cloud} are the data rate at the fog node and cloud as

$$R_{fog} = Wlog(1 + SINR_{fog}) \tag{2}$$

and

$$R_{cloud} = Wlog(1 + SINR_{cloud}) \tag{3}$$

where W is the bandwidth, $SINR_{fog}$ and $SINR_{cloud}$ are the SINR at the fog node and cloud, respectively. More specifically, the SINR of the i^{th} end device for $i = 1, 2, \cdots, n_0$ at the fog node becomes

$$SINR_{fog}(i) = \frac{P_i h_i x_i^{-\alpha}}{\sigma^2 + I_{fog}} \tag{4}$$

where P_i is the transmission power of the i^{th} end device, h_i is the channel power coefficient, x_i is the distance between the end device and the fog node as shown in Fig. 2, α is the path loss coefficient, σ^2 is the noise variance and I_{fog} is the residual interference power at the fog node after some interference mitigation techniques whose detailed discussion are out of scope for this paper. Notice that $I_{fog} = 0$ in the idealized case, i.e., if the interference is perfectly mitigated. Similarly, the SINR due to the j^{th} fog node for $j = 1, 2, \cdots, n_1$ at the cloud can be written as

$$SINR_{cloud}(j) = \frac{P_j h_j y_j^{-\alpha}}{\sigma^2 + I_{cloud}} \tag{5}$$

[2]All of the links are assumed to be wireless.

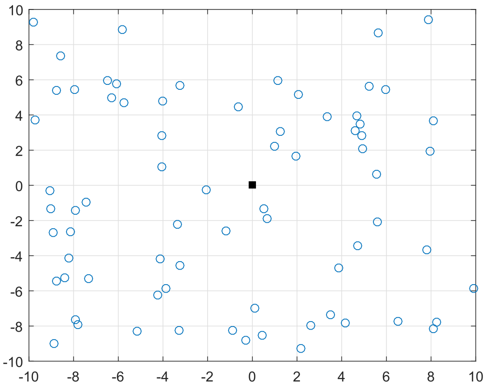


FIGURE 3. The sample distribution of nodes for $a = 10$.

where P_j is the transmission power of the j^{th} fog node, h_j is the channel power coefficient, y_j represents the distance between the j^{th} fog node and the cloud, which is depicted in Fig. 2 as well. I_{cloud} is the residual interference power at the cloud. Similarly, if one makes the assumption of perfect interference mitigation, I_{cloud} becomes 0.

Consider the simple network structure that demonstrates the nodes given in Fig. 3. Here, circles represent the ordinary nodes, some of which will be upgraded to the fog nodes and the square denotes the cloud. The distance between the circle that will not be upgraded as a fog node and be upgraded as a fog node is $\{x_i\}$ and the distance between a circle, i.e., the circle that will be upgraded to a fog node which is not known as *a priori* and found after the optimization, and the square is $\{y_j\}$. Bearing in mind this network structure, the objective function can be written as (6), as shown at the bottom of this page, in terms of SINR that can maximize the data rate or minimize the τ_{trans} for the i^{th} end device that sends packets to the cloud through the j^{th} fog node, because W, M, $M-K$ are constant values. For the sake of simplicity, propagation delay is omitted, since the optimization that minimizes the right-hand side of (6) with respect to the distances can automatically minimize the propagation delay.

Since the logarithm is a monotone function, (6) is equivalent to

$$\hat{J}_{ij} = \min \left(\frac{\sigma^2 + I_{fog}}{P_i h_i x_i^{-\alpha}} + \frac{\sigma^2 + I_{cloud}}{P_j h_j y_j^{-\alpha}} \right). \tag{7}$$

Taking the expected value of (7) produces

$$\hat{J}_{ij}(avg) = \min \left(\frac{\sigma^2 + I_{fog}}{P_i} E\left[\frac{x_i^{\alpha}}{h_i}\right] + \frac{\sigma^2 + I_{cloud}}{P_j} E\left[\frac{y_j^{\alpha}}{h_j}\right] \right) \tag{8}$$

for given $P_i, P_j, \sigma^2, I_{fog}$ and I_{cloud}, which may be either given as *a priori* information or estimated at the receiver, and thus they do not impress the optimization. Also, channel power coefficients are independent from distances that lead to

$$\tilde{J}_{ij}(avg) = \min \left(E[x_i^{\alpha}]E\left[\frac{1}{h_i}\right] + E[y_j^{\alpha}]E\left[\frac{1}{h_j}\right] \right) \tag{9}$$

where $E[1/h_i] = c_i$ and $E[1/h_j] = c_j$ such that c_i and c_j are constant values. This yields

$$J_{\alpha}^{single} = \min \left(E[x_i^{\alpha}] + E[y_j^{\alpha}] \right). \tag{10}$$

Solving (10) gives one end device for one fog node that minimizes the transmission delay by maximizing the average data rate while a packet is sent from an end device to the cloud through a fog node. Since there are n_0 number of end devices and n_1 number of fog nodes, the objective function is defined as

$$J_{\alpha} = \min_{p} \left(\sum_{i=1}^{n_0} E[x_i^{\alpha}] + \sum_{j=1}^{n_1} E[y_j^{\alpha}] \right) \tag{11}$$

assuming that packets coming from the end devices to the fog nodes are aggregated, partially processed and relayed to the cloud. Since $n_0 = n(1-p)$ and $n_1 = np$, (11) is optimized with respect to p, i.e., the value of p that minimizes (11) gives the number of fog nodes that will be upgraded from the ordinary nodes, which are randomly spatially distributed within the area of interest.

