

Article

A Latency-Aware Offloading Strategy over Fiber-Wireless (FiWi) Infrastructures for Tactile Internet Services

Qinglong Dai ^{1,*}, Jin Qian ², Guangjun Qin ¹, Jianwu Li ³ and Jun Zhao ⁴

¹ Smart City College, Beijing Union University, Beijing 100100, China; zhtguangju@buu.edu.cn

² Dareway Software Co., Ltd., Jinan 250101, China; jinjian@dareway.org.cn

³ Advanced Technology Research Institute, Beijing Institute of Technology, Beijing 100811, China; jianwuli@163.com

⁴ China Telecom Beijing Research Institute, Beijing 100035, China; junzhao@163.com

* Correspondence: xxtqinglong@buu.edu.cn

Abstract: With the emergence of the tactile internet, low-latency, even, real-time data transmission is indispensable for human-agent–robot teamwork. Offloading is considered a feasible approach. Determining the offloading solution according to the dynamic network circumstance is attractive. In this paper, we investigate the resource management issue in a three-tier, heterogeneous, fiber-wireless (FiWi) network with offloading. Based on the model of the wireless link, the fiber link, the data rate, and the offloading, a mixed-integer, non-linear problem is formulated to obtain the minimum total latency for tactile internet services. Through constraint relaxation, MINLP is converted to a linear problem (LP). A Lagrange multiplier method with Karush–Kuhn–Tucker (KKT) conditions is used to solve LP. Using the numerical simulation, the superiority of our work is evaluated and compared with the previous work.



Citation: Dai, Q.; Qian, J.; Qin, G.; Li, J.; Zhao, J. A Latency-Aware Offloading Strategy over Fiber-Wireless (FiWi) Infrastructures for Tactile Internet Services. *Appl. Sci.* **2022**, *12*, 6417. <https://doi.org/10.3390/app12136417>

Academic Editors: George Kalfas, Marios Gatzianas and Agapi Mesodiakaki

Received: 28 April 2022

Accepted: 21 June 2022

Published: 24 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rapid advancement of information and communications technology (ICT) causes the data exchange paradigm to evolve, from human-to-human (H2M) to human-to-machine (H2M), and even to machine-to-machine (M2M) [1]. Correspondingly, the latency requirement is also getting higher. This very low-latency communication gives rise to the tactile internet [2,3]. The tactile is considered the next evolution of the Internet of Things (IoT), supporting the real-time haptic data transmission (e.g., touch, action) for the remote object/machine control. For the tactile internet, low-latency, accurate, and collaborative human-agent–robot teamwork (HART) is important.

The tactile internet involves enormous, wide-area-distributed sensors/machines. Wire and wireless access networks are essential [4]. The fiber-wireless (FiWi) network is one of them, as shown in Figure 1, which integrates the wired fiber backhaul and wireless fronthaul. In the wired fiber backhaul, a tree-like topology with the IEEE 802.3 ah/av 1/10 G passive optical network (PON) protocol that connects the central optical line terminal (OLT) to four optical network units (ONUs) is adopted. To provide fiber-to-the-home/building/road (FTTx) services to fixed subscribers, an ONU is located at subscriber communities (i.e., the first sub-network). This fixed method is not relevant to the main content of this work. The second sub-network is equipped with a based station (BS)-aided ONU (i.e., ONU-BS). The mesh portal point (MPP)-aided ONU (ONU-MPP), the mesh point (MP), and the mesh access point (MAP), are deployed to host IoT services in the third sub-network. Note that multi-access edge computing (MEC) can be used on ONUs. Finally, some computing-centric tasks can be offloaded onto MEC for the tactile services. This scenario forms the background of our work.

As the physical underlying infrastructure of the tactile internet, the FiWi network has attracted much attention. An efficient FiWi network means good support of the tactile internet service. The efficient indicator can be latency, throughput, energy consumption, etc. The seamless integration of wired fiber backhaul and wireless fronthaul is the key issue. Surrounding this issue, the energy-efficient frame aggregation [5], throughput optimization [6], and medium access control (MAC) protocol design [7] were conducted. These three works focused on the improvement of the FiWi network's performance, neglecting the uniqueness of the tactile internet.

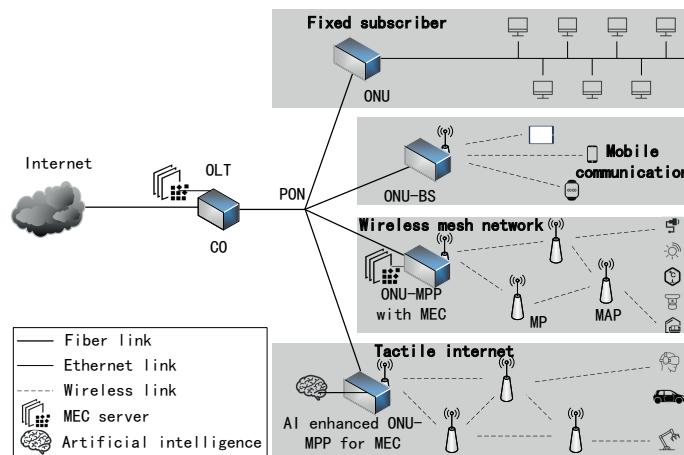


Figure 1. The FiWi architecture.

The tactile internet is a kind of network that can transfer not only pictures and voice, but also touch, sense, scent, etc. Human experience, service data, and tactile internet are interact with each other online and with feedback, through joysticks, digital gloves, robots, augmented reality (AR)/virtual reality (VR)/mixed reality (MR), and other instruments. The online interaction in this instance is that a man can operate a remoted object without any latency or the sense of lag. The increase in service data dimensions (such as pictures, voice, touch sense, etc.) leads to an increase in the amount of data in tactile services.

The tactile internet has a latency requirement of 1–10 milliseconds, which is much more demanding than current networks. HART empowers humans with artificial intelligence (AI) based on agile computing. The tactile internet complement humans rather than providing a substitute for them [8]. The human is a member of HART, not just a conventional user. For the tactile internet, its characteristics of low latency, enormous access, heterogeneity, and intelligence, are in line with the evolution direction beyond 5th generation (B5G)/6th generation (6G) communications.

MEC is a promising way to reduce latency by physically or logically moving data-centric computing to the infrastructure for the near-end user [9]. By 2025, the number of connected IoT devices will be 41.6 billion. They will generate 79.4 zettabytes of data for computing [10]. In the form of offloading, the computing task can immigrate from the central cloud far from humans to the distributed edge (e.g., access points). This directly decreases the latency. However, depending on service distribution and characteristics, deciding where a computing task should be executed remains still a problem.

In a FiWi network for tactile internet services, offloading computing workload to edge nodes via MEC is feasible. This offloading can reduce the data transmission path. Therefore, the latency is declined. The offloading destination place, i.e., gateway/ONU, OLT, and the internet, provide different kinds of latency reductions. The decision to offload a tactile internet service to gateway/ONU, OLT, and the internet, becomes the point.

To solve this problem, a latency-aware offloading strategy over FiWi infrastructures for tactile internet services is proposed in this paper. The contributions of our work are listed as follows.

- The FiWi network with MEC is modeled, including the wireless link, the fiber link, the data rate, and the offloading. Note that the multiple-hop communication, not directed, one-single-hop communication, and three kinds of offloading are taken into consideration. The interference of all the other users for a specific user is modeled.
- To obtain total minimum latency, a mixed-integer non-linear programming (MINLP) problem is formulated. This MINLP is further converted to a linear programming (LP) problem using the constraint condition relaxation. The Lagrange multiplier method is used to solve the converted LP problem.
- Through numerical simulation, our work is compared with the previous work regarding aspects of the data transmission rate, the signal strength, and the total latency.

The rest of this paper is organized as follows. In Section 2, works on low-latency tactile services regarding FiWi infrastructures are analyzed. Section 3 provides a systematic modeling of the wireless link, the fiber link, the data rate, and the offloading, and proposed a minimized total latency MINLP problem. A two-step solution to the proposed problem is represented in Section 4. Section 5 is devoted to presenting the characteristics of our work via numerical simulation. Finally, Section 6 concludes our work and looks ahead to future research.

2. Related Work

Many works have been completed, ranging from those focusing on the underlying FiWi infrastructures to the top-level, time-sensitive services in the tactile internet.

Yang et al. held that different applications demand different levels of intelligence and efficiency in data-processing [11]. Multi-tier computing for real-time, data-streaming processing, which integrates cloud-, fog-, and edge-computing, associated with offloading, was required for intelligent, tactile internet applications. Mondal et al. derived the exact expressions of a typical Voronoi area of a multi-tier Poisson network with a generalized association rule [12].

Chen et al. surveyed the future 6G vision of IoT and its convergence with the radio-over-fiber (RoF) system. They deemed that the RoF-based FiWi network was one of the most promising enablers of 6G IoT [13]. To achieve 6G, highly efficient data training or computing was absolutely necessary.

Maier et al. pointed out that the internet of everything in 5G would transform into the internet of no-things and provided an experiment looking at how the internet of no-things served as a useful stepping stone toward the far-reaching vision of future 6G networks [14]. The internet of no-things that they mentioned referred to the internet's unnoticed and silent integration into people's lives. From the perspective of network evolution, low latency is significant.

To cope with the problem that IoT devices do not have enough available capacity to collaboratively co-host newly emerged tasks, Li et al. proposed a novel resource allocation and service co-placement (RaSP) algorithm [15]. These tasks were latency-aware, online, and re-configurable. Mondal et al. proved that the latency was a stubborn bottleneck for tactile internet applications. They thought that the edge-based AI forecasting algorithm could alleviate this dilemma [16].

Wu et al. concentrated on minimizing the overall latency under the quality of service (QoS) in the indoor, multimode, radio-frequency-optical, heterogeneous network, without considering the outdoor environment [17]. Guo et al. conducted a software-defined, time-sensitive networking experiment [18]. They demonstrated that the time synchronization of cross-domain interconnection was helpful to reduce latency. However, the underlying network infrastructure was neglected.

Peréz et al. studied the different delay requirements in a FiWi-enhanced, heterogeneous, cloud radio-access network (H-CRAN) with MEC and offloading. A decentralized algorithm that maximized the utility of all users while meeting the delay requirement of different services was proposed [19]. Only wireless mobile communication was taken into consideration in their work.

Pan et al. investigated the delay minimization problem for task offloading in a hierarchical, fog-computing C-RAN network [20]. In the formulated MINLP, they minimized the total delay in the computing tasks by optimizing the receive beamforming vectors, task allocation, each server's computing speed for offloading tasks, and the bandwidth of fronthaul links. Note that the wireless communication in the C-RAN network was realized via the single-hop link. Their work is most relevant to ours. In our work, the wireless part was the multi-hop link. The wire part was a fiber network, while their work had no specifications for the wire part. They also did not consider the application of tactile internet services.

3. System Model

3.1. The Key Notations

The key notations are listed in Table 1. Symbols not mentioned are explained at the time of first use.

Table 1. Key Notations.

Symbol	Description
i	user/human/machine i
k	k hops
$Neb(i, k)$	Set of MAP/MP/ONUs that are connected to user/human/machine i
\mathbb{I}	Set of user/human/machine
\mathbb{J}	Set of ONU
\mathbb{R}	The set of OLT
\mathbb{C}	The Internet
y_j	The received signal vector for MAP
p_0	Transmission power
h_{ji}	wireless channel from i to j
s_i	Transmission signal
n	Gaussian noise
α	Coefficient of signal attenuation
y_k	Received signal vector for ONU
y_r	Received signal vector for OLT
β	Fiber attenuation coefficient
L_{rj}	The distance from ONU to OLT
$SINR_i$	Received signal to interference plus noise ratio
h_{ki}	The wireless channel from i to k
σ^2	the noise power of the Gaussian noise
R_i^X	The data rate of user/human/machine i via wireless links
B_i^X	The bandwidth of wireless links

3.2. Basic Description

The FiWi network in our work consists of a wavelength division multiplexing (WDM)-based fiber network (i.e., WDM-based PON) and an 802.11 protocol-based wireless network. At the locations of OLT and ONUs, the cache is deployed for MEC, enabling traffic offloading. Note that the links from ONU to OLT and those from OLT to Internet are both fiber. However, the underlying technologies, such as carrier wavelength, are different. Each user/human/machine has a single transmit antenna. Each MAP/MP/ONU is equipped with M -receiving antennas. The user/human/machine is served by its nearest MAP/MP/ONU. Through one single or multiple hops in the wireless network, each user/human/machine is connected to one ONU. The set of MAP/MP/ONUs that are connected to user/human/machine i via k hops is represented by $Neb(i, k)$.