IV. THE OPTIMUM NUMBER OF FOG NODES

The number of fog nodes for each fog network can be optimized with respect to the objective function in (11). In the model, it is assumed that there are n nodes within the area of interest including n_0 end devices and n_1 fog nodes so that $n = n_0 + n_1$. To find the relationship among n_0 and n_1, assume that the probability of being a fog node is p for all n nodes. This produces $n_0 = n(1-p)$, $n_1 = np$ end devices and fog nodes, respectively. Here, the critical point is the determination of p. Accordingly, first the objective function (11), will be derived as a closed-form expression in terms of p. Next, the objective function is optimized with respect to p which determines the optimum values of n_0 and n_1. Notice that $p = 0$ refers to the fact that there are no fog node whereas p=1 means that all nodes are fog nodes.

The values of n_0 and n_1 are optimized for $\alpha = 1$, $\alpha = 2$ and $\alpha = 4$ in this paper. Although the analysis of $\alpha = 1$ is an approximation when the nodes are connected to a cable, this is physically meaningless for wireless connections. The main reason for analyzing the case for $\alpha = 1$ is to better specify the relation between the optimum number of fog nodes and the path loss coefficient. Following that, the analysis is given for a free space path loss, i.e., $\alpha = 2$. Lastly, a more practical case is considered for $\alpha = 4$ accounting for the impact of shadowing and fading.

$$J_{ij} = \min \left(\frac{1}{log(1 + SINR_{fog}(i))} + \frac{1}{log(1 + SINR_{cloud}(j))} \right) \tag{6}$$

A. HYPOTHETICAL PATH LOSS

The objective function in (11) is first obtained as a closed-form expression in case of $\alpha = 1$, which is a physically meaningless, but a mathematically meaningful quantity for wireless channels. This yields the following objective function

$$J_1 = \min_p (x + y) \tag{12}$$

where

$$y = \sum_{j=1}^{n_1} E[y_j] \tag{13}$$

and

$$x = \sum_{i=1}^{n_0} E[x_i]. \tag{14}$$

All nodes are assumed to be independently and uniformly distributed in a given square area of side $2a$ for 2-dimensional Euclidean space with coordinates (i_x, i_y). The expected distance of a fog node from the cloud can be expressed as

$$E[y_j] = \frac{1}{4a^2} \int_{-a}^{a} \int_{-a}^{a} \sqrt{i_x^2 + i_y^2} d_{i_x} d_{i_y} = 0.765a. \tag{15}$$

Based on (15), the total distance between the fog nodes and the cloud given in (13) can be written for $n_1 = np$ fog nodes as

$$y = \sum_{j=1}^{n_1} E[y_j] = 0.765npa. \tag{16}$$

The average distance between two arbitrarily located points in a BPP is required to find (14). Specifically, the mean distance between a fog node and an end device is needed to find the value of x_i. A recently derived formula specifies the mean distance between two points for an isotropic BPP [7], as

$$E[x_i] = \frac{ri^{1/2}}{(N+1)^{1/2}} \tag{17}$$

where r is the maximum range of the fog node, and N is the total number of end devices controlled by a fog node, which becomes $N = (n - np)/np$. This gives

$$E[x_i] = \frac{ri^{1/2}}{((n-np)/np+1)^{1/2}}. \tag{18}$$

The maximum range r can be specified by considering that each fog node, which is located at a center of a 2-dimensional ball $b(o, r)$, has identical range and constitutes non-overlapping partitions without any loss of generality. This leads to

$$r = \frac{\pi R}{np} \tag{19}$$

where R denotes the radius of the circular mesh fog network. Based on above, the sum distance between a fog node and end devices can be written as

$$\sum_{i=1}^{n_0/n_1} E[x_i] = \sum_{i=1}^{n_0/n_1} \frac{\pi R i^{1/2}}{(np)((n-np)/np+1)^{1/2}} \tag{20}$$

and the total distances due to having np fog nodes is given by

$$x = \sum_{i=1}^{n_0/n_1} \frac{\pi R i^{1/2}}{((n-np)/np+1)^{1/2}}. \tag{21}$$

After some mathematical operations, (21) can be simplified to

$$x = \frac{\pi R(n-np)}{np}. \tag{22}$$

The objective function J_1 in (12) can be found, given (16) and (22) as

$$J_1 = \frac{\pi R(n-np)}{np} + 0.765npa. \tag{23}$$

Lemma 1: There is a unique optimum global value of p that minimizes (23).