As shown in Figure 2, four kinds of location are involved in the FiWi network scenario. These four kinds of location, in order of the distance from users, are user/human/machine, ONU with MEC, OLT with MEC, and the Internet. The ONU with MEC is seen as one-tier. The OLT with MEC is seen as two-tier. Internet is seen as three-tier. The computing services

traditionally occur on the Internet. The task data have to be transmitted in sequence through user/human/machine, ONU, OLT, and the Internet. After data computing, results or response contents return from the Internet to the user/human/machine. With the application of MEC, data computing can migrate from the Internet to ONU with MEC or OLT with MEC, i.e., the task offloading in Figure 2. The path that involved service reductions from user/human/machine-ONU-OLT-Internet to user/human/machine-ONU or user/human/machine-ONU-OLT. Latency is reduced due to the migration of data computing (i.e., task offloading) from the Internet to ONU with MEC or OLT with MEC. The MEC in ONU (i.e., one-tier) is denoted as MEC-1. The MEC in OLT (i.e., two-tier) is denoted as MEC-2. The original data computing on Internet (i.e., three-tier) is denoted as MEC-3.

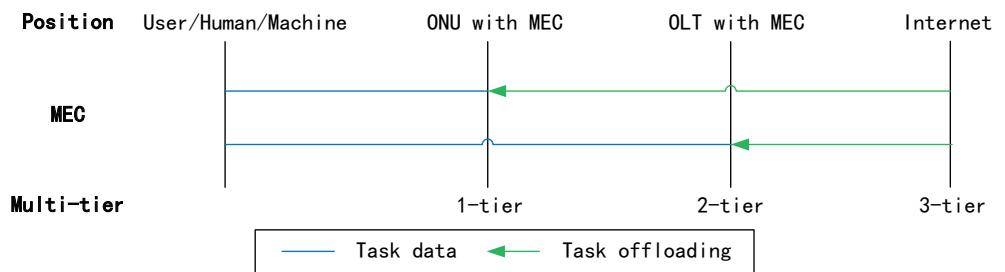


Figure 2. The task and offloading in the FiWi network equipped with three tiers of MEC services.

The set of user/human/machine is \mathbb{I} . User/human/machine $i \in \mathbb{I}$. The set of ONU is \mathbb{J} . ONU $j \in \mathbb{J}$. The set of OLT is \mathbb{R} . OLT $r \in \mathbb{R}$. The Internet is \mathbb{C} . Because that there is only one Internet in the world, \mathbb{C} needs not further refinement.

3.3. The Wireless Link Model

The received signal vector of MAP/MP/ONU j from user/human/machine i via one single hop is \mathbf{y}_j , $j \in \text{Neb}(i, 1)$.

$$\mathbf{y}_j = \sqrt{p_0} \sum_{j \in \text{Neb}(i, 1)} (\mathbf{h}_{ji} * s_i) + \mathbf{n} \quad (1)$$

where p_0 is the transmit power of user/human/machine i , s_i is the transmit signal satisfying $E[s_i^2] = 1$, \mathbf{h}_{ji} is the wireless transmit channel from user/human/machine i to MAP/MP/ONU j via one single hop. \mathbf{h}_{ji} is also a $M \times 1$ vector, and \mathbf{n} is the Gaussian noise-obeying $N(0, \sigma^2)$. s_i is a tactile service form user/human/machine i . The tactile service, such as remote digital gloves or robots, always requires the long-term transmission of interactive data between user/human/machine and task-computing cloud. s_i is considered to be related to the data transfer rate. Note that the signals of any two different users, s_i and s_l , are independent from each other, i.e., $E[s_i, s_l] = 0, \forall i \neq l$.

For the multi-hop, the signal is only relayed by the intermedia MAPs and MPs, without de-modulation and re-modulation. The received signal vector of MAP/MP/ONU j from user/human/machine i via a K -hop path ($K \geq 2$) is

$$\mathbf{y}_j = \alpha * \mathbf{y}_k + \mathbf{n} \quad (2)$$

where $j \in \text{Neb}(i, K)$, $k \in \text{Neb}(i, K - 1)$, and α is the coefficient of signal attenuation ($0 < \alpha \leq 1$).

Combining Equations (1) and (2), no matter the value of K is, \mathbf{y}_j is

$$\mathbf{y}_j = \alpha^{K-2} \sqrt{p_0} \sum_{j \in \text{Neb}(i, K)} (\mathbf{h}_{ji} * s_i) + \frac{1 - \alpha^{K-1}}{1 - \alpha} \mathbf{n} \quad (3)$$

3.4. The Fiber Link Model

Due to the absorption and scatter attenuation, the signal in the fiber link decays. The link from the ONU to the OLT or the Internet is fiber. The received signal vector of the OLT or the Internet, \mathbf{y}_r , from ONU j via fiber links is

$$\mathbf{y}_r = \left(\frac{\beta}{L_{rj}} \right) * \mathbf{y}_j \quad (4)$$

where β is the fiber attenuation coefficient, L_{rj} is the distance from ONU r to the OLT or the Internet.

3.5. The Data Rate

A receiver receives multiple signals from different source users/humans/machines simultaneously. For the target-estimated signal, the Gaussian noise and other non-target signals are both background noise.

The attenuation in the fiber link is linear, which can be seen in Equation (4). Therefore, whether data computing occurs at the place of ONUs, OLT, or Internet, the received signal to interference plus noise ratio (SINR) for user/human/machines i is

$$SINR_i = \frac{\alpha^2 p_0 \sum_j^K |\mathbf{h}_{ki}|^2}{\alpha^2 p_0 \sum_{k \neq j} |\mathbf{h}_{ki}|^2 + \left(\frac{1-\alpha^{K-1}}{1-\alpha} \right)^2 \sigma^2} \quad (5)$$

where σ^2 is the noise power of the Gaussian noise, \mathbf{n} , in Equation (1).

Let B_i^X be the bandwidth assigned to user/human/machine i via X link. $X \in \{W, F, Int\}$. W is the wireless link from user/human/machine to ONU (MEC-1). F is the fiber link from ONU to OLT (MEC-2). Int is the fiber link from OLT to Internet (MEC-3). Note that the carrier wavelengths of F and Int are different. The data rate of user/human/machine i via wireless links is

$$R_i^X = B_i^X \log(1 + SINR_i) \quad (6)$$

Each user/human/machine has a task to execute. The task of the user/human/machine i is denoted as task i . The data size of task i is denoted as D_i , and the CPU frequency cycles of task i data computing that are required by ONU, OLT, or Internet, are proportional to D_i with the coefficient η_i . The CPU frequency cycles required for data computing for task i are $\eta_i D_i$.

3.6. The Offloading Model

To represent the offloading decision in the FiWi network, the binary variables x_{1i} , x_{2i} , and x_{3i} are used. x_{1i} indicates the MEC offloading upon the ONU (1-tier). x_{2i} indicates the MEC offloading upon the OLT (2-tier). x_{3i} indicates the original data-computing in the Internet (three-tier).

$$C1 : x_{1i}, x_{2i}, x_{3i} \in \{0, 1\} \quad (7)$$

For different offloading, the latency is different.

(1) MEC-1 offloading: MEC-1 offloading reduces the latency, via offloading data-computing from the Internet to ONU. The transmission latency of task i via the wireless link is $T_i^W = D_i / R_i^W$. The data-computing latency of ONU is $T_{ji}^{ONU} = \frac{\eta_i D_i}{f_{ji}^{ONU}}$. f_{ji}^{ONU} is the task i data-computing speed of servers at the place of ONU j . The total latency of MEC-1 offloading is denoted as T_{1i} .

$$T_{1i} = T_i^W + T_{ji}^{ONU} \quad (8)$$

(2) MEC-2 offloading: MEC-2 offloading migrates data-computing from Internet to OLT. The transmission path for task i includes user/human/machine-ONU and ONU-OLT.

The transmission latency via the wireless link from user/human/machine to ONU is still T_i^W . The transmission latency via the fiber link from ONU j to OLT r is $T_{ri}^F = D_i/R_{ri}^F$. The data-computing latency of OLT r is $T_{ri}^{OLT} = \frac{\eta_i D_i}{f_{ri}^{OLT}}$. f_{ri}^{OLT} is the task i data-computing speed of servers at OLT r . The total latency of MEC-2 offloading is T_{2i}

$$T_{2i} = T_i^W + T_{ri}^F + T_{ri}^{OLT} \quad (9)$$

(3) MEC-3 offloading: If task i is handled in the Internet, the data-computing latency can be calculated by $T_{Ci}^{Int} = \frac{\eta_i D_i}{f_i^{Int}}$. f_i^{Int} is the computing frequency of the server in Internet for task i . The transmission path for task i includes user/human/machine-ONU, ONU-OLT, and OLT-Internet. The transmission latency via the fiber link from OLT to Internet is $T_{Ci}^{Int} = D_i/R_{Ci}^{Int}$. The transmission latency for task i from user/human/machine to ONU via the wireless link is T_i^W . The transmission latency for task i from ONU j to OLT r via the fiber link is T_{ri}^F . The total latency of MEC-3 offloading is T_{3i} .

$$T_{3i} = T_i^W + T_{ri}^F + T_{Ci}^{Int} + T_{Ci}^{Int} \quad (10)$$

3.7. Problem Statement

For task i , the data-computing can be offloading onto ONU, OLT, and the Internet. Only one place can be selected for task i data-computing. This place constraint is

$$C2 : x_{1i} + x_{2i} + x_{3i} = 1 \quad (11)$$

The computing capacities of MEC-1, MEC-2, and MEC-3 are different. The sum of the required computing CPU frequency cycles for all the tasks can not exceed the CPU frequency cycles of the servers in ONU (MEC-1), OLT (MEC-2), and Internet (MEC-3). The CPU frequency cycles constraints are:

$$C3 : \sum_{i \in Neb(j,K)} \sum_{k=1}^K x_{1i} * f_{ji}^{ONU} \leq F_j^{ONU} \quad (12)$$

$$C4 : \sum_j \sum_{i \in Neb(j,K)} \sum_{k=1}^K x_{2i} * f_{ri}^{OLT} \leq F_r^{OLT} \quad (13)$$

$$C5 : \sum_r \sum_j \sum_{i \in Neb(j,K)} \sum_{k=1}^K x_{3i} * f_i^{Int} \leq F_C^{Int} \quad (14)$$

Equation (12) means that the occupied data-computing frequency on ONU j cannot exceed the physical frequency of ONU j . Equation (13) means that the occupied data-computing frequency on OLT r cannot exceed the physical frequency of OLT r . Equation (14) means that the occupied data-computing frequency on the Internet cannot exceed the physical frequency of the Internet. The links' abilities are limited. For links, the data rate constraints are:

$$C6 : \sum_r \sum_j \sum_{i \in Neb(j,K)} \sum_{k=1}^K x_{3i} * f_i^{Int} \leq F_C^{Int} \quad (15)$$

$$C7 : \sum_{\mathbb{C}} x_{3i} * R_{Ci}^F \leq R^{Int} \quad (16)$$

Equation (15) means that, for each ONU-OLT link, the sum of the MEC-1 offloading data rate and MEC-2 offloading data rate cannot exceed the physical data rate of this link. Equation (16) means that, for each OLT-Internet link, the MEC-3 offloading data rate cannot exceed the physical data rate of this link.

Then, the objective function of this optimization problem can be formulated as

$$\begin{aligned} \min & \quad \left[\sum_r \sum_j \sum_k (x_{1i} * T_{1i} + x_{2i} * T_{2i} + x_{3i} * T_{3i}) \right] \\ \text{s.t.} & \quad C1 - C7 \end{aligned} \quad (17)$$

4. Problem Solution and Analysis

Problem 17 is an MINLP problem, which is non-convex and not easy to directly solve. $\{x_{1i}, x_{2i}, x_{3i}, f_{ji}^{ONU}, f_{ri}^{OLT}, f_i^{Int}, R_{ri}^F, R_{Ci}^F\}$ are the variables to be determined. Note that each of $\{x_{1i}, x_{2i}, x_{3i}\}$ is an integer of either 0 or 1. Considering the co-existence of integer variables and real variables, Equation (17) can be solved in two steps. First, Equation (17) is converted into an LP problem by relaxing the constraints. Second, the Lagrange multiplier method is used to solve the converted LP problem.

4.1. The Constraint Relaxation

For the C1 constraint in Equation (7), the value of $x_{1i}/x_{2i}/x_{3i}$ is either 0 or 1. The constraint of x can also be denoted as

$$\begin{aligned} x_{1i}(x_{1i} - 1) &= 0 \\ x_{2i}(x_{2i} - 1) &= 0 \\ x_{3i}(x_{3i} - 1) &= 0 \end{aligned} \quad (18)$$

Selecting Equation (18) as penalty functions for variable x , Equation (17) is converted to

$$\begin{aligned} \min & \quad \left[\sum_r \sum_j \sum_k (x_{1i} * T_{1i} + x_{2i} * T_{2i} + x_{3i} * T_{3i}) \right. \\ & \quad + \sum_i \gamma_{1i} (x_{1i}^2 - x_{1i}) + \sum_i \gamma_{2i} (x_{2i}^2 - x_{2i}) \\ & \quad \left. + \sum_i \gamma_{3i} (x_{3i}^2 - x_{3i}) \right] \\ \text{s.t.} & \quad C2 - C7 \end{aligned} \quad (19)$$

where γ_{1i} , γ_{2i} , and γ_{3i} are, respectively, the penalty parameters for x_{1i} , x_{2i} , and x_{3i} . By adding penalty functions, the integer variable constraint is relaxed and the MINLP is converted to an LP, i.e., Equation (19).