Proof: The first and second derivative of (23) with respect to p is

$$\frac{\partial J_1}{\partial p} = \frac{153anp^2 - 200\pi R}{200p^2} \tag{24}$$

and

$$\frac{\partial^2 J_1}{\partial p^2} = \frac{2\pi R}{p^3}. \tag{25}$$

Since the second derivative of (25) is greater than 0, this means that (23) is strictly convex function and the value of p that makes (24) 0 is a global minimum point and unique, if it exists. Then, the question is whether a real p exists or not. After some straightforward calculations, p can be approximately found as

$$p = \left(\frac{200\pi R}{153an}\right)^{1/2} \tag{26}$$

and hence it is a global minimum point and unique. □

Due to Lemma 1, the optimum number of fog nodes for a given total n number of nodes can be calculated as

$$n_1 = \left(\frac{200\pi Rn}{153a}\right)^{1/2}.$$

Alternatively, one can determine the optimum number of end devices in two steps if the number of fog nodes n_1 is known. First the optimum value of p is found as

$$p = \frac{200\pi R}{153an_1}.$$

Second, the number of end devices is calculated as

$$n_0 = \frac{n_1}{p} - n_1.$$

B. FREE SPACE PATH LOSS

For wireless channels the signal power falls off with path loss exponents of $\alpha > 1$. In free space the path loss $\alpha = 2$ and to find the optimum number of fog nodes, the objective function can be specified as

$$J_2 = \min_p(\tilde{x} + \tilde{y}) \tag{27}$$

where

$$\tilde{y} = \sum_{j=1}^{n_1} E[y_j^2] \tag{28}$$

and

$$\tilde{x} = \sum_{i=1}^{n_0} E[x_i^2]. \tag{29}$$

The closed-form derivation of (28) can be obtained as

$$E[y_j^2] = \frac{1}{4a^2}\int_{-a}^{a}\int_{-a}^{a}(i_x^2 + i_y^2)d_{i_x}d_{i_y} = 2a^2/3. \tag{30}$$

Generalizing (30) for np fog nodes produces

$$\tilde{y} = \sum_{j=1}^{n_0} E[y_j^2] = 2npa^2/3. \tag{31}$$

On the other hand, the second moment of the distance between two arbitrarily located nodes in a BPP is derived in [7] as

$$E[x_i^2] = \frac{r^2 i}{(N+1)}. \tag{32}$$

Integration of (32) into our formulation gives

$$\sum_{i=1}^{n_0/n_1} E[x_i^2] = \sum_{i=1}^{n_0/n_1} \frac{\pi^2 R^2 i}{(np)^2((n-np)/np+1)} = \frac{(n-np)\pi^2 R^2}{2(np)^3}. \tag{33}$$

Generalizing (33) for np fog nodes results in

$$\tilde{x} = \frac{(n-np)\pi^2 R^2}{2(np)^2}. \tag{34}$$

Using (31) and (34), the objective function in (27) becomes

$$J_2 = \frac{(n-np)\pi^2 R^2}{2(np)^2} + \frac{2npa^2}{3} \tag{35}$$

Lemma 2: There is a unique optimum global value of p that minimizes (35).

Proof: One can easily show that (35) is a concave upward function by inspecting its second derivative. For this purpose, the first and second derivative of (35) is consecutively written as

$$\frac{\partial J_2}{\partial p} = \frac{4a^2n^2p^3 + 3\pi^2R^2p - 6\pi^2R^2}{6np^3} \tag{36}$$

and

$$\frac{\partial^2 J_2}{\partial p^2} = \frac{\pi^2 R^2(3-p)}{np^4}. \tag{37}$$

It is clear from (37) that the acquired objective function in (35) is strictly convex for $0 < p < 1$. This means that any real root of (36) minimizes the objective function if such kind of a p exists. After some mathematical operations, it can be shown that the optimum value is nearly equal to

$$p = \left(\frac{6\pi^2R^2}{4a^2n^2}\right)^{1/3}. \tag{38}$$

Hence, (38) is the unique global minimum point, which minimizes (35). □

Suppose that there are n number of nodes in an area. According to Lemma 2, the optimum fog number for this area becomes

$$n_1 = \left(\frac{6\pi^2R^2 n}{4a^2}\right)^{1/3}.$$

Analogously, if the number of fog nodes is known as *a priori* information, then one can easily find n_0 as

$$n_0 = \frac{n_1}{p} - n_1$$

where

$$p = \frac{6\pi^2R^2}{4a^2n_1^2}.$$

C. SHADOWING AND FADING

In wireless channels, shadowing and fading are other factors that affect the path loss exponent in addition to free space path loss. Accounting for the impact of free space path loss, shadowing and fading, it is reasonable to take $\alpha = 4$ [18] and formulate the objective function accordingly. More rigorously,

$$J_4 = \min_p(\hat{x} + \hat{y}) \tag{39}$$

where

$$\hat{y} = \sum_{j=1}^{n_1} E[y_j^4] \tag{40}$$

and

$$\hat{x} = \sum_{i=1}^{n_0} E[x_i^4]. \tag{41}$$

The expression in (40) can be written similar to (15) and (30) as

$$E[y_j^4] = \frac{1}{4a^2}\int_{-a}^{a}\int_{-a}^{a}(i_x^2 + i_y^2)^2 d_{i_x}d_{i_y} = 0.62a^4 \tag{42}$$

which results in

$$\hat{y} = \sum_{j=1}^{n_0} E[y_j^2] = 0.62npa^4. \tag{43}$$

The fourth moment of a distance that belongs to two points in a BPP can be specified as [7]

$$E[x_i^4] = \frac{r^4 i^2}{(N+1)^2}. \tag{44}$$

In the problem at hand, (44) can be interpreted as

$$\sum_{i=1}^{n_0/n_1} E[x_i^4] = \sum_{i=1}^{n_0/n_1} \frac{\pi^4 R^4 i^2}{(np)^4((n-np)/np+1)^2}. \quad (45)$$