4.2. The Solution of the Converted LP

To solve Equation (19), the Lagrange multiplier method is used. The Lagrange function is

$$\begin{aligned} L(x, f, R, \gamma, \delta, \zeta, \theta, \lambda, \mu, \rho) &= \sum_r \sum_j \sum_k (x_{1i} * T_{1i} + x_{2i} * T_{2i} + x_{3i} * T_{3i}) + \sum_i \gamma_{1i} (x_{1i}^2 - x_{1i}) + \sum_i \gamma_{2i} (x_{2i}^2 - x_{2i}) \\ &+ \sum_i \gamma_{3i} (x_{3i}^2 - x_{3i}) + \sum_i \delta_i (x_{1i} + x_{2i} + x_{3i} - 1) + \sum_i \zeta_i \left(\sum_{i \in Neb(j, K)} \sum_{k=1}^K x_{1i} * f_{ji}^{ONU} - F_j^{ONU} \right) \\ &+ \sum_i \theta_i \left(\sum_j \sum_{i \in Neb(j, K)} \sum_{k=1}^K x_{2i} * f_{ri}^{OLT} - F_r^{OLT} \right) + \sum_i \lambda_i \left(\sum_r \sum_j \sum_{i \in Neb(j, K)} \sum_{k=1}^K x_{3i} * f_i^{Int} - F_C^{Int} \right) \\ &+ \sum_i \mu_i \left(\sum_r \sum_j \sum_{i \in Neb(j, K)} \sum_{k=1}^K x_{3i} * f_i^{Int} - F_C^{Int} \right) + \sum_i \rho_i \left(\sum_{\mathbb{C}} x_{3i} * R_{Ci}^F - R^{Int} \right) \end{aligned} \quad (20)$$

The first-order derivatives of the Lagrange function with respect to $x, f, R, \gamma, \delta, \zeta, \theta, \lambda, \mu, \rho$, are easy to get. Let these first-order derivatives equal 0. By solving these equations,

their optimal values can be obtained. The optimal values are denoted as $(x^*, f^*, R^*, \gamma^*, \delta^*, \zeta^*, \theta^*, \lambda^*, \mu^*, \rho^*)$.

$$\begin{cases} \frac{\partial L}{\partial x} = 0, \frac{\partial L}{\partial f} = 0, \frac{\partial L}{\partial R} = 0, \frac{\partial L}{\partial \gamma} = 0, \frac{\partial L}{\partial \delta} = 0 \\ \frac{\partial L}{\partial \zeta} = 0, \frac{\partial L}{\partial \theta} = 0, \frac{\partial L}{\partial \lambda} = 0, \frac{\partial L}{\partial \mu} = 0, \frac{\partial L}{\partial \rho} = 0 \end{cases} \quad (21)$$

as Equation (19) is an LP with inequality constraint conditions. The result $(x^*, f^*, R^*, \gamma^*, \delta^*, \zeta^*, \theta^*, \lambda^*, \mu^*, \rho^*)$ should meet the Karush–Kuhn–Tucker (KKT) conditions, as follows.

$$\begin{aligned} \delta^*, \zeta^*, \theta^*, \lambda^*, \mu^*, \rho^* &\geq 0 \\ \delta^*(x_{1i} + x_{2i} + x_{3i} - 1) &= 0 \\ \zeta^*\left(\sum_{i \in Neb(j,K)} \sum_{k=1}^K x_{1i} * f_{ji}^{ONU} - F_j^{ONU}\right) &= 0 \\ \theta^*\left(\sum_j \sum_{i \in Neb(j,K)} \sum_{k=1}^K x_{2i} * f_{ri}^{OLT} - F_r^{OLT}\right) &= 0 \\ \lambda^*\left(\sum_r \sum_j \sum_{i \in Neb(j,K)} \sum_{k=1}^K x_{3i} * f_i^{Int} - F_C^{Int}\right) &= 0 \\ \mu^*\left(\sum_r \sum_j \sum_{i \in Neb(j,K)} \sum_{k=1}^K x_{3i} * f_i^{Int} - F_C^{Int}\right) &= 0 \\ \rho^*\left(\sum_{\mathbb{C}} x_{3i} * R_{Ci}^F - R^{Int}\right) &= 0 \end{aligned} \quad (22)$$

5. Numerical Simulation

To validate the performance of our work, a numerical simulation is executed in this section. The common parameters in this simulation are shown in Table 2. If extra or varied parameters are used in a specific scenario, they will be separately described in the following discussion.

Table 2. Common parameters.

Parameter	Value
p_0 , transmit power of users	35 dBm
σ , noise power density	-50 dBm/Hz
M, the number of receiving antennas	10
f^{ONU} , the data-computing speed of servers at ONU	3000 GHz
f^{OLT} , the data-computing speed of servers at OLT	2000 GHz
f^{Int} , the data-computing speed of servers at Intnet	1000 GHz
The number of OLTs	2
The number of ONUs for each OLT	16

To display the interaction of ONUs, MPPs, and MAPs, two fiber networks are concurrently connected to the Internet, forming the simulation scenario. Each OLT has 16 ONUs, via a 1:16 splitter. The fiber backhaul is a WDM-based PON. In this PON, the bandwidth is allocated in a static way. This means that the bandwidth for each ONU is fixed. The OLT would allocate a fixed bandwidth for each ONU, according to the ONU's regular service-level agreement (i.e., bandwidth, latency, etc.) The user/human/machine connects to ONUs through the multi-hop wireless link. The data-computing speed of servers at ONUs, OLTs, and the Internet, are different. The values of some constant parameters, such as p_0 , σ , and M, are also shown in Table 1.

The work of [20] is selected as a control group. In this control group, a three-tier network is also considered. The backhaul is not the fiber network with signal attenuations, but a fixed network with an ideal bandwidth (i.e., no signal attenuation). The wireless communication in the control group is completed via the single-hop wireless link, not the multi-hop wireless

link used in our work. Note that the solution to the control group's formulated problem is realized by the commutation method, not the constraint condition relaxation.

The data transmission rates of our work and the control group in the wireless sub-network are shown in Figure 3. The bandwidth of the wireless link is 1000 Hz. According to the 802.11 protocol family, the number of hops varies from 1 to 15. The traffic in this simulation is simplified as a data stream with a certain rate.

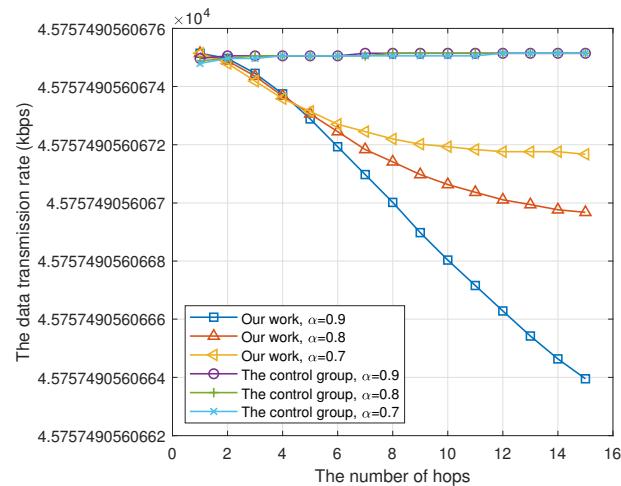


Figure 3. The data transmission rate vs. the number of hops.

For the control group, the data transmission rates with different α values are almost the same. For our work, data transmission rates with different α values are presented as three clearly differentiated curves. The signal attenuation in the control group only has an influence on the initial user/human/machine signal. The signal attenuation in our work affects the initial user/human/machine signal and noise at the same time.

With the increase in hop number, the transmission rate of our work decreases. This is the result of the multiple hops' multiplication effect. For the same reason, the larger α is, the stronger the downward trend of the data transmission rate. For each situation in Figure 3, the smallest value is 99.9999% of the largest value. The difference is negligible.

The signal strength received at the place of OLTs is shown in Figure 4. The distance here is from an ONU to its connected OLT. Assume that the signal strength at the place of ONUs is 0.045 mW, according to the calculation in Figure 3. The signal strength is inversely proportional to the distance. The further the distance, the more the signal fades in the fiber network and the weaker the signal strength. This can be also extrapolated, based on Equation (4).

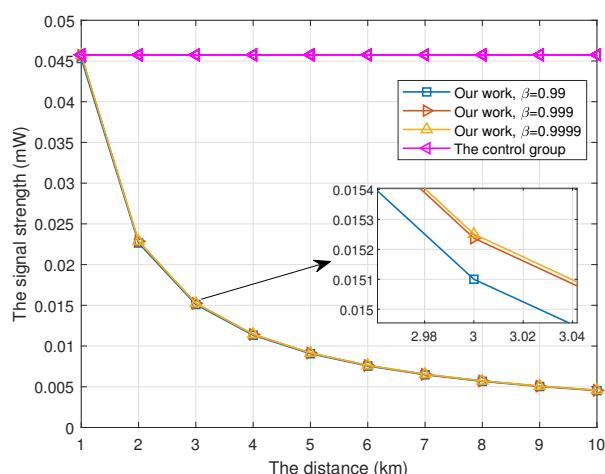


Figure 4. The signal strength vs. the distance.

For each distance in our work, the situation with the greatest β has the largest signal strength. In addition, the signal strength of our work is always than that of the control group, because the fiber link in the control group is ideal.

In Figures 4 and 5, the hop number of wireless communication is 1. The distance from ONUs to OLTs is 10 km. The distance from OLTs to the Internet is 70 km. In Figure 4, the maximum available data-computing speed of OLTs and the Internet are both 1.6 GHz. The data-computing speed of servers at the place of ONU varies from 0.2 GHz to 3.2 GHz. In Figure 5, the maximum available data-computing speed of ONUs and the Internet are both 1.6 GHz. The data-computing speed of servers at the place of OLT varies from 0.2 GHz to 3.2 GHz. To show the difference between MEC-1 offloading and MEC-2 offloading, the total latency of all tasks was selected as an indicator. This is in accordance with Equation (17).

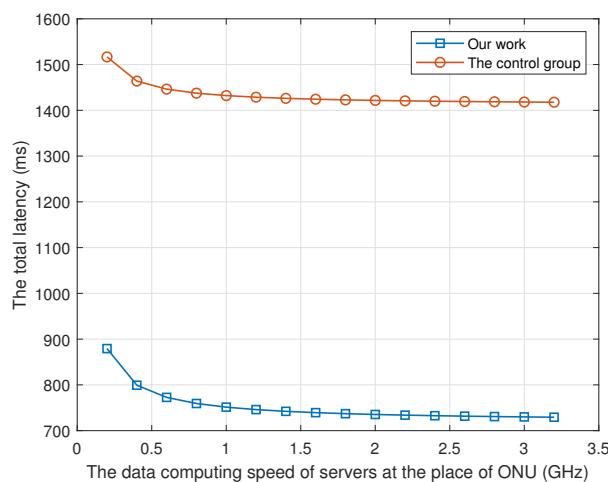


Figure 5. The total latency with MEC-1 offloading.

In Figure 5, the total latency of our work is always smaller than that of the control group. The wireless sub-network is established upon multi-hop links. Through multiple hops, a user/human/machine can choose the most suitable path to the server via its nearest ONU or sub-nearest ONU. Meanwhile, for the control group, there is only one choice, i.e., the connected ONU. With the increase in the data-computing speed of servers at the place of ONU, the total latency decreases. When the data-computing speed increases from 0.2 to 1.0, the total latency decreases sharply. When the data-computing speed increases from 1.0 to 3.2, the total latency slightly decreases.

In Figure 6, the total latency of our work has a crossover point with that of the control group. The data-computing speed of servers at the place of OLT is 1.6 GHz when two lines cross. The reason for this is that MEC-2 offload can provide a trade-off between shorting paths and increasing offloading locations. Before this point (≤ 1.6), the total latency of our work is larger than that of the control group. After this point (≥ 1.6), the total latency of our work is smaller than that of the control group. In addition, whether our work or the control group, the total latency decreases with the increase in data-computing speed.

Combining Figures 5 and 6, the MEC-2 offloading may be more capable of MEC-1 offloading with the parameters we set in this simulation.

In the control group, the service is modeled as a uniformly generated computing task in the range of [5, 30] Mbits. However, in the real production environment, the tactile internet service, such as remote robots, lasts for a long-term period with a relatively stable data rate. The amount of data in the tactile internet service is greater than that of the traditional service, due to the existence of high-definition picture or video transmissions. The same situation exists for the data transmission rate of tactile internet services. Each tactile service is seen as a long-lasting data stream with a uniformly distributed data rate in the range of [200, 500] Mbit/s.