This leads to

$$\hat{x} = \sum_{i=1}^{n_0/n_1} \frac{\pi^4 R^4 i^2}{(np)^3((n-np)/np+1)^2} \approx \frac{\pi^4 R^4 n(1-p)}{(np)^4}. \quad (46)$$

Due to (43) and (46), J_4 can be given by

$$J_4 = \frac{\pi^4 R^4 n(1-p)}{(np)^4} + 0.62npa^4. \quad (47)$$

Lemma 3: There is a unique optimum global value of p that minimizes (47).

Proof: The first and second derivative of (47) becomes

$$\frac{\partial J_4}{\partial p} = \frac{31a^4n^4p^5 + 150\pi^4R^4p - 200\pi^4R^4}{50n^3p^5} \quad (48)$$

and

$$\frac{\partial^2 J_4}{\partial p^2} = \frac{4\pi^4R^4(5-3p)}{n^3p^6} \quad (49)$$

respectively. Since (49) is greater than 0 for $0 < p < 1$, (47) is strictly convex. Then, the real root of (48), which is nearly equal to

$$p = \left(\frac{200\pi^4R^4}{31a^4n^4}\right)^{1/5} \quad (50)$$

is the unique minimum. □

As a consequence, the optimum number of fog nodes is equal to

$$n_1 = \left(\frac{200\pi^4R^4n}{31a^4}\right)^{1/5}.$$

Let's assume that the number of fog nodes is given as *a priori* information, i.e., n_1 is known. Then, the optimum number of n_0 can be calculated as

$$p = \frac{200\pi^4R^4}{31a^4n_1^4}$$

so that

$$n_0 = \frac{n_1}{p} - n_1.$$

V. VALIDATION OF ANALYSES

To develop more insights about the optimum number of fog nodes in a given area for different path loss exponents, the derived closed-form objective functions in (23), (35), (47) are numerically analyzed. In particular, these functions are plotted with respect to p and the optimum values of p that minimize the objective functions are found. The numerically derived p values are compared with the values that are found analytically (with an approximation) for different path loss exponents in (26), (38) and (50). Following that, the optimum number of fog nodes and the average number of end devices that are controlled by a single fog node are specified for a given n. Lastly, the improvement of the average SINR due to the optimized number of fog nodes is quantified in terms of data rate that can trivially affect the latency considering the transmission delay.

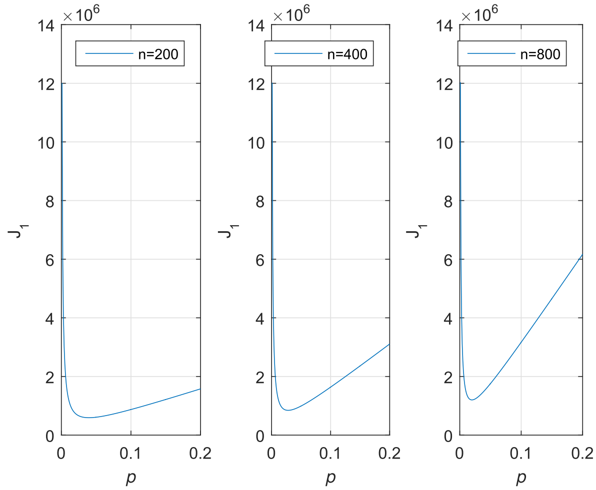


FIGURE 4. The objective function for $\alpha = 1$ in terms of p.

TABLE 1. The comparison of analytical and numerical results and the optimum number of fog nodes with average end devices for $\alpha = 1$.

	Analytical value of p in (26)	Numerical value of p	Average number of end devices	Average optimum number of fog nodes
$n = 200$	0.0396	0.04	24.25	7.92
$a = 400$	0.028	0.028	34.71	11.2
$a = 800$	0.0198	0.02	49.5	15.84

In our network model, a cloud covers a large area, and thus a is selected as 50 km, which corresponds to an area of 10000 km^2. On the other hand, each local fog network is responsible for a relatively small area. The radius of the circular fog network is taken as $R = 0.0765a$ without any loss of generality, which corresponds to one-tenth of the average distance between the fog node and cloud, that covers an area of roughly 45 km^2. The analysis is repeated for a total number of 200, 400 and 800 nodes for one fog network and $\alpha = 1$, $\alpha = 2$ and $\alpha = 4$.