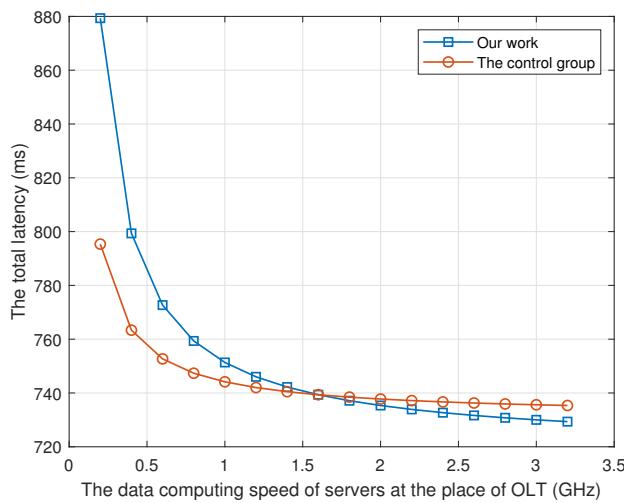


Figure 6. The total latency with MEC-2 offloading.

The latency with different tactile services is shown in Figure 7. According to the 802.11 protocol cluster, the hop number in a wireless mesh network cannot be greater than 16. To demonstrate the variation in the latencies of multiple hop numbers, [1, 7, 15] hops were selected. The distance of each hop is 200 m. The distance between ONUs and OLT is 15 km. The distance between OLT and the Internet is 50 km.

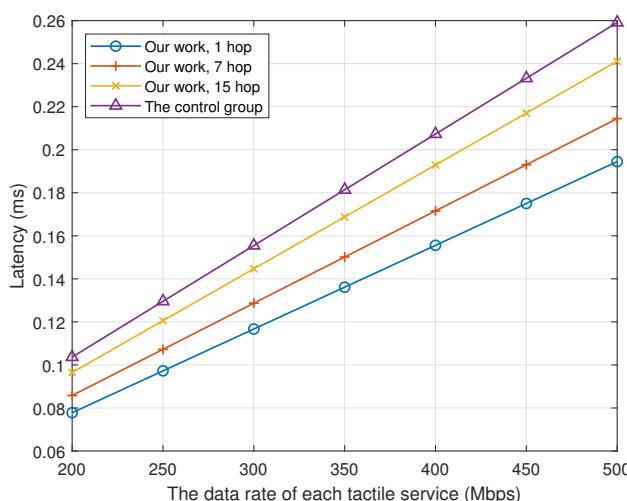


Figure 7. The latency with different tactile services.

With the growth in the data rate of each tactile service, the latencies of different situations always increases. The control group has the largest values. The larger the hop number, the greater the latency of our work. The multiple hops in the wireless mesh network form a complex network with redundant paths. The existence of communication among ONUs makes the data transmission path short, from user-ONU-OLT-ONU-user to user-ONU-another ONU-user. This is the reason that the latency of our work with 1 hop is still smaller than that of the control group.

The capacity of our work is shown in Figure 8. The network between ONUs and OLT is WDM-based. The bandwidth of the fiber link from ONUs to OLT is 5 Gbps. The link from OLT to the Internet is the core network. Its bandwidth is 100 Gbps. The required bandwidth of each tactile internet service varies from 100 Mbps to 500 Mbps, with a step of 100 Mbps.

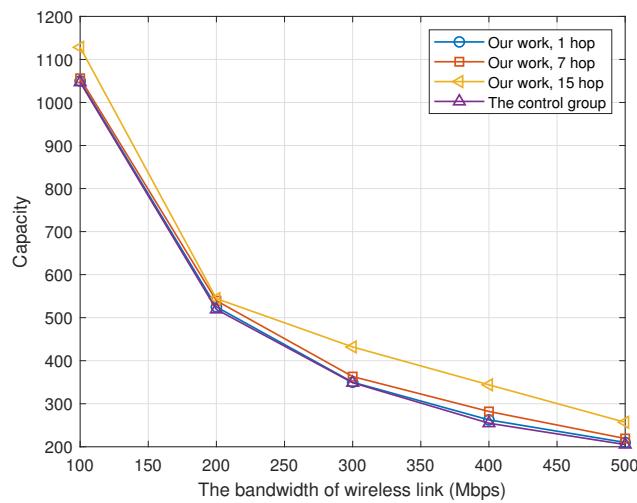


Figure 8. The capacity vs. the bandwidth of wireless link.

In Figure 8, our work with 15 hops showed the greatest capacity compared with our work (1 hop), our work (7 hops), and the control group. A greater hop means that the wireless mesh network has a more redundant path. By offloading the computing task for tactile internet services from the Internet, or OLT to the ONUs closer to users, the capacity of the entire network can grow. When the bandwidth of wireless links increases, the capacities of different situations all decrease. The total bandwidth of the entire network is limited; the bandwidth of a tactile internet service leads to small capacity. For a fixed bandwidth in the wireless link, such as 400 Mbps, our work with one hop still has a higher capacity compared with that of the control group.

The capacity with different fiber bandwidths from ONUs to OLT is shown in Figure 9. In the WDM-based PON network, the bandwidth of the ONU-OLT link is [1.25, 2.50, 3.75, 5.00, 6.25, 7.50, 8.75, 10.00] Gbps. The bandwidth of each tactile internet service is 300 Mbps. The control group has the lowest capacity and a relatively stable increase rate. The reason for this phenomenon is that the wireless network of the control group only has a single hop. This means that the tactile internet service distribution has no choice of candidates. By contrast, our works with different hops all have more fluctuating curves. As the hop number grows, the candidate paths in the wireless mesh network increase. Our work with the largest hop has the most fluctuating curve. These fluctuations in our work are caused by the variety of offloading schemes.

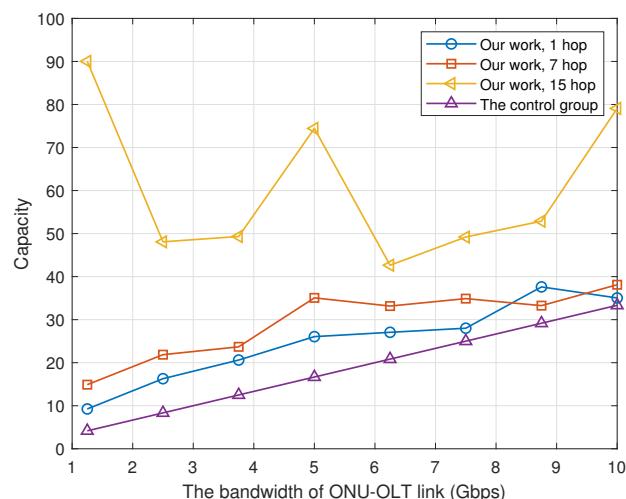


Figure 9. The capacity vs. the bandwidth of ONU-OLT link.