In the first case, the objective function for $\alpha = 1$ in (23) is obtained for $n = 200$, $n = 400$ and $n = 800$ as depicted in Fig. 4. Notice that as proven in Lemma 1, each curve has a unique minimum point, and this point specifies the optimum number of fog nodes. Note that the ratio of fog nodes to the total number of nodes decreases with increasing n. This suggests that more devices should be handled by fog nodes in ultra-densely deployed networks. To be more specific, Table 1 gives the optimum number of fog nodes and specifies the average end devices in each fog node for different values of n. Furthermore, the minimum value of p found in (26) for $\alpha = 1$ is compared with that obtained numerically in Table 1.

A simulation is performed to compare the average data rate for the optimized and unoptimized number of fog nodes

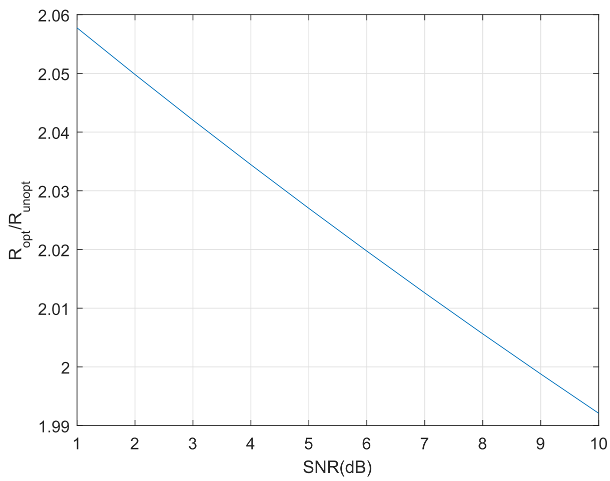


FIGURE 5. **The gain with the optimized number of fog nodes in terms of the average data rate when $\alpha = 1$.**

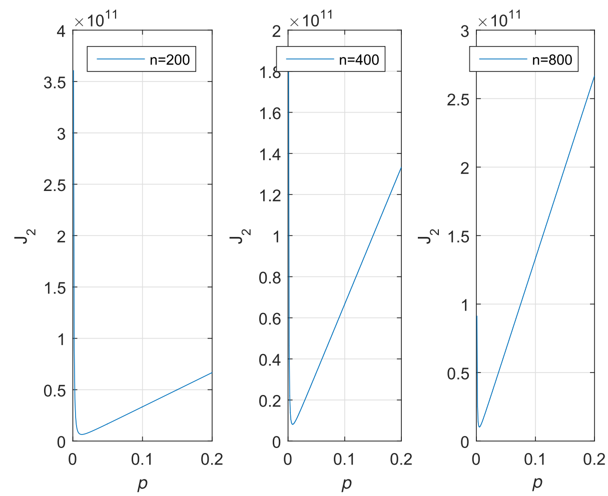


FIGURE 6. **The objective function for $\alpha = 2$ in terms of p.**

TABLE 2. **The comparison of analytical and numerical results and the optimum number of fog nodes with average end devices for $\alpha = 2$.**

	Analytical value of p in (38)	Numerical value of p	Average number of end devices	Average optimum number of fog nodes
$n = 200$	0.0129	0.013	76.51	2.58
$a = 400$	0.0082	0.008	120.95	3.28
$a = 800$	0.0051	0.005	195.07	4.08

based on the values in Table 1. It is assumed that there are 200 nodes within the area of interest without any loss of generality. For the former case, i.e., the optimized number of fog nodes, there are nearly 8 nodes at the fog layer and 192 nodes at the thing layer. For the latter unoptimized case, we randomly generate different number of fog nodes and take their average. Then, the ratio of the average data rate for the optimized number of fog nodes R_{opt} and the average data rate for the unoptimized number of fog nodes R_{unopt} becomes as illustrated in Fig. 5. As one can observe, the ratio slightly decreases for higher signal-to-noise ratio (SNR) values, however, there is still significant advantage of optimizing the number of fog nodes. Note that this benefit will grow as the bandwidth increases. For instance, if the bandwidth is 100 MHz, the difference in throughput due to the optimization becomes approximately 200 Mbps or if the bandwidth is equal to 200 MHz, the throughput enhancement becomes nearly 400 Mbps. Additionally, it is straightforward to see this effect on latency considering the transmission delay, which is obtained by dividing the packet size by the data rate.

The same experiment is repeated for a path loss exponent of $\alpha = 2$ that represents the free space path loss, whose closed-form objective function is given by (35). The results for $\alpha = 2$ reveals similar characteristics as with $\alpha = 1$ and is depicted in Fig. 6. Here, the numerically determined p is nearly the same as the analytical result in (38) as shown in Table 2, which also displays the optimum number of fog nodes in one fog network and the average number of end devices in one fog node. One more important point is that the optimum number of fog nodes inside a fog network decreases with respect to the previous case, i.e., $\alpha = 1$ for the same number of n. For nodes whose signal power falls more rapidly, the fog nodes that are further from the cloud with respect to the other fog nodes will significantly decrease the performance. This means that it may be more advantageous to send the packets to the closest fog node instead of becoming a fog node that decreases the overall optimum number of fog nodes. In this case, the average number of end devices controlled by a fog node increases. This result indicates that more computational power is necessary for the fog nodes in case of channels with high path loss exponents.

The ratio of the average data rate between the optimized and unoptimized number of fog nodes are evaluated for $\alpha = 2$ as well in Fig. 7 when the total number of nodes is selected as 200. Although the ratio of R_{opt}/R_{unopt} shows little decreases with incremental SNR, it is higher than the case of $\alpha = 1$. In fact, the data rate approximately doubles once the number of fog nodes is optimized, which is quite important in future wireless networks considering the expected increase in user demand.

Lastly, channels that are subject to shadowing and fading are considered to determine the optimum number of fog nodes to maximize the average data rate. Here, the path loss exponent is selected as $\alpha = 4$ as in the derived objective function (47). The results for this case are illustrated in Fig. 8 and Table 3. Compared to $\alpha = 1$ and $\alpha = 2$, the fewer number of fog nodes, each of which has higher number of end devices, are needed for $\alpha = 4$. It can be deduced that providing services to the end devices by a fog node is more challenging for channels with higher path loss exponents, because the increase in the number of end devices complicates the data processing and resource allocation.

When it comes to specifying the average data rate for the optimized number of fog nodes, there is a considerable enhancement with respect to the unoptimized one as depicted

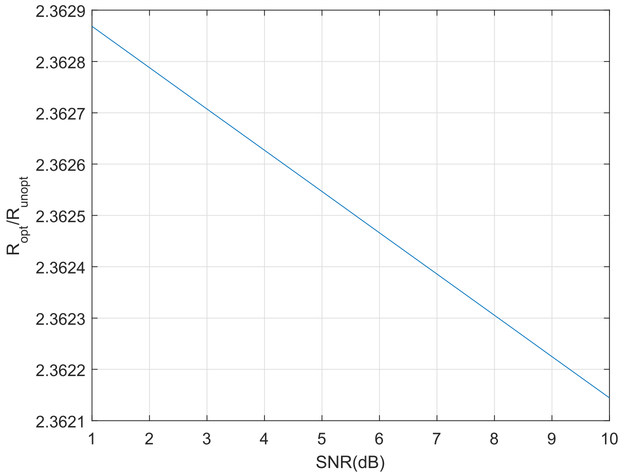


FIGURE 7. The gain with the optimized number of fog nodes in terms of the average data rate when $\alpha = 2$.

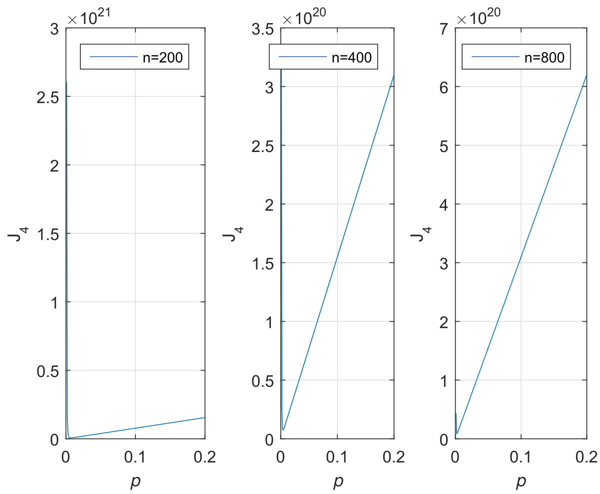


FIGURE 8. The objective function for $\alpha = 4$ in terms of p.

TABLE 3. The comparison of analytical and numerical results and the optimum number of fog nodes with average end devices for $\alpha = 4$.

	Analytical value of p in (50)	Numerical value of p	Average number of end devices	Average optimum number of fog nodes
$n = 200$	0.0067	0.007	148.25	1.34
$a = 400$	0.0038	0.004	262.15	1.52
$a = 800$	0.0022	0.002	453.54	1.76

in Fig. 9. More specifically, the average data rate increases almost be an order of magnitude if the number of fog nodes is optimized. Notice that the improvement is nearly the same for different SNR values. This emphasizes that it is much more important to optimize the number of nodes in the fog layer and the thing layer for channels with higher path loss exponents.

The overall results imply that the boost in the number of total nodes leads to a moderate increase in the number of fog nodes, whereas there is a significant increase in the number of end devices. Furthermore, fewer fog nodes are sufficient in case of severe fading. This indicates that each fog node must have more resources to manage the needs of users when there are many nodes in channels with high path loss exponents. This result also indicates the importance of cooperation and virtualization in fog networks as we discuss in the next section. A limiting situation can be to have a large number of path loss exponent channels that result in a single fog node with many end devices. As a result, the channels with smaller path loss exponent have a tendency to have more fog nodes implying more distributed networking while the channels with higher path loss exponents tend to a more centralized solution to maximize the data rate.