6. Conclusions

To obtain the minimum latency over FiWi infrastructures for tactile internet services, a latency-aware offloading strategy is proposed in this paper. Three kinds of offloading are investigated and modeled, as well as the wireless link, the fiber link, and the data rate. Based on these models, an MINLP problem is formulated. A two-step solution, including the constraint relaxation and the Lagrange multiplier method, is put forward. Through numerical simulation, the performance of our work is validated. In future work, the mobility of user/human/machine and the handover mechanism may be of interest.

Author Contributions: Conceptualization, Q.D.; Formal analysis, J.L.; Investigation, J.Q.; Methodology, G.Q.; Validation, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by R&D Program of Beijing Municipal Education Commission (Research on Optical and Wireless converged Access Network Networking Technology in Smart Traffic, No. KM202111417010), China Computer Federation (CCF) Opening Project of Information System (Research on Massive Event Flow oriented Stream Computing Framework, No. CCFIS2019-01-01). And thanks to Li.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Chowdhury, M.; Maier, M. Toward Dynamic HART-Centric Task Offloading over FiWi Infrastructures in the Tactile Internet Era. *IEEE Commun. Mag.* **2019**, *57*, 123–128. [[CrossRef](#)]
- Maier, M. Toward 6G: A new era of convergence. In *Optical Fiber Communication Conference*; Optical Society of America: Washington, DC, USA, 2021; p. F4H-1.
- Lo Bello, L.; Steiner, W. A Perspective on IEEE Time-Sensitive Networking for Industrial Communication and Automation Systems. *Proc. IEEE* **2019**, *107*, 1094–1120. [[CrossRef](#)]
- Maier, M. The Tactile Internet: Where Do We Go From Here? In Proceedings of the 2018 Asia Communications and Photonics Conference (ACP), Hangzhou, China, 26–29 October 2018; pp. 1–3. [[CrossRef](#)]
- Zhang, H.; Hu, Y.; Wang, R.; Li, Z.; Zhang, P.; Xu, R. Energy-Efficient Frame Aggregation Scheme in IoT Over Fiber-Wireless Networks. *IEEE Internet Things J.* **2021**, *8*, 10779–10791. [[CrossRef](#)]
- Gu, Z.; Lu, H.; Zhu, Z. On Throughput Optimization and Bound Analysis in Cache-Enabled Fiber-Wireless Networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 9068–9082. doi: 10.1109/TVT.2020.3000487. [[CrossRef](#)]
- Mesodiakaki, A.; Maniotis, P.; Gatzianas, M.; Vagionas, C.; Pleros, N.; Kalfas, G. A Gated Service MAC Protocol for Sub-Ms Latency 5G Fiber-Wireless mmWave C-RANs. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 2502–2515. [[CrossRef](#)]
- Beniiche, A.; Ebrahimbzadeh, A.; Maier, M. The Way of the DAO: Toward Decentralizing the Tactile Internet. *IEEE Netw.* **2021**, *35*, 190–197. [[CrossRef](#)]
- Zhou, Y.; Zhang, D. Near-End Cloud Computing: Opportunities and Challenges in the Post-Cloud Computing Era. *Chin. J. Comput.* **2019**, *42*, 677–700.
- Qi, Q.; Chen, X.; Zhong, C.; Zhang, Z. Integration of Energy, Computation and Communication in 6G Cellular Internet of Things. *IEEE Commun. Lett.* **2020**, *24*, 1333–1337. [[CrossRef](#)]
- Yang, Y. Multi-tier computing networks for intelligent IoT. *Nat. Electron.* **2019**, *2*, 4–5. [[CrossRef](#)]
- Mondal, W.U.; Das, G. On Exact Distribution of Poisson-Voronoi Area in K-Tier HetNets with Generalized Association Rule. *IEEE Commun. Lett.* **2020**, *24*, 2142–2146. [[CrossRef](#)]
- Chen, N.; Okada, M. Toward 6G Internet of Things and the Convergence With RoF System. *IEEE Internet Things J.* **2021**, *8*, 8719–8733. [[CrossRef](#)]
- Maier, M.; Ebrahimbzadeh, A.; Rostami, S.; Beniiche, A. The Internet of No Things: Making the Internet Disappear and “See the Invisible”. *IEEE Commun. Mag.* **2020**, *58*, 76–82. [[CrossRef](#)]
- Li, X.; Zhou, Z.; Zhu, C.; Shu, L.; Zhou, J. Online Reconfiguration of Latency-Aware IoT Services in Edge Networks. *IEEE Internet Things J.* **2021**, *1*–1. [[CrossRef](#)]
- Mondal, S.; Ruan, L.; Maier, M.; Larrabeiti, D.; Das, G.; Wong, E. Enabling Remote Human-to-Machine Applications With AI-Enhanced Servers Over Access Networks. *IEEE Open J. Commun. Soc.* **2020**, *1*, 889–899. [[CrossRef](#)]
- Wu, Z.Y.; Ismail, M.; Serpedin, E.; Wang, J. Data-Driven Link Assignment With QoS Guarantee in Mobile RF-Optical HetNet of Things. *IEEE Internet Things J.* **2020**, *7*, 5088–5102. [[CrossRef](#)]

18. Guo, M.; Shou, G.; Xue, J.; Hu, Y.; Liu, Y.; Guo, Z. Cross-domain Interconnection with Time Synchronization in Software-defined Time-Sensitive Networks. In Proceedings of the 2020 Asia Communications and Photonics Conference (ACP) and International Conference on Information Photonics and Optical Communications (IPOC), Beijing, China, 24–27 October 2020; pp. 1–3.
19. Pérez, G.O.; Ebrahimzadeh, A.; Maier, M.; Hernández, J.A.; López, D.L.; Veiga, M.F. Decentralized Coordination of Converged Tactile Internet and MEC Services in H-CRAN Fiber Wireless Networks. *J. Light. Technol.* **2020**, *38*, 4935–4947. [[CrossRef](#)]
20. Pan, Y.; Jiang, H.; Zhu, H.; Wang, J. Latency Minimization for Task Offloading in Hierarchical Fog-Computing C-RAN Networks. In Proceedings of the ICC 2020—2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6. [[CrossRef](#)]