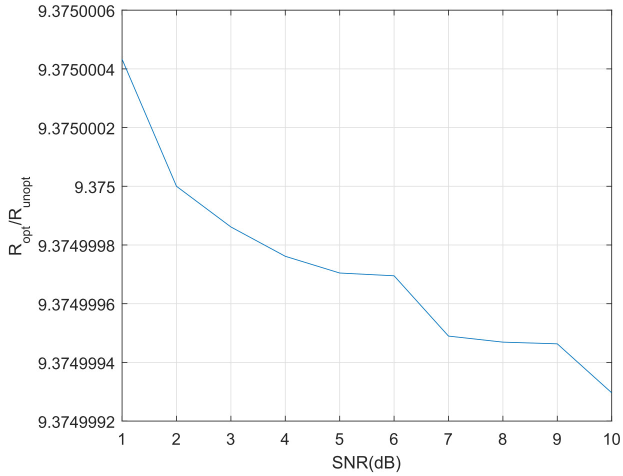


FIGURE 9. The gain with the optimized number of fog nodes relative to the average data rate when $\alpha = 4$.

VI. ADDITIONAL BENEFITS AND FUTURE WORK

Optimizing the number of fog nodes brings with it numerous benefits for the design of *cloud-fog-thing* networks. Regarding the cloud layer of this architecture, whose primary advantage comes from virtualization, the problem of underutilized or overutilized virtual machines is one of the problems that decreases the efficiency of virtualization. In general, to balance the loads in virtual machines that run on physical machines, prediction based algorithms that observe the past statistics are employed [19]. However, if the load dynamically changes, i.e., it is time-varying, these sorts of algorithms do not give accurate estimates. In this situation, the number of dynamically optimized number of fog nodes may be used to assist in the design of the virtual machines, which can be viewed as fog-aided cloud virtualization. The basic idea here is to associate the virtual machines in the cloud with different fog nodes whose optimum number is determined according to the total number of nodes in the area of interest, and each node has a balanced average load. That is, if virtual machines were created such that each virtual machine became responsible for an equal number of fog nodes, all virtual machines would have more balanced loads. This could minimize the number of physical machines and lead to more efficient green computing as well. Notice that each fog node needs to communicate with a virtual machine in the cloud for many reasons, e.g., to update its cache, manage its resources more efficiently, or send some portion of data coming from end devices to the virtual machines for further processing. The details of this subject will be handled in future work.

Fog layer design requires not only knowing the optimum number of fog nodes but also finding the locations of fog nodes. That is, which nodes in the network are upgraded as fog nodes, among the many alternatives has to be determined. Clustering algorithms can be used to determine the optimum locations of the fog nodes. One of the widely used clustering algorithm in machine learning is the K-means clustering algorithm based on the principle of minimizing inter-cluster distances [20]. Accordingly, the geographical locations of the potential fog nodes can be considered as a data set that can be clustered with a K-means clustering algorithm so that the closest nodes to the center of each cluster, i.e., cluster-heads, can be upgraded to fog nodes. Despite its simplicity and efficiency, the major drawback of this algorithm is in the determination of the value of K. It is not clear how one should select K, and there are only heuristics instead of mathematical analysis [21], [22]. As a promising solution, the stochastic geometry analysis given in this paper can fulfill this gap. Specifically, the value of K, which is the optimum number of fog nodes, can be analytically obtained as $K = np$ where n is given as *a priori* information and p is derived as a closed-form expression in (26), (38) and (50). The details of this subject will be explored in future work as well.

QoE for users in the thing layer is highly related to the efficient caching mechanism in the fog layer. A recent paper reveals that the performance of caching depends on both the capacity of the front-haul network between the fog and the cloud, as well as the caching resources in the fog nodes for *cloud-fog-thing* network [23]. This means that even upgrading all of the nodes to fog nodes to exploit the unused resources for the sake of caching is not sufficient to have better caching performance. Therefore, the optimized number of fog nodes that can improve the average data rate within the network will enhance the front-haul capacity and affect the caching. It is worth noting that evaluating the caching performance quantitatively in terms of the number of fog nodes is a good research problem.

VII. CONCLUSIONS

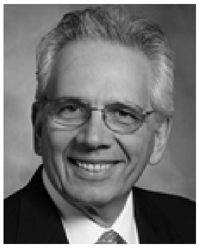
Determining the optimum number of fog nodes, which has been one of the open questions in the *cloud-fog-thing* architecture, is found using the tools of stochastic geometry. This optimization enhances the average data rate and minimizes the transmission delay. It is quite meaningful and important to maximize the average data rate especially for the networks that require substantial data processing. Optimizing the number of fog nodes significantly improves the average data rate. To illustrate, the average data rate doubles and increases by almost an order of magnitude for the free-space path loss channels, and shadowing and fading channels, respectively. Indeed, having more than or less than the optimum number of fog nodes degrades the average data rate, and its effect becomes greater for the channels with high path loss exponents. Furthermore, the optimum number of fog nodes decreases for high path loss exponents channels indicating that fog nodes must be carefully selected among the nodes that have the highest computational power for these channels. The results presented in this paper provide guidelines for the translation of theoretical results on fog networking to practical network implementation. Our results may be quite useful in the design of cloud virtualization, while future work can be directed towards determining the optimum locations of fog nodes and enhancing caching performance.

REFERENCES

[1] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. 1st ACM Wksp. Mobile Cloud Comput.*, 2012, pp. 13–16.
[2] M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 854–864, Dec. 2016.
[3] B. Tang, Z. Chen, G. Hefferman, T. Wei, H. He, and Q. Yang, "A hierarchical distributed fog computing architecture for big data analysis in smart cities," in *Proc. ASE Big Data Social Informat. (ASE BDSI)*, 2015, Art. no. 28.
[4] T. H. Luan, L. Gao, Z. Li, Y. Xiang, G. Wei, and L. Sun. (2015). "Fog computing: Focusing on mobile users at the edge." [Online]. Available: https://arxiv.org/abs/1502.01815
[5] S. Yi, C. Li, and Q. Li, "A survey of fog computing: Concepts, applications and issues," in *Proc. Workshop Mobile Big Data*, 2015, pp. 37–42.
[6] S. A. Banani, A. W. Eckford, and R. S. Adve, "Analyzing the impact of access point density on the performance of finite-area networks," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 5143–5161, Dec. 2015.
[7] S. Srinivasa and M. Haenggi, "Distance distributions in finite uniformly random networks: Theory and applications," *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 940–949, Feb. 2010.
[8] M. Peng, S. Yan, K. Zhang, and C. Wang, "Fog-computing-based radio access networks: Issues and challenges," *IEEE Netw.*, vol. 30, no. 4, pp. 46–53, Jul. 2016.
[9] S.-C. Hung, H. Hsu, S.-Y. Lien, and K.-C. Chen, "Architecture harmonization between cloud radio access networks and fog networks," *IEEE Access*, vol. 3, pp. 3019–3034, Dec. 2015.
[10] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. Hawaii Int. Conf. Syst. Sci.*, Jan. 2000, p. 10.
[11] V. Loscri, G. Morabito, and S. Marano, "A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH)," in *Proc. 62nd IEEE Veh. Technol. Conf. (VTC-Fall)*, Sep. 2008, pp. 1809–1813.
[12] O. Younis and S. Fahmy, "Distributed clustering in ad-hoc sensor networks: A hybrid, energy-efficient approach," in *Proc. IEEE INFOCOM*, Mar. 2004, p. 640.
[13] A. Youssef, M. Younis, M. Youssef, and A. Agrawala, "Distributed formation of overlapping multi-hop clusters in wireless sensor networks," in *Proc. 49th IEEE Global Commun. Conf. (Globecom)*, Nov./Dec. 2006, pp. 1–6.
[14] M. Youssef, A. Youssef, and M. Younis, "Overlapping multihop clustering for wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 20, no. 12, pp. 1844–1856, Dec. 2009.
[15] S. Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *Proc. INFOCOM*, Mar./Apr. 2003, pp. 1713–1723.
[16] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief. (2017). "A survey on mobile edge computing: The communication perspective." [Online]. Available: https://arxiv.org/abs/1701.01090
[17] M. Haenggi and R. K. Ganti, "Interference in large wireless networks," *Found. Trends Netw.*, vol. 3, no. 2, pp. 127–248, 2009.
[18] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2004
[19] Z. Xiao, W. Song, and Q. Chen, "Dynamic resource allocation using virtual machines for cloud computing environment," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 6, pp. 1107–1117, Jun. 2013.

[20] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, Aug. 2006.
[21] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.
[22] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *J. Roy. Statist. Soc., B (Statist. Methodol.)*, vol. 63, no. 2, pp. 411–423, 2001.
[23] R. Tandon and O. Simeone, "Cloud-aided wireless networks with edge caching: Fundamental latency trade-offs in fog radio access networks," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2029–2033.
[24] E. Balevi and R. D. Gitlin, "Unsupervised machine learning in 5G networks for low latency communications," in *Proc. IEEE Int. Perform. Comput. Commun. Conf. (IPCCC)*, Dec. 2017, pp. 1–2.
[25] E. Balevi and R. D. Gitlin, "A clustering algorithm that maximizes throughput in 5G heterogeneous F-RAN networks," in *Proc. IEEE Int. Conf. Commun.*, May 2018, pp. 1–6.

EREN BALEVI received the B.S., M.S., and Ph.D. degrees in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 2008, 2010, and 2016, respectively. He is currently a Post-Doctoral Scholar with the Electrical Engineering Department, University of South Florida. His current research interests are in the general areas of 5G wireless systems and the Internet of Things in addition to the general areas of molecular communications, multi-user wireless communications, and signal processing.



RICHARD D. GITLIN (S'67–M'69–SM'76–F'86–LF'09) has over 45 years of leadership in the communications and networking industry. He was at Bell Labs/Lucent Technologies for 32 years performing and leading pioneering research and development in digital communications, broadband networking, and wireless systems. He was the Senior Vice President for Communications and Networking Research, Bell Labs, and later the CTO of Lucent's Data Networking Business Unit. After retiring from Lucent, he was a Visiting Professor of electrical engineering at Columbia University. He is a Bell Laboratories Fellow. He is a member of the National Academy of Engineering and a Charter Fellow of the National Academy of Inventors. He was a co-recipient of the 2005 Thomas Alva Edison Patent Award and the S.O. Rice Prize. He is currently a State of Florida 21st Century World Class Scholar, a Distinguished University Professor, and the Agere Systems Chaired Distinguished Professor of electrical engineering at the University of South Florida